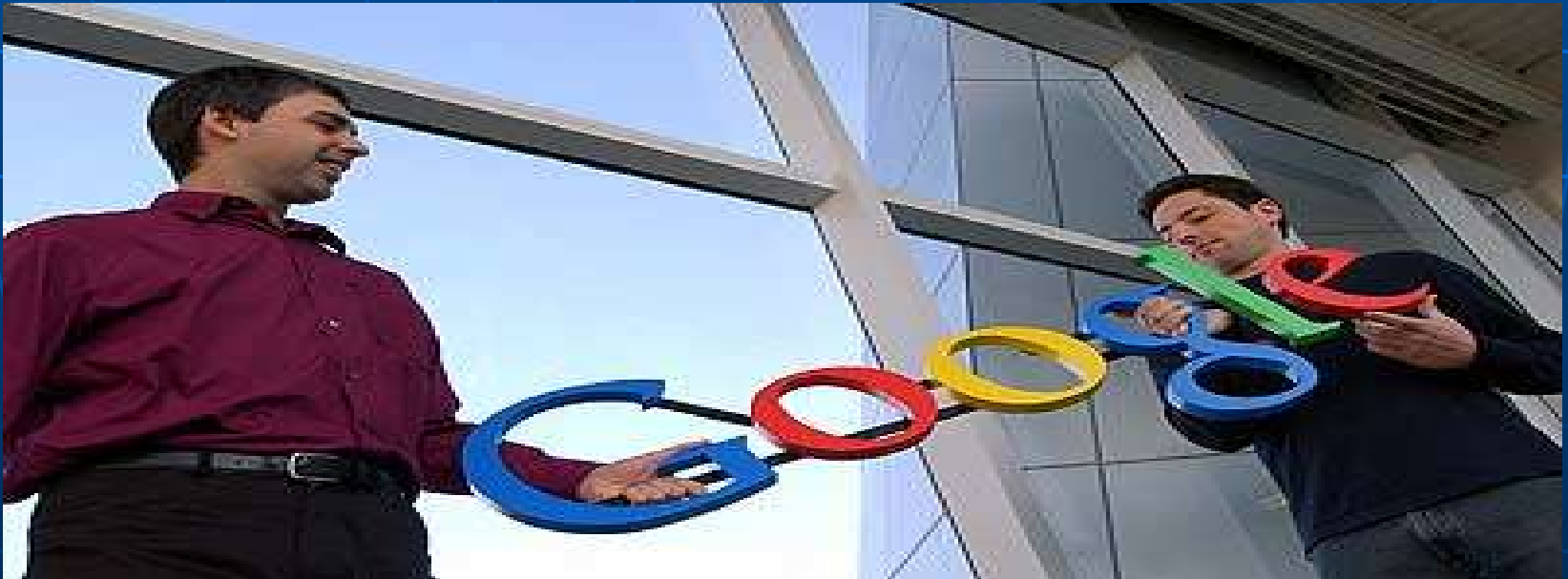




INTRODUCCION

Google Inc. es una compañía fundada en septiembre de 1998 por Larry Page y Sergey Brin, dos estudiantes de la Universidad de Stanford (EEUU).

Desde 1996 llevaban desarrollando un buscador de Internet, que con el tiempo se ha convertido en el más efectivo.



INTRODUCCION

- En otoño de 1995, Larry y Sergey comienzan a trabajar en el 'Digital Library Project'
- El nombre que Larry Page da a esta tecnología fue 'PageRank'.
- Enero 1996: Comienzan a desarrollar un buscador llamado 'BackRub'.
- Backrub está escrito en Java y Python



- Google indexa más de 3 mil millones de páginas web
- Hay varios 'rastreadores': el general (una vez al mes), que busca en la mayoría de la WWW
- El 'Fresh', que rastrea en las páginas que se actualizan frecuentemente; y el de noticias, que rastrea cada 10 minutos.



- Google analiza más de 100 factores para determinar la relevancia de una página web.
- Para calcular el valor del PageRank, Google utiliza la teoría de grafos, mediante una matriz de 30 billones de nodos. Cada uno de estos nodos tiene 10 arcos (o aristas) diferentes.



- Los sitios importantes, de alta calidad reciben un PageRank más alto, que Google recuerda cada vez que conduce búsqueda.
- Google no vende la colocación dentro de los resultados ellos mismos (es decir, nadie puede comprar un PageRank más alto).



Hay 4 tipos de servidores en el clúster de Google, situados en paralelo del servidor web

1. Los servidores índice.
2. Los servidores de documentos.
3. Los servidores correctores de deletreo.
4. Los servidores AdWords. .

PageRank™

- PageRank es un valor numérico que representa lo importante que es una página en la web.
- Google calcula la importancia de una página a partir de los votos que obtiene.
- Para calcular el PageRank de una página, todos sus enlaces entrantes se tienen en cuenta.

PageRank™

$$PR(A) = (1-d) + d(PR(t1)/C(t1) + \dots + PR(tn)/C(tn))$$

Esta es la ecuación que calcula el PageRank de una página.

Es el original que fue publicado cuando PageRank estaba siendo desarrollado, y es probable que Google use una variación de él.

PageRank™

El Pagerank de Google se mide de 1 a 10.

Puedes realizar una estimación de cual es el valor de tu web en esta página:

www.webrankinfo.com/english/tools/pagerank.php

PageRank™

¿Para qué vale el Pagerank?

Google lo utiliza, o eso se supone, para calcular las posiciones que ocupan las webs en las búsquedas. Tener un Pagerank más alto que las webs de los competidores se convierte entonces, en un factor muy importante.

PageRank™

Sitio Web – Enlaces

- 1.- Adobe Reader - Download 767000
- 2.- Google 534000
- 3.- World Wide Web Consortium 510000
- 4.- Yahoo! 386000
- 5.- MSN.com 347000
- 6.- U.S. Government s Official Web Portal 209000
- 7.- Microsoft Corporation 137000
- 8.- Adobe Systems Incorporated 117000
- 9.- National Institutes of Health (NIH) 117000
- 10.- Blogger 89200

CONOCER EL PAGERANK DE LAS PÁGINAS QUE VISITAMOS

Si queremos conocer el valor de PageRank de las páginas que visitamos, aunque sólo sea por curiosidad, podemos descargar la barra de herramientas de Google.

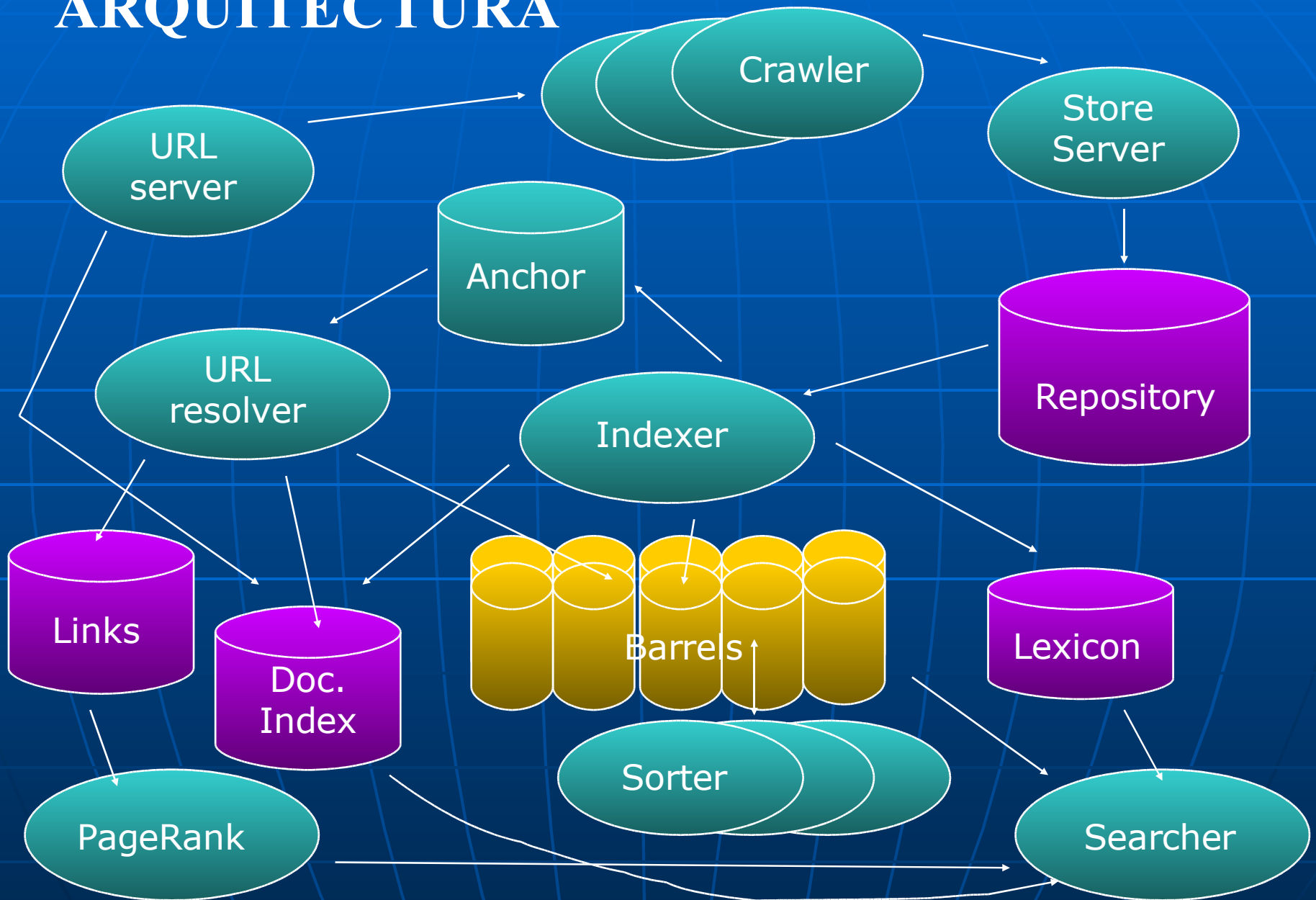


Vínculos en caché

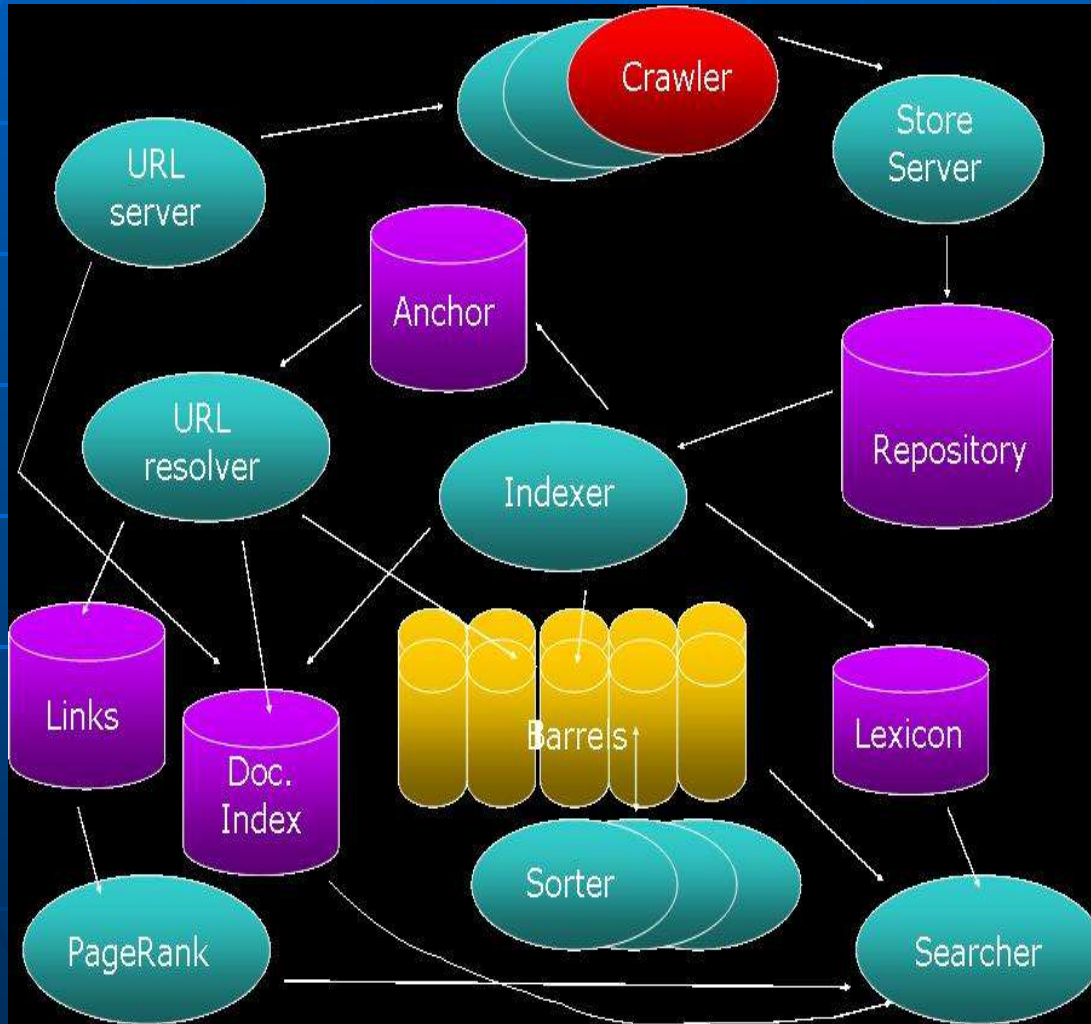
- Google toma una instantánea de cada página examinada
- El contenido caché es el contenido que Google usó para juzgar si la página es una buena coincidencia para su búsqueda.

ARQUITECTURA

ARQUITECTURA



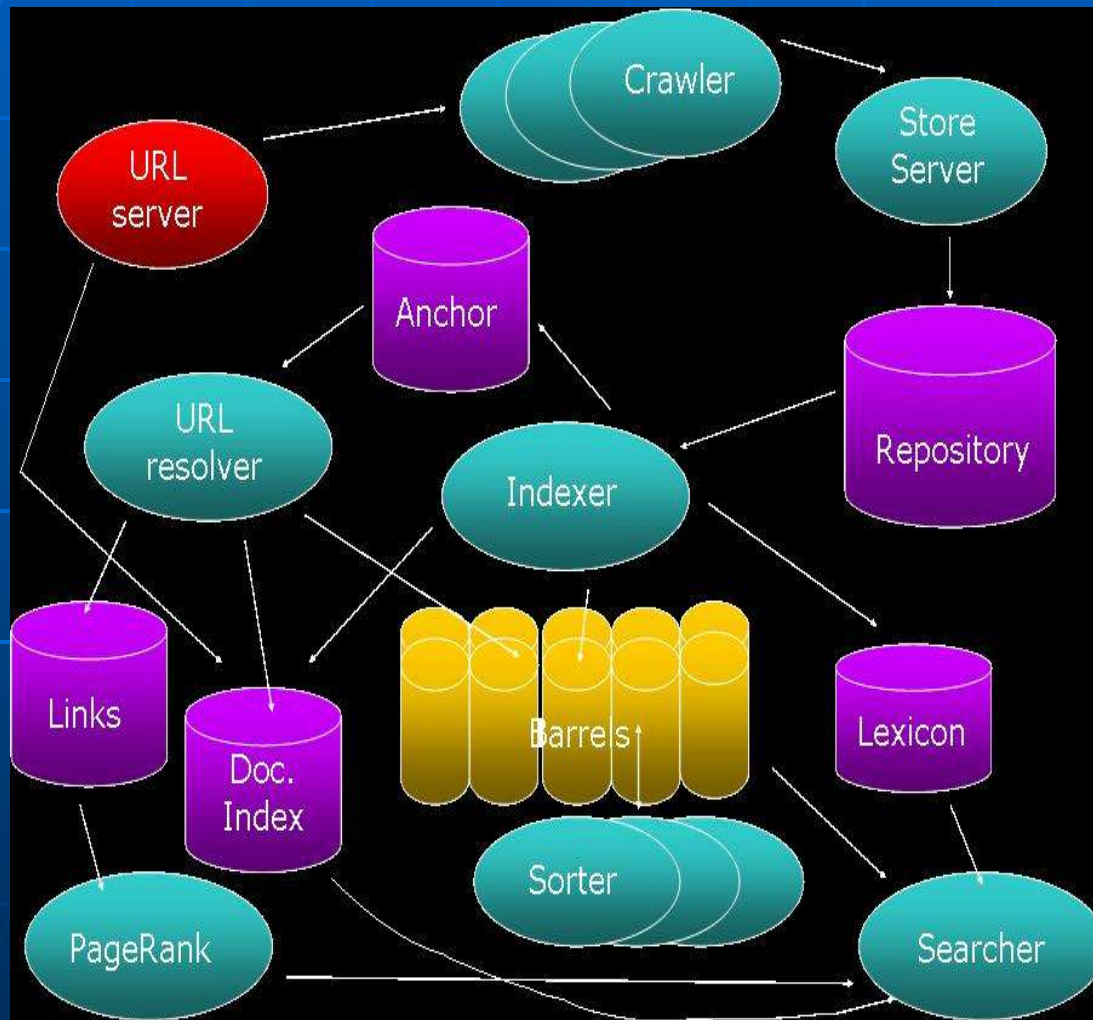
ARQUITECTURA



CRAWLERS DISTRIBUIDOS

Permiten realizar el crawling o rastreo de los servidores web con el fin de indexar la información que contienen

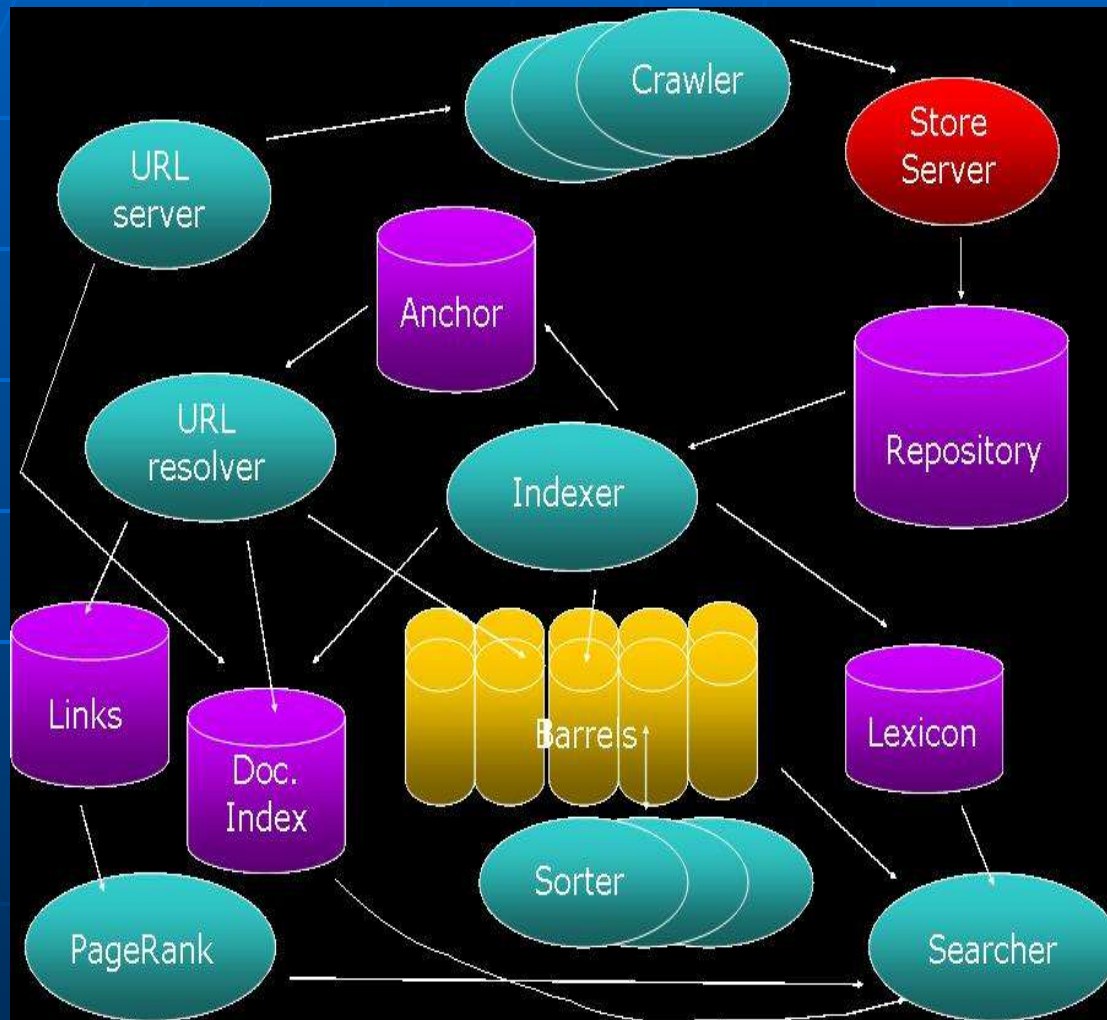
ARQUITECTURA



SERVIDOR DE URLs

Envía a los
crawlers una lista
con los URLs que
deben ser
rastreados

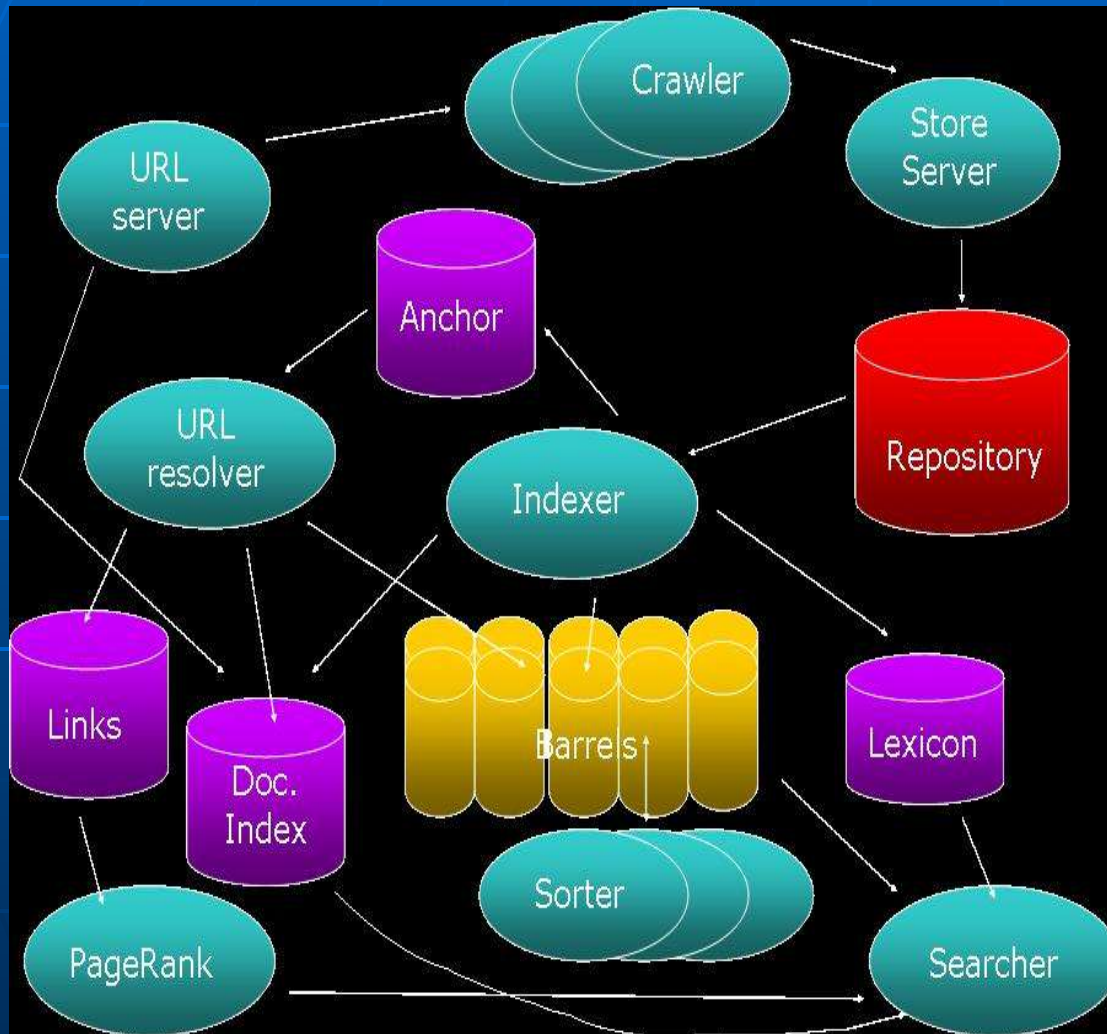
ARQUITECTURA



SERVIDOR DE ALMACENAMIENTO

Comprime y
almacena las
paginas
descargadas por
los crawlers

ARQUITECTURA

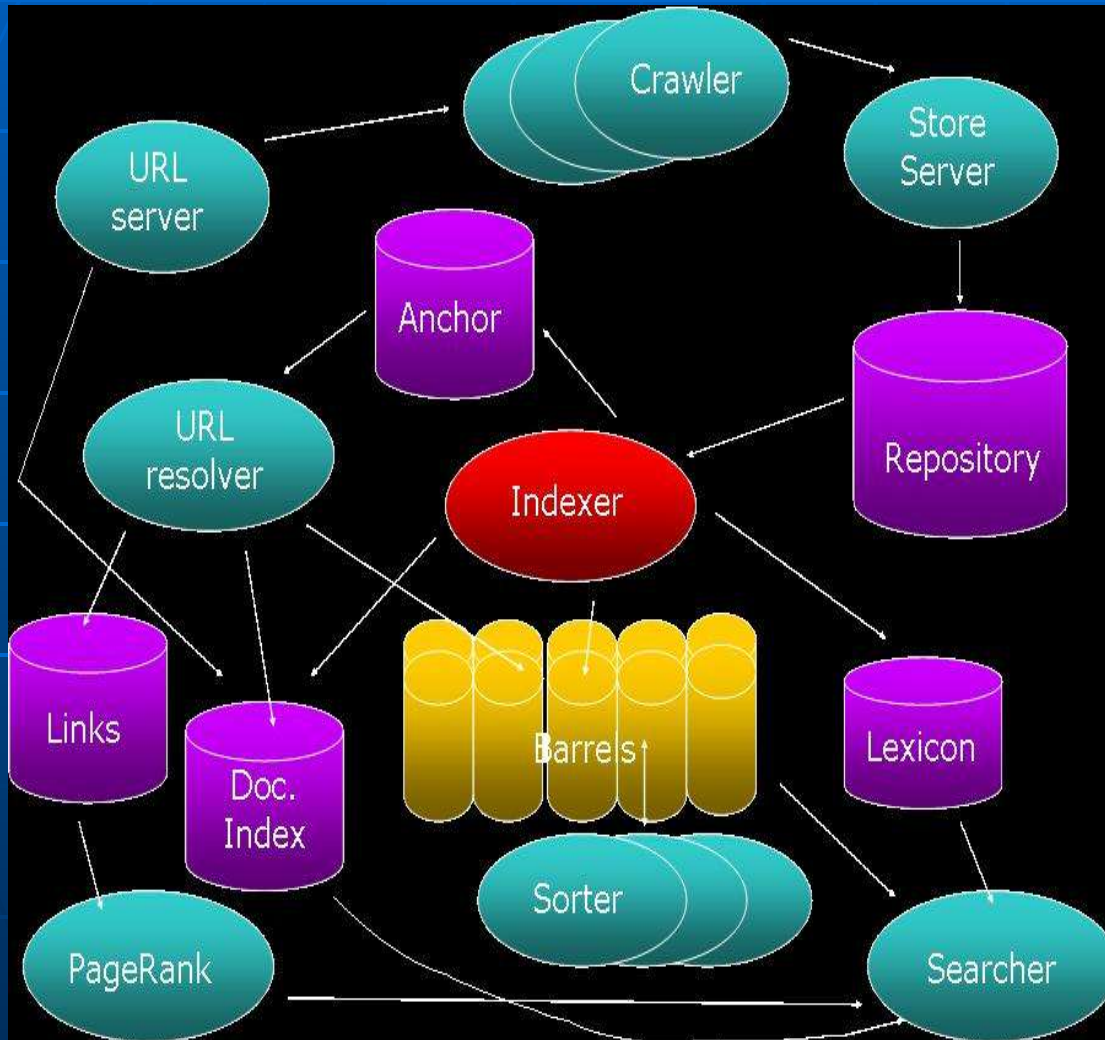


ALMACEN

Lugar físico en el que se almacenan los documentos comprimidos por el servidor de almacenamiento

A cada pagina se le asigna un numero denominado docID

ARQUITECTURA



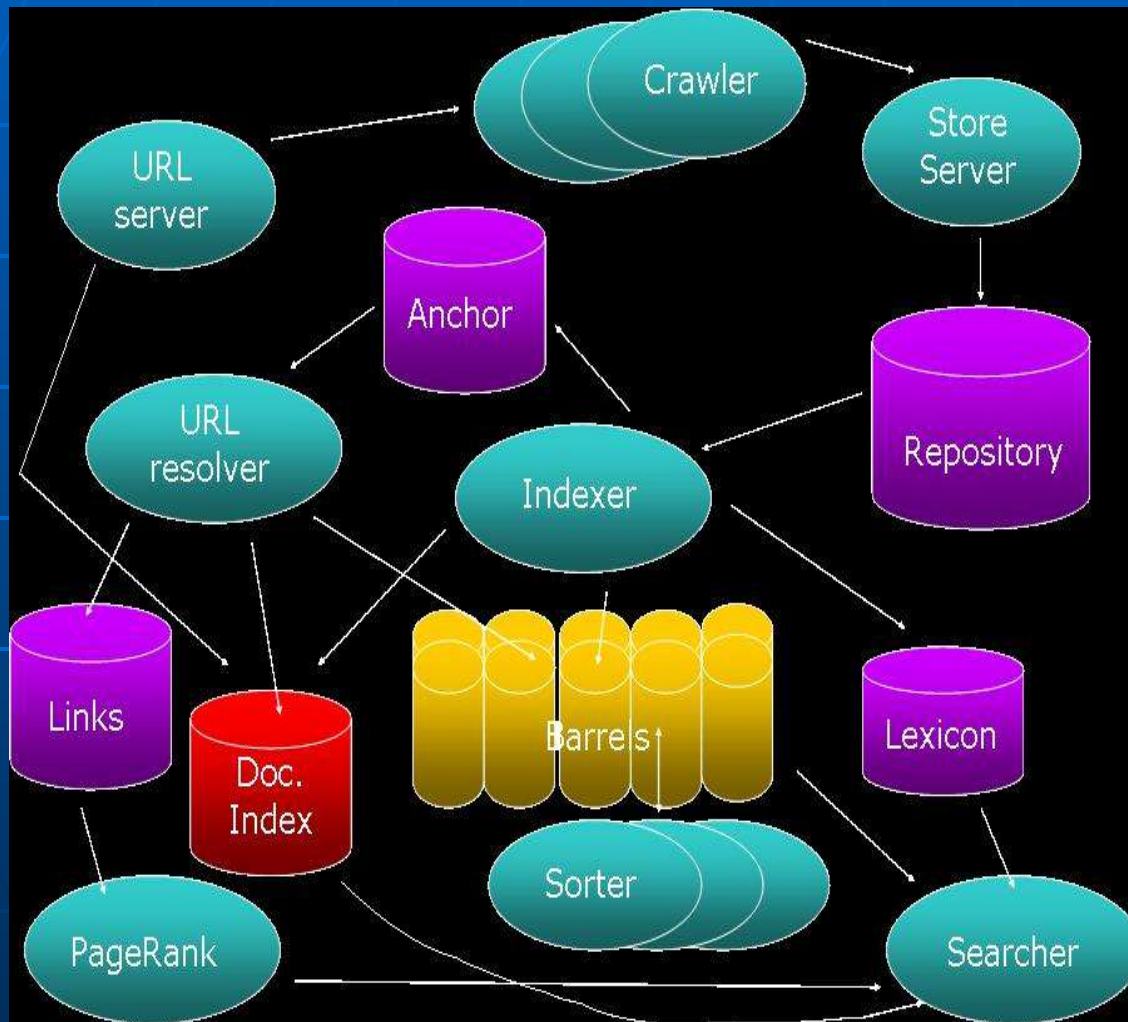
INDEXADOR

Lee y descomprime los documentos existentes en el almacén:

1.- Realiza el parsing. Cada documento se convierte en un conjunto de Hits (palabra, tamaño de fuente, capitalización)

2.- Extrae información importante de los links de cada documento. Esta información se almacena en un "fichero de anclas"

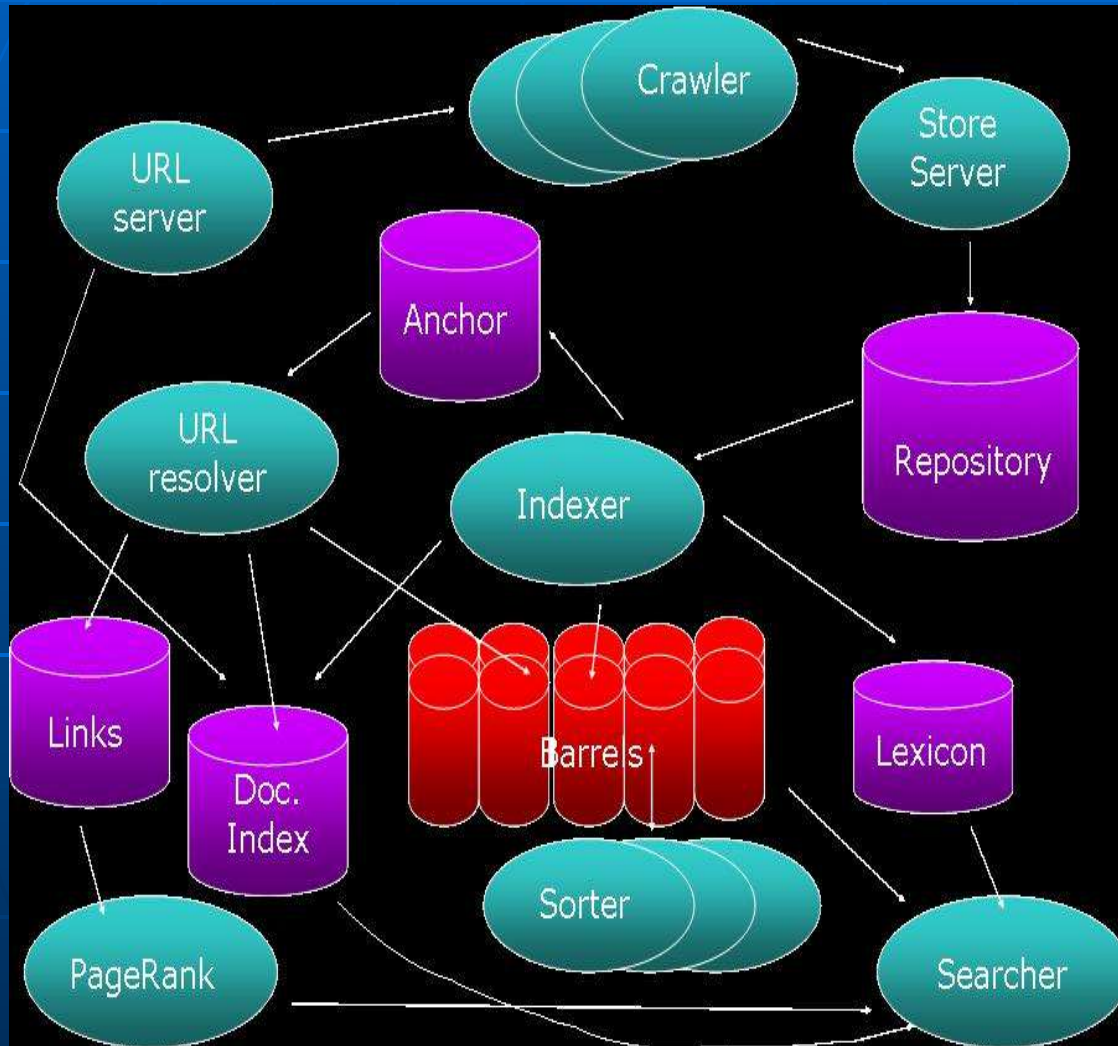
ARQUITECTURA



INDICE DE DOCUMENTOS

Base de datos de almacenamiento de los docsIDs

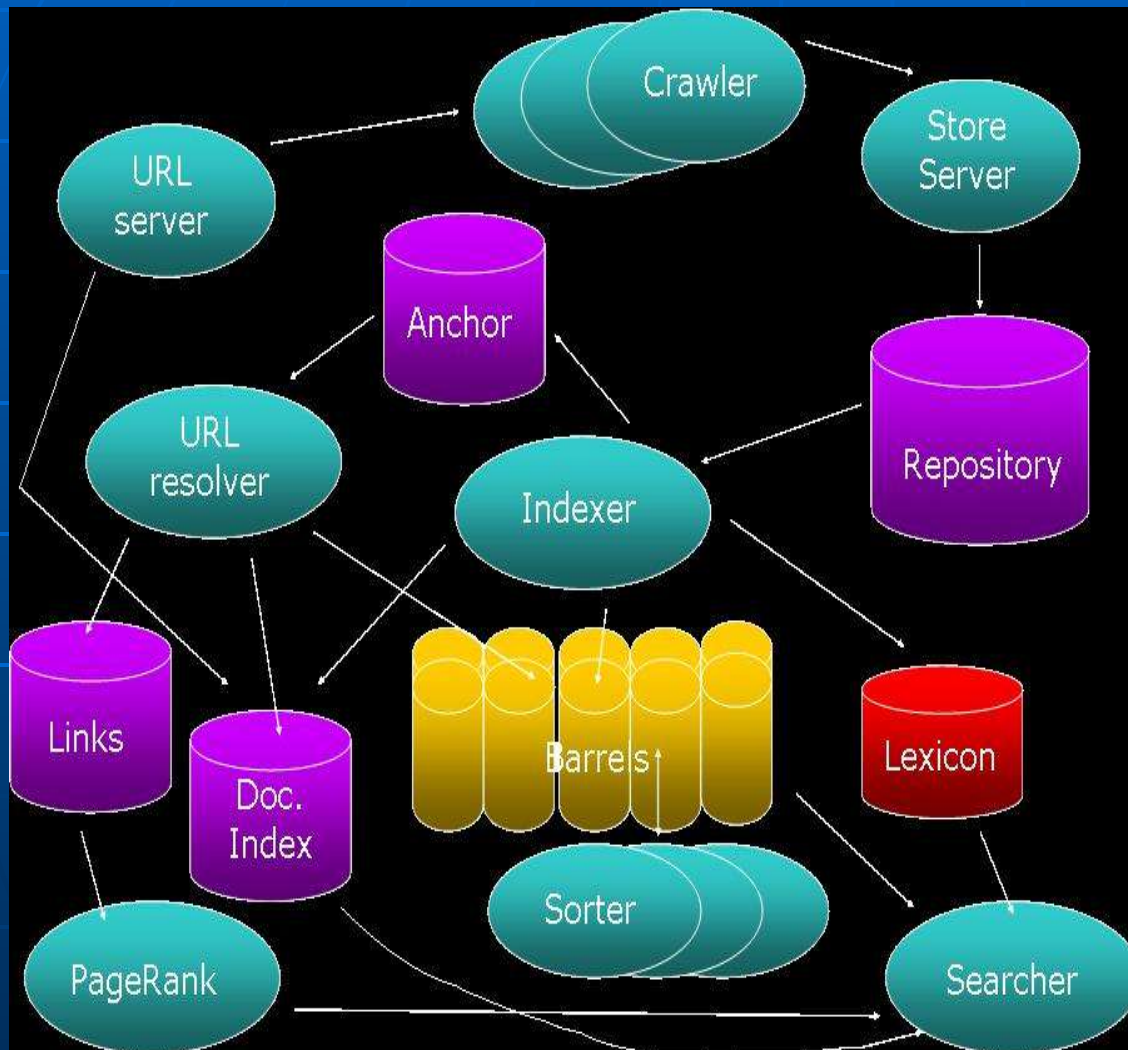
ARQUITECTURA



BARRELS

Conjunto de mini-almacenes en los que se guardan los hits generados por el indexador

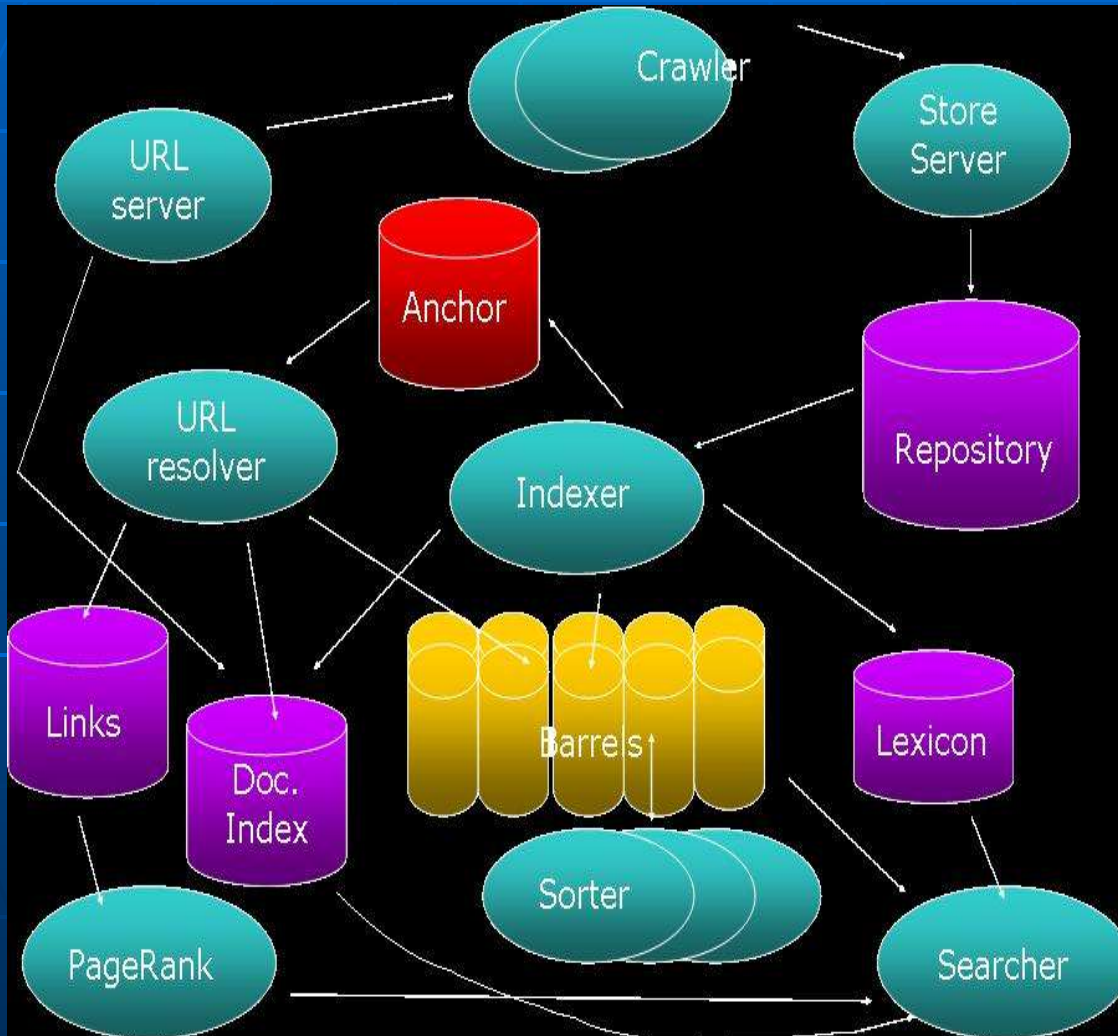
ARQUITECTURA



DICCIONARIO

Almacena términos junto con sus identificadores. Es utilizado por el indexador para construir los hits

ARQUITECTURA

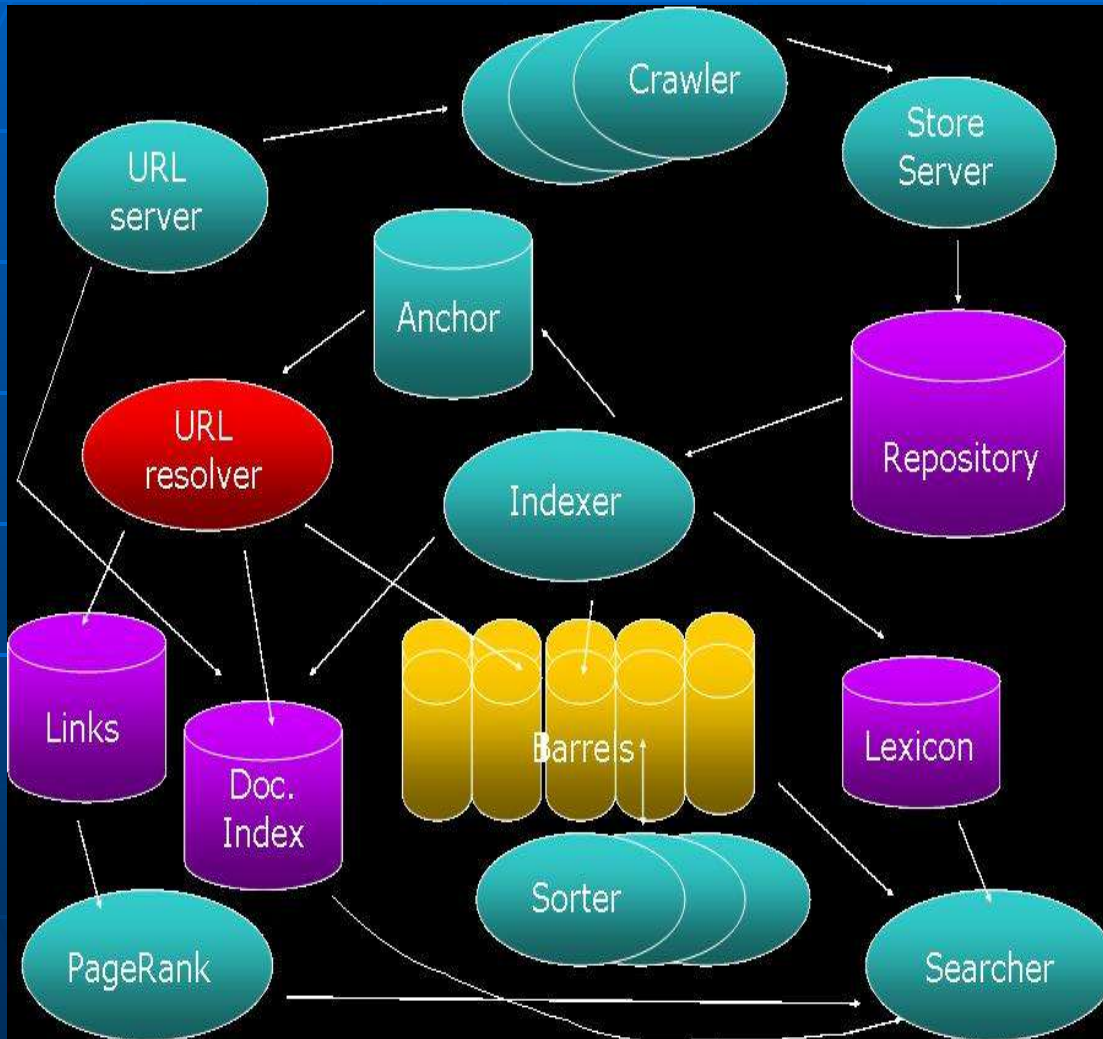


ANCLAS

Almacena la información generada por el indexador:

Desde y a donde apuntan los enlaces y si texto

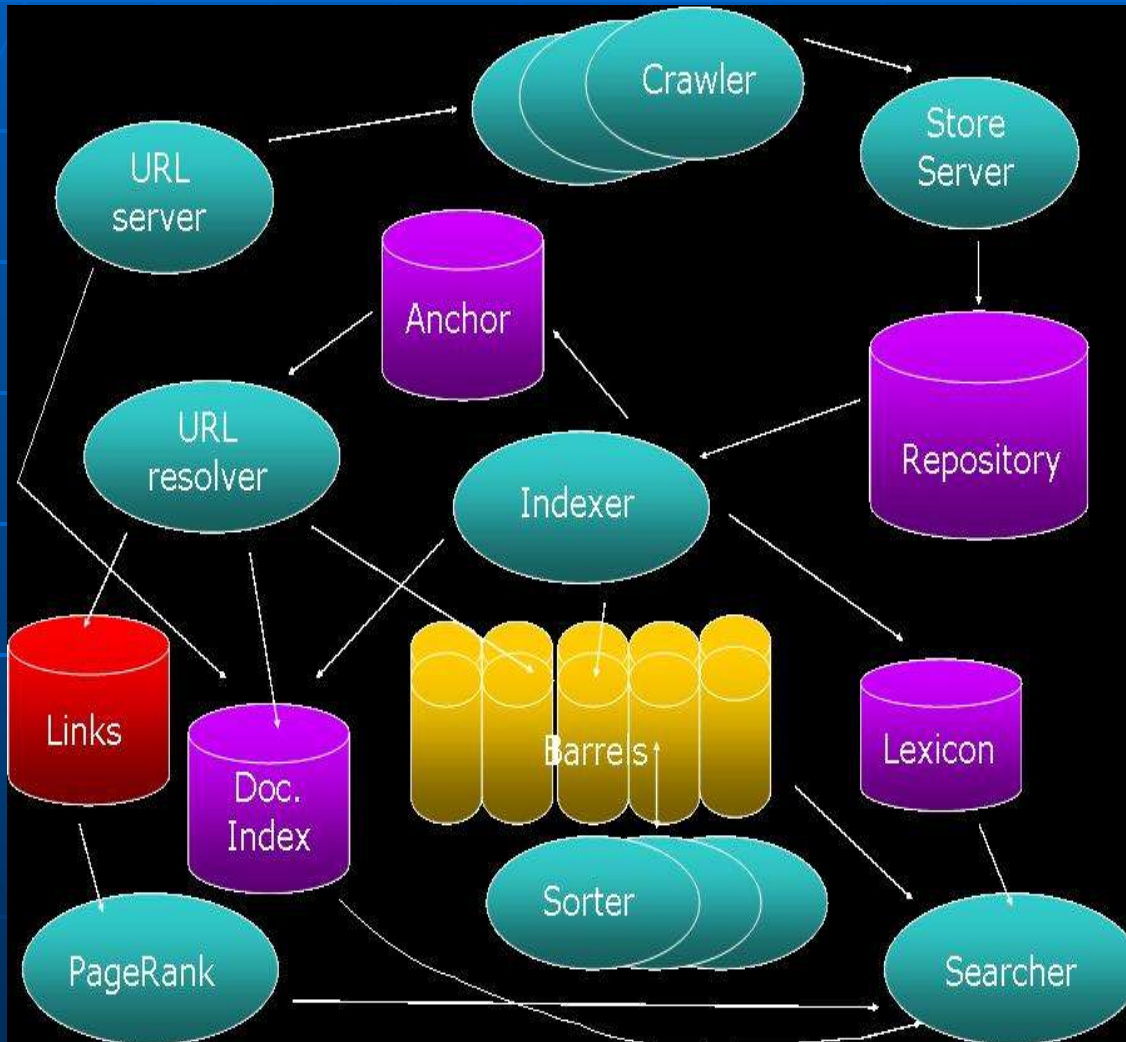
ARQUITECTURA



RESOLVEDOR DE URLs

Lee los "ficheros de enlaces" y conveniente los URLs relativos en absolutos. Además les asigne un docID y los almacene

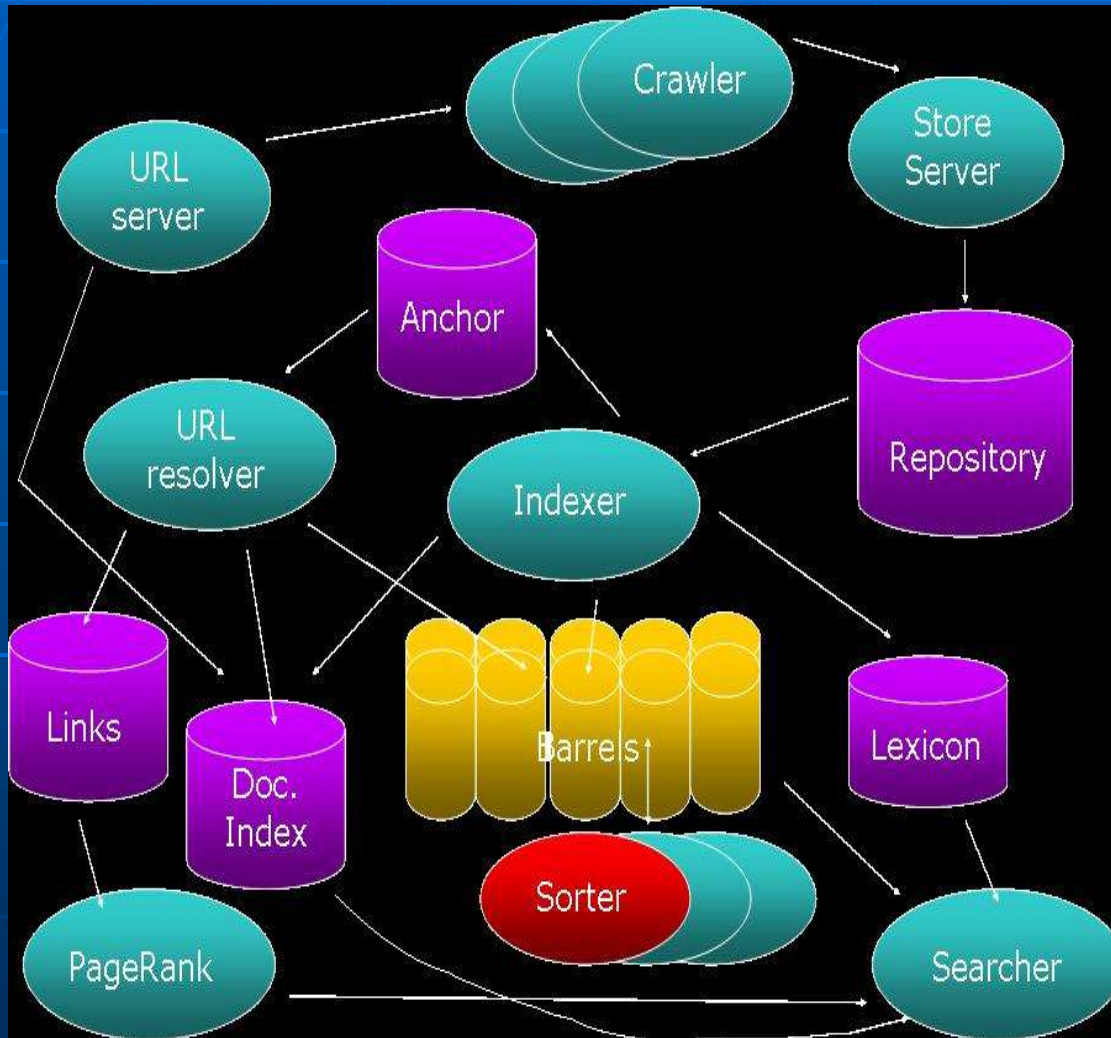
ARQUITECTURA



ENLACES

Base de datos en los que se almacenan pares (enlaces, docID) Esta información será utilizada para calcular el PageRank de cada documento en las búsquedas

ARQUITECTURA



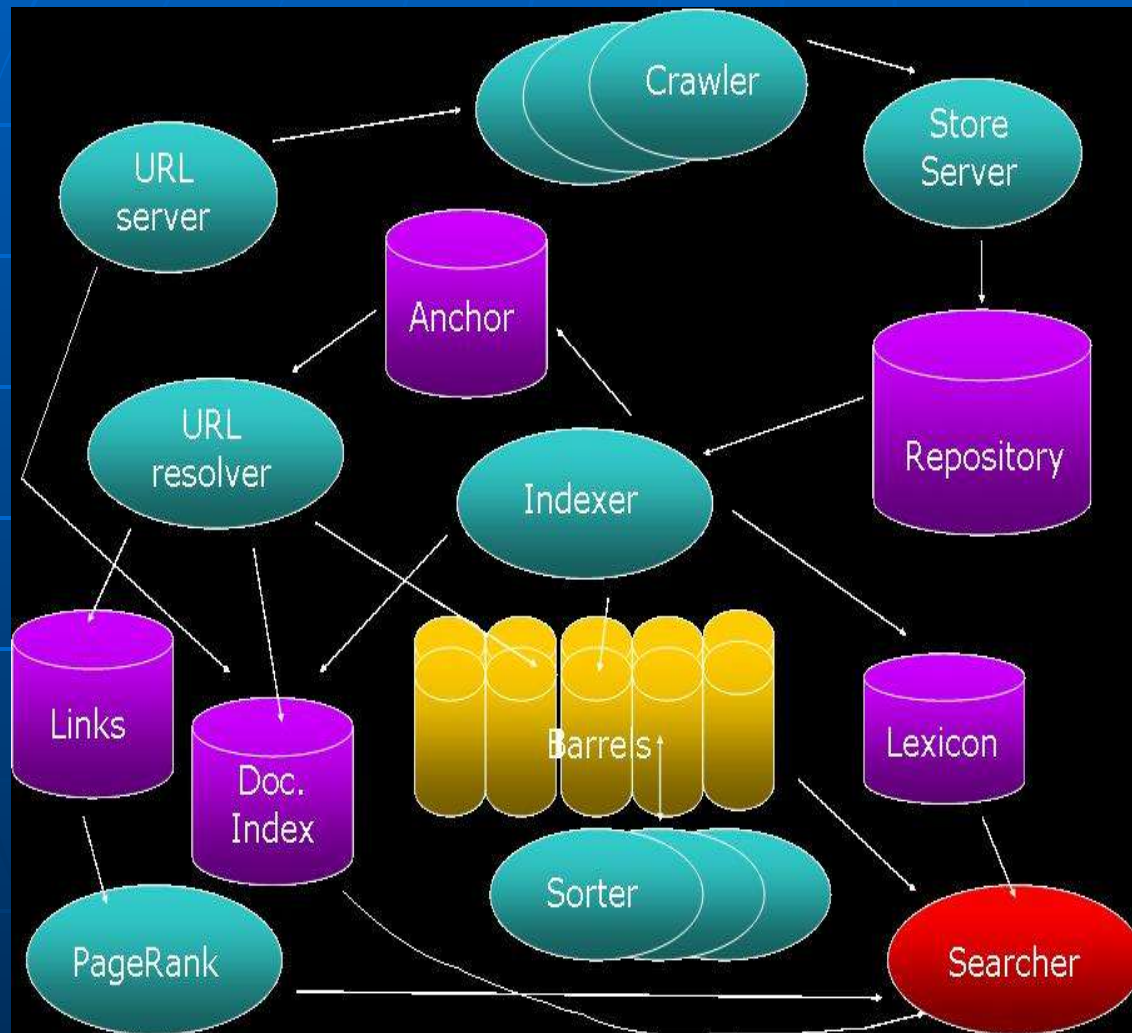
CLASIFICADOR

Genera un índice invertido de los hits existentes en los barrels.

1.- Asigna a cada término existente un identificador wordID.

2.- Convierte el índice de docIDs en uno de wordIDs

ARQUITECTURA

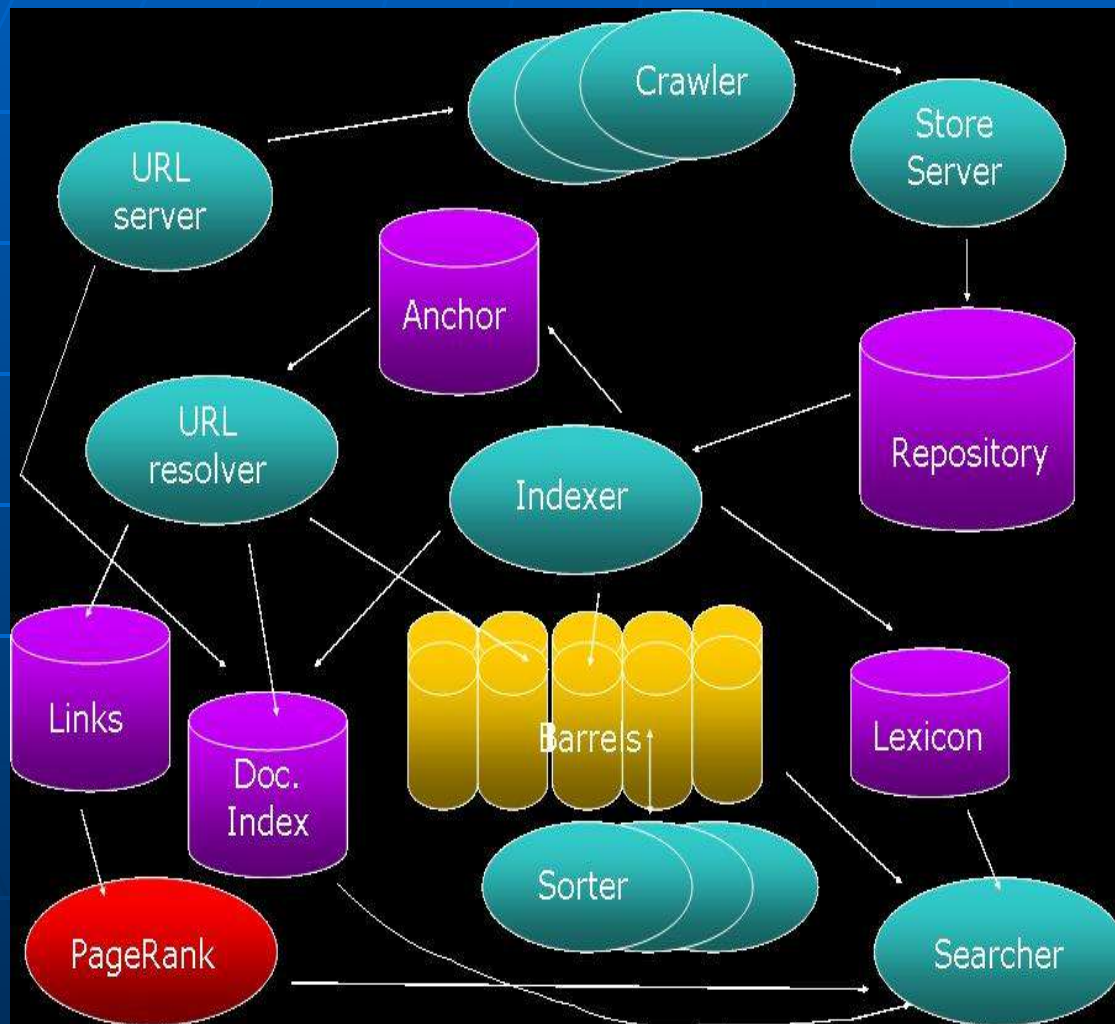


BUSCADOR

Se ejecuta en un servidor web y es el encargado de realizar las búsquedas para responder a las peticiones de los clientes

Utiliza el léxico generado por el indexador, el índice invertido generado por el clasificador y la información sobre el PageRank para calcular la relevancia de cada documento

ARQUITECTURA



PAGERANK

Encargado de calcular el Page Rank de cada documento que cumple las restricciones impuestas de una búsqueda.

Este calculo se realiza basándose en la información almacenada en la base de datos de Enlaces

COMO REALIZA LA BÚSQUEDA?

- 1.- Análisis sintáctico de la consulta
- 2.- Convertir términos es wordIDs
- 3.- Busca los doclist en los barrels correspondientes
- 4.- Examinar cada doclist hasta encontrar un documento que cumpla con los requisitos de la búsqueda
- 5.- Calcular el ranking del documento
- 6.- Asegurar que se hace el paso 4 en todos los barrels
- 7.- Ordenar los documentos por ranking y los devuelve al usuario

Que viene?

