



Information Retrieval II

David Hawking

30 Sep 2010

Machine Learning Summer School, ANU

Session Outline

- ▶ Ranking documents in response to a query
- ▶ Measuring the quality of such rankings
- ▶ Case Study: Tuning 40 parameters at the London School of Economics
- ▶ *Coffee Break*
- ▶ Web SearchEngineering
- ▶ Field Work: how do Web search engines really work?
- ▶ *Stretch Break*
- ▶ Discussion: Other IR problems for machine learning
- ▶ Historical context

The Web Search Scene (Sep 10)

1. Google
2. Yahoo!
3. BING - Bing Is Not Google

-
- ▶ These have the vast majority of the market in the USA but they're not alone.

Baidu - leader in China

 [新闻](#) [网页](#) [贴吧](#) [知道](#) [MP3](#) [图片](#) [视频](#) [词典](#)

[帮](#)

把百度设为首页 百度一下，找到相关网页约1,350,000篇，用时0.095秒

提示：您可以在支持英汉互译的百度词典中查看olympic games的解释

[2008 Beijing Olympic Games video PCauto](#)
Olympic Games 张怡宁夺得奥运百金 2004 Athens Olympic Games 陈艳青举起第七金 2004 Athens Olympic Games 刘翔平... Olympic Games 张宁夺雅典奥运羽毛球首金 2004 Athens Olympic Games 女柔洗东妹为国揽金 2004 Athens Olympic Games...
www1.pcauto.com.cn/zt/beijing-olympic/ 80K 2008-9-6 - [百度快照](#)

[国际奥委会](#)
国际奥委会官方网站。
www.olympic.org/ 1K 2008-9-3 - [百度快照](#)

[残奥会 - 北京2008年残奥会官方网站](#)
北京2008奥运会官方网站,提供奥组委介绍、新闻、奥运视听及相关活动。
www.beijing2008.cn/ 25K 2008-9-6 - [百度快照](#)
www.beijing2008.cn/ 上的更多结果

[Olympic Games News Post Topic 爆强沙画：全球华人大团结迎2008...](#)
爆强沙画:全球华人大团结迎2008北京奥运(Beijing 2008 Olympic Games) Published in 2008 Beijing Olympic Games...Games, Beijing Olympic Games, Olympic Games 爆强沙画:全球华人大团结迎2008北京奥运(Beijing 2008 Olympic ...
olympicgamesnews.org/ 鏊增己埋槽散锈氢吏闾 ... 95K 2008-6-11 - [百度快照](#)
olympicgamesnews.org/ 上的更多结果

[Art Hotel Olympic优惠预订](#)
在线Art Hotel Olympic预订,上Booking网! Booking提供50000家从...
www.booking.com

[pi pant expert Olympic ..](#)
Zibo Olympic Fashion Co., Ltd. Can devel op, design, produce, man...
www.alshizhuang.com

[找olympic games在阿里巴巴](#)
[找本地olympic games上百姓网](#)
[访问通用网址olympic games](#)
[更多olympic games在慧聪](#),快查看
[看olympic games小说在起点中文网](#)

[发布/查看关于奥运会的留言 370694 篇](#)

[发布/查看关于games的留言 166 篇](#)

Naver - leader in Korea

The screenshot displays the Naver homepage interface. At the top, there are navigation links for '메일 카페 블로그 지식IN 쇼핑' and the Naver logo with the slogan '독도를 지켜요!'. Below this is a green search bar with the text '4800만의 문화생활! 검색블로그 네이버 OpenAPI' and a search button. The search bar also includes a dropdown menu for '통합검색' and a search history section showing '지식: 유니언책이 국기에 포함된 나라 | Windows XP 종료 쉽게 하는 법' and '실시간 급상승 검색어: 이진욱'.

On the left side, there is a login section with fields for '아이디' and '비밀번호', and buttons for '로그인' and '회원가입'. Below this is a 'HOT' section with links for 'PC그린', '교통정보', and '모바일'. Further down is a '오늘의 뮤직' section featuring '뮤지션's 초미스 김창완 (산울림)'.

The main content area features a '뉴스홈' section for '9. 8. [월]' with a '패럴림픽 스포츠 연예 날씨 편성표' and a '+설정' button. Below this are news headlines such as '美 금융구제책 발표...코스피 4% 폭등 연중 최대·환율 급락' and '1인 가구'가 한국을 바꾼다 | "성격에 따라 좋아하는 음악 다르다". A featured article includes a photo of two people and the headline '美 "빅2 모기지업체" 역사적 구제금융...영향은?'.


On the right side, there is a '날씨' section with links for '달력', '시계', and '문자'. Below this is a '소핑' section with links for '핫세일', '뜨는소핑', and '럭키투데이'.

Yandex - leader in Russia

[Сделать Яндекс стартовой страницей](#) [Помощь](#) [Настройка](#) [Войти в почту](#)

Новости 03:41

1. Райкконен сложил [с себя чемпионские полномочия](#)
2. Открылась [станция метро «Славянский бульвар»](#)
3. Фильм Алексея Германа-младшего «Бумажный солдат» [получил награды](#)
4. Саркози и Медведев обсудят [урегулирование ситуации на Кавказе](#)
5. На Кубу обрушился [новый ураган «Айк»](#)


 **Я.Онлайн**
А ты?

Яндекс
Найдётся всё

Например, [три закона робототехники](#) [расширенный поиск](#)

Везде [Новости](#) [Маркет](#) [Карты](#) [Словари](#) [Блоги](#) [Картинки](#) [Все службы...](#)

Почта
логин
пароль
 запомнить меня


[Забыли пароль?](#)
[Завести почтовый ящик](#)



Каталог сайтов
[Развлечения](#)
[Спорт](#) и [отдых](#)
[Работа](#) и [учеба](#)
[Технологии](#)
[Бизнес](#)

Маркет
[купить фотоаппарат](#)

Авто
поиск по объявлениям

Мой Круг
[новые вакансии](#)

Австралия. 8 сентября, понедельник
Погода: Канберра  **+5** днем
Телепрограмма



Google revenue: US\$23.65B in 2009!

Let's get a share of all this
money!

Let's build a Search Engine.

*** IMAGE ***

My design parameters for new Web SE

- ▶ 500 million queries/day - say peak of 10000 per second.
- ▶ Average page revisit frequency - once per month
- ▶ 50 billion Web pages
 - ▶ Average page size is 25KB (2008)
 - ▶ Average URL length is approx 100 characters (5 terabytes!)
- ▶ 1.25 petabyte of data (Plus, at least 100% network overhead)
 - ▶ 2.5 million gigabytes of network traffic per crawl.
- ▶ hundreds of millions of distinct words
- ▶ 25 trillion word occurrences

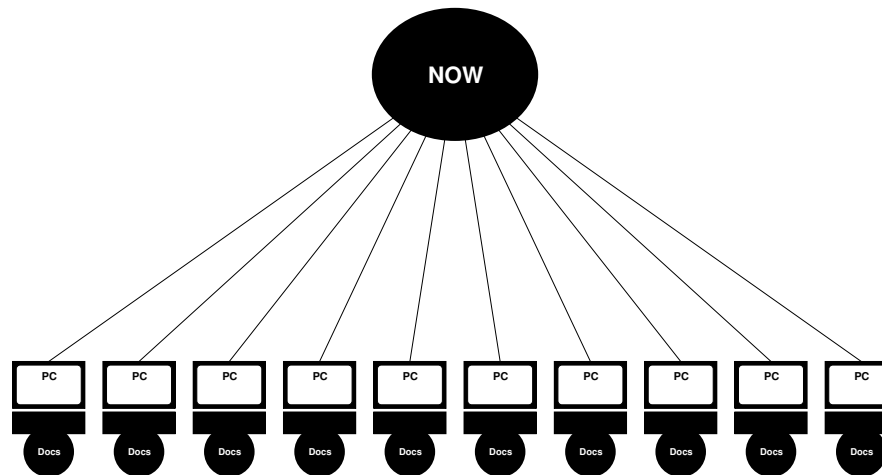
*** IMAGE ***

Design issues for the spider

- ▶ Data bandwidth = 9.6 gigabits/sec.
 - ▶ better not focus the spider on one server (or one country)!
 - ▶ need a huge degree of parallelism
 - ▶ a major load balancing problem
- ▶ Cache needs 5 terabytes to store just URLs!
 - ▶ need very efficient lookup methods

Parallel approach to Indexing and Query Processing

- ▶ Divide the data across a cluster of PCs
- ▶ Node = 2.8 GHz i7 PC, 8 gB RAM, 2 x 1 tB disk,
 - ▶ \$2k per year to lease and operate
- ▶ Replicate clusters and use load-balancing network front-end



Query processing

- ▶ Each node is responsible for 50 million documents (1 tB)
 - ▶ Therefore need 1000 nodes to handle full collection,
- ▶ If each node can process 10 queries/sec
 - ▶ Need 10000 queries/sec, therefore 1000 clusters.
 - ▶ Need $1000 \times 1000 = 1,000,000$ PCs
 - ▶ \$2.0B/year lease cost.

