

Las Matemáticas en Google

Javier Tordable
Ingeniero de Software

Índice

1. Los comienzos de Google
2. PageRank
3. Repertorio de Matemáticas
4. Preguntas



Los comienzos de Google

Backrub

<http://www.google.es/intl/es/about/corporate/company/history.html>

- **1995:** Larry Page y Sergey Brin se conocen en la Universidad de Stanford. Larry tiene 22 años y ha finalizado sus estudios en la Universidad de Michigan. Se plantea estudiar en Stanford, y Sergey, de 21 años, es el encargado de enseñarle el campus
- **1996:** Larry y Sergey participan en un programa de posgrado en Informática en Stanford y empiezan a colaborar en el desarrollo de un motor de búsqueda llamado BackRub. BackRub se utiliza en los servidores de Stanford durante más de un año, pero finalmente la Universidad deja de emplear este motor porque requiere demasiado ancho de banda
- **1997:** Larry y Sergey se dan cuenta de que el motor de búsqueda BackRub necesita un nuevo nombre. Tras una sesión de lluvia de ideas, se deciden por Google
- **1998:** En Septiembre Google fija su lugar de trabajo en el garaje de Susan Wojcicki e inicia los trámites de constitución de la sociedad en California el 4 de septiembre



Búsqueda en la web (1)

- La web considerada como una colección de documentos. La manera estándar de buscar un término es a partir de un índice
- Un índice es una tabla invertida, en la que para cada término tenemos la lista de documentos que contienen el término.
- Usando el índice buscar consiste únicamente en mostrar la intersección de las listas para cada documento
- El problema ahora es cómo ordenar la lista



Búsqueda en la web (2)

<http://www.uno.com> (1)

documento
numero uno

<http://www.dos.com> (2)

documento
numero dos

<http://www.tres.com> (3)

y otra
página más

documento	1,2
numero	1,2
uno	1
dos	2
y	3
...	

Google descarga
documentos y
construye un índice

documento

Buscar con Google

<http://www.uno.com>

<http://www.dos.com>

Al buscar una palabra
se muestran las
entradas del índice
correspondientes



Búsqueda en la web (3)

- Simplemente mostrar todos los documentos no es posible
- Mostrar documentos ordenados en base a criterios sencillos, como fecha, o número de ocurrencias de una palabra da malos resultados
- La idea de Larry Page y Sergey Brin es usar los enlaces entre páginas como señal de la calidad de un documento. De manera similar a citaciones entre artículos científicos



PageRank

Google

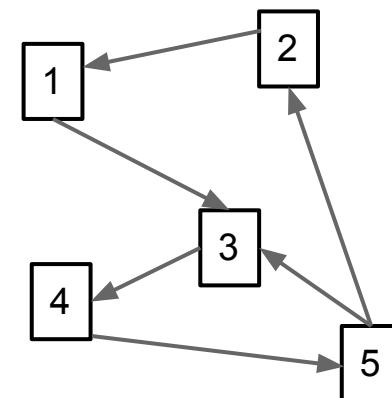
La Web como un grafo

<http://uno.com/>

```
<h1>  
Página web 1  
</h1>  
<p>  
Esta es mi web.  
</p>  
<p>  
Mi otra página es  
<a href="http://dos.com/">  
esta  
</a>  
<p>
```

<http://dos.com/>

```
<h1>  
Página web 2  
</h1>  
<p>  
Esta es mi otra web.  
</p>  
<p>  
Mi otra página es  
<a href="http://uno.com/">  
esta  
</a>  
<p>
```

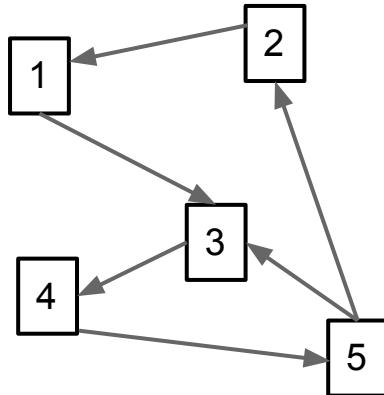


La idea iterativa de PageRank (1)

- PageRank es una aproximación a la probabilidad de llegar a una página web de manera aleatoria siguiendo enlaces
- Ejemplo: si una persona está en la página i con probabilidad π_i , que tiene enlaces a: $\{j, k\}$ la probabilidad de llegar a j es $1/2 * \pi_i$ y la probabilidad de llegar a k es también $1/2 * \pi_i$
- Si una página no tiene enlaces, asumimos que enlaza con todas las demás
- Inicialmente consideramos que todas las páginas tienen igual probabilidad



La idea iterativa de PageRank (2)



$$p_{1k} = p_{2_{k-1}}$$

$$p_{2k} = 1/2 * p_{5_{k-1}}$$

$$p_{3k} = p_{1_{k-1}} + 1/2 * p_{5_{k-1}}$$

$$p_{4k} = p_{3_{k-1}}$$

$$p_{5k} = p_{4_{k-1}}$$

$$\begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{bmatrix}_k = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 \\ 1 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{bmatrix}_{k-1}$$

- En cada etapa, la probabilidad (PageRank) se calcula a partir de la probabilidad en la etapa anterior
- Construimos A la matrix que tiene en la posición (i,j) 0 si la página j no tiene un enlace a la página i, o $1/k$ si la página j tiene k enlaces, y uno de ellos es hacia la página i
- La etapa inicial da igual probabilidad a todas las páginas. La siguiente etapa se calcula como $p_k = A * p_{k-1}$
- En general después de un cierto número de iteraciones obtendremos una aproximación razonable para el PageRank