'Google Personalized' (en fase beta). En él, el usuario puede establecer una serie de temas en los que está interesado (sociedad, deportes, programación, ...), y la herramienta devuelve una serie de resultados que varían en función del grado de personalización.

Images.google.com

Google Search: paisajes - Opera

Archivo Editar Ver Navegación Marcadores Correo Ventanas Ayuda

Opera Opera Community OperaMail Buscar Super Buscar Amazon.com

Google Search: paisajes

http://images.google.com/images?q=paisajes&hl=en&btnG=Google+Search








Google Images

Web Images Groups News Froogle Local ^{New!} more »

paisajes Search Advanced Image Search Preferences

Moderate SafeSearch is on

Images Results 1 - 20 of about 66,700 for paisajes. (0.08 seconds)
Show: All sizes - Large - Medium - Small

 <p>paisajes_virtuales_wall.jpg 800 x 600 pixels - 74k webmail.openbox.ivisp.cl/ Downloads/ paisajes_v...</p>	 <p>manashahuemana.jpg 500 x 483 pixels - 112k www.cima-cordilleraazul.org.pe/ images/ manasha...</p>	 <p>choco3.jpg 513 x 441 pixels - 19k www.geocities.com/ Baja/Ravine/1464/</p>	 <p>tresrios.jpg 268 x 420 pixels - 55k www.diariogratis.com/834</p>
 <p>Paisajes. Jardines Andalu... 1704 x 2272 pixels - 1472k</p>	 <p>DREAMINGDAYS.com - Paisaj... 1024 x 768 pixels - 296k</p>	 <p>choco4.jpg 509 x 413 pixels - 14k</p>	 <p>cabal.JPG 312 x 205 pixels - 29k</p>

Home | Internet Explorer | Google | Micros... | Sin tit... | Docu... | Googl... | Dibujo... | 10:53 p.m.

Google Web Services

The screenshot shows the Opera browser window titled "Google Web APIs - Home - Opera". The address bar displays "http://www.google.com/apis/". The page content includes the Google logo, a navigation menu, and a main heading "Develop Your Own Applications Using Google". The main text describes the Google Web APIs service, which allows developers to query more than 8 billion web pages. It lists three steps: 1. Download the developer's kit, 2. Create a Google Account, and 3. Write your program using your license key. A sidebar on the left contains links for "Overview", "Download", "Create Account", "Getting Help", "API Terms", "FAQs", "Reference", and "Release Notes". A search box is also present. On the right, there is an illustration of a person at a computer and a box titled "Program ideas" with five suggestions: auto-monitoring the web, gathering market research, inventing a game, creating a novel UI, and adding spell-checking.

Google Web APIs (beta)

Develop Your Own Applications Using Google

With the Google Web APIs service, software developers can query more than 8 billion web pages directly from their own computer programs. Google uses the SOAP and WSDL standards so a developer can program in his or her favorite environment - such as Java, Perl, or Visual Studio .NET.

To start writing programs using Google Web APIs:

- 1 Download the developer's kit**
The Google Web APIs developer's kit provides documentation and example code for using the Google Web APIs service. The [download](#) includes Java and .NET programming examples and a WSDL file for writing programs on any platform that supports web services.
- 2 Create a Google Account**
To access the Google Web APIs service, you must [create a Google Account](#) and obtain a license key. Your Google Account and license key entitle you to 1,000 automated queries per day.
- 3 Write your program using your license key**
Your program must include your license key with each query you submit to the Google Web APIs service. Check out our [Getting Help](#) page or read the [FAQs](#) for more information.

Google Web APIs are a free beta service and are available for non-commercial use only. Please see our [terms of service](#).

With Google Web APIs, your computer can do the searching for you.

Program ideas

- ▶ Auto-monitor the web for new information on a subject
- ▶ Glean market research insights and trends over time
- ▶ Invent a catchy online game
- ▶ Create a novel UI for searching
- ▶ Add Google's spell-checking to an application

Maps.google

The image shows a screenshot of the Google Maps website as it appeared in the Opera browser. The browser window title is "Google Maps - Opera". The address bar shows the URL "http://maps.google.com/". The page features the Google Maps logo, navigation links for "Local Search" and "Directions", and a search input field. The main content is a map of the Washington, D.C. area, showing major highways like I-495, I-270, and I-66, and cities such as Baltimore and Washington. The map includes various landmarks and parks. On the right side of the map, there are links for "Print", "Email", and "Link to this page". The bottom of the browser window shows the Windows taskbar with several open applications and the system clock displaying "11:17 p.m.".

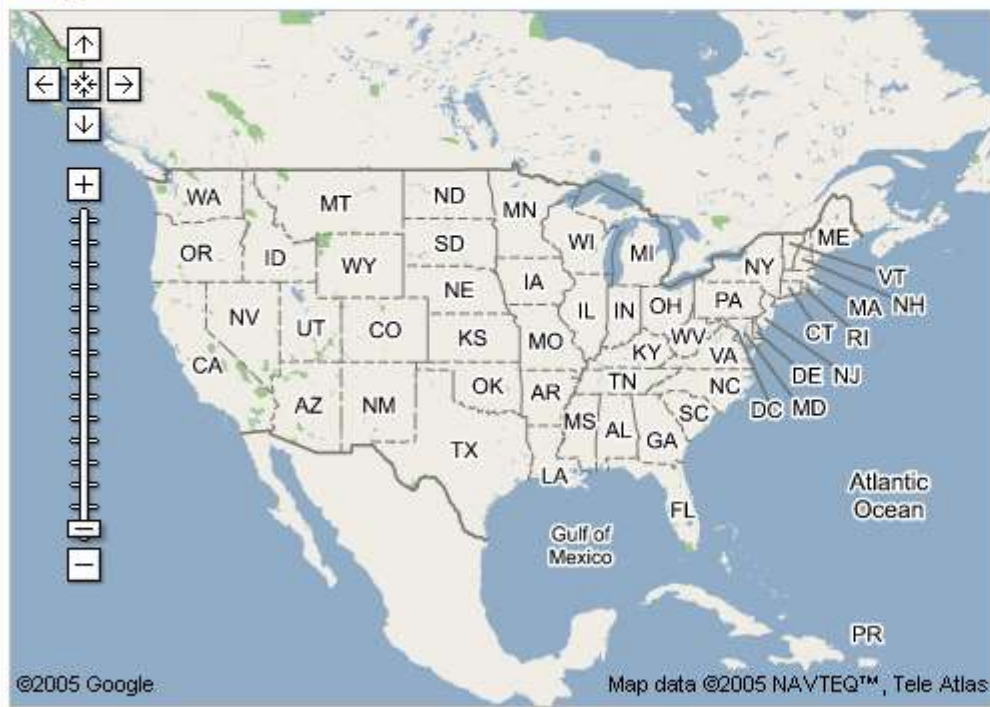


Maps Local Search Directions

Search

[Help](#)
[Send Feedback](#)

Maps



[Print](#) [Email](#) [Link to this page](#)

Drag the map with your mouse, or double-click to center.

Example searches:

Go to a location

- "kansas city" [try it](#)
- "10 market st, san francisco" [try it](#)

Find a business

- "hotels near lax" [try it](#)
- "pizza" [try it](#)

Get directions

- "jfk to 350 5th, new york, ny" [try it](#)
- "seattle to 98109" [try it](#)

[Take a tour >](#)



[Preferences](#) [Google Video Help](#)

Search recent TV programs online.

[Google Labs](#) - [Discuss](#) - [Terms of use](#) - [About Google Video](#)

©2005 Google

Google - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

Atrás Búsqueda Favoritos Multimedia

Dirección <http://www.google.com/webhp?complete=1&hl=en> Ir Vínculos »

Google™

Suggest BETA

Web Images Groups News Froogle Local ^{New!} more »

As you type, Google suggests:

page rank	
page rank	14,400,000 results
page ranking	19,200,000 results
page redirect	3,680,000 results
page rank checker	197,000 results
page refresh	3,460,000 results
page rank calculator	242,000 results
page rank tool	2,650,000 results
page refresher	514,000 results
page rank google	2,350,000 results
page redirection	1,190,000 results

Advanced Search
Preferences
Language Tools

sults. [Learn more](#)

Listo Internet

Inicio Microsoft PowerPoint - ... Adobe Reader - [googl... Google - Microsoft... 11:56 p.m.



Search [Advanced Scholar Search](#)
[Scholar Preferences](#)
[Scholar Help](#)

Stand on the shoulders of giants

[Google Home](#) - [About Google](#)

©2005 Google



To view personalized results, first create a profile of your interests.

[Create Profile](#)

[Feedback](#) - [Discuss](#) - [Terms of Use](#) - [FAQ](#)

©2004 Google - Searching 8,058,044,651 web pages

Google Sets - Microsoft Internet Explorer

Archivo Edición Ver Favoritos Herramientas Ayuda

Atrás Búsqueda Favoritos Multimedia

Dirección <http://labs.google.com/sets> Ir Vínculos

[Feedback](#) [Discuss](#) [Terms of Use](#)

Google™ Sets

Automatically create sets of items from a few examples.

Enter a few items from a set of things. ([example](#))
Next, press *Large Set* or *Small Set* and we'll try to predict other items in the set.

-
-
-
-
-

(clear all)

Listo Internet

Inicio Microsoft PowerPoint - ... Adobe Reader - [googl... Google Sets - Micr... 11:52 p.m.