- ▶ Say query result page is 25kB long
 - ▶ 10000 queries/sec - 2.5 gigabits/sec

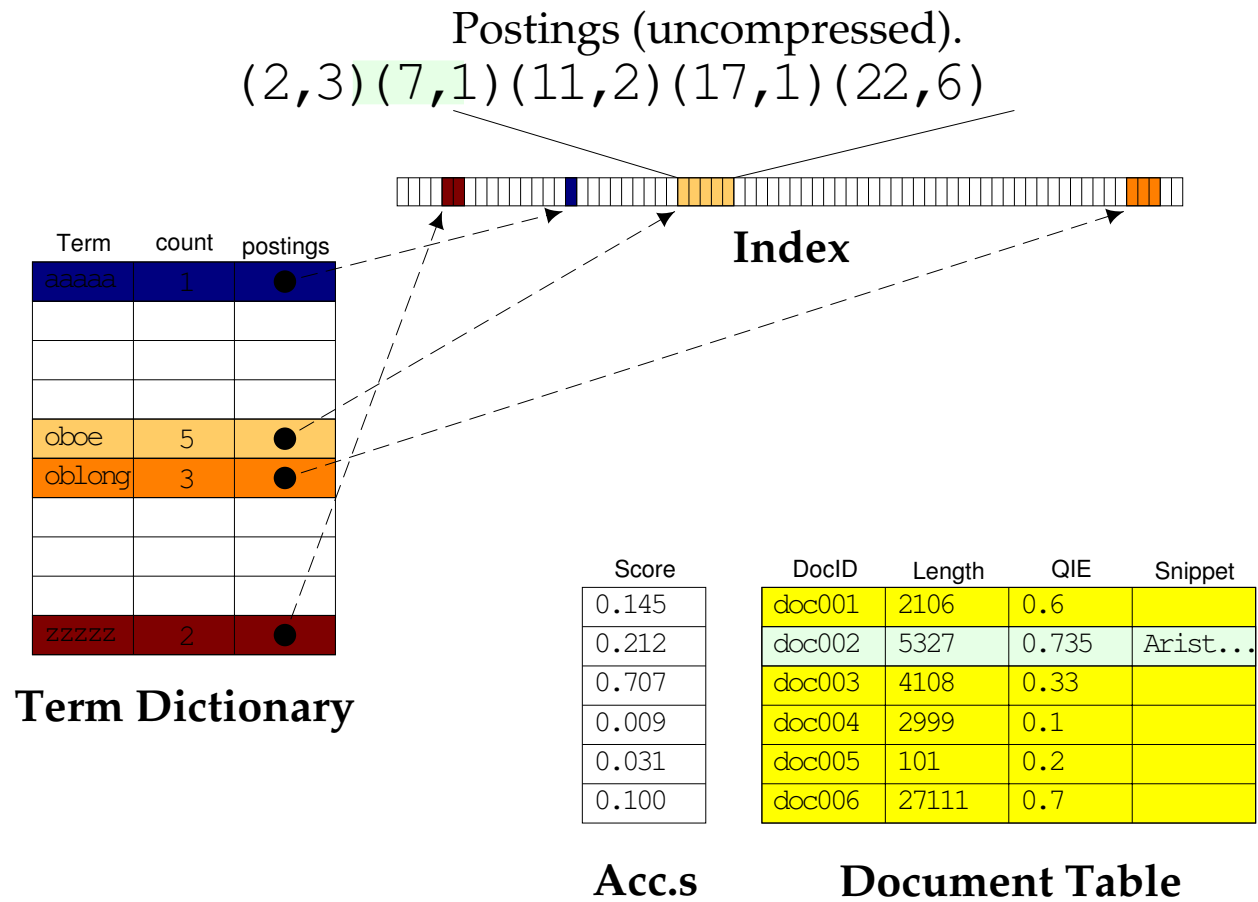
*** IMAGE ***

(Thanks to Wei-Ying Ma (MSRA) and Mark Sanderson, Sheffield University)

Electricity

- ▶ Say each node consumes 250Watts of electricity.
- ▶ Total power draw for 1,001,000 machines = 250 megaWatts
- ▶ Plus airconditioning at 1:3 = 83 megaWatts
- ▶ Total power = **333 megaWatts**

Index structure



Data structure sizes (per node)

- ▶ Term Dictionary – each entry: 12 byte string, 4 byte count, 4 byte offset
 - ▶ total size: $10M \times 20 = 200 \text{ MB}$
- ▶ Doc Table – each entry: 50 byte string (name), 4 byte length, 4 byte score, 100 byte string (snippet)
 - ▶ total size: $50M \times 158 = 8.0 \text{ gB}$
- ▶ Inverted File – each entry: 4-byte docno, 4-byte count
 - ▶ total size: $6.5B \times 8 = 53 \text{ gB}$
- ▶ Compressed text for document summaries
 - ▶ 100GB?
- ▶ It fits easily.
- ▶ What about link, anchor text and user behaviour indexes?
- ▶ UK-2007 (115M pages) has 4.1 billion links!

Indexing - Use one cluster

- ▶ Index 50M pages in about 2.5 days.
- ▶ Dirt cheap

Two ways of query processing: TAAT and DAAT

Query: **lpg vehicle subsidy**

1. [**lpg**; 27,000] - (2,7) (2,9) (2,100) (173,5) (2005, 19)
(2005,178) (9999, 1) ...
2. [**vehicle**; 112,000] - (13,31) (18,25) (173,6) (5006,88) (9999, 18)
3. [**subsidy**; 11,000] - (99,108) (173,7) (173,99) (7798,13)
(9999,205) ...

-
- ▶ TAAT (Term at a time)
 - ▶ Process all postings for lpg, then all for vehicle, then subsidy.
 - ▶ Non-zero scores for every doc containing one or more of the terms
 - ▶ DAAT (Document at a time)
 - ▶ Scan three postings lists in 'parallel'
 - ▶ Only count matches which meet a criterion. e.g. AND or WAND (Broder et al, CIKM, 2003)

(Use skipping)

DAAT Ranking

- ▶ Fewer results to sort (full matches only)
 - ▶ Can stop when we have required number of matches (or so)
- ▶ Assign document numbers in order of descending static score.
 - ▶ Compression techniques relying on doc-number diffs still work
 - ▶ Ensure that unscanned matches are not likely to be good.
 - ▶ Use limited document accumulators (better memory residency)
- ▶ Can apply local ranking heuristics to candidate set.
 - ▶ Terms in proximity, terms in headings etc.

ClueWeb09 Demo

Cutting costs

- ▶ Answer Caching
 - ▶ Pre-compute answers to popular queries (manually if nec.)
 - ▶ Top X queries account for 30% of query load
 - ▶ One PC can serve answers for all X queries.
 - ▶ Save 300 clusters, \$600M per year!
- ▶ Process Queries on Partial Data, eg. document titles, anchor text.
- ▶ Shorten results page
- ▶ Make spidering adaptive, continuous
- ▶ Avoid crawling spam and low value content
- ▶ Make query processing more efficient
 - ▶ more data per node
 - ▶ less time per query
 - ▶ (but 10 queries/sec is already good going)

General Efficiency Observations

- ▶ Two types of efficiency technique:
 - ▶ A. No loss of effectiveness
 - ▶ fast algorithms
 - ▶ B. Trade-off efficiency/effectiveness
 - ▶ use heuristics to reduce work
- ▶ Memory management is critical – Can reduce memory demand by
 - ▶ careful design/layout of data structures
 - ▶ compression
- ▶ Disk accesses very slow = 5M instructions
 - ▶ Avoid as much as possible
- ▶ Disk transfers negligible except for very common terms
 - ▶ compression pays off in reduced transfer costs

Hardware Issues

1. Choose cheap, low-heat-output, reliable servers
2. Build in fault tolerance. Can't allow a whole cluster to be down because one node has failed!
3. Efficient fault diagnosis and repair procedures.

Speeding up query processing

1. Use a limited number of score accumulators
2. Process term lists in parallel, not sequentially
3. Assign document numbers according to descending "value"
4. Stop early
5. Seek perfect load balance across clusters. The cluster is as slow as the slowest node.

Result Presentation

- ▶ For each top-ranked doc, print DOCID, snippet etc.
- ▶ How to generate summaries
 - ▶ Canned summaries - query independent
 - ▶ Query biased summaries - quite a computational challenge

Other Issues

- ▶ Privacy
 - ▶ David Hawking
 - ▶ 687 897 255
 - ▶ 02 6161 7777
 - ▶ 5163 9999 9999 9999
 - ▶ 85 Smith St, Nowhere 2586
- ▶ Masses of query and click data – what can we use it for?
 - ▶ Search quality tuning – machine learning
 - ▶ Ranking evidence
 - ▶ Spelling corrections.
 - ▶ Query suggestions
- ▶ Legal issues with advertisers, copyright owners, ...

Doing More with Queries

[Web](#) [Images](#) [Videos](#) [Maps](#) [News](#) [Shopping](#) [Gmail](#) [more](#) ▼



kathy griffiths

Search

About 4,300,000 results (0.27 seconds)

[Advanced search](#)

Everything

Images

Videos

More

All results

[Related searches](#)

[Sites with images](#)



[More search tools](#)

[Kathy Griffin](#)

Well, Kathy's specials are famous for their names and we are happy to announce this one is **Kathy Griffin** Does the Bible Belt. But we've seen last minute ...

[Tour dates](#) - [TV Schedule](#) - [Pics](#) - [Contacts](#)

www.kathygriffin.net/ - [Cached](#) - [Similar](#)

[Tour info - Kathy Griffin - The Official Site of Kathy Griffin](#)

Catch double Emmy award winning **Kathy Griffin** live in her true element. She has more stories, ... **kathy griffin** tickets Tennessee Theatre in Knoxville, TN ...

www.kathygriffin.net/tour.php - [Cached](#) - [Similar](#)

[Kathy Griffin - Wikipedia, the free encyclopedia](#)

Kathleen Mary "**Kathy**" **Griffin** (born November 4, 1960) is an American actress, stand-up comedian, media personality and LGBT rights advocate. ...

[Early life](#) - [Career](#) - [Personal life](#) - [Filmography](#)

en.wikipedia.org/wiki/Kathy_Griffin - [6 hours ago](#) - [Cached](#) - [Similar](#)

- Maintaining Quality - Are we really Better?

Side-by-side comparisons

ipod **GC Generalised Comparison** I prefer: because

<http://search.live.com/results.aspx?q=ipod> <http://www.google.com/search?q=ipod>

Live Search | MSN | Windows Live

Live Search ipod




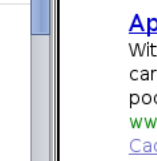
Web 1-10 of 16,400,000 results · [Advanced](#)
See also: [Images](#), [Video](#), [News](#), [Maps](#), [More](#) ▾

Ipod - www.ebay.com Live Search cashback
Buy **Ipod**. You may get 8% off with PayPal if eligible.

[Official iPod® Store](http://www.apple.com/store) - www.apple.com/store
Buy an Apple **iPod** today with Free shipping and engraving.

[Powerbuy Electronics](http://www.powerbuyelectronics.com) - www.powerbuyelectronics.com
iPod Nano and Classic 160GB starting at 109.00! Low Prices on all

➤ [Research and shop for ipod](#) Helpful? [Yes](#) | [No](#)

			
Apple iPod nano digital...	Apple iPod digital player...	Apple iPod digital player...	Apple iPod digital player...

[MP3 Player Buying Guide](#) HowStuffWorks.com
[MP3 player buying guide](#) CNET.com
[Buying Guide: Flash-Based MP3 Players](#) PCMag.com

[Apple - iPod + iTunes](#)

Web [Images](#) [Maps](#) [News](#) [Shopping](#) [Gmail](#) [more](#) ▾ [Sign in](#)

ipod

Web [News](#) [Video](#) Results 1 - 10 of about **383,000,000** for ipod

Official Apple® Store Sponsored Link
store.apple.com/australia Buy direct in Australia from Apple. Free shipping on orders over A\$90.

Apple - iPod + iTunes
Learn about **iPod**, Apple TV, and accessories. Download iTunes software free and purchase iTunes Gift Cards. Check out the most popular TV shows, movies, ...
www.apple.com/itunes/ - 38k - [Cached](#) - [Similar pages](#)

Apple - iPod classic
With up to 160GB of storage, **iPod** classic lets you carry your entire music and video collection in your pocket: up to 40000 songs or up to 200 hours of ...
www.apple.com/ipodclassic/ - 17k - [Cached](#) - [Similar pages](#)
[More results from www.apple.com »](#)

iPod - Wikipedia, the free encyclopedia
iPod is a popular brand of portable media players designed and marketed by Apple Inc. and launched on 23 October 2001. As of 2008, the current product ...
en.wikipedia.org/wiki/iPod - 181k - [Cached](#) - [Similar pages](#)

http://search.live.com/ results.aspx?q=sankaranarayana

Live Search | MSN | Windows Live

Live Search

Web 1-10 of 6,160 results · [Advanced](#)
 See also: [Images](#), [Video](#), [News](#), [Maps](#), [More](#) ▾

Valangiman Sankaranarayana Srinivasa Sastri - Wiki
 ...
 Valangiman Sankarana-rayana Srinivasa Sastri CH (1869 – 1946) was an Indian statesman and founder of the Indian Liberal Federation. Sastri, of a Brahmin background, started his career as a school teacher. In 1915 he became the Secretary of the Servants of India Society. He served as a member of the Madras Legislative Council, the central legislature and the council of state established under the Government of India Act 1919.
en.wikipedia.org/wiki/Valangiman_Sankaranarayana_Srinivasa_Sastri
[page](#)

Srinivasa Sastri - Wikipedia, the free encyclopedia
 Rt. Hon'ble V. S. Srinivasa Sastri was an educator, statesman, diplomat and ambassador from India. He was especially renowned for his expertise in the command of the English language and remarkable oratorical skills. He was born on September 22, 1869. Born in the village of Valangaiman, near Madurai, to the poor Vaidik **Sankaranarayana** Sastri, Sastri had a brilliant education at the Native High School, Kumbakonam, and then in the government school in the same town.
en.wikipedia.org/wiki/Srinivasa_Sastri · [Cached page](#)
 Show more results from en.wikipedia.org

The Mathematics Genealogy Project - Ramesh Sankaranarayana
 Advisor: Lorna Stewart. No students known. If you have additional information or corrections regarding this mathematician, please use the update form at www.genealogy.math.ndsu.nodak.edu/id.php?id=72434 · [Cached page](#)

http://www.google.com/ search?q=sankaranarayana

Web [Images](#) [Maps](#) [News](#) [Shopping](#) [Gmail](#) [more](#) ▾

[Sign in](#)

Google

Web [Books](#) Results 1 - 10 of about 39,500 for **sankaranarayana**

Ramesh Sankaranarayana's Home Page
 Ramesh **Sankaranarayana**. Welcome to my home page. I am a Lecturer in the Department of Computer Science at the Australian National University
cs.anu.edu.au/~ramesh/ - 2k - [Cached](#) - [Similar pages](#)


DBLP: Ramesh S. Sankaranarayana
 Ramesh S. **Sankaranarayana**. List of publications from the DBLP Bibliography Server - ... 3, Ramesh S. **Sankaranarayana**: A generalization of Favaron's theorem. ...
www.informatik.uni-trier.de/~ley/db/indices/a-tree/s/Sankaranarayana:Ramesh_S.html - 7k - [Cached](#) - [Similar pages](#)

Sankara Narayana - Wikipedia, the free encyclopedia
 [edit] References. O'Connor, John J. & Robertson, Edmund F., "**Sankara Narayana**", MacTutor History of Mathematics archive ...
en.wikipedia.org/wiki/Sankara_Narayana - 18k - [Cached](#) - [Similar pages](#)

Sankara summary
Sankara Narayana (about 840-about 900) ... **Sankara Narayana** was an Indian astronomer and mathematician. He wrote a commentary on the work of Bhaskara I. ...

http://search.live.com/ results.aspx?q=bank

Live Search | MSN | Windows Live

 Live Search

Web 1-10 of 293,000,000 results · [Advanced](#)
See also: [Images](#), [Video](#), [News](#), [Maps](#), [More](#) ▼

[Rates Above 3.50%](#) - [www.OnlineSavingsAccount.org](#)
Compare Rates from Several **Banks** in Minutes


[Bank - Wikipedia, the free encyclopedia](#)
A banker or **bank** is a financial institution whose primary activity is to act as a payment agent for customers to borrow and lend. The first modern bank was founded in Italy in Genoa in 1406, its name was Banco di San Giorgio (St. George). Many other financial activities were added over time.
[Traditional banking ...](#) · [Definition](#) · [Accounting for bank ...](#)
[en.wikipedia.org/wiki/Bank](#) · [Cached page](#)

[Bank of Greece - Wikipedia, the free encyclopedia](#)
The **Bank of Greece** (Greek: "Τράπεζα της Ελλάδος") is the national bank of Greece, located in Athens (Venizelos Avenue) and founded in 1928. Its operations started officially in 1928.
[Introduction](#) · [Governors](#) · [Deputy Governors](#)
[en.wikipedia.org/wiki/Bank_of_Greece](#) · [Cached page](#)
[Show more results from en.wikipedia.org](#)

[The Finest Men's Clothing at JoS. A. Bank Clothiers -](#)
...
Since 1905, the expert in men's suits, dress shirts, neckties, shoes, sportcoats, trousers, sweaters, sportshirts, pants, underwear, formal polos
[www.josbank.com](#) · [Cached page](#)

http://www.google.com/ search?q=bank

Web [Images](#) [Maps](#) [News](#) [Shopping](#) [Gmail](#) [more](#) ▼ [Sign in](#)



Web [News](#) Results 1 - 10 of about 658,000,000 for **bank**

Sponsored Link
[St. George Bank Account](#)

Sponsored Links

[Online Banking 24/7 with AMP eASYCash Management](#), No min balance. Apply online now
[www.amp.com.au/easycash](#)

[BankWest Savings Account](#)
bankwest tddirect. Up to 8.1%p.a. Fixed rate term deposit. No fees!
[www.bankwest.com.au/savings](#)

[ANZ Online Saver Account](#)
Start saving with our high interest savings account. Apply today!
[ANZ.com/OnlineSaver](#)

[Citibank Plus Account](#)
A great every day account for only a low \$5 a month. Apply online Now
[www.Citibank.com.au/Plus](#)

[Savings Account](#)
7.80% 1 yr, No Monthly Fees Or Min Balance. Apply For Suncorp eOptions
[www.Suncorp.com.au](#)

Web [Images](#) [Maps](#) [News](#) [Shopping](#) [Gmail](#) [more](#) [Sign in](#)

Google enterprise search

Web [News](#) [Blogs](#) Results 1 - 10 of about 42,300,000 for [enterprise search](#)

Enterprise Search from Microsoft
Learn more about extending your Microsoft **enterprise search** solution to connect ... This week, Microsoft made its latest move into the **enterprise search** ...

Sponsored Links

Coveo Enterprise Search
Download the full version today
30 day all inclusive free trial!
www.coveo.com

Enterprise Search
Thinking about **enterprise search**?
Read what **search** experts say first.
SearchDoneRight.com

Google Search Mini
One Box allows specific **search**
Search upto 300,000 documents
Google.Devnet.com.au


HEDLOC & Universal Search
Aust Leaders in **Enterprise Search**
Google **Search** Appliance & Mini
www.hedloc.com.au

Enterprise Search
Top Ranked Software Solutions.
Find and Compare Software.
www.capterra.com

Enterprise Search
Add hosted or server based **search**
to your public website and intranet
www.funnelback.com

Yahoo! My Yahoo! Mail

Web [Images](#) [Video](#) [Local](#) [Shopping](#) [more](#)

enterprise search 

1 - 10 of 1,090,000,000 for [enterprise search](#) (About)

Also try: [enterprise search engine](#), [enterprise search sur](#)

Affordable Enterprise Search Software
Bring the time saving value of **search** software to your organization. ...
www.isys-search.com

Free Enterprise Search from IBM and Yahoo!
[Yahoo! Shortcut](#) - [About](#)

Enterprise search - Wikipedia, the free encyclopedia
Enterprise search is the practice of identifying and enabling specific content
Enterprise Search Suite ...
en.wikipedia.org/wiki/Enterprise_search - [Cached](#)

Enterprise Search from Microsoft
Enterprise Search from Microsoft. Empower people to find information. FF
within your business, they need **enterprise search**. ...
www.microsoft.com/enterprisesearch - [Cached](#)

Enterprise Search
Enterprise Search solutions from Microsoft are a collection of related
Deploying and Supporting **Enterprise Search** with SharePoint Serv
www.microsoft.com/sharepoint/capabilities/es/overview.aspx - [Cached](#)

Enterprise Rent-A-Car: Car Rental Locations in US, Canada
Find car rental locations throughout the US, Canada, UK, Ireland and Germ
Rent-A-Car. **Search** over 6,000 car rental locations to start your car ...
www.enterprise.com/car_rental/worldLocationSearch.do - [Cached](#)

Enterprise Rent-A-Car - Rental Cars at Low Rates

Summary of this Section

- ▶ The Web search engine business model is very risky
- ▶ Starting a new Web search engine requires:
 - ▶ Lots of hardware (intelligently chosen and operated.)
 - ▶ Huge capital
 - ▶ Very large bandwidth network connections
 - ▶ Clever engineering (CE)
 - ▶ save huge amounts of network traffic during spidering
 - ▶ save vast amounts of query processing hardware
- ▶ CE during QP relies on
 - ▶ caching / pre-computation
 - ▶ clever parallelism
 - ▶ memory residency / avoidance of disk accesses
 - ▶ smart algorithms

BTW - What's the role of a Chief Scientist?