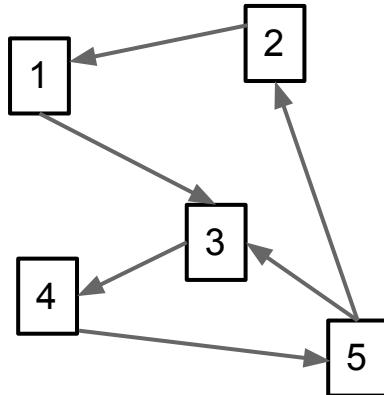


La idea algebraica de PageRank (1)

- Considerar las páginas web nodos, los enlaces vértices, y la web un grafo dirigido
- PageRank como estimación de la importancia de cada nodo del grafo
- Si una página tiene k enlaces, a las páginas P_1, \dots, P_k , consideramos cada uno de estos enlaces un "voto" para P_k
- El PageRank de la página P_k , p_k es la suma de todos los votos para esta página. Donde cada voto proveniente de P_i está multiplicado por el PageRank de P_i



La idea algebraica de PageRank (2)



$$\begin{aligned} p_1 &= p_2 \\ p_2 &= 1/2 * p_5 \\ p_3 &= p_1 + 1/2 * p_5 \\ p_4 &= p_3 \\ p_5 &= p_4 \end{aligned}$$

$$\begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/2 \\ 1 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{bmatrix}$$

- Si tomamos $p = G \cdot p$, el vector de PageRank es un autovector de autovalor 1
- G es una matriz estocástica. Todos los elementos son positivos, y la suma de cada columna es igual a 1
- La columna i contiene $1/k$ para cada uno de los k enlaces que parten del nodo i
- Si un nodo no tiene enlaces salientes, se asume que tiene un enlace con todos los demás nodos. Esto es necesario para que la matriz sea estocástica
- En estas condiciones la matriz siempre tiene el autovalor 1



PageRank (1)

- Los algoritmos anteriores tienen problemas cuando el grafo no es conexo. Bien porque no es posible alcanzar una determinada página a través de enlaces (para el método iterativo) bien porque hay varios autovectores para el autovalor 1 (en el método algebraico)
- La solución está en añadir $\lambda * \frac{1}{n} * \mathbf{1}$, donde $\mathbf{1}$ es una matriz con unos en todas las posiciones y n es el número de nodos (Normalmente $\lambda = 0.15$)

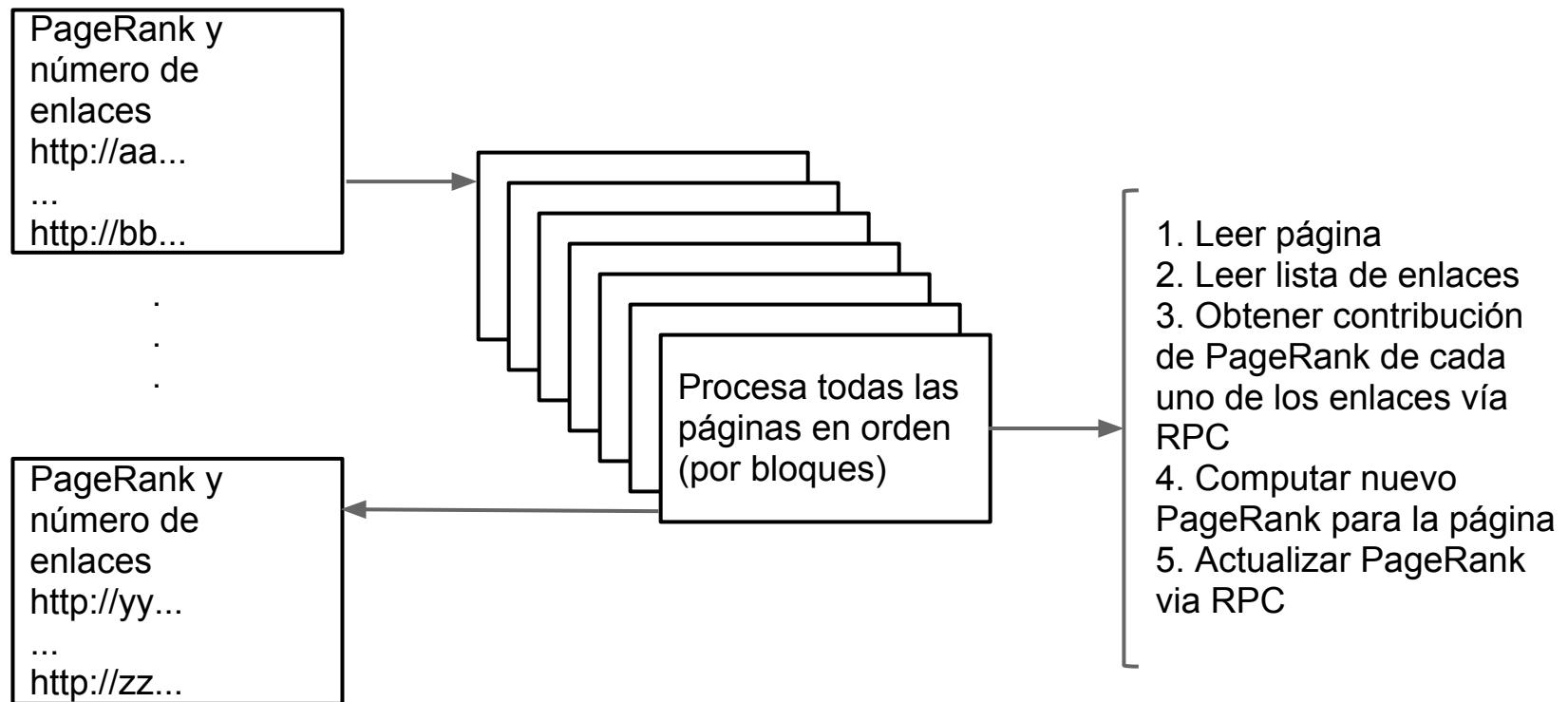


PageRank (2)

- La *Matriz de Google* es:
$$G = (1 - \lambda) A + \lambda \frac{1}{n} \mathbf{1}$$
- Esta matriz también es estocástica y todos los elementos son estrictamente positivos
- Por el Teorema de Perron-Frobenius G tiene el autovalor 1 y el correspondiente autovalor tiene multiplicidad 1
- Usando el método de las potencias para G se puede encontrar este mismo autovector de forma iterativa



Ejemplo de implementación



La resolución del sistema lineal es difícil de parallelizar. Aunque el método iterativo sea lento, es más rápido para obtener una aproximación



PageRank. Enlaces

Artículo original sobre Google de Sergey Brin y Larry Page:

<http://infolab.stanford.edu/~backrub/google.html>

Presentación sobre PageRank en Cornell University:

<http://www.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture3/lecture3.html>

PageRank:

<http://es.wikipedia.org/wiki/PageRank>

Teorema de Perron-Frobenius:

http://en.wikipedia.org/wiki/Perron%E2%80%93Frobenius_theorem

Método de las potencias:

http://es.wikipedia.org/wiki/M%C3%A9todo_de_las_potencias



Repertorio de Matemáticas



Gmail (1)

Gmail Calendar Documents Photos Reader Web more ▾

Hiking Fan

Gmail by Google SEARCH MAIL SEARCH THE WEB Show search options Create a filter

Mail Archive Report spam Delete Move to ▾ Labels ▾ More ▾ C 1 - 15 of 15

COMPOSE MAIL

Inbox (3)

- Jason Cornwell » Please return my stapler - Hi, You seem to have taken my stapler. Please, 1:10 pm
- Paul McDonald » Fun Hike Yesterday! - Thanks for the great hike yesterday, it was awesome! 1:06 pm
- Arielle Reinstein » July 4th weekend - Hi there: I heard you'll be around this weekend and I'd like to go out 1 Jun 28

Drafts (2)

- JS Bach » Tonhalle concert Friday - Hey man, there's a great concert this Friday evening! Jun 22
- Christine Chiu » Hi Hiking, Looking for opinion on my diet/fitness app - Hi Hiking, I bumped into you at the park Jun 9

Hiking (3)

- Urgent!
- 12 more ▾

Chat

Search, add, or invite

Hiking Fan Set status here ▾

Call phone

Arielle

Emily

Jason

Michael

Paul

Spammer spammer@spammer.com Mar 12

to me ▾

Be careful with this message. It might contain a virus or a malicious link. [Learn more](#)

* How are you? - Hey there, We haven't spoken in a while. How are you? Wow Mar 24

* VW Auction in Tacoma - Hi, I was doing a search on Google for VW's in Tac Mar 6

* New voicemail from (610) 810-5507 at 5:10 AM - /voicemail from (610) 810-5507 Mar 6



Gmail (2)

- La detección de Spam es un problema clásico de clasificación usando aprendizaje automático (el ordenador aprende el algoritmo a partir de los datos) en particular supervisado (con ejemplos ya clasificados)
- En esencia el aprendizaje automático tiene dos fases, la fase de aprendizaje (obtener el modelo de clasificación) y la fase de clasificación (usar el modelo para clasificar instancias)



Gmail (3)

- Clasificar consiste en extraer características de la instancia y después aplicar el modelo a esas características
- Las características de una instancia es en general un vector en un espacio euclídeo n -dimensional para un n muy grande (100-1000 dimensiones es normal, pero hay ejemplos con 1M-10M)
- El modelo es en general un subespacio de dimensión $n-1$ que divide el espacio original en dos espacios disjuntos

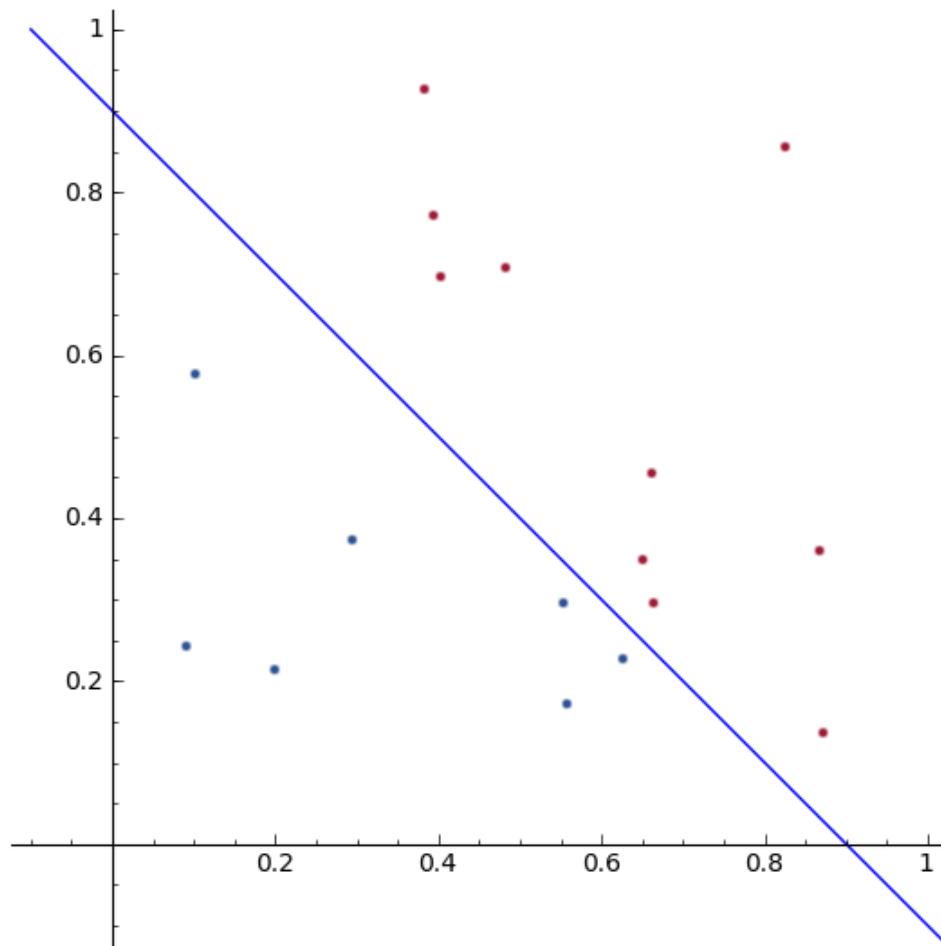


Gmail (4)

- Un ejemplo sencillo
- En un email, podemos tomar características como: longitud del email, número de mayúsculas, remitente en el libro de contactos, etc.
- Un modelo de clasificación sencillo es un hiperplano en el espacio de características. Instancias a un lado del hiperplano son clasificadas como mensajes válidos e instancias al otro lado como spam



Gmail (5)



Gmail (6)

Ejemplos más complicados de modelos:

- Árboles de decisión (funciones a escalones)
- Redes neuronales (un nodo de la red es la composición de una función, normalmente logística, con una combinación lineal de sus entradas. Una red es una combinación de múltiples niveles de nodos)
- Máquinas de vectores de soporte con una función kernel (funciones lineales en un espacio transformación del espacio original)



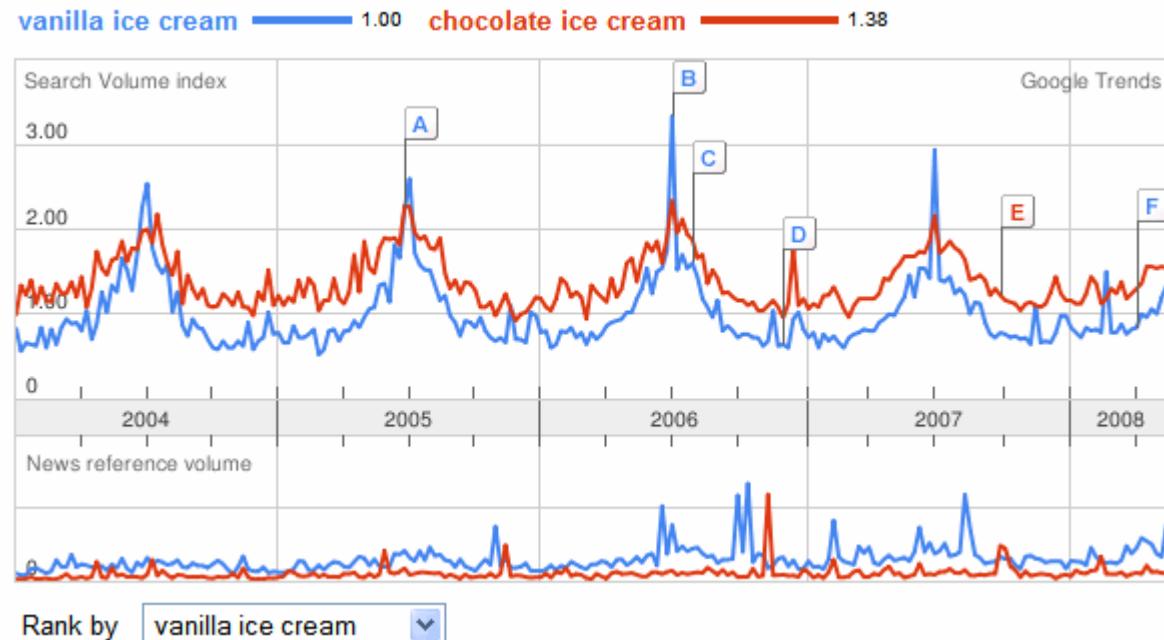
Gmail (7)

Enlaces:

- The War Against Spam: A report from the front line
<http://research.google.com/pubs/pub36954.html>
- The Learning Behind Gmail Priority Inbox
research.google.com/pubs/archive/36955.pdf
- Publications by Googlers in Artificial Intelligence and Machine Learning
<http://research.google.com/pubs/ArtificialIntelligenceandMachineLearning.html>



Google trends (1)



TM

Google trends (2)

- El procesamiento de series temporales es un ejemplo clásico de aplicación matemática. Las técnicas van desde regresión a análisis de Fourier, modelos ocultos de Markov o autocorrelación
- Se utiliza por ejemplo para predecir el número de búsquedas, número de usuarios, ingresos, etc. para una gran variedad de productos (miles de análisis cada día)

Large-Scale Parallel Statistical Forecasting Computations in R

<http://research.google.com/pubs/pub37483.html>



Búsqueda por voz (1)



Búsqueda por voz (2)

- El procesamiento del habla automático (ASR) tiene dos partes fundamentales
 - Primero procesamiento de la señal de sonido, dividir en partes pequeñas, aplicar FT, tomar los coeficientes más significativos
 - Segundo, el habla se modela usando un modelo oculto de Markov, en el los estados ocultos son texto, y la secuencia de eventos es la señal de sonido. El algoritmo de Viterbi se puede usar para obtener la secuencia de estados de mayor probabilidad
- Google Search by Voice: A case study
<http://research.google.com/pubs/archive/36340.pdf>



Google books (1)

Google

Books

[GET PRINT BOOK](#)

No eBook available

[Amazon.com](#)
[Barnes&Noble.com](#)
[Books-A-Million](#)
[IndieBound](#)

[Find in a library](#)
[All sellers »](#)



New! Shop for Books on Google Play

Browse the world's largest eBookstore and start reading today on the web, tablet, phone, or ereader.

[Go to Google Play Now »](#)

My library
My History
Books on Google Play

El ingenioso hidalgo Don Quijote de la Mancha

Alonso Fernández de Avellaneda  

★★★★★ 1 Review

Biblioteca Nueva, 2000 - [Literary Criticism](#) - 789 pages



From inside the book

100 pages matching el quijote de la mancha in this book

Page 10

de Vega. Pero lo cierto es que Avellaneda nunca tuvo en su mira hacer lo mismo que ellos. De hecho, evitó mencionar intencionadamente la continuación más cercana en el tiempo, pero también la más problemática y próxima a la

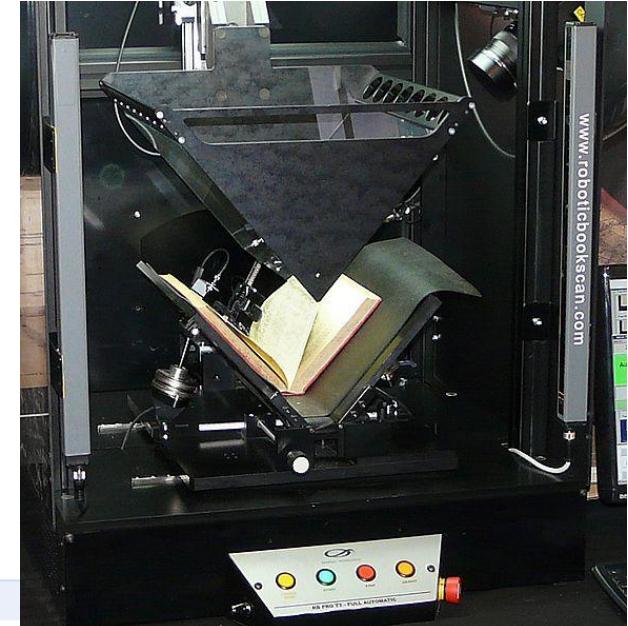


Imagen: Wikimedia. Vadaro

Google books (2)

- OCR es en general un problema de procesamiento de imagen (obtención de imágenes de caracteres individuales, con resolución, orientación y contraste adecuados) y aprendizaje automático (clasificación del carácter)
- Por ejemplo:

An Overview of the Tesseract OCR Engine

<http://research.google.com/pubs/archive/33418.pdf>

Low Cost Correction of OCR Errors Using Learning in a Multi-Engine Environment

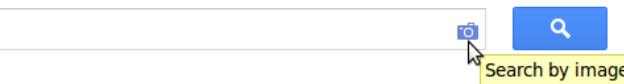
<http://research.google.com/pubs/archive/35525.pdf>

Translation-Inspired OCR

<http://research.google.com/pubs/pub37260.html>



Búsqueda por imágenes (1)



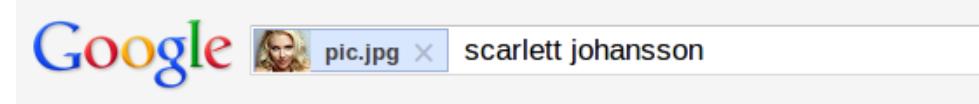
at image results with new related search previews. [Learn more.](#)



Search by image
Search Google with an image instead of text.

Paste image URL | Upload an image [?](#)

No file chosen



Search

About 3,450 results (0.85 seconds)

Everything

Images

Maps

Videos

News

Shopping

More



Image size:
1600 × 1200

Find other sizes of this image:
[All sizes](#) - [Medium](#) - [Large](#)

Best guess for this image: [scarlett johansson](#)

[**Scarlett Johansson - Wikipedia, the free encyclopedia**](#)

en.wikipedia.org/wiki/Scarlett_Johansson

Scarlett Johansson (born November 22, 1984) is an American actress, singer. Johansson made her film debut in *North* (1994) and was later ...

[Demo](#)



Búsqueda por imágenes (2)

- La búsqueda por imágenes es un ejemplo de recuperación de información en base al contenido (colores, formas, texturas, etc.)
Content-based Multimedia Information Retrieval: State of the Art and Challenges
<http://www.liacs.nl/home/mlew/mir.survey16b.pdf>
- El fundamento es una medida de similaridad entre imágenes. Por ejemplo la diferencia entre los histogramas de color. O en general la diferencia entre los vectores de características de las imágenes
Tour the World: building a web-scale landmark recognition engine
<http://research.google.com/pubs/archive/35291.pdf>
(Image search) Web-scale Image Annotation
<http://research.google.com/pubs/archive/34669.pdf>



Picasa (1)

The screenshot shows the Picasa Web Albums interface. At the top, there's a navigation bar with the Picasa logo, Home, My Photos, Explore, Upload, and a Search bar. Below the navigation is a 'Featured Photos' grid and a 'Recent Photos' slider.

Popular Tags:

- wedding
- trip
- day
- new
- pics
- christmas
- party
- family
- photos
- birthday
- pictures
- vacation
- park
- year
- beach
- summer
- people
- city
- fotos
- camera
- photography
- lake
- house
- cruise
- nature
- show
- photo
- haloween
- tour
- visit
- october
- winter
- valley
- bday
- water
- south
- december
- fall
- landscape
- flowers
- Locations
- europe
- italy
- usa
- pers
- china
- france
- india
- new york
- hawaii

Recent Photos: Slideshow

People: Unnamed people

Unacademy

©2011 Google. Terms · Privacy Policy · De

A large red arrow points from the 'Add a name' input field in the 'People' overlay to the 'Add a name' input fields for each of the six unnamed people listed in the sidebar.

TM

Picasa (2)

- En esencia una imagen es un conjunto de tres matrices, una para cada color primario
- El procesamiento digital de imágenes, en particular la aplicación de filtros consiste en ejecutar una convolución en estas matrices

<http://lodev.org/cgtutor/filtering.html>

http://www.emt.jku.at/education/Inhalte/se_moderne_methoden/WS0304/Haim-Mathematics_in_Imaging.pdf

- Una funcionalidad nueva en Picasa es la detección automática de caras. En general la detección de caras es un problema complejo de procesamiento de imagen y de aprendizaje automático

Handbook of Face Recognition

<http://research.google.com/pubs/archive/36368.pdf>

Large-Scale Manifold Learning

<http://research.google.com/pubs/pub34395.html>



YouTube (1)

YouTube

Explorar | Películas | Subir video

Vídeos Música Películas Programas En directo Deportes Educación Noticias

Todas las categorías

Recomendaciones personalizadas

Motor

Comedia

Ocio

Cine y animación

Juegos

Consejos y estilo

ONG y activismo

★ Más vistos hoy »

Reveal Trailer - Official Call of Duty: Black Ops 2
Visto 14049349 veces | hace 4 días
CALLOFDUTY

CHEESEBURGER! - Ray William Johnson
Visto 4168215 veces | hace 4 días
RayWilliamJohnson

KATE UPTON IS YO CAT DADDY!!
Visto 2344191 veces | hace 5 días
sxephil

Minecraft - Mind Control Hat - Explosives+ Mod Spotlight
Visto 1433136 veces | hace 5 días
BlueXephos

Black Ops 2 - Official Gameplay Reveal Trailer
Visto 1102291 veces | hace 4 días
TmarTn

Dear YouTube
Visto 816481 veces | hace 4 días
RoosterTeeth



YouTube (2)

- Hay una multitud de ramas de las máticas que se pueden relacionar con un servicio complejo como YouTube. Por ejemplo:
- El vídeo en YouTube está comprimido (http://en.wikipedia.org/wiki/Data_compression). Los algoritmos de compresión (http://en.wikipedia.org/wiki/Rate%E2%80%93distortion_theory) provienen de la teoría de la información, códigos, etc.
- Otro problema es detectar eventos. Por ejemplo, para clasificar el vídeo, o para crear un snippet
YouTubeEvent: On Large-Scale Video Event Classification
<http://research.google.com/pubs/archive/37392.pdf>
YouTubeCat: Learning to Categorize Wild Web Videos
<http://research.google.com/pubs/archive/36387.pdf>



Traductor de Google (1)

The screenshot shows the Google Translate interface. At the top, there's a navigation bar with links like '+Tú', 'Búsqueda', 'Imágenes', 'Maps', 'Play', 'YouTube', 'Noticias', 'Gmail', 'Docs', 'Calendar', and 'Más'. Below the bar is the Google logo and a 'Iniciar sesión' button. The main area is titled 'Traductor' and has language selection dropdowns set to 'Del: inglés - detectado' and 'Al: español'. A large blue button labeled 'Traducir' is next to these. Below the dropdowns are three tabs: 'inglés' (which is selected), 'español', and 'francés'. On the left, a text input box contains the English phrase 'hello everybody'. On the right, the translated Spanish phrase 'Hola a todos' is displayed. There are small icons for a speaker and a checkmark below the Spanish text. A note at the bottom says '¡Nuevo! Haz clic en las palabras anteriores para editar y ver traducciones alternativas. [Descartar](#)'.

Traductor de Google para empresas: [Google Translator Toolkit](#) [Traductor de sitios web](#) [Global Market Finder](#)



Traductor de Google (2)

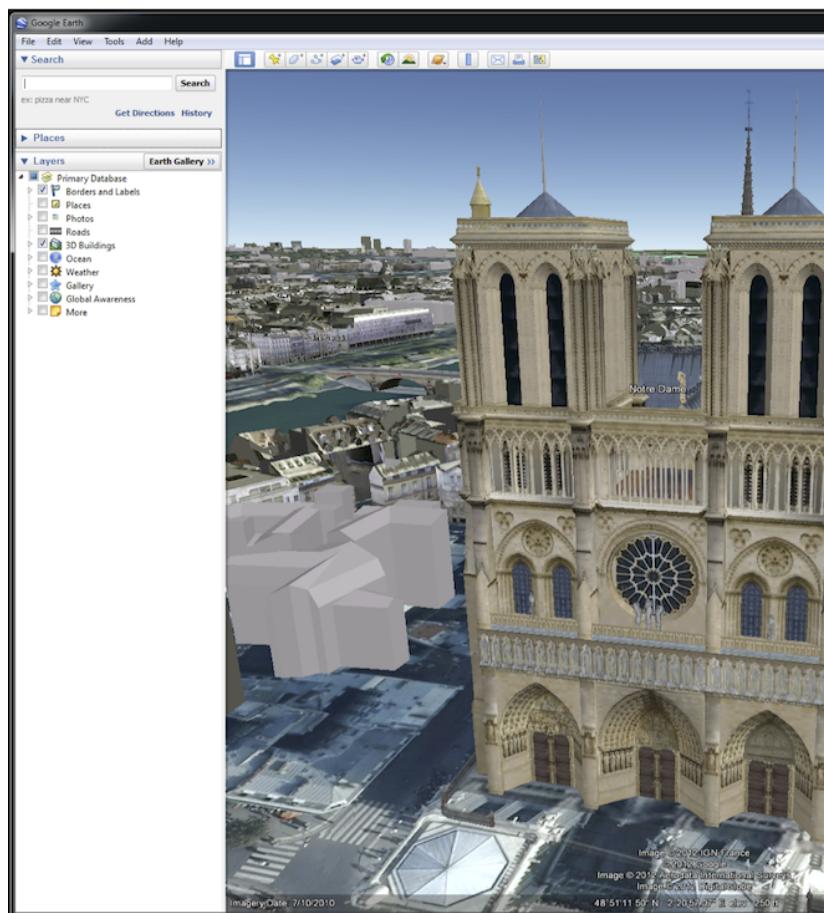
- Hay varias maneras de realizar traducción. Una manera clásica consiste en procesar el texto en una representación abstracta que implica conocimiento del lenguaje (NLP) y transformar esta representación en el lenguaje de destino
- El método usado en Google usa una cantidad enorme de texto para construir un **modelo estadístico**

Large Language Models in Machine Translation

<http://research.google.com/pubs/archive/33278.pdf>



Google Earth (1)



Google Earth (2)

- Los fundamentos están en la geometría euclídea en 3 dimensiones, topografía y fotogrametría, fusión de información 2D y 3D, etc. áreas que han sido estudiadas profundamente
- Las mayores aportaciones en Google normalmente consisten en atacar los problemas a escala Web



AdWords (1)

Google best laptops 2012 Sign in

Search About 199,000,000 results (0.19 seconds) 

Everything Ad related to best laptops 2012 Why this ad?

Images [The new MacBook Air - The future of the notebook | apple.com](#)

Maps [www.apple.com/macbookair](#)

Videos Available now. Learn more.

News 211 people +1'd this page

Shopping [↳ Why you'll love a Mac - OS X Lion - Great Mac apps](#)

More

Seattle, WA [Laptop Computer 2012 | Best Laptop Computers | Compare Laptop ...](#)

Change location [computers.totenreviews.com > Computers](#)

Any time [Compare the best laptop computers. Side-by-side comparisons of features and prices of top rated laptops. Easily see which laptop computer stands above the ...](#)

Past hour [↳ Dell XPS 15z - Laptop Configurator - HP Pavilion dm4x - Sony S Series](#)

[Best & Worst Laptop Brands 2012](#)

[www.laptopmag.com/mobile-life/best-brands-2012.aspx](#)

Mar 14, 2012 – The biggest laptop-makers ranked from software to support. Find out which brand makes the best laptops.

[↳ Overall Scorecard - Acer/Gateway - ASUS - Dell/Alienware](#)

Ads - Why these ads?

[Windows Recommended PCs](#)

[windows.microsoft.com/](#)

Answer 5 Quick Questions & We'll Recommend the **Best** PCs For You.

[Best-Selling Laptops](#)

[www.bestbuy.com/Laptops](#)

Get Connected with a **Laptop**. Free Shipping + In-Store Pickup! 330 NE Northgate Way, Seattle, WA

[New 2012 Dell Laptops](#)

[www.dell.com/Laptops](#)

dell.com is rated ★★★★☆

Find the Latest **2012**, Performance **Laptops** w/ Intel® Core™ at Dell! 1,724 people +1'd Dell

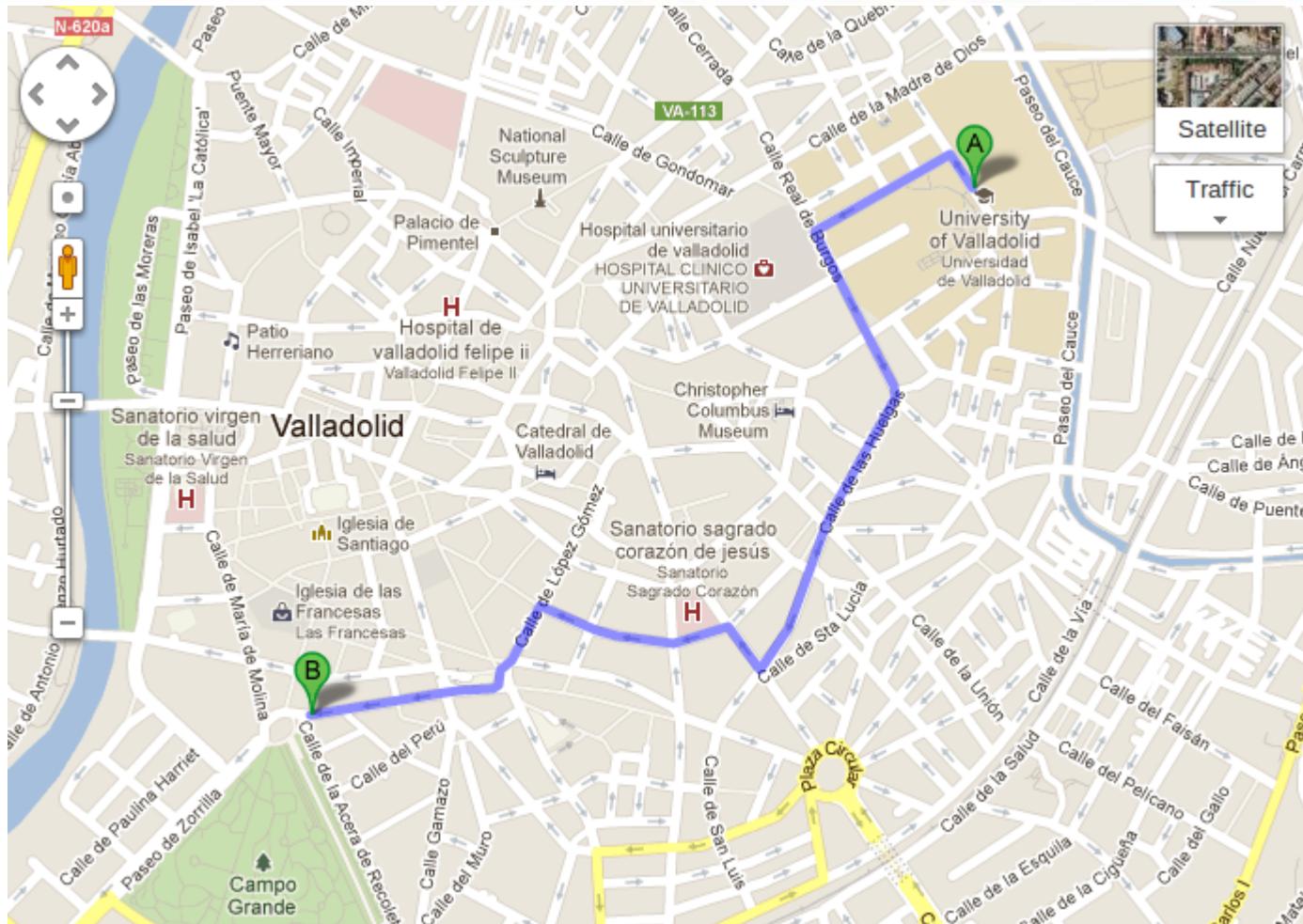


AdWords (2)

- AdWords funciona con un mecanismo de subasta. Los anunciantes hacen propuestas por cada espacio de anuncios
Hal Varian. Online Ad Auctions:
<http://people.ischool.berkeley.edu/~hal/Papers/2009/online-ad-auctions.pdf>
- La teoría de subastas estudia las distintas estrategias posibles y su eficacia. Es una rama aplicada de la teoría de juegos
- Adwords usa una subasta generalizada de segundo precio
http://en.wikipedia.org/wiki/Generalized_second-price_auction
Adwords, An Algorithmic Perspective
<http://paul.rutgers.edu/~mangesh/cs514/notes/pres3.pdf>



Google Maps (1)



Google

Google Maps (2)

- Muchos de los algoritmos en Google Maps provienen de teoría de grafos básica. Por ejemplo, encontrar el camino más corto entre dos nodos de un grafo (Dijkstra) al buscar direcciones
- Un problema es que los grafos usados en mapas tienen millones de nodos. Pero es importante poder calcular direcciones muy rápido. Por ejemplo usando jerarquías:

<http://algo2.iti.kit.edu/schultes/hwy/esaHwyHierarchies.pdf>



Redes y Sistemas distribuidos (1)



Imagen: Wikimedia. Midom



Redes y Sistemas distribuidos (2)

- Técnicas estadísticas para modelar la disponibilidad de sistemas de computación, de manera similar a estudios de control de calidad en otras industrias
- Por ejemplo, modelos ocultos de Markov:
Availability in Globally Distributed Storage Systems
<http://research.google.com/pubs/pub36737.html>
Designs, Lessons and Advice from Building Large Distributed Systems
<http://www.cs.cornell.edu/projects/ladis2009/talks/dean-keynote-ladis2009.pdf>
- La teoría de colas se puede usar para modelar la ejecución de trabajos en un sistema distribuido



Redes y Sistemas distribuidos (3)

- Un ejemplo clásico es la teoría de grafos aplicada a conexiones entre centros de datos, o redes de ordenadores
- Una red es modelada como un grafo en la que los enlaces pueden fallar con una determinada probabilidad
- Es interesante estudiar cuando un grafo proporciona la mayor resistencia a fallos o ancho de banda (conectividad) o latencia (diametro) por el menor costo



Redes y Sistemas distribuidos (4)

- Un tema especialmente interesante de teoría de grafos con multitud de aplicaciones en computación (no necesariamente en Google) son los Grafos de Ramanujan
- Los grafos de Ramanujan son un ejemplo de grafos de expansión, que se usan para obtener topologías de redes con propiedades especiales
- También se usan para construir sorting networks (AKS) capaces de ordenar n entradas en tiempo $\log(n)$
- Los Grafos de Ramanujan satisfacen la Hipótesis de Riemann para la función zeta de Ihara

http://en.wikipedia.org/wiki/Ihara_zeta_function



Preguntas



Más enlaces

Publicaciones de Googlers

<http://research.google.com/pubs/papers.html>

Curso sobre el grafo de la Web

<http://www.math.ryerson.ca/~abonato/webgraph.html>



Gracias!