Buscadores Web

(segunda parte).

José David Flores Peñaloza

IIMAS-IMATE

UNAM

8/11/2005

Contenido

- Recordatorio
- Respuestas pendientes de la clase pasada
- PageRank
  - Definición simplificada
  - Ejemplo
  - Modelo de visitante aleatorio
  - Convergencia
  - Ventajas
  - Desventajas
- Vistazo (muy) rapido a Swoogle
- Conclusiones

Recordatorio

HITS

- Basado en el análisis de ligas de la red.
- Otorga calificaciones de hub y de autoridad a cada página de un conjunto de documentos relevantes a una consulta.
- Conjunto vecindad para cada consulta.
- Se resuelve un eigenproblema por consulta.

Mi tarea

Link: liga

Mi tarea [cont.]

Aplicación de HITS sobre Google.

Obtendremos resultados posiblemente similares (en autoridades), pero ahora tendremos tambien un conjunto de buenos hubs.

Mi tarea [cont.]

Agrupar con HITS documentos por tema.

- Los vectores de hub y autoridad dan la colección ligada más *densa* de hubs y autoridades de la subgráfica G_σ definida por una cadena de consulta σ.
- Uno puede estar interesado en encontrar varias colecciones densas de hubs y autoridades.
- Cada una de estas colecciones puede ser potencialmente relevante al tema de la consulta, pero pueden estar separadas entre si por una variedad de razones...

Mi tarea [cont.]

- La cadena de consulta puede tener varios significados diferentes (Ej. “Jaguar”).
- Puede aparecer como un término en el contexto de múltiples comunidades técnicas. (Ej. “Randomized algorithms”).
- Puede referirse a un tema muy popular, involucrando a grupos que dificilmente van a ligarse entre si (Ej. “abortion”)

Mi tarea [cont.]

Los eigenvectores no-principales de las matrices de hubs y autoridad, ofrecen una manera natural de extraer colecciones densamente ligadas del conjunto base.

A diferencia del principal, los otros eigenvectores tienen coordenadas positivas y negativas.

Cada uno de estos vectores induce dos conjuntos densamente ligados de hubs y autoridades: El de las páginas que corresponden a los valores más positivos, y el de las páginas de los valores más negativos.

Ejemplo 1: consulta “Jaguar”

(jaguar*) Autoridades: 2o vector no principal, calificaciones más positivas.

.255	http://www.jaguarsnfl.com/	Official Jacksonville Jaguars NFL Website
.137	http://www.nando.net/SportServer/football/nfl/jax.html	Jacksonville Jaguars Home Page
.133	http://www.ao.net/;brett/jaguar/index.html	Brett’s Jaguar Page
.110	http://www.usatoday.com/sports/football/sfn/sfn30.htm	Jacksonville Jaguars

(jaguar*) Autoridades: 3er vector no principal, calificaciones más positivas.

.227	http://www.jaguarvehicles.com/	Jaguar Cars Global Home Page
.227	http://www.collection.co.uk/	The Jaguar Collection—Official Web site
.211	http://www.moran.com/sterling/sterling.html	
.211	http://www.coys.co.uk/	

Ejemplo 2: consulta “randomized algorithms”

(“randomized algorithms”) Authorities: 1st nonprincipal vector, positive end

.125	http://theory.lcs.mit.edu/~goemans/	Michel X. Goemans
.122	http://theory.lcs.mit.edu/~spielman/	Dan Spielman’s Homepage
.122	http://www.nada.kth.se/~johanh/	Johan Hastad
.122	http://theory.lcs.mit.edu/~rivest/	Ronald L. Rivest: HomePage

(“randomized algorithms”) Authorities 1st nonprincipal vector, negative end

-.00116	http://lib.stat.cmu.edu/	StatLib Index
-.00115	http://www.geo.fmi.fi/prog/tela.html	Tela
-.00107	http://gams.nist.gov/	GAMS: Guide to Available Mathematical Software
-.00107	http://www.netlib.org	Netlib

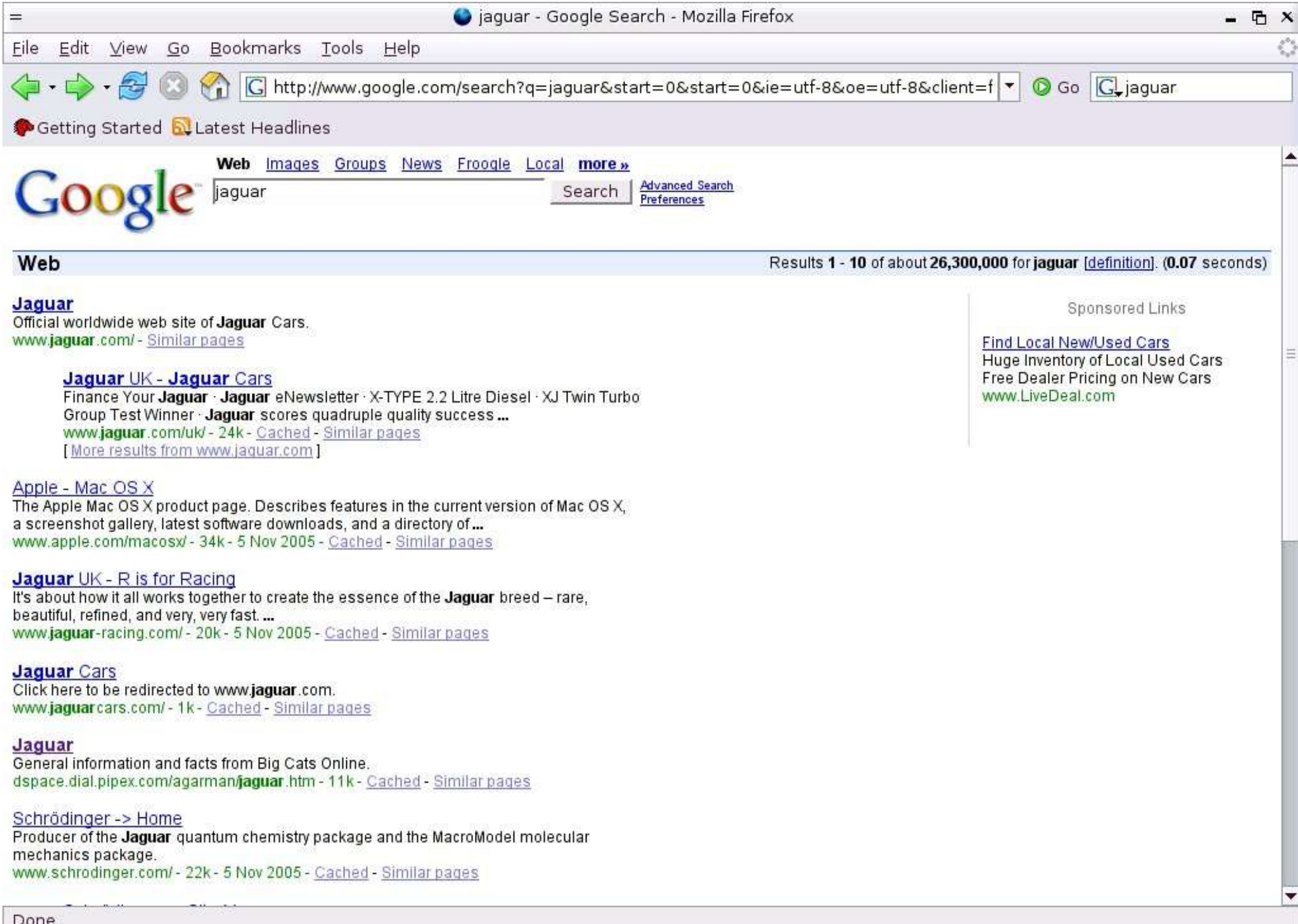
(“randomized algorithms”) Authorities 4th nonprincipal vector, negative end

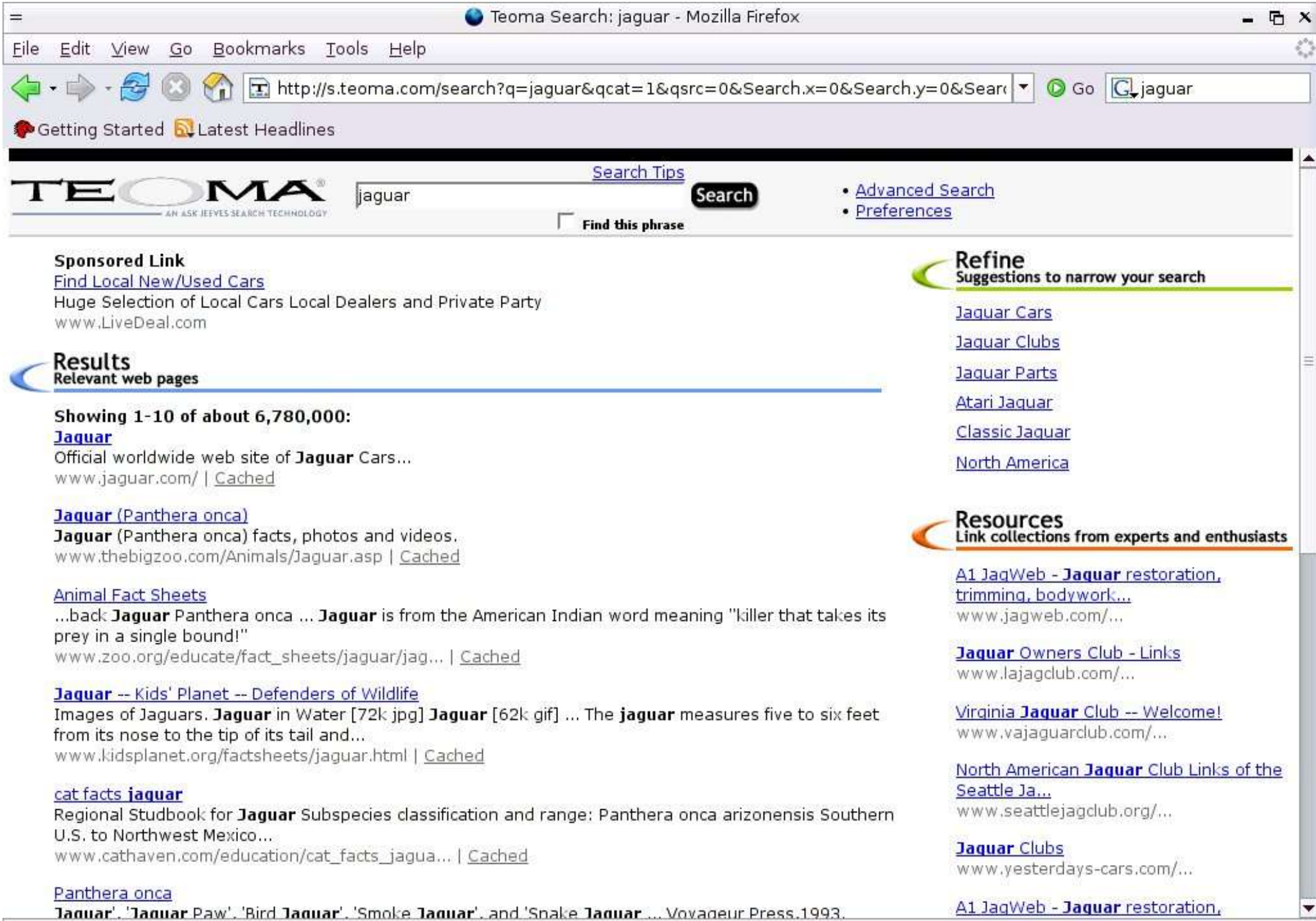
-.176	http://www.amara.com/current/wavelet.html	Amara’s Wavelet Page
-.172	http://www-ocean.tamu.edu/;baum/wavelets.html	Wavelet sources
-.161	http://www.mathsoft.com/wavelets.html	Wavelet Resources
-.143	http://www.mat.sbg.ac.at/;uhl/wav.htm	l Wavelets

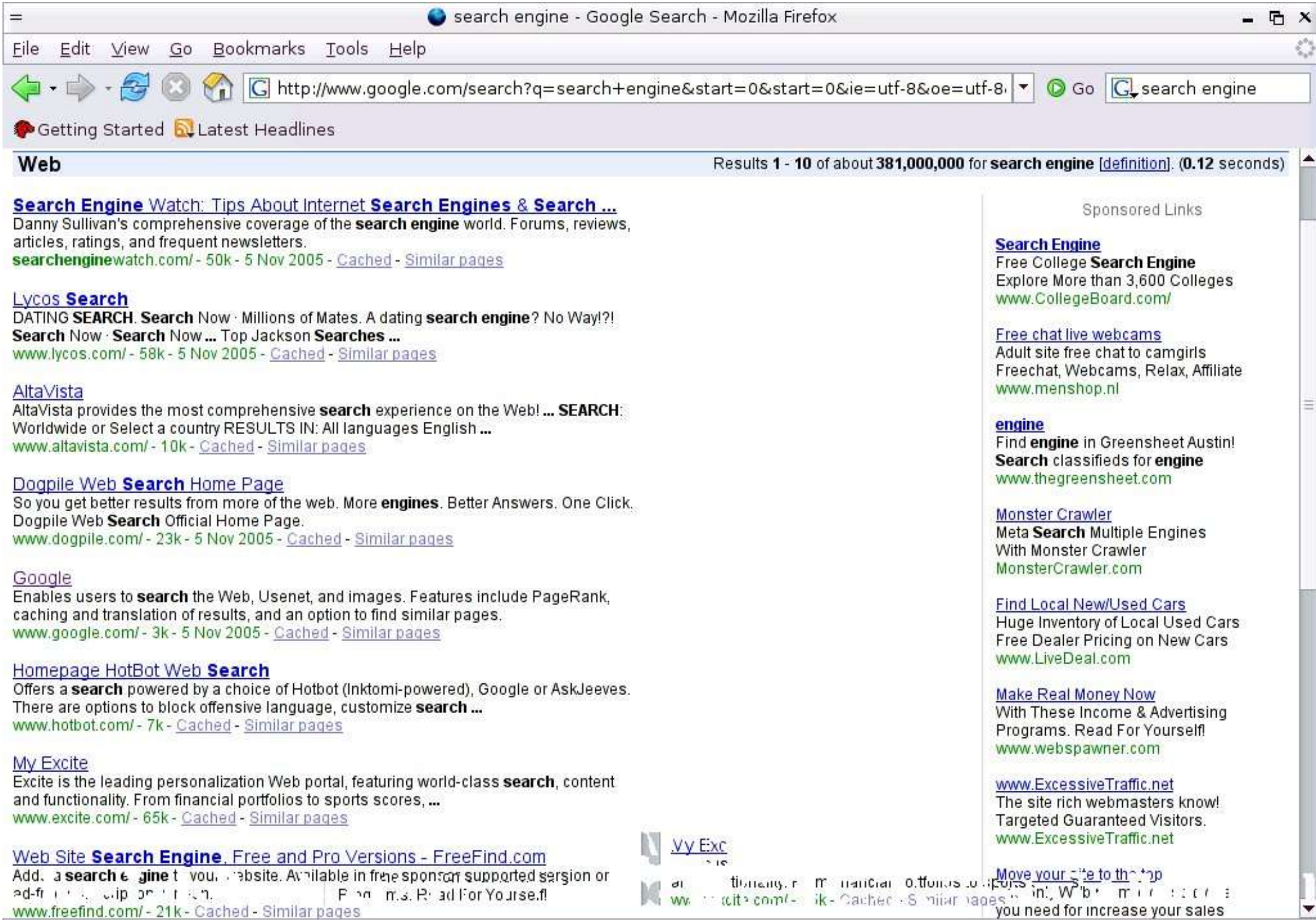
Ejemplo 3: consulta “abortion”

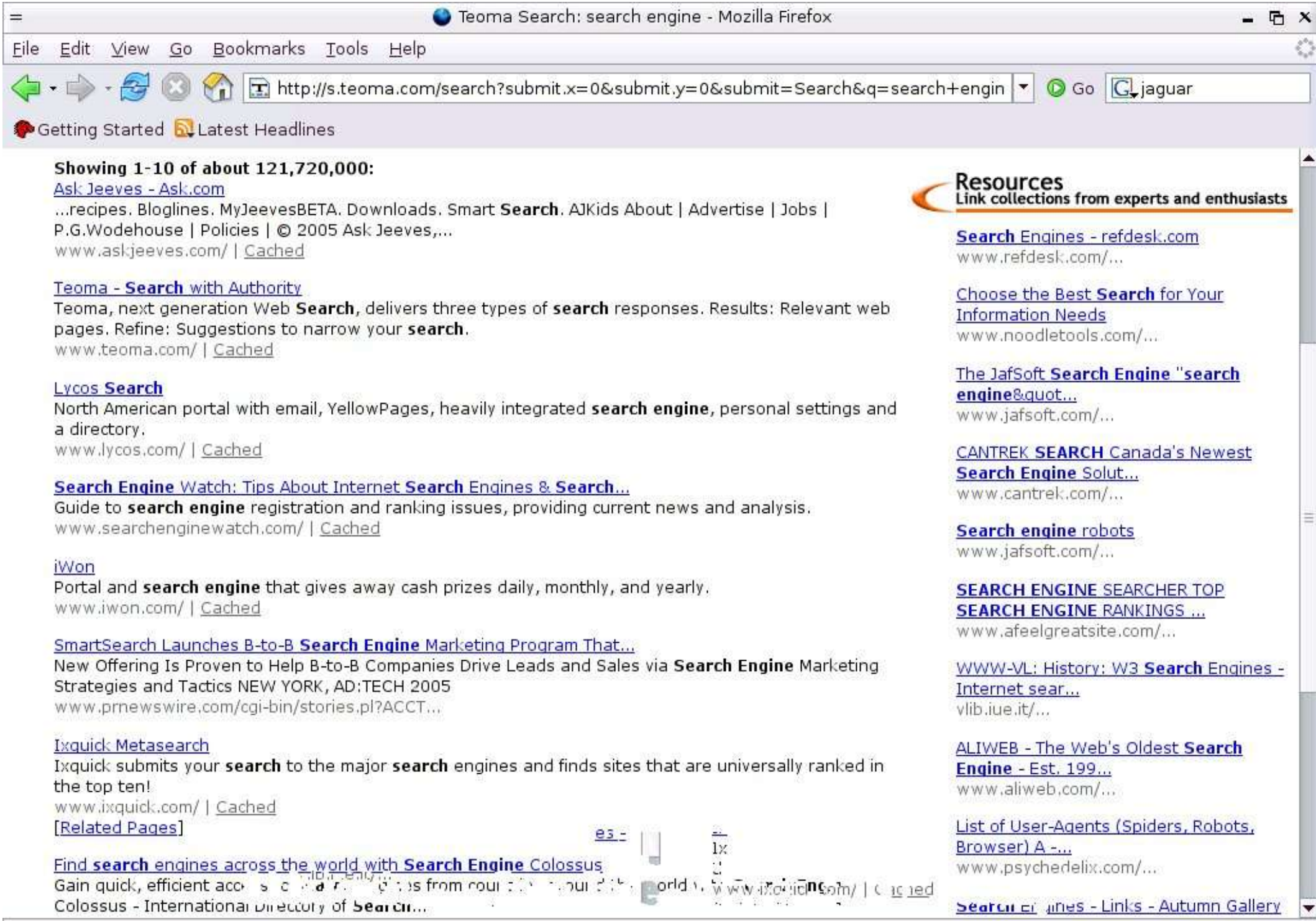
(abortion) Authorities: 2nd nonprincipal vector, positive end
.321 http://www.caral.org/abortion.html Abortion and Reproductive Rights Internet Resources
.219 http://www.plannedparenthood.org/ Welcome to Planned Parenthood
.195 http://www.gynpages.com/ Abortion Clinics OnLine
.172 http://www.oneworld.org/ippf/ IPPF Home Page
.162 http://www.prochoice.org/naf/ The National Abortion Federation
.161 http://www.lm.com/;lmann/feminist/abortion.html

(abortion) Authorities: 2nd nonprincipal vector, negative end
-.197 http://www.awinc.com/partners/bc/commpass/lifenet/lifenet.htm LifeWEB
-.169 http://www.worldvillage.com/wv/square/chapel/xwalk/html/peter.htm Healing after Abortion
-.164 http://www.nebula.net/;maeve/lifelink.html
-.150 http://members.aol.com/pladvocate/ Pro-Life Advocate
-.144 http://www.clark.net/pub/jeffd/factbot.html The Right Side of the Web
-.144 http://www.catholic.net/HyperNews/get/abortion.html


jaguar - Google Search - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://www.google.com/search?q=jaguar&start=0&start=0&ie=utf-8&oe=utf-8&client=f
Go
jaguar
Getting Started
Latest Headlines
Google
Web Images Groups News Froogle Local more »
jaguar
Search
Advanced Search
Preferences
Web
Results 1 - 10 of about 26,300,000 for jaguar [definition]. (0.07 seconds)
Jaguar
Official worldwide web site of Jaguar Cars.
www.jaguar.com/ - Similar pages
Jaguar UK - Jaguar Cars
Finance Your Jaguar · Jaguar eNewsletter · X-TYPE 2.2 Litre Diesel · XJ Twin Turbo Group Test Winner · Jaguar scores quadruple quality success ...
www.jaguar.com/uk/ - 24k - Cached - Similar pages
[More results from www.jaguar.com]
Apple - Mac OS X
The Apple Mac OS X product page. Describes features in the current version of Mac OS X, a screenshot gallery, latest software downloads, and a directory of ...
www.apple.com/macosx/ - 34k - 5 Nov 2005 - Cached - Similar pages
Jaguar UK - R is for Racing
It's about how it all works together to create the essence of the Jaguar breed – rare, beautiful, refined, and very, very fast. ...
www.jaguar-racing.com/ - 20k - 5 Nov 2005 - Cached - Similar pages
Jaguar Cars
Click here to be redirected to www.jaguar.com.
www.jaguarcars.com/ - 1k - Cached - Similar pages
Jaguar
General information and facts from Big Cats Online.
dspace.dial.pipex.com/agarman/jaguar.htm - 11k - Cached - Similar pages
Schrödinger -> Home
Producer of the Jaguar quantum chemistry package and the MacroModel molecular mechanics package.
www.schrodinger.com/ - 22k - 5 Nov 2005 - Cached - Similar pages
Sponsored Links
Find Local New/Used Cars
Huge Inventory of Local Used Cars
Free Dealer Pricing on New Cars
www.LiveDeal.com
Done


Teoma Search: jaguar - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://s.teoma.com/search?q=jaguar&qcat=1&qsrc=0&Search.x=0&Search.y=0&Sear
Go
jaguar
Getting Started
Latest Headlines
TEOMA
AN ASK JEEVES SEARCH TECHNOLOGY
Search Tips
jaguar
Search
Find this phrase
Advanced Search
Preferences
Sponsored Link
Find Local New/Used Cars
Huge Selection of Local Cars Local Dealers and Private Party
www.LiveDeal.com
Results
Relevant web pages
Showing 1-10 of about 6,780,000:
Jaguar
Official worldwide web site of Jaguar Cars...
www.jaguar.com/ | Cached
Jaguar (Panthera onca)
Jaguar (Panthera onca) facts, photos and videos.
www.thebigzoo.com/Animals/Jaguar.asp | Cached
Animal Fact Sheets
...back Jaguar Panthera onca ... Jaguar is from the American Indian word meaning "killer that takes its prey in a single bound!"
www.zoo.org/educate/fact_sheets/jaguar/jag... | Cached
Jaguar -- Kids' Planet -- Defenders of Wildlife
Images of Jaguars. Jaguar in Water [72k jpg] Jaguar [62k gif] ... The jaguar measures five to six feet from its nose to the tip of its tail and...
www.kidsplanet.org/factsheets/jaguar.html | Cached
cat facts jaguar
Regional Studbook for Jaguar Subspecies classification and range: Panthera onca arizonensis Southern U.S. to Northwest Mexico...
www.cathaven.com/education/cat_facts_jagua... | Cached
Panthera onca
Jaguar', 'Jaguar Paw', 'Bird Jaguar', 'Smoke Jaguar', and 'Snake Jaguar ... Voyageur Press,1993.
Refine
Suggestions to narrow your search
Jaguar Cars
Jaguar Clubs
Jaguar Parts
Atari Jaguar
Classic Jaguar
North America
Resources
Link collections from experts and enthusiasts
A1 JagWeb - Jaguar restoration, trimming, bodywork...
www.jagweb.com/...
Jaguar Owners Club - Links
www.lajagclub.com/...
Virginia Jaguar Club -- Welcome!
www.vajaguarclub.com/...
North American Jaguar Club Links of the Seattle Ja...
www.seattlejagclub.org/...
Jaguar Clubs
www.yesterdays-cars.com/...
A1 JagWeb - Jaguar restoration,
Done


search engine - Google Search - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://www.google.com/search?q=search+engine&start=0&start=0&ie=utf-8&oe=utf-8
Go
search engine
Getting Started
Latest Headlines
Web
Results 1 - 10 of about 381,000,000 for search engine [definition]. (0.12 seconds)
Search Engine Watch: Tips About Internet Search Engines & Search ...
Danny Sullivan's comprehensive coverage of the search engine world. Forums, reviews, articles, ratings, and frequent newsletters.
searchenginewatch.com/ - 50k - 5 Nov 2005 - Cached - Similar pages
Lycos Search
DATING SEARCH. Search Now · Millions of Mates. A dating search engine? No Way!?! Search Now · Search Now ... Top Jackson Searches ...
www.lycos.com/ - 58k - 5 Nov 2005 - Cached - Similar pages
AltaVista
AltaVista provides the most comprehensive search experience on the Web! ... SEARCH: Worldwide or Select a country RESULTS IN: All languages English ...
www.altavista.com/ - 10k - Cached - Similar pages
Dogpile Web Search Home Page
So you get better results from more of the web. More engines. Better Answers. One Click. Dogpile Web Search Official Home Page.
www.dogpile.com/ - 23k - 5 Nov 2005 - Cached - Similar pages
Google
Enables users to search the Web, Usenet, and images. Features include PageRank, caching and translation of results, and an option to find similar pages.
www.google.com/ - 3k - 5 Nov 2005 - Cached - Similar pages
Homepage HotBot Web Search
Offers a search powered by a choice of Hotbot (Inktomi-powered), Google or AskJeeves. There are options to block offensive language, customize search ...
www.hotbot.com/ - 7k - Cached - Similar pages
My Excite
Excite is the leading personalization Web portal, featuring world-class search, content and functionality. From financial portfolios to sports scores, ...
www.excite.com/ - 65k - Cached - Similar pages
Web Site Search Engine, Free and Pro Versions - FreeFind.com
www.freefind.com/ - 21k - Cached - Similar pages
Sponsored Links
Search Engine
Free College Search Engine
Explore More than 3,600 Colleges
www.CollegeBoard.com/
Free chat live webcams
Adult site free chat to camgirls
Freechat, Webcams, Relax, Affiliate
www.menshop.nl
engine
Find engine in Greensheet Austin!
Search classifieds for engine
www.thegreensheet.com
Monster Crawler
Meta Search Multiple Engines
With Monster Crawler
MonsterCrawler.com
Find Local New/Used Cars
Huge Inventory of Local Used Cars
Free Dealer Pricing on New Cars
www.LiveDeal.com
Make Real Money Now
With These Income & Advertising
Programs. Read For Yourself!
www.webspawner.com
www.ExcessiveTraffic.net
The site rich webmasters know!
Targeted Guaranteed Visitors.
www.ExcessiveTraffic.net
Move your site to the top
you need for increase your sales
Done


Teoma Search: search engine - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://s.teoma.com/search?submit.x=0&submit.y=0&submit=Search&q=search+engin
Go
jaguar
Getting Started
Latest Headlines
Showing 1-10 of about 121,720,000:
Ask Jeeves - Ask.com
...recipes. Bloglines. MyJeevesBETA. Downloads. Smart Search. AJKids About | Advertise | Jobs | P.G.Wodehouse | Policies | © 2005 Ask Jeeves,...
www.askjeeves.com/ | Cached
Teoma - Search with Authority
Teoma, next generation Web Search, delivers three types of search responses. Results: Relevant web pages. Refine: Suggestions to narrow your search.
www.teoma.com/ | Cached
Lycos Search
North American portal with email, YellowPages, heavily integrated search engine, personal settings and a directory.
www.lycos.com/ | Cached
Search Engine Watch: Tips About Internet Search Engines & Search...
Guide to search engine registration and ranking issues, providing current news and analysis.
www.searchenginewatch.com/ | Cached
iWon
Portal and search engine that gives away cash prizes daily, monthly, and yearly.
www.iwon.com/ | Cached
SmartSearch Launches B-to-B Search Engine Marketing Program That...
New Offering Is Proven to Help B-to-B Companies Drive Leads and Sales via Search Engine Marketing Strategies and Tactics NEW YORK, AD:TECH 2005
www.prnewswire.com/cgi-bin/stories.pl?ACCT...
Ixquick Metasearch
Ixquick submits your search to the major search engines and finds sites that are universally ranked in the top ten!
www.ixquick.com/ | Cached
[Related Pages]
Find search engines across the world with Search Engine Colossus
Gain quick, efficient acc
s from
Colossus - International Directory of Search...
Resources
Link collections from experts and enthusiasts
Search Engines - refdesk.com
www.refdesk.com/...
Choose the Best Search for Your Information Needs
www.noodletools.com/...
The JafSoft Search Engine "search engine&quot...
www.jafsoft.com/...
CANTREK SEARCH Canada's Newest Search Engine Solut...
www.cantrek.com/...
Search engine robots
www.jafsoft.com/...
SEARCH ENGINE SEARCHER TOP SEARCH ENGINE RANKINGS ...
www.afeelgreatsite.com/...
WWW-VL: History: W3 Search Engines - Internet sear...
vlib.iue.it/...
ALIWEB - The Web's Oldest Search Engine - Est. 199...
www.aliweb.com/...
List of User-Agents (Spiders, Robots, Browser) A -...
www.psychedelix.com/...
Search Engines - Links - Autumn Gallery
Done

PageRank

PageRank

Propuesto por Larry Page y Sergei Brin para RI en web mientras eran estudiantes de doctorado en Stanford (1998).

PageRank [cont.]

Idea original de Geller, N. en 1978 para su uso en bibliometría[3].

Curiosamente ese articulo no es citado por Page-Brin en sus artículos [4,5].

Pone énfasis en normalización de pesos de ligas y navegación web basada en modelos de caminatas aleatorias.

PageRank [cont.]

- Es un método para asignar una calificación **a cada página**, basado en la estructura de la gráfica de la web.
- Tiene aplicaciones en búsqueda, navegación y estimación de trafico.
- “El corazón de nuestro software es PageRank... provee la base para todas nuestras herramientas de busqueda” [Google]
- Ha sido responsable de posicionar a Google como el buscador más usado del mundo.

PageRank [cont.]

- Es un método para asignar una calificación **a cada página**, basado en la estructura de la gráfica de la web.
- Tiene aplicaciones en búsqueda, navegación y estimación de trafico.
- "El corazón de nuestro software es PageRank... provee la base para todas nuestras herramientas de busqueda" [Google]
- Ha sido responsable de posicionar a Google como el buscador más usado del mundo.

PageRank [cont.]

- Es un método para asignar una calificación **a cada página**, basado en la estructura de la gráfica de la web.
- Tiene aplicaciones en búsqueda, navegación y estimación de trafico.
- "El corazón de nuestro software es PageRank... provee la base para todas nuestras herramientas de busqueda" [Google]
- Ha sido responsable de posicionar a Google como el buscador más usado del mundo.

PageRank [cont.]

- Es un método para asignar una calificación **a cada página**, basado en la estructura de la gráfica de la web.
- Tiene aplicaciones en búsqueda, navegación y estimación de trafico.
- "El corazón de nuestro software es PageRank... provee la base para todas nuestras herramientas de busqueda" [Google]
- Ha sido responsable de posicionar a Google como el buscador más usado del mundo.

PageRank [cont.]

- Idea parecida a la de HITS..
- La idea es que una liga de A a B se considera una recomendación de A hacia B.
- “Una página es importante si es referenciada por páginas importantes”.
- Definición recursiva...

PageRank [cont.]

- Idea parecida a la de HITS..
- La idea es que una liga de A a B se considera una recomendación de A hacia B.
- "Una página es importante si es referenciada por páginas importantes".
- Definición recursiva...

PageRank [cont.]

- Idea parecida a la de HITS..
- La idea es que una liga de A a B se considera una recomendación de A hacia B.
- “Una página es importante si es referenciada por páginas importantes”.
- Definición recursiva...

PageRank [cont.]

- Idea parecida a la de HITS..
- La idea es que una liga de A a B se considera una recomendación de A hacia B.
- “Una página es importante si es referenciada por páginas importantes”.
- Definición recursiva...

PageRank [cont.]

No todas las recomendaciones son igual de importantes. Tiene más valor una recomendación de Mozilla.org que una de mi página web.

Pero si Mozilla tiene una liga hacia mi página no se debe considerar que mi página es igual de importante!.

El valor del voto de A hacia B es la importancia de A, dividida entre el numero de enlaces de A hacia otras páginas... La importancia de A se distribuye entre todas las páginas a las que recomienda.

PageRank [cont.]

No todas las recomendaciones son igual de importantes. Tiene más valor una recomendación de Mozilla.org que una de mi página web.

Pero si Mozilla tiene una liga hacia mi página no se debe considerar que mi página es igual de importante!.

El valor del voto de A hacia B es la importancia de A, dividida entre el numero de enlaces de A hacia otras páginas... La importancia de A se distribuye entre todas las páginas a las que recomienda.

PageRank [cont.]

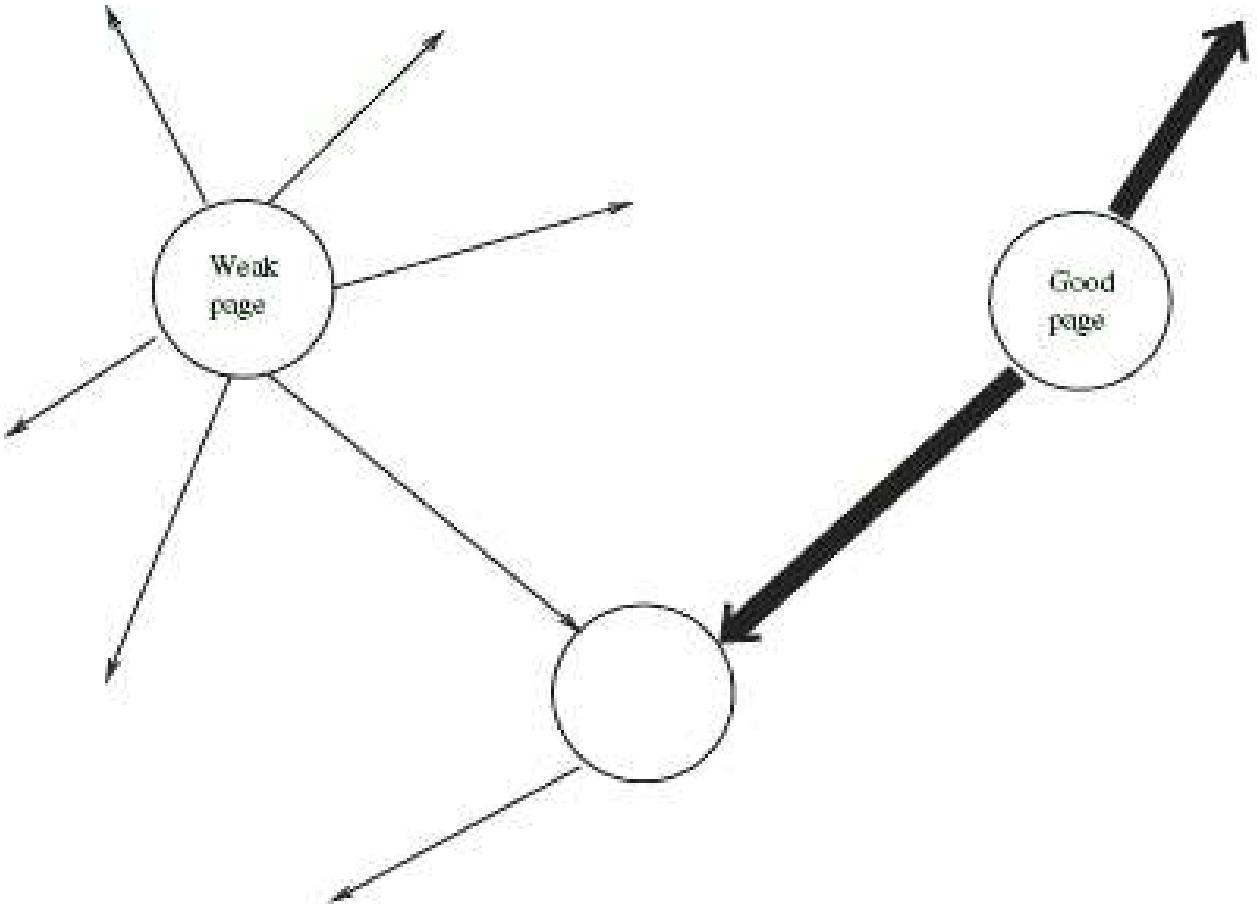
No todas las recomendaciones son igual de importantes. Tiene más valor una recomendación de Mozilla.org que una de mi página web.

Pero si Mozilla tiene una liga hacia mi página no se debe considerar que mi página es igual de importante!.

El valor del voto de A hacia B es la importancia de A, dividida entre el numero de enlaces de A hacia otras páginas... La importancia de A se distribuye entre todas las páginas a las que recomienda.

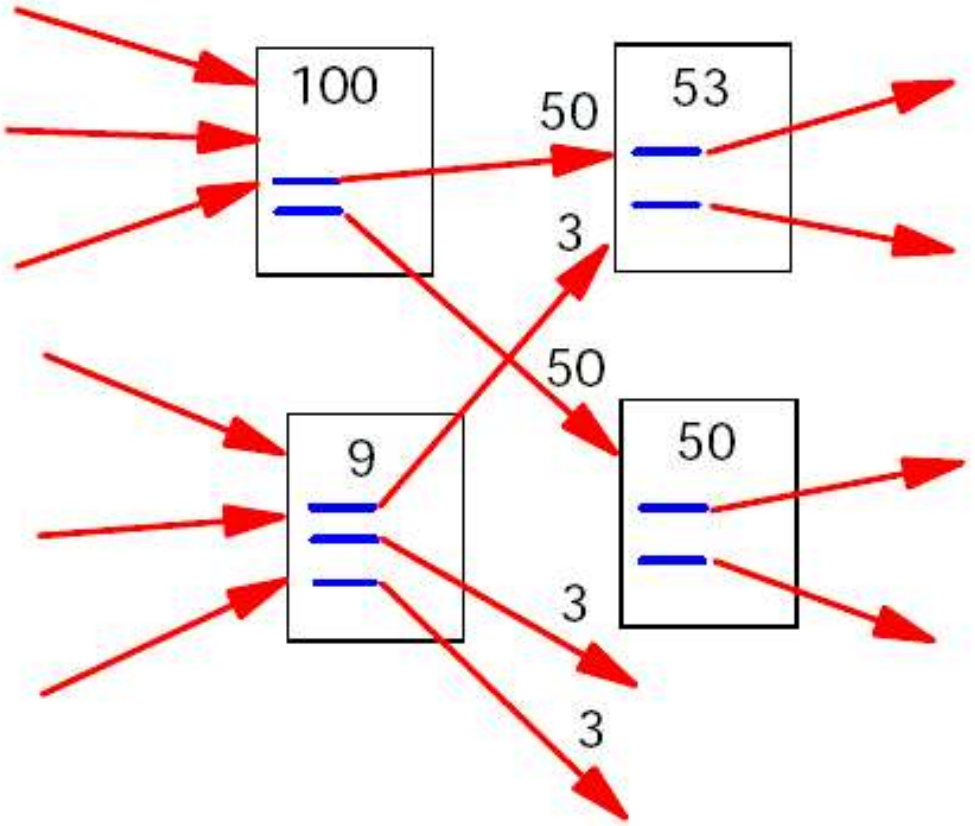
PageRank [cont.]

Ejemplo de los pesos ponderados.



PageRank [cont.]

El ranking se propaga a travez de las páginas.



PageRank [cont.]

Una formulación simplificada de PageRank es:

- Sea u una página web.
- F_u el conjunto de páginas a las que u apunta.
- B_u el conjunto de páginas que apuntan a u.
- $N_u = |F_u|$ el número de enlaces desde u.

$$R(u) = \sum_{v \in B_u} \frac{R(v)}{N_v}$$

PageRank [cont.]

$$R(u) = \sum_{v \in B_u} \frac{R(v)}{N_v}$$

Aplicamos esta expresión a todas las páginas para, de forma iterativa, encontrar un valor estable.

Esto se logra haciendo $\pi_j^T = (r_j(P_1), r_j(P_2), \ldots, r_j(P_n))$ e iterativamente calculando

$$\pi_j^T = \pi_{j-1}^T \mathbf{P}$$

Donde P es la matriz con $p_{ij} = \begin{cases} 1/|P_i| \text{ si } P_i \text{ enlaza a } P_j, \\ 0 \text{ de otra forma.} \end{cases}$

PageRank [cont.]

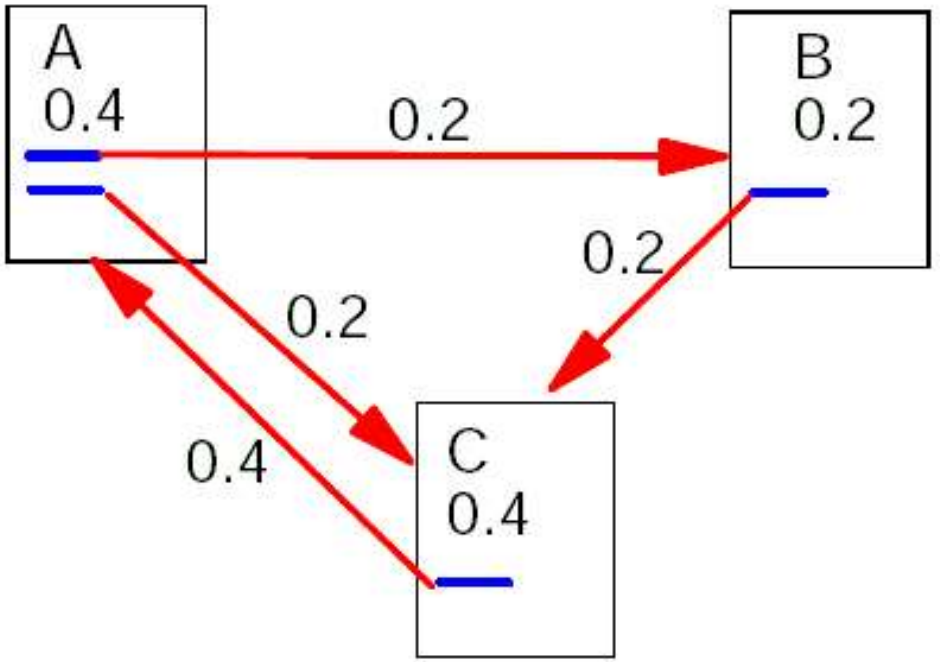
$$\pi_j^T = \pi_{j-1}^T \mathbf{P}$$

¡El método de las potencias nuevamente!

P se conoce como la matriz de Google.

PageRank [cont.]

Ejemplo de R(u) estable (PageRank).



PageRank [cont.]

$$\pi_j^T = \pi_{j-1}^T \mathbf{P}$$

Donde P es la matriz con $p_{ij} = \begin{cases} 1/|P_i| \text{ si } P_i \text{ enlaza a } P_j, \\ 0 \text{ de otra forma.} \end{cases}$

Hay algunos problemas con esa definición sencilla.

- Hay algunas páginas que no tienen enlaces de salida, y su peso se pierde del sistema (nodos colgantes). Una cuarta parte del total de nodos en la web son de este tipo[1].

PageRank [cont.]

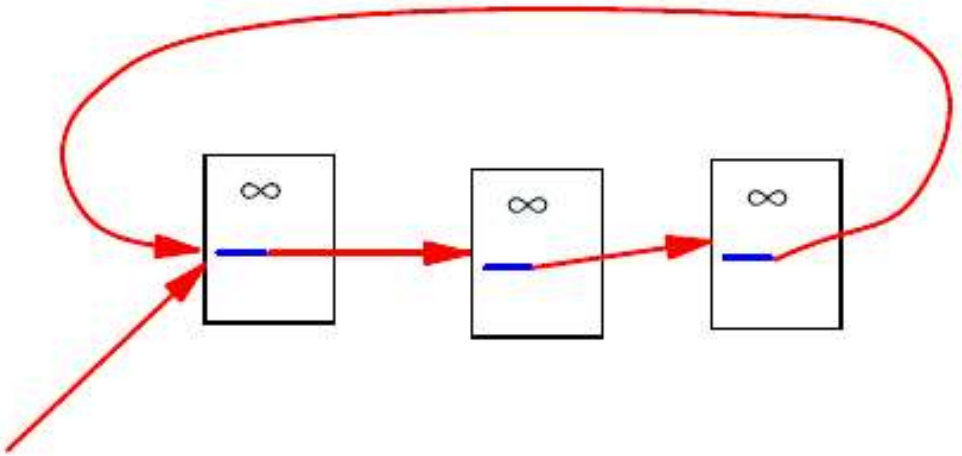
$$\pi_j^T = \pi_{j-1}^T \mathbf{P}$$

Donde P es la matriz con $p_{ij} = \begin{cases} 1/|P_i| \text{ si } P_i \text{ enlaza a } P_j, \\ 0 \text{ de otra forma.} \end{cases}$

- Consideremos dos páginas que se enlazan entre si, pero que no tienen enlaces hacia una tercera página. Aún más, que hay una página externa con una liga hacia una de ellas. Entonces durante la iteración, este lazo acumula calificación, pero no la distribuye, por que no hay enlaces hacia afuera. A estos lazos se les conoce como *desagüe* de calificación.

PageRank [cont.]

Ejemplo de un lazo que actua como desagüe de calificación.



PageRank [cont.]

Uso de PageRank en Google

- PageRank califica la *importancia* de una página. Pero no dice nada respecto a la relevancia de esa página respecto a alguna consulta. La calificación que PageRank da a una página es estatica.

- PageRank es solo una parte de Google, de hecho PageRank se combina con otras heuristicas para formar una calificación de una página respecto a una consulta.

PageRank [cont.]

Uso de PageRank en Google

- PageRank califica la *importancia* de una página. Pero no dice nada respecto a la relevancia de esa página respecto a alguna consulta. La calificación que PageRank da a una página es estatica.

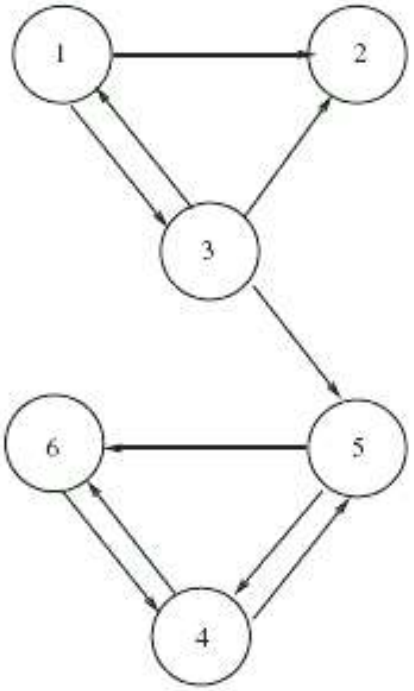
- PageRank es solo una parte de Google, de hecho PageRank se combina con otras heuristicas para formar una calificación de una página respecto a una consulta.

PageRank [cont.]

Un escenario (muy simple) de uso de PageRank es el siguiente.

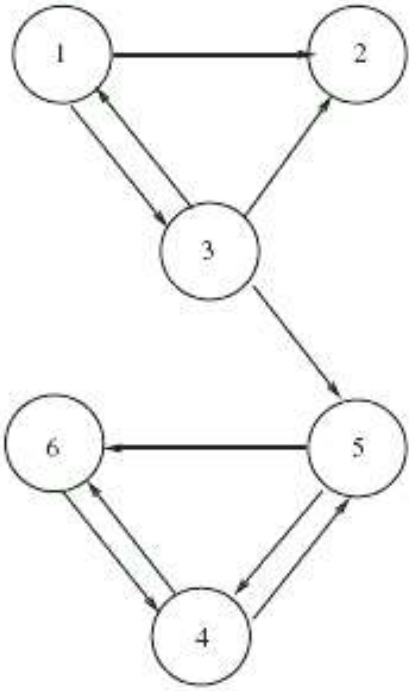
Consideremos la microred de la figura.

Primero calculamos la matriz *cruda* de Google....



PageRank [cont.]

$$\mathbf{P} = \begin{array}{c} \\ d_1 \\ d_2 \\ d_3 \\ d_4 \\ d_5 \\ d_6 \end{array} \begin{array}{c} \begin{array}{cccccc} d_1 & d_2 & d_3 & d_4 & d_5 & d_6 \end{array} \\ \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \end{array}$$

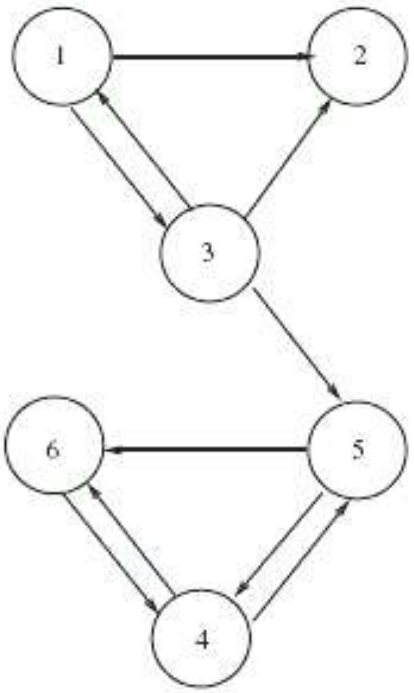


La segunda fila solo contiene ceros por que no hay enlaces de salida desde la segunda página.

Lo arreglamos añadiendo 1/6 a cada elemento de esa fila...

PageRank [cont.]

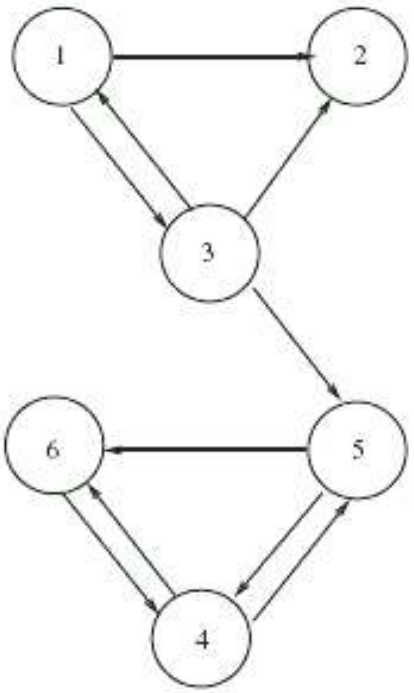
$$\bar{\mathbf{P}} = \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$



Esta matriz tiene “desagues” de calificación.
Para evitarlo “ligamos” cada página a cualquier otra con poco peso (10%)

PageRank [cont.]

$$= \begin{pmatrix} 1/60 & 7/15 & 7/15 & 1/60 & 1/60 & 1/60 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 19/60 & 19/60 & 1/60 & 1/60 & 19/60 & 1/60 \\ 1/60 & 1/60 & 1/60 & 1/60 & 7/15 & 7/15 \\ 1/60 & 1/60 & 1/60 & 7/15 & 1/60 & 7/15 \\ 1/60 & 1/60 & 1/60 & 11/12 & 1/60 & 1/60 \end{pmatrix}$$



Esta matriz es adecuada,
y su vector estacionario (y el vector de PageRank) es:

$\pi^T = (.03721\ .05396\ .04151\ .3751\ .206\ .2862\ )$

PageRank [cont.]

$\pi^T = (.03721\ .05396\ .04151\ .3751\ .206\ .2862\ )$

Estos PageRank's son independientes de la consulta.

Supongamos que se hace una consulta conteniendo los términos 1 y 2.

Accedemos a un indice invertido término-documento con las sig. entradas:

term1 -> doc 1, doc 4, doc 6
term2 .> doc 1, doc 3

....

....

PageRank [cont.]

$\pi^T = (.03721\ .05396\ .04151\ .3751\ .206\ .2862\ )$

Estos PageRank's son independientes de la consulta.

Supongamos que se hace una consulta conteniendo los términos 1 y 2.

Accedemos a un indice invertido término-documento con las sig. entradas:

term1 -> doc 1, doc 4, doc 6
term2 .> doc 1, doc 3

....

....

PageRank [cont.]

$\pi^T = (.03721\ .05396\ .04151\ .3751\ .206\ .2862\ )$

Estos PageRank's son independientes de la consulta.

Supongamos que se hace una consulta conteniendo los términos 1 y 2.

Accedemos a un indice invertido término-documento con las sig. entradas:

term1 -> doc 1, doc 4, doc 6
term2 .> doc 1, doc 3

....

....

PageRank [cont.]

Consulta: términos 1 y 2

$\pi^T = (.03721\ .05396\ .04151\ .3751\ .206\ .2862\ )$

term1 -> doc 1, doc 4, doc 6
term2 .> doc 1, doc 3
....

El conjunto de documentos relevantes es {1,3,4,6}.
Comparamos los PageRank's de esos documentos para determinar cuales de estos cuatro documentos relevantes son los más importantes, lo que da:

$\pi_4 = .3751,\ \pi_6 = .2862,\ \pi_3 = 0.4151,\ \pi_1 = .03721$

PageRank [cont.]

Consulta: términos 1 y 2

$\pi^T = (.03721\ .05396\ .04151\ .3751\ .206\ .2862\ )$

term1 -> doc 1, doc 4, doc 6
term2 .> doc 1, doc 3
....

El conjunto de documentos relevantes es {1,3,4,6}. Comparamos los PageRank's de esos documentos para determinar cuales de estos cuatro documentos relevantes son los más importantes, lo que da:

$\pi_4 = .3751,\ \pi_6 = .2862,\ \pi_3 = 0.4151,\ \pi_1 = .03721$

PageRank [cont.]

$\pi_4 = .3751, \pi_6 = .2862, \pi_3 = 0.4151, \pi_1 = .03721$

En consecuencia el documento 4 es el más relevante de los documentos relevantes, seguido de los documentos 6, 3, y 1.

PageRank [cont.]

$\pi_4 = .3751, \pi_6 = .2862, \pi_3 = 0.4151, \pi_1 = .03721$

En consecuencia el documento 4 es el más relevante de los documentos relevantes, seguido de los documentos 6, 3, y 1.

PageRank [cont.]

Modelo de Markov de la web

Si suponemos por un momento que no hay nodos colgados (o hacemos un arreglo artificial para evitarlos), entonces P es una matriz estocástica, lo que significa que la iteración de PageRank representa la evolución de una cadena de Markov.

Más precisamente, esta cadena de Markov es una caminata aleatoria en la gráfica definida por la estructura de las ligas de la red.

"We compare PageRank to an idealized web surfer"[5]

PageRank [cont.]

Modelo de Markov de la web

Si suponemos por un momento que no hay nodos colgados (o hacemos un arreglo artificial para evitarlos), entonces P es una matriz estocástica, lo que significa que la iteración de PageRank representa la evolución de una cadena de Markov.

Más precisamente, esta cadena de Markov es una caminata aleatoria en la gráfica definida por la estructura de las ligas de la red.

"We compare PageRank to an idealized web surfer"[5]

PageRank [cont.]

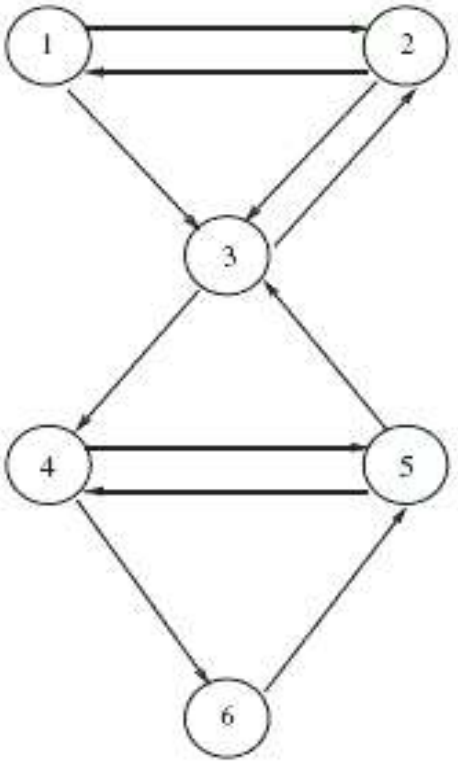
Modelo de Markov de la web

Si suponemos por un momento que no hay nodos colgados (o hacemos un arreglo artificial para evitarlos), entonces P es una matriz estocástica, lo que significa que la iteración de PageRank representa la evolución de una cadena de Markov.

Más precisamente, esta cadena de Markov es una caminata aleatoria en la gráfica definida por la estructura de las ligas de la red.

“We compare PageRank to an idealized web surfer”[5]

PageRank [cont.]

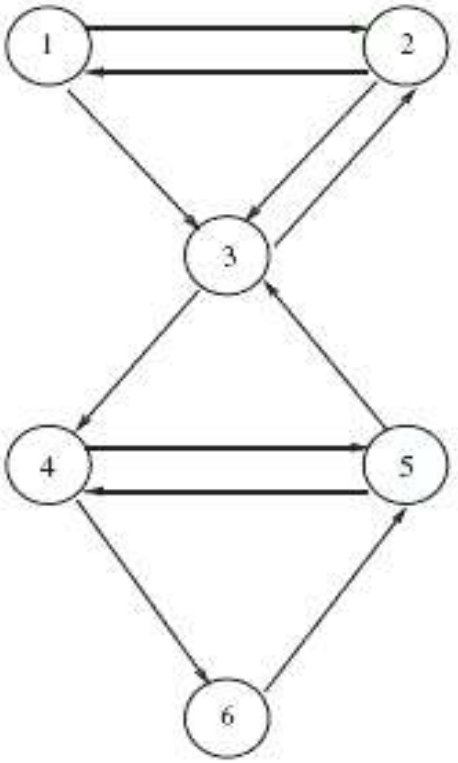


Ejemplo: Consideremos esta estructura de enlaces.

El modelo de Markov representa la gráfica dirigida como una matriz cuadrada de probabilidades de transición cuyo elemento p_{ij} es la probabilidad de moverse del estado i (pag. i) al estado j (pag. j) en un paso (click).

$$\mathbf{P} = \begin{pmatrix} 0 & .5 & .5 & 0 & 0 & 0 \\ .5 & 0 & .5 & 0 & 0 & 0 \\ 0 & .5 & 0 & .5 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & .5 \\ 0 & 0 & .5 & .5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

PageRank [cont.]



Otra manera de interpretar el PageRank de una página es como la fracción promedio de tiempo que el navegador aleatorio estará visitando esa pagina, dado un tiempo infinito.

$$\mathbf{P} = \begin{pmatrix} 0 & .5 & .5 & 0 & 0 & 0 \\ .5 & 0 & .5 & 0 & 0 & 0 \\ 0 & .5 & 0 & .5 & 0 & 0 \\ 0 & 0 & 0 & 0 & .5 & .5 \\ 0 & 0 & .5 & .5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

PageRank [cont.]

En general, el eigenvalor dominante para cada matriz estocastica es $\lambda = 1$.

Si el vector PageRank converge, lo hara al eigenvector normalizado que satisface:

$$\pi^T = \pi^T \mathbf{P}, \quad \pi^T e = \mathbf{1} \text{ (e es una columna de unos)}$$

El problema de calcular PageRank se reduce entonces al de encontrar el eigenvector dominante, o equivalentemente resolver el sistema lineal homogeneo

$$\pi^T \mathbf{(I-P)} = 0 \text{ con } \pi^T e = \mathbf{1}$$

PageRank [cont.]

En general, el eigenvalor dominante para cada matriz estocastica es $\lambda = 1$.

Si el vector PageRank converge, lo hara al eigenvector normalizado que satisface:

$$\pi^T = \pi^T \mathbf{P}, \quad \pi^T e = \mathbf{1} \text{ (e es una columna de unos)}$$

El problema de calcular PageRank se reduce entonces al de encontrar el eigenvector dominante, o equivalentemente resolver el sistema lineal homogeneo

$$\pi^T (\mathbf{I-P}) = 0 \text{ con } \pi^T e = \mathbf{1}$$

PageRank [cont.]

En general, el eigenvalor dominante para cada matriz estocastica es $\lambda = 1$.

Si el vector PageRank converge, lo hara al eigenvector normalizado que satisface:

$$\pi^T = \pi^T \mathbf{P}, \quad \pi^T e = \mathbf{1} \text{ (e es una columna de unos)}$$

El problema de calcular PageRank se reduce entonces al de encontrar el eigenvector dominante, o equivalentemente resolver el sistema lineal homogeneo

$$\pi^T \mathbf{(I-P)}=0 \text{ con } \pi^T e = \mathbf{1}$$

PageRank [cont.]

El sistema anterior es muy sencillo, pero su solución no es trivial por el tamaño de las matrices..

Actualmente P es una matriz de mas de 64,000,000,000,000,000 términos!

De hecho se dice que calcular PageRank es “La mayor computación matricial del mundo”.

El hecho de que sea una matriz extremadamente rala es una razón para usar el metodo de las potencias.

PageRank [cont.]

El sistema anterior es muy sencillo, pero su solución no es trivial por el tamaño de las matrices..

Actualmente P es una matriz de mas de 64,000,000,000,000,000,000 términos!

De hecho se dice que calcular PageRank es “La mayor computación matricial del mundo”.

El hecho de que sea una matriz extremadamente rala es una razón para usar el metodo de las potencias.

PageRank [cont.]

El sistema anterior es muy sencillo, pero su solución no es trivial por el tamaño de las matrices..

Actualmente P es una matriz de mas de 64,000,000,000,000,000,000 términos!

De hecho se dice que calcular PageRank es “La mayor computación matricial del mundo”.

El hecho de que sea una matriz extremadamente rala es una razón para usar el metodo de las potencias.

PageRank [cont.]

El sistema anterior es muy sencillo, pero su solución no es trivial por el tamaño de las matrices..

Actualmente P es una matriz de mas de 64,000,000,000,000,000,000 términos!

De hecho se dice que calcular PageRank es “La mayor computación matricial del mundo”.

El hecho de que sea una matriz extremadamente rala es una razón para usar el metodo de las potencias.

PageRank [cont.]

Ajuste de la matriz de Google

P es la matriz con $p_{ij} = \begin{cases} 1/|P_i| \text{ si } P_i \text{ enlaza a } P_j, \\ 0 \text{ de otra forma.} \end{cases}$

La matriz de Google (P) asi definida tiene dos graves problemas que evitan calcular PageRank de manera directa:

- Existen muchas filas de solo 0's (nodos colgantes).
- Hay desagües de calificación.

Para solucionarlos hay que modificar un poco a P...

PageRank [cont.]

- Existen muchas filas de solo 0's (nodos colgantes).

Esto se soluciona facilmente cambiando cada renglón de P que contiene solo 0's, con el renglon (1/n,1/n,...,1/n).

Entonces trabajamos ahora con la matriz P' que tiene dichos cambios.

PageRank [cont.]

- Hay desagües de calificación.

Esto se soluciona haciendo a cada página alcanzable desde cualquier otra.

La justificación intuitiva de hacerlo es la tendencia de un usuario de brincar de manera aleatoria de una pagina a cualquier otra de la red, aunque no haya una liga de la primera a la segunda. Esto sucede p.e. cuando un usuario introduce una dirección en la barra del navegador.

Al hacer cualquier pag. alcanzable desde cualquier otra, hacemos irreducible a la matriz estocastica.

PageRank [cont.]

- Hay desagües de calificación.

Esto se soluciona haciendo a cada página alcanzable desde cualquier otra.

La justificación intuitiva de hacerlo es la tendencia de un usuario de brincar de manera aleatoria de una pagina a cualquier otra de la red, aunque no haya una liga de la primera a la segunda. Esto sucede p.e. cuando un usuario introduce una dirección en la barra del navegador.

Al hacer cualquier pag. alcanzable desde cualquier otra, hacemos irreducible a la matriz estocastica.

PageRank [cont.]

- Hay desagües de calificación.

Esto se soluciona haciendo a cada página alcanzable desde cualquier otra.

La justificación intuitiva de hacerlo es la tendencia de un usuario de brincar de manera aleatoria de una pagina a cualquier otra de la red, aunque no haya una liga de la primera a la segunda. Esto sucede p.e. cuando un usuario introduce una dirección en la barra del navegador.

Al hacer cualquier pag. alcanzable desde cualquier otra, hacemos irreducible a la matriz estocastica.

PageRank [cont.]

La solución a los problema anterior lleva al replanteamiento de la matriz de google.

$$P'' = \alpha P' + (1-\alpha)E$$

Donde α es un escalar entre 0 y 1 y representa la probabilidad de que un navegador siga una liga de la página actual.

$(1-\alpha)$ es entonces la probabilidad de "teletransportarse" de la página actual a cualquier otra usando la barra de direcciones. E es una matriz del mismo tamaño que P' que representa la distribucion de probabilidades para este salto aleatorio.

PageRank [cont.]

La solución a los problema anterior lleva al replanteamiento de la matriz de google.

$$P'' = \alpha P' + (1-\alpha)E$$

Donde α es un escalar entre 0 y 1 y representa la probabilidad de que un navegador siga una liga de la página actual.

$(1-\alpha)$ es entonces la probabilidad de "teletransportarse" de la página actual a cualquier otra usando la barra de direcciones. E es una matriz del mismo tamaño que P' que representa la distribucion de probabilidades para este salto aleatorio.

PageRank [cont.]

Convergencia del cálculo de PageRank[1]

Ajustando α se puede controlar la velocidad de convergencia del cálculo.

Valores pequeños hacen que converja rapidamente, pero entonces se trabaja con una matriz P" muy diferente de la estructura de la web real. Diferentes valores de α pueden producir PageRank's muy diferentes.

Como se forza a la matriz a ser irreducible con la introducción de E, no hay problemas con la unicidad de la solución. Cualquier vector de probabilidad positiva puede usarse como valor inicial.

PageRank [cont.]

Convergencia del cálculo de PageRank[1]

Ajustando α se puede controlar la velocidad de convergencia del cálculo.

Valores pequeños hacen que converja rapidamente, pero entonces se trabaja con una matriz P" muy diferente de la estructura de la web real. Diferentes valores de α pueden producir PageRank's muy diferentes.

Como se forza a la matriz a ser irreducible con la introducción de E, no hay problemas con la unicidad de la solución. Cualquier vector de probabilidad positiva puede usarse como valor inicial.

PageRank [cont.]

Convergencia del cálculo de PageRank[1]

Ajustando α se puede controlar la velocidad de convergencia del cálculo.

Valores pequeños hacen que converja rapidamente, pero entonces se trabaja con una matriz P" muy diferente de la estructura de la web real. Diferentes valores de α pueden producir PageRank's muy diferentes.

Como se forza a la matriz a ser irreducible con la introducción de E, no hay problemas con la unicidad de la solución. Cualquier vector de probabilidad positiva puede usarse como valor inicial.

PageRank [cont.]

Convergencia del cálculo de PageRank[1]

- Brin y Page reportan una convergencia a niveles“razonablemente tolerables” en aprox. 52 iteraciones para una base de datos de 332 millones de enlaces.[5]
- Sugieren tambien que el factor de escala es casi lineal a log n.

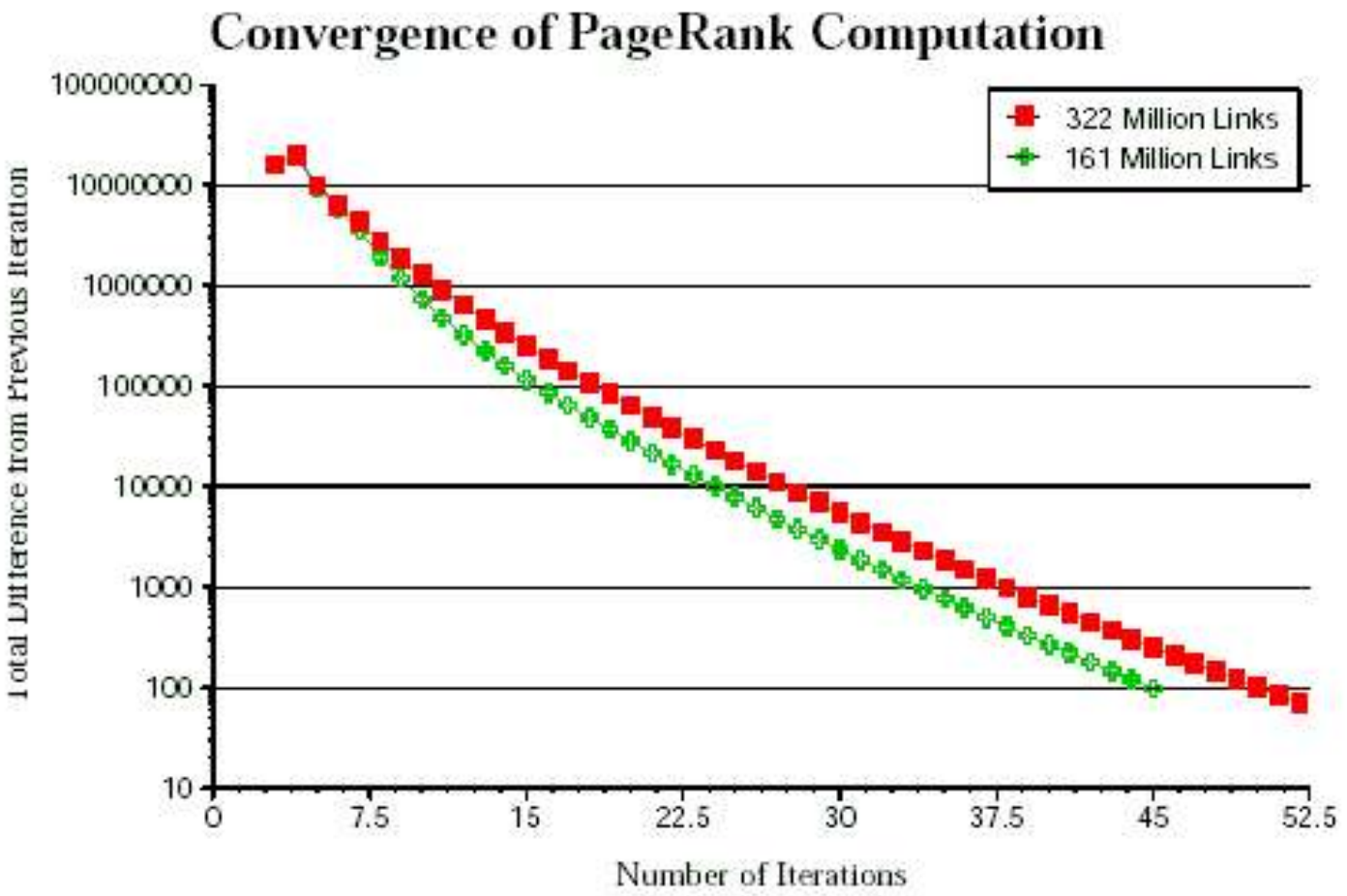
PageRank [cont.]

Convergencia del cálculo de PageRank[1]

- Brin y Page reportan una convergencia a niveles“razonablemente tolerables” en aprox. 52 iteraciones para una base de datos de 332 millones de enlaces.[5]
- Sugieren tambien que el factor de escala es casi lineal a log n.

PageRank [cont.]

[5]



PageRank [cont.]

Actualización de PageRank [1]

- La actualización del vector PageRank es muy costosa.
- Google ha reportado hacerlo cada ciertas semanas.
- El vector anterior es practicamente inutil para cálcular el nuevo.
- Google comienza practicamente de cero cada vez que actualiza PageRank

PageRank [cont.]

Actualización de PageRank [1]

- La actualización del vector PageRank es muy costosa.
- Google ha reportado hacerlo cada ciertas semanas.
- El vector anterior es practicamente inutil para cálcular el nuevo.
- Google comienza practicamente de cero cada vez que actualiza PageRank

PageRank [cont.]

Actualización de PageRank [1]

- La actualización del vector PageRank es muy costosa.
- Google ha reportado hacerlo cada ciertas semanas.
- El vector anterior es practicamente inutil para cálcular el nuevo.
- Google comienza practicamente de cero cada vez que actualiza PageRank

PageRank [cont.]

Actualización de PageRank [1]

- La actualización del vector PageRank es muy costosa.
- Google ha reportado hacerlo cada ciertas semanas.
- El vector anterior es practicamente inutil para cálcular el nuevo.
- Google comienza practicamente de cero cada vez que actualiza PageRank

PageRank [cont.]

Ventajas de PageRank

- El uso de importancia, más que relevancia para calificar una página, es la clave del exito de Google. Midiendo la importancia, la dependencia de la consulta (el principal problema de HITS) no es ya un asunto de importancia.
- En tiempo de consulta, solo es necesario hacer una rápida busqueda de documentos relevantes, y ordenarlos de aucerdo a su PageRank.
- Es más resistente a spamming.

PageRank [cont.]

Ventajas de PageRank

- El uso de importancia, más que relevancia para calificar una página, es la clave del exito de Google. Midiendo la importancia, la dependencia de la consulta (el principal problema de HITS) no es ya un asunto de importancia.
- En tiempo de consulta, solo es necesario hacer una rápida busqueda de documentos relevantes, y ordenarlos de aucerdo a su PageRank.
- Es más resistente a spamming.

PageRank [cont.]

Ventajas de PageRank

- El uso de importancia, más que relevancia para calificar una página, es la clave del exito de Google. Midiendo la importancia, la dependencia de la consulta (el principal problema de HITS) no es ya un asunto de importancia.
- En tiempo de consulta, solo es necesario hacer una rápida busqueda de documentos relevantes, y ordenarlos de aucerdo a su PageRank.
- Es más resistente a spamming.

PageRank [cont.]

Ventajas de PageRank

- Flexibilidad en la selección de E, el factor de desvio. Con esto, Google puede alterar los rankings de manera predecible. Otorgando de esta manera ventajas a algunos sitios o “castigar” a quienes tratan de engañar a Google.

PageRank [cont.]

Desventajas de PageRank

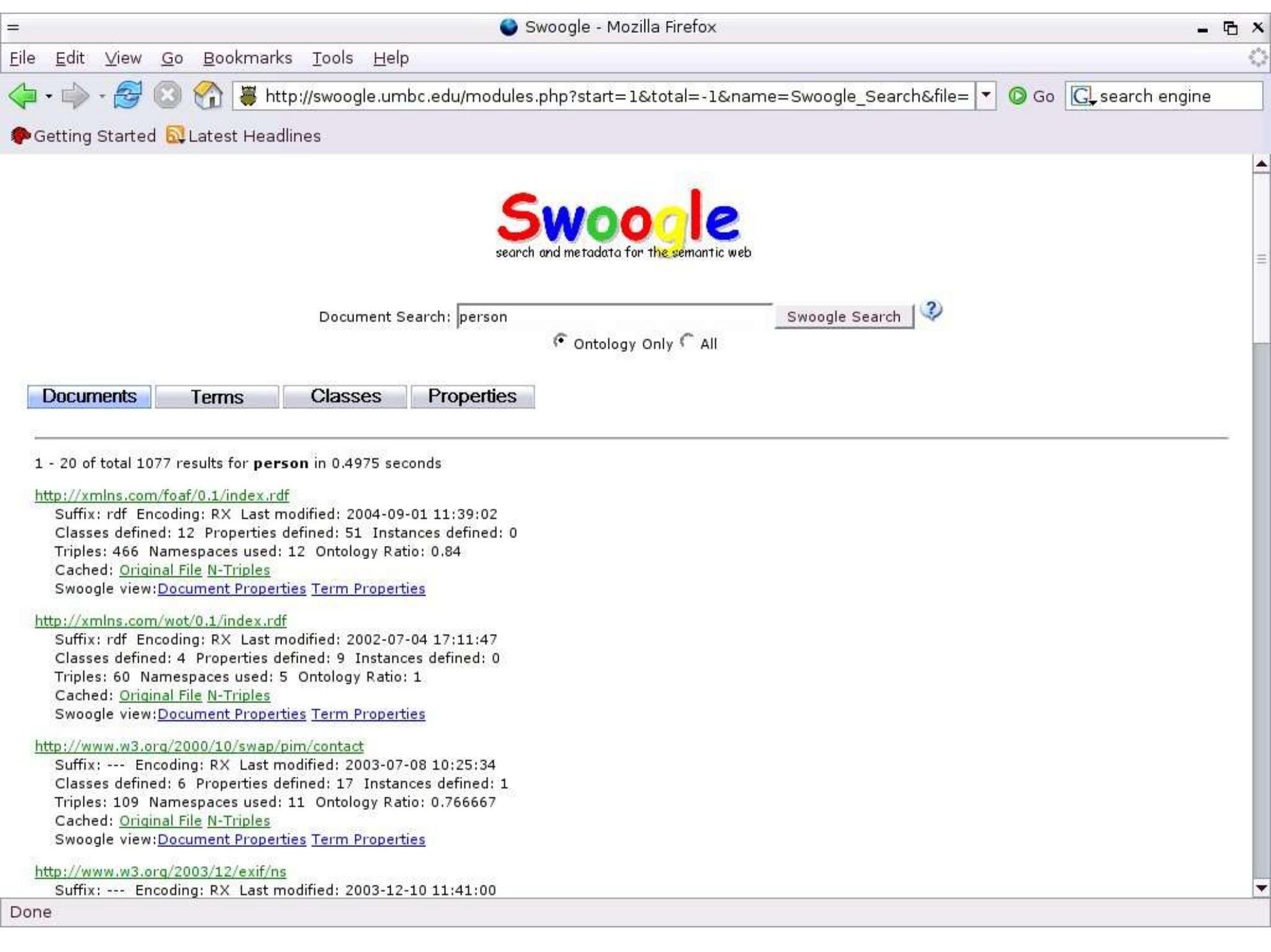
La necesidad de aplicar extensivamente heuristicas extras para determinar páginas realmente relevantes, de otra manera regresar páginas importantes no sirve de nada si estan fuera del tema.

“Como PageRank es independiente de la consulta, no puede distinguir entre páginas que son autoritativas en general, y las que los son en el tema particular de la consulta”.

Búsqueda en la Web Semántica.

Swoogle

Para ayudar a los usuarios humanos y a los agentes de software, Swoogle descubre, indexa y analiza las ontologias y hechos que estan presentes en la web semantica.


Swoogle - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://swoogle.umbc.edu/modules.php?start=1&total=-1&name=Swoogle_Search&file=
Go
search engine
Getting Started
Latest Headlines
Swoogle
search and metadata for the semantic web
Document Search: person
Swoogle Search
Ontology Only
All
Documents
Terms
Classes
Properties
1 - 20 of total 1077 results for person in 0.4975 seconds
http://xmlns.com/foaf/0.1/index.rdf
Suffix: rdf Encoding: RX Last modified: 2004-09-01 11:39:02
Classes defined: 12 Properties defined: 51 Instances defined: 0
Triples: 466 Namespaces used: 12 Ontology Ratio: 0.84
Cached: Original File N-Triples
Swoogle view:Document Properties Term Properties
http://xmlns.com/wot/0.1/index.rdf
Suffix: rdf Encoding: RX Last modified: 2002-07-04 17:11:47
Classes defined: 4 Properties defined: 9 Instances defined: 0
Triples: 60 Namespaces used: 5 Ontology Ratio: 1
Cached: Original File N-Triples
Swoogle view:Document Properties Term Properties
http://www.w3.org/2000/10/swap/pim/contact
Suffix: --- Encoding: RX Last modified: 2003-07-08 10:25:34
Classes defined: 6 Properties defined: 17 Instances defined: 1
Triples: 109 Namespaces used: 11 Ontology Ratio: 0.766667
Cached: Original File N-Triples
Swoogle view:Document Properties Term Properties
http://www.w3.org/2003/12/exif/ns
Suffix: --- Encoding: RX Last modified: 2003-12-10 11:41:00
Done

Swoogle [Cont.]

La búsqueda en la web semantica difiere de la búsqueda en web convencional por las siguientes razones:

- El contenido semantico de la WS está diseñado para su publicación por máquinas, y para máquinas.
- El conocimiento codificado en los lenguajes de la WS, como RDF, difiere de el muy desestructurado texto libre encontrado en la mayor parte de la web, y de la información altamente estructurada en las bases de datos

Swoogle [Cont.]

- Los documentos de la web semantica tienden a ser una mezcla de hechos concretos, definiciones de clases y propiedades, restricciones lógicas y metadatos. Entender por completo estos documentos puede requerir un razonamiento sustancial.

Swoogle [Cont.]

- La estructura de la gráfica de una colección de documentos de la WS difiere significativamente de la estructura que aparece en una colección de documentos HTML. Esta diferencia influencía sobremanera las estrategias para descubrir y rankear los documentos semánticos.

Swoogle [Cont.]

Descubriendo y reindexando documentos.

SWDs son como agujas en un pajar.

- Búsqueda exhaustiva no es opción.
- No conviene comenzar con un conjunto semilla pequeño, por que la WS no es tan densa ni tan conexa como la www.

Swoogle [Cont.]

Descubriendo y reindexando documentos.

Un enfoque es:

- Usar buscadores convencionales para encontrar un conjunto grande de semillas potenciales. Swoogle usa Google para esta etapa.
- Validar estos documentos con un parser semántico.
- Filtrar URL's obtenidas de este documento.
- La decisión de que tan seguido revisitar un documento, es la misma que para la www.

Swoogle [Cont.]

Procesando consultas.

Varios niveles de granularidad en las respuestas:

- Base de datos RDF: busquedas en tripletas.
- Documentos de WS completos.
- Términos de vocabulario de WS (URIrefs). Análogos a palabras en lenguaje natural.

Swoogle [Cont.]

Ranking: OntoRank y TermRank.

- Ordenan documentos y términos semánticos, respectivamente.
- Extienden la idea de PageRank de un modelo de visitante “aleatorio”.

Swoogle [Cont.]

Ranking: OntoRank y TermRank.

Crean un modelo de visitante “racional”. Es racional en el sentido de que sigue ligas de acuerdo a la semántica de la liga: Cuando se encuentre con un SWD , el visitante racional importará (transitivamente) las ontologías “oficiales” que definen las propiedades y clases a las que hace referencia.

Conclusiones

- Los búscadores web son una infraestructura clave de la www.
- Los búscadores de clase industrial requieren una cantidad enorme de recursos.
- Su tecnólogia es propietaria y cerrada por motivos económicos.
- El éxito de los buscadores modernos se debe al uso de técnicas de análisis de ligas, combinado con métodos de RI estandar.
- Las técnicas de análisis de ligas estan basadas en álgebra lineal

Conclusiones

- Las técnicas de análisis de ligas estan basadas en álgebra lineal.
- PageRank y HITS son los algoritmos de A.L. más usados.
- Ambos fueron publicados y desarrollados inicialmente en ambientes académicos.
- La web semántica presenta nuevos retos para el desarrollo de buscadores de SWD efectivos. Swoogle es el abanderado en esta área.

Referencias

[1] A Survey of Eigenvector Methods for Web Information Ratrieval. Amy N. Langville, Carl D. Meyer. 2005. SIAM Review Vo. 47, No. 1. pp. 135-161

[2]The Use of the Linear Algebra by Web Search Engines. Amy N. Langville, Carl D. Meyer. 2004.

[3]Authoritative Sources in a Hiperlinked Environment. Jon M. Kleinberg. Journal of the ACM, vol. 46, No. 5, 1999, pp. 604-632

[4]The Anatomy of a Large-Scale Hypertextual Web Search Engine. S Brin, L Page - WWW7 / Computer Networks, 1998 - kulturinformatik.uni-lueneburg.de

[5]The pagerank citation ranking: Bringing order to the web. L Page, S Brin, R Motwani, T Winograd - 1998 – dbpubs.stanford.edu

[6]Web Information Resource Discovery: Past, Present, and Future. G Oezsoyoglu, A Al-Hamdani - Yazici, Adnan – art.cwru.edu

[7]Web search engines. C Schwartz - Journal of the American Society for Information Science 49(11), 1998.

[8]Hyperlink analysis for the Web. MR Henzinger - IEEE Internet Computing, 2001 – ieeexplore.ieee.org

[9]The indexable web is more than 11.5 billion pages. A Gulli, A Signorini – Proceedings of the 14th international conference on World Wide Web. ACM 2005.

[10]Breadth-first search crawling yields high-quality pages. M Najork, JL Wiener - Proceedings of the 10th International World Wide Web …, 2001

[11] [Ding et al., 2004] Li Ding, Tim Finin, and Anupam Joshi. Swoogle: A search and metadata engine for the semantic web. In Proceedings of the Thirteenth ACM Conference on Information and Knowledge Management, pages 58–61, Washington DC, November 2004.

[12]Aplicación de las factorizaciones QR y SVD a motores de búsqueda. Humberto M. de la Vega, I. Delia G. Calvillo. Universidad Autónoma de Coahuila.

[13]Eigenvalue algorithm. http://en.wikipedia.org/wiki/Eigenvalue_algorithm

[14]Eigenvalue, eigenvector and eigenspace. http://en.wikipedia.org/wiki/Eigenvalue

¡Gracias de nuevo!