



Introduction to the Use of

# Link Analysis

by

**Google**

Amy Langville

Carl Meyer

Department of Mathematics  
North Carolina State University  
Raleigh, NC

SIAM Annual Meeting 7/15/04



# Outline

- Introduction to Information Retrieval (IR)
- Link Analysis
- PageRank Algorithm



# Short History of IR

IR = search within doc. coll. for particular info. need (query)

B. C.	cave paintings
7-8th cent. A.D.	Beowulf
12th cent. A.D.	invention of paper, monks in scriptoria
1450	Gutenberg's printing press
1700s	Franklin's public libraries
1872	Dewey's decimal system
1900s	Card catalog
1940s-1950s	Computer
1960s	Salton's SMART system
1989	Berner-Lee's WWW

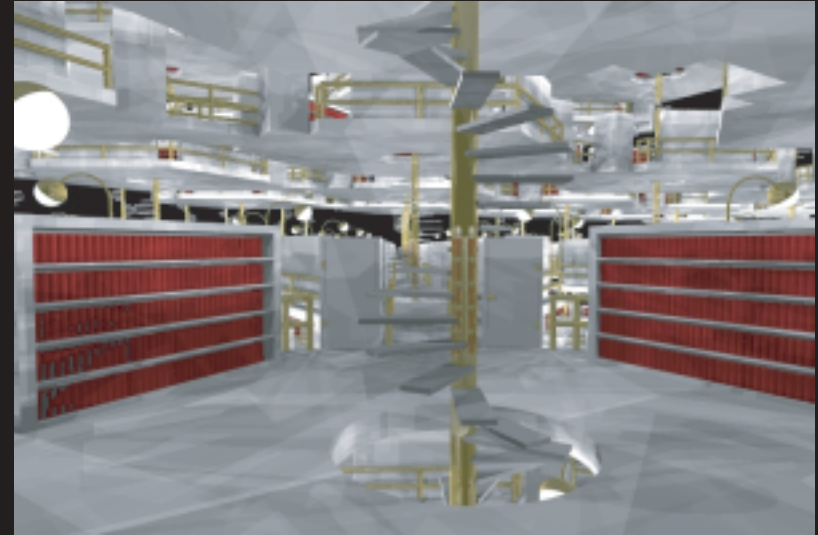
# the pre-1998 Web

## Yahoo

- hierarchies of sites
- organized by humans

## Best Search Techniques

- word of mouth
- expert advice



## Overall Feeling of Users

- Jorge Luis Borges' 1941 short story, *The Library of Babel*

When it was proclaimed that the Library contained all books, the first impression was one of extravagant happiness. All men felt themselves to be the masters of an intact and secret treasure. There was no personal or world problem whose eloquent solution did not exist in some hexagon.

... As was natural, this inordinate hope was followed by an excessive depression. The certitude that some shelf in some hexagon held precious books and that these precious books were inaccessible, seemed almost intolerable.





# 1998 ... enter Link Analysis

## Change in User Attitudes about Web Search

### Today

- “It’s not my homepage, but it might as well be. I use it to ego-surf. I use it to read the news. Anytime I want to find out anything, I use it.” - **Matt Groening, creator and executive producer, The Simpsons**
- “I can’t imagine life without Google News. Thousands of sources from around the world ensure anyone with an Internet connection can stay informed. The diversity of viewpoints available is staggering.” - **Michael Powell, chair, Federal Communications Commission**
- “Google is my rapid-response research assistant. On the run-up to a deadline, I may use it to check the spelling of a foreign name, to acquire an image of a particular piece of military hardware, to find the exact quote of a public figure, check a stat, translate a phrase, or research the background of a particular corporation. It’s the Swiss Army knife of information retrieval.” - **Garry Trudeau, cartoonist and creator, Doonesbury**



# Web Information Retrieval

IR before the Web = traditional IR

IR on the Web = **web IR**



# Web Information Retrieval

IR before the Web = traditional IR

IR on the Web = **web IR**

How is the Web different from other document collections?



# Web Information Retrieval

IR before the Web = traditional IR

IR on the Web = **web IR**

**How is the Web different from other document collections?**

- It's huge.
  - over 10 billion pages, average page size of 500KB
  - 20 times size of Library of Congress print collection
  - Deep Web - 550 billion pages



# Web Information Retrieval

IR before the Web = traditional IR

IR on the Web = **web IR**

## How is the Web different from other document collections?

- It's huge.
  - over 10 billion pages, average page size of 500KB
  - 20 times size of Library of Congress print collection
  - Deep Web - 550 billion pages
- It's dynamic.
  - content changes: 40% of pages change in a week, 23% of .com change daily
  - size changes: billions of pages added each year



# Web Information Retrieval

IR before the Web = traditional IR

IR on the Web = **web IR**

## How is the Web different from other document collections?

- It's huge.
  - over 10 billion pages, average page size of 500KB
  - 20 times size of Library of Congress print collection
  - Deep Web - 550 billion pages
- It's dynamic.
  - content changes: 40% of pages change in a week, 23% of .com change daily
  - size changes: billions of pages added each year
- It's self-organized.
  - no standards, review process, formats
  - errors, falsehoods, link rot, and spammers!



# Web Information Retrieval

IR before the Web = traditional IR

IR on the Web = **web IR**

## How is the Web different from other document collections?

- It's huge.
  - over 10 billion pages, average page size of 500KB
  - 20 times size of Library of Congress print collection
  - Deep Web - 550 billion pages
- It's dynamic.
  - content changes: 40% of pages change in a week, 23% of .com change daily
  - size changes: billions of pages added each year
- It's self-organized.
  - no standards, review process, formats
  - errors, falsehoods, link rot, and spammers!

**A Herculean Task!**

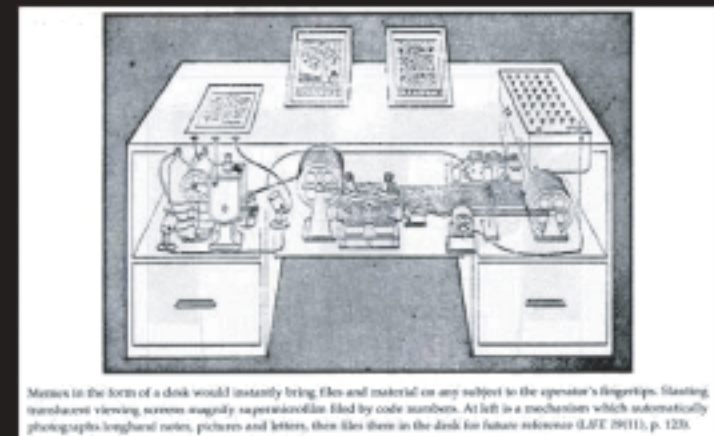
# Web Information Retrieval

IR before the Web = traditional IR

IR on the Web = **web IR**

## How is the Web different from other document collections?

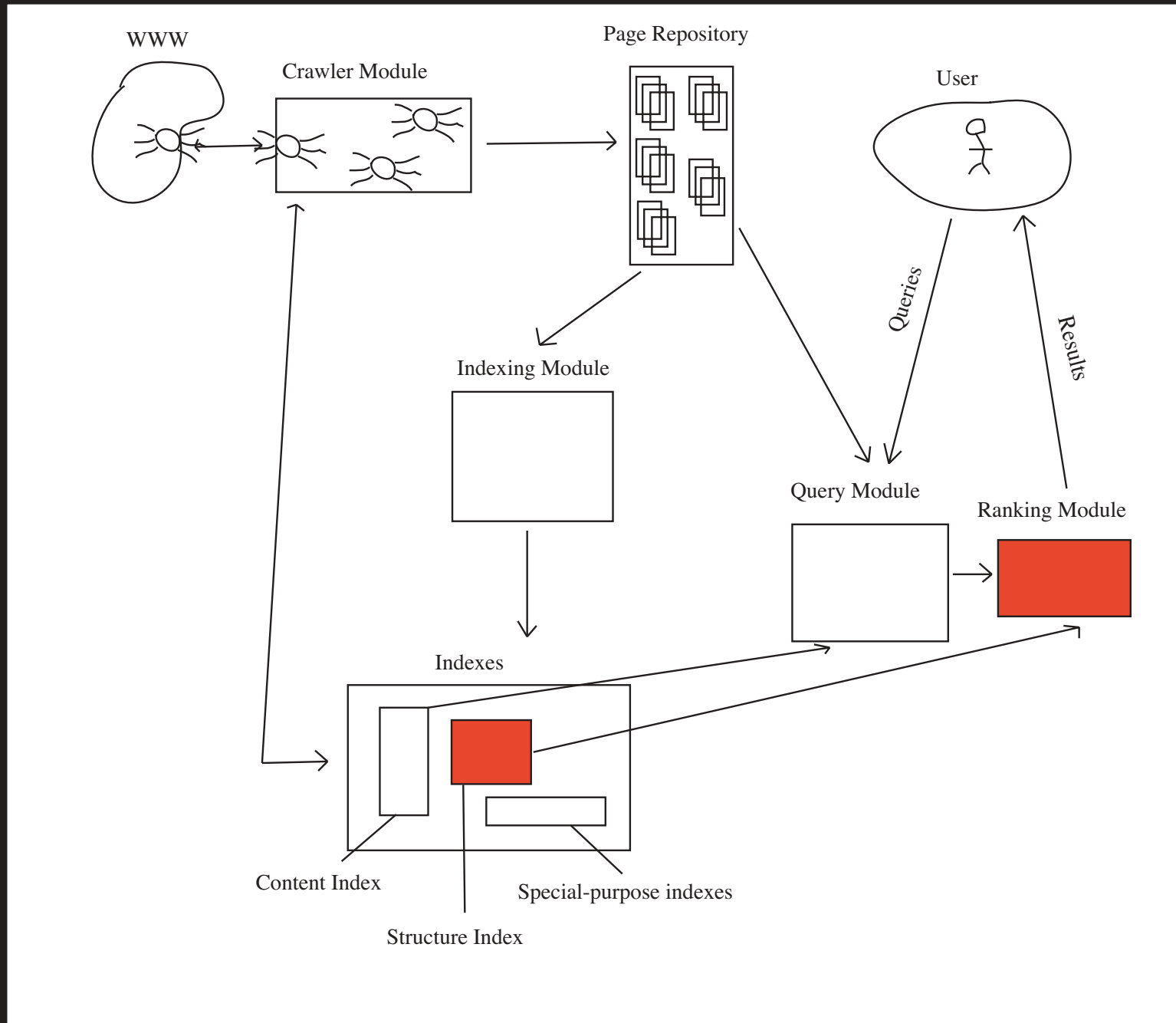
- It's huge.
  - over 10 billion pages, each about 500KB
  - 20 times size of Library of Congress print collection
  - Deep Web - 550 billion pages
- It's dynamic.
  - content changes: 40% of pages change in a week, 23% of .com change daily
  - size changes: billions of pages added each year
- It's self-organized.
  - no standards, review process, formats
  - errors, falsehoods, link rot, and spammers!
- Ah, but it's hyperlinked !
  - Vannevar Bush's 1945 memex







# Elements of a Web Search Engine





# Query Processing

**Step 1:** User enters query, i.e., aztec baby

**Step 2:** Inverted file consulted

- term 1 (aardvark) - 3, 117, 3961
- $\vdots$
- term 10 (aztec) - 3, 15, 19, 101, 673, 1199
- term 11 (baby) - 3, 31, 56, 94, 673, 909, 11114, 253791
- $\vdots$
- term  $m$  (zymurgy) - 1159223

**Step 3:** Relevant set identified, i.e. (3, 673)

**Simple traditional engines stop here.**



# Modification to Inverted File

- add more features to inverted file by appending vector to each page identifier, i.e., [in title?, in descrip.?, # of occurrences]
- Modified inverted file

- term 1 (aardvark) - 3 [0,0,3], 117 [1,1,10], 3961 [0,1,4]  
⋮
- term 10 (aztec) - 3 [1, 1, 27], 15 [0,0,1], 19 [1,1,21], 101 [0,1,7], 673 [0, 0, 3], 1199 [0,0,3]
- term 11 (baby) - 3 [1, 1, 10], 31 [0,0,2], 56 [0,1,3], 94 [1,1,11], 673 [1, 1, 14], 909 [0,0,2], 11114 [1,1,22], 253791 [0,1,6]  
⋮
- term  $m$  (zymurgy) - 1159223 [1,1,9]

- IR score computed for each page in relevant set.

$$\text{EX: IR score (page 3)} = (1 + 1 + 27) \times (1 + 1 + 10) = 348$$

$$\text{IR score (page 673)} = (0 + 0 + 3) \times (1 + 1 + 14) = 48$$

**Early web engines stop here.**

**Problem = Ranking by IR score is not good enough.**

# Link Analysis

- uses hyperlink structure to focus the relevant set
- combine IR score with popularity or importance score

PageRank - Brin and Page  $\Rightarrow$

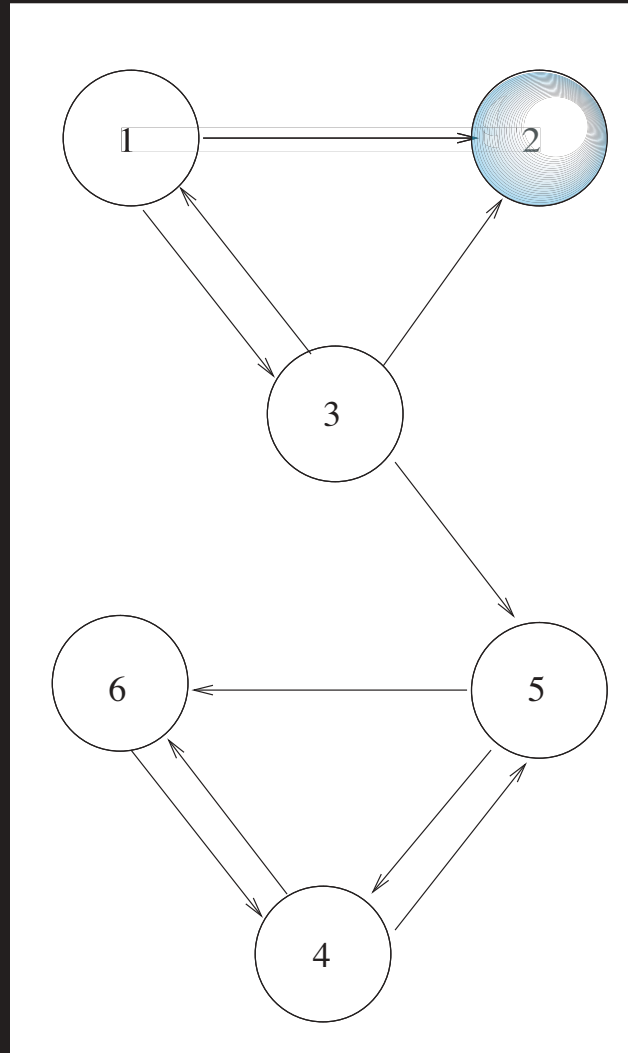


HITS - Kleinberg  $\Rightarrow$





# The Web as a Graph



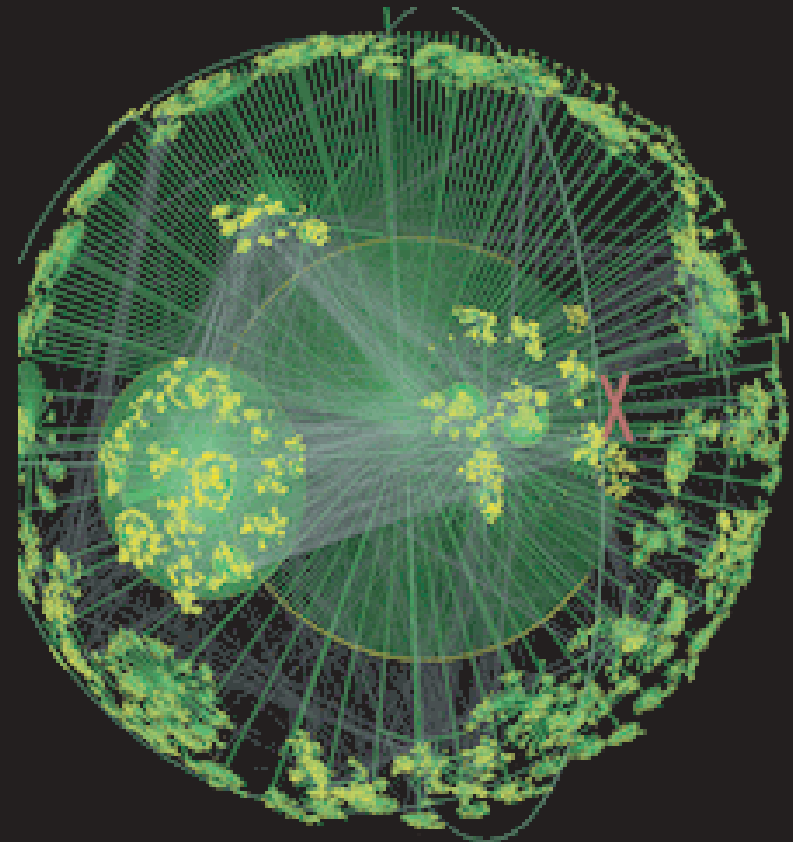
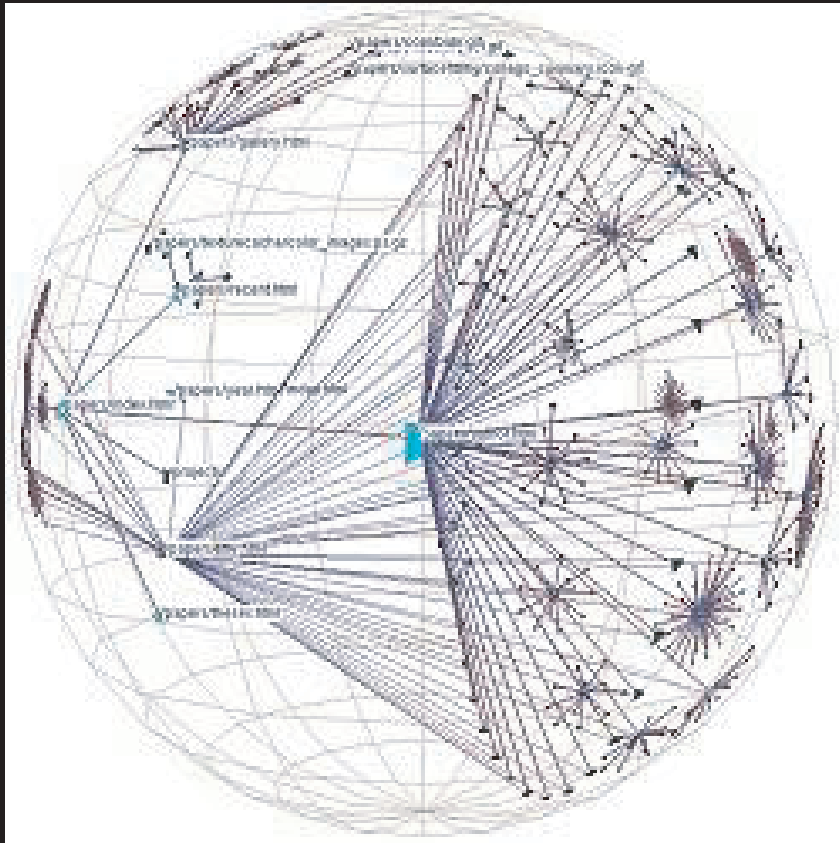
Nodes = webpages

Arcs = hyperlinks

# Web Graphs

## CSC and MATH problems here:

- store adjacency matrix
- update adjacency matrix
- visualize web graph
- locate clusters in graph





# How to Use Web Graph for Search

Hyperlink = Recommendation

- page with 20 recommendations (inlinks) must be more important than page with 2 inlinks.
- but status of recommender matters.  
EX: letters of recommendation: 1 letter from Trump vs. 20 from unknown people
- but what if recommender is generous with recommendations?  
EX: suppose Trump has written over 40,000 letters.
- each inlink should be weighted to account for status of recommender and # of outlinks from that recommender



# How to Use Web Graph for Search

Hyperlink = Recommendation

- page with 20 recommendations (inlinks) must be more important than page with 2 inlinks.
- but status of recommender matters.  
EX: letters of recommendation: 1 letter from Trump vs. 20 from unknown people
- but what if recommender is generous with recommendations?  
EX: suppose Trump has written over 40,000 letters.
- each inlink should be weighted to account for status of recommender and # of outlinks from that recommender

**PAGERANK** - importance/popularity score given to each page





## Our Search: Google Technology

[Home](#)

[All About Google](#)

[Help Central](#)

[Google Features](#)

[Services & Tools](#)

Our Technology

► [Why Use Google](#)  
[Benefits of Google](#)

*Find on this site:*

**Google searches more sites more quickly, delivering the most relevant results.**

### Introduction

Google runs on a unique combination of advanced hardware and software. The speed you experience can be attributed in part to the efficiency of our search algorithm and partly to the thousands of low cost PC's we've networked together to create a superfast search engine.

The heart of our software is PageRank™, a system for ranking web pages developed by our founders [Larry Page](#) and [Sergey Brin](#) at Stanford University. And while we have dozens of engineers working to improve every aspect of Google on a daily basis, PageRank continues to provide the basis for all of our web search tools.

### PageRank Explained

PageRank relies on the uniquely democratic nature of the web by using its



# PageRank

## The PageRank Idea

(Sergey Brin & Lawrence Page 1998)

- Ranking is preassigned (An off-line calculation)
- Your page  $P$  has some rank  $r(P)$
- Adjust  $r(P)$  higher or lower depending on ranks of pages that point to  $P$
- Importance is not just number, but *quality* of in-links
  - role of outlinks relegated
  - much less sensitive to spamming



# PageRank

## The Definition

- $r(P) = \sum_{P \in \mathcal{B}_P} \frac{r(P)}{|P|}$  —  $\mathcal{B}_P = \{\text{all pages pointing to } P\}$   
—  $|P| = \text{number of out links from } P$

## Successive Refinement

- Start with  $r_0(P_i) = 1/n$  for all pages  $P_1, P_2, \dots, P_n$
- Iteratively refine rankings for each page

$$\text{— } r_1(P_i) = \sum_{P \in \mathcal{B}_{P_i}} \frac{r_0(P)}{|P|}$$

$$\text{— } r_2(P_i) = \sum_{P \in \mathcal{B}_{P_i}} \frac{r_1(P)}{|P|}$$

⋮

$$\text{— } r_{j+1}(P_i) = \sum_{P \in \mathcal{B}_{P_i}} \frac{r_j(P)}{|P|}$$



# In Matrix Notation

After Step  $j$

$$\pi_j^T = [r_j(P_1), r_j(P_2), \dots, r_j(P_n)]$$

$$\pi_{j+1}^T = \pi_j^T \mathbf{P} \quad \text{where} \quad p_{ij} = \begin{cases} 1/|P_i| & \text{if } i \rightarrow j \\ 0 & \text{o.w.} \end{cases}$$

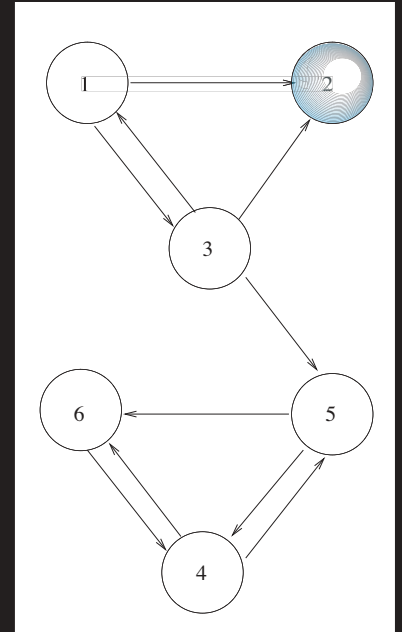
# In Matrix Notation

After Step  $j$

$$\pi_j^T = [r_j(P_1), r_j(P_2), \dots, r_j(P_n)]$$

$$\pi_{j+1}^T = \pi_j^T \mathbf{P} \quad \text{where} \quad p_{ij} = \begin{cases} 1/|P_i| & \text{if } i \rightarrow j \\ 0 & \text{o.w.} \end{cases}$$

$$\mathbf{P} = \begin{matrix} & \begin{matrix} p_1 & p_2 & p_3 & p_4 & p_5 & p_6 \end{matrix} \\ \begin{matrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \end{matrix} & \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \end{matrix}$$





# In Matrix Notation

After Step  $j$

$$\pi_j^T = [r_j(P_1), r_j(P_2), \dots, r_j(P_n)]$$

$$\pi_{j+1}^T = \pi_j^T \mathbf{P} \quad \text{where} \quad p_{ij} = \begin{cases} 1/|P_i| & \text{if } i \rightarrow j \\ 0 & \text{o.w.} \end{cases}$$

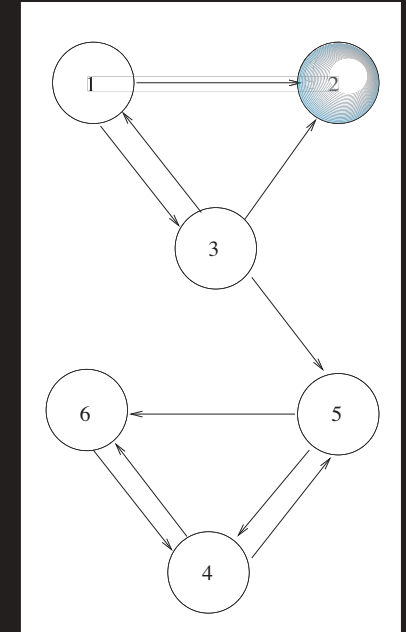
$$\mathbf{P} = \begin{matrix} & \begin{matrix} p_1 & p_2 & p_3 & p_4 & p_5 & p_6 \end{matrix} \\ \begin{matrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \end{matrix} & \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \end{matrix}$$

$$\text{PageRank} = \lim_{j \rightarrow \infty} \pi_j^T = \pi^T$$

(provided limit exists)

It's Almost a Markov Chain

$\mathbf{P}$  has row sums = 1 for ND nodes, row sums = 0 for D nodes





# In Matrix Notation

## It's Almost a Markov Chain

- $\mathbf{P}$  has row sums = 1 for ND nodes, row sums = 0 for D nodes



# In Matrix Notation

## It's Almost a Markov Chain

- $\mathbf{P}$  has row sums = 1 for ND nodes, row sums = 0 for D nodes

Stochasticity Fix:  $\bar{\mathbf{P}} = \mathbf{P} + \mathbf{a}\mathbf{v}^T$ . ( $a_i=1$  for  $i \in D$ , 0, o.w.)





# In Matrix Notation

## It's Almost a Markov Chain

- $\mathbf{P}$  has row sums = 1 for ND nodes, row sums = 0 for D nodes

Stochasticity Fix:  $\bar{\mathbf{P}} = \mathbf{P} + \mathbf{a}\mathbf{v}^T$ . ( $a_i=1$  for  $i \in D$ , 0, o.w.)

$$\bar{\mathbf{P}} = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \text{ where } \mathbf{a} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \mathbf{v}^T = 1/6 \mathbf{e}^T$$



# In Matrix Notation

## It's Almost a Markov Chain

- $\mathbf{P}$  has row sums = 1 for ND nodes, row sums = 0 for D nodes

Stochasticity Fix:  $\bar{\mathbf{P}} = \mathbf{P} + \mathbf{a}\mathbf{v}^T$ . ( $a_i=1$  for  $i \in D$ , 0, o.w.)

$$\bar{\mathbf{P}} = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \text{ where } \mathbf{a} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \mathbf{v}^T = 1/6 \mathbf{e}^T$$

- Each  $\pi_j^T$  is a probability distribution vector ( $\sum_i r_j(P_i) = 1$ )
- $\pi_{j+1}^T = \pi_j^T \bar{\mathbf{P}}$  is random walk on the graph defined by links
- $\pi^T = \lim_{j \rightarrow \infty} \pi_j^T$  = stationary probability distribution



# Random Surfer

## Web Surfer Randomly Clicks On Links

(Back button not a link)

Long-run proportion of time on page  $P_i$  is  $\pi_i$

## Problems



# Random Surfer

## Web Surfer Randomly Clicks On Links

(Back button not a link)

Long-run proportion of time on page  $P_i$  is  $\pi_i$

## Problems

Dead end page (nothing to click on)

( $\pi^T$  not well defined)

Could get trapped into a cycle ( $P_i \rightarrow P_j \rightarrow P_i$ ) (No convergence)



# Random Surfer

## Web Surfer Randomly Clicks On Links

(Back button not a link)

Long-run proportion of time on page  $P_i$  is  $\pi_i$

## Problems

Dead end page (nothing to click on)

( $\pi^T$  not well defined)

Could get trapped into a cycle  $(P_i \rightarrow P_j \rightarrow P_i)$  (No convergence)

## Convergence

Markov chain must be irreducible and aperiodic



# Random Surfer

## Web Surfer Randomly Clicks On Links

(Back button not a link)

Long-run proportion of time on page  $P_i$  is  $\pi_i$

## Problems

Dead end page (nothing to click on)

( $\pi^T$  not well defined)

Could get trapped into a cycle ( $P_i \rightarrow P_j \rightarrow P_i$ ) (No convergence)

## Convergence

Markov chain must be irreducible and aperiodic

DEFN: a chain is *irreducible* if every page is reachable from every other page.

DEFN: every *reducible* chain can be permuted to the form  $\begin{bmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{0} & \mathbf{Z} \end{bmatrix}$ .



# Random Surfer

## Bored Surfer Enters Random URL

Irreducibility Fix:  $\bar{\bar{\mathbf{P}}} = \alpha \bar{\mathbf{P}} + (1 - \alpha) \mathbf{E} \quad e_{ij} = 1/n \quad \alpha \approx .85$

$$\bar{\bar{\mathbf{P}}} = \alpha \mathbf{P} + \alpha \mathbf{a} \mathbf{v}^T + (1 - \alpha) \mathbf{E} \quad (\text{trivially irreducible})$$

- $\pi^T$  is now guaranteed to exist and be unique and power method is guaranteed to converge to  $\pi^T$ .



# Random Surfer

## Bored Surfer Enters Random URL

Irreducibility Fix:  $\bar{\bar{\mathbf{P}}} = \alpha \bar{\mathbf{P}} + (1 - \alpha) \mathbf{E} \quad e_{ij} = 1/n \quad \alpha \approx .85$

$$\bar{\bar{\mathbf{P}}} = \alpha \mathbf{P} + \alpha \mathbf{a} \mathbf{v}^T + (1 - \alpha) \mathbf{E} \quad (\text{trivially irreducible})$$

- $\pi^T$  is now guaranteed to exist and be unique and power method is guaranteed to converge to  $\pi^T$ .
- Different  $\mathbf{E} = \mathbf{e} \mathbf{v}^T$  and  $\alpha$  allow customization & speedup, yet rank-one update maintained;  $\bar{\bar{\mathbf{P}}} = \alpha \mathbf{P} + (\alpha \mathbf{a} + (1 - \alpha) \mathbf{e}) \mathbf{v}^T$

$$\bar{\bar{\mathbf{P}}} = \alpha \bar{\mathbf{P}} + (1 - \alpha) \mathbf{E} = \begin{bmatrix} 1/60 & 7/15 & 7/15 & 1/60 & 1/60 & 1/60 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 19/60 & 19/60 & 1/60 & 1/60 & 19/60 & 1/60 \\ 1/60 & 1/60 & 1/60 & 1/60 & 7/15 & 7/15 \\ 1/60 & 1/60 & 1/60 & 7/15 & 1/60 & 7/15 \\ 1/60 & 1/60 & 1/60 & 11/12 & 1/60 & 1/60 \end{bmatrix}$$





# Computing $\pi^T$

## A Big Problem

$$\text{Solve } \pi^T = \pi^T \bar{\mathbf{P}}$$

(stationary distribution vector)

$$\pi^T (\mathbf{I} - \bar{\mathbf{P}}) = 0$$

(too big for direct solves)

Google's PageRank is an eigenvector of a matrix of order 2.7 billion.

One of the reasons why Google is such an effective search engine is the PageRank™ algorithm, developed by Google's founders, Larry Page and Sergey Brin, when they were graduate students at Stanford University. PageRank is determined entirely by the link structure of the Web. It is recomputed about once a month and does not involve any of the actual content of Web pages or of any individual query. Then, for any particular query, Google finds the pages on the Web that match that query and lists those pages in the order of their PageRank.

Imagine surfing the Web, going from page to page by randomly choosing an outgoing link from one page to get to the next. This can lead to dead ends at pages with no outgoing links, or cycles around cliques of interconnected pages. So, a certain fraction of the time, simply choose a random page from anywhere on the Web. This theoretical random walk of the Web is a *Markov chain* or *Markov process*. The limiting probability that a dedicated random surfer visits any particular page is its PageRank. A page has high rank if it has links to and from other pages with high rank.

Let  $W$  be the set of Web pages that can be reached by following a chain of hyperlinks starting from a page at Google and let  $n$  be the number of pages in  $W$ . The set  $W$  actually varies with time, but in May 2002,  $n$  was about 2.7 billion. Let  $G$  be the  $n$ -by- $n$  connectivity matrix of

BY CLEVE MOLER

It tells us that the largest eigenvalue of  $A$  is equal to one and that the corresponding eigenvector, which satisfies the equation

$$x = Ax,$$

exists and is unique to within a scaling factor. When this scaling factor is chosen so that

$$\sum_i x_i = 1$$

then  $x$  is the state vector of the Markov chain. The elements of  $x$  are Google's PageRank.

If the matrix were small enough to fit in MATLAB, one way to compute the eigenvector  $x$  would be to start with a good approximate solution, such as the PageRanks from the previous month, and simply repeat the assignment statement

$$x = Ax$$

until successive vectors agree to within specified tolerance. This is known as the power method and is about the only possible approach for very large  $n$ . I'm not sure how Google actually computes PageRank, but one step of the power method would require one pass over a database of Web pages, updating weighted reference counts generated by the hyperlinks between pages.



# Computing $\pi^T$

## A Big Problem

Solve  $\pi^T = \pi^T \bar{\mathbf{P}}$  (stationary distribution vector)

$\pi^T (\mathbf{I} - \bar{\mathbf{P}}) = 0$  (too big for direct solves)

Start with  $\pi_0^T = \mathbf{e}/n$  and iterate  $\pi_{j+1}^T = \pi_j^T \bar{\mathbf{P}}$  (power method)



# Power Method to compute PageRank

$$\pi_0^T = \mathbf{e}^T / n$$

until convergence, do

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

(dense computation)

end



# Power Method to compute PageRank

$$\pi_0^T = \mathbf{e}^T / n$$

until convergence, do

✗  $\pi_{j+1}^T = \pi_j^T \bar{\mathbf{P}}$

(dense computation)

•  $\pi_{j+1}^T = \alpha \pi_j^T \bar{\mathbf{P}} + (1 - \alpha) \pi_j^T \mathbf{e} \mathbf{v}^T$

(sparser computation)

end



# Power Method to compute PageRank

$$\pi_0^T = \mathbf{e}^T / n$$

until convergence, do

✗  $\pi_{j+1}^T = \pi_j^T \bar{\mathbf{P}}$  (dense computation)

✗  $\pi_{j+1}^T = \alpha \pi_j^T \bar{\mathbf{P}} + (1 - \alpha) \pi_j^T \mathbf{e} \mathbf{v}^T$  (sparser computation)

•  $\pi_{j+1}^T = \alpha \pi_j^T \mathbf{P} + (\alpha \pi_j^T \mathbf{a} + (1 - \alpha)) \mathbf{v}^T$  (even less computation)

end

- $\mathbf{P}$  is very, very sparse with about 3-10 nonzeros per row.
- $\Rightarrow$  one vector-matrix mult. is  $O(\text{nnz}(\mathbf{P})) \approx O(n)$ .



# Convergence

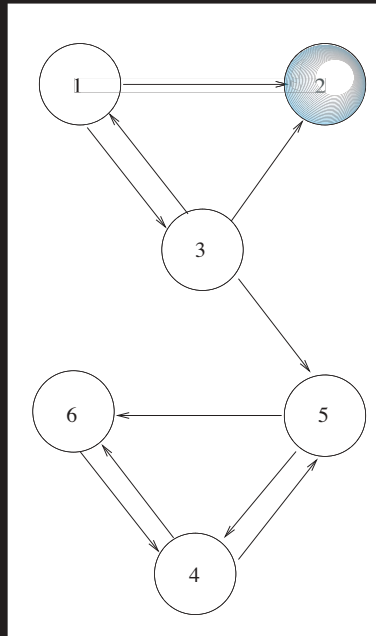
Can prove  $\lambda_2(\bar{\mathbf{P}}) = \alpha$

( $\Rightarrow$  asymptotic rate of convergence of PageRank method is rate at which  $\alpha^k \rightarrow 0$ )

Google

- uses  $\alpha = .85$  (5/6, 1/6 interpretation)
- report 50-100 iterations til convergence
- still takes days to converge

# PageRank Example



$$\pi^T = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ .03721 & .05396 & .04151 & .3751 & .206 & .2862 \end{pmatrix}$$

**Global** ranking of pages = [4 6 5 2 3 1]

**Query-independent** way of ranking relevant set





# Enhancements to the PR power method

- Kamvar et al. Extrapolation
- Kamvar et al. Adaptive PageRank
- Kamvar et al. BlockRank
- Lee et al. Lumpability of Dangling Nodes
- Langville/Meyer: Updating PageRank
- Ipsen/Kirkland: more theory for Langville/Meyer



# PageRank Issues

## Spamming

- Link Farms



# THE WALL STREET JOURNAL.

© 2003 Dow Jones & Company. All Rights Reserved

WEDNESDAY, FEBRUARY 26, 2003 ~ VOL. CCXLI NO. 39 ~ ★★★ \$1.00

WSJ.com

## What's News—

### Business and Finance

**NEWS CORP.** and Liberty are no longer working together on a joint offer to take control of Hughes, with News Corp. proceeding on its own and Liberty considering an independent bid. The move threatens to cloud the process of finding a new owner for the GM unit.

(Article on Page A3)

**The SEC signaled** it may file civil charges against Morgan Stanley, alleging it doled out IPO shares based partly on investors' commitments to buy more stock.

(Article on Page C1)

**Ahold's problems deepened** as U.S. authorities opened inquiries into accounting at the Dutch company's U.S. Foodservice unit.

**Fleming said** the SEC upgraded to a formal investigation an inquiry into the food wholesaler's trade practices with suppliers.

(Articles on Page A2)

**Consumer confidence fell** to its lowest level since 1993, hurt by energy costs, the terrorism threat and a stagnant job market.

(Article on Page A3)

**The industrials rebounded** on rumors of a peaceful solution to

### World-Wide

**BUSH IS PREPARING** to present Congress a huge bill for Iraq costs.

The total could run to \$95 billion depending on the length of the possible war and occupation. As horse-trading began at the U.N. to win support for a war resolution, the president again made clear he intends to act with or without the world body's imprimatur. Arms inspectors said Baghdad provided new data, including a report of a possible biological bomb. Gen. Franks assumed command of the war-operations center in Qatar. Allied warplanes are aggressively taking out missile sites that could threaten the allied troop buildup. (Column 4 and Pages A4 and A6)

*Turkey's parliament debated legislation to let the U.S. deploy 62,000 to open a northern front. Kurdish soldiers lined roads in a show of force as U.S. officials traveled into Iraq's north for an opposition conference.*

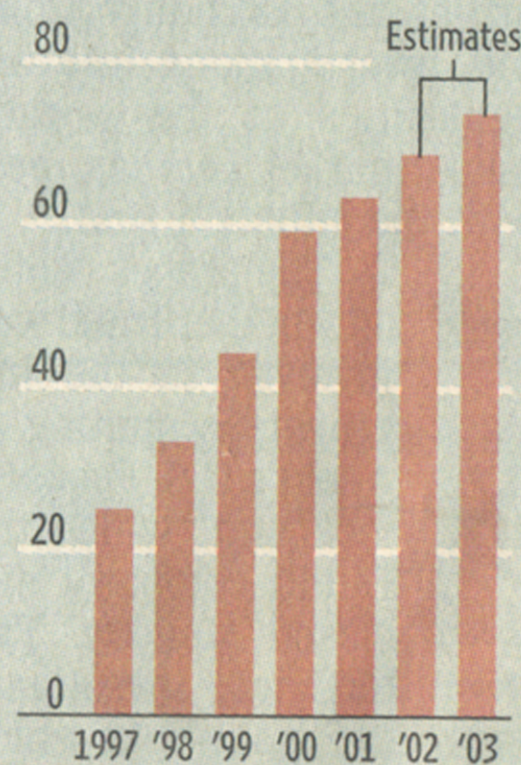
**Powell said** North Korea hasn't restarted a reactor and plutonium-processing facility at Yongbyon, hinting such forbearance might constitute an overture. But saber rattling continued a day after a missile test timed for the inauguration in Seoul. Pyongyang accused U.S. spy planes of violating its airspace and told its army to prepare for U.S. attack. (Page A14)

**The FBI came under** withering bipartisan criticism in a Senate Judiciary report in which Sen. Specter

## Web Master

### As the Web spreads...

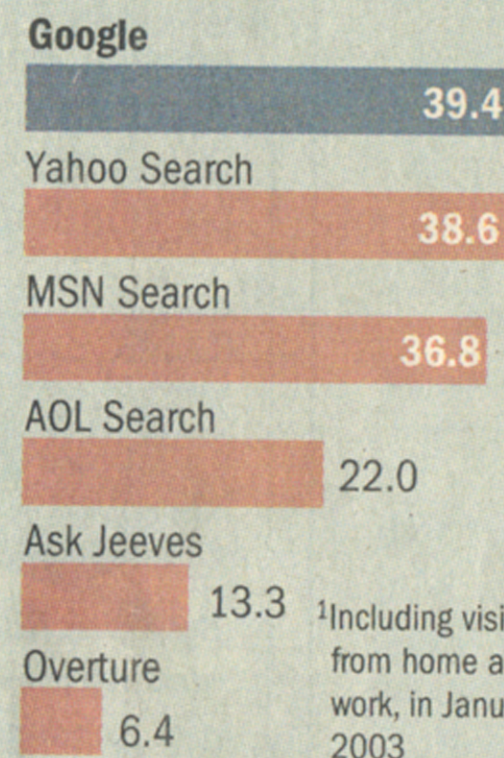
Total Internet users, by household, in millions



Sources: Forrester Research; Nielsen NetRatings

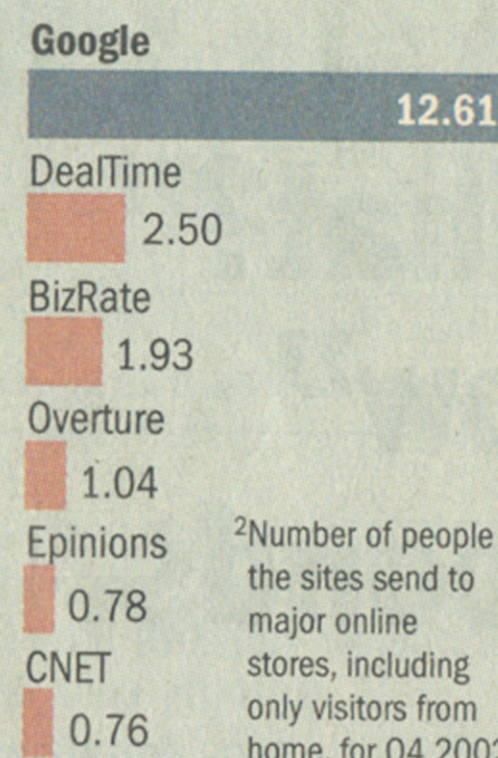
### Google's U.S. presence expands

Top search engines, in millions of unique visitors<sup>1</sup>



<sup>1</sup>Including visitors from home and work, in January 2003

Top shopping-referral sites, in millions of referrals<sup>2</sup>



<sup>2</sup>Number of people the sites send to major online stores, including only visitors from home, for Q4 2002

## Bush to Seek up to \$95 Billion To Cover Costs of War on Iraq

By GREG JAFFE  
And JOHN D. MCKINNON

WASHINGTON—The Bush administration is preparing supplemental spending requests totaling as much as \$95 billion for a war with Iraq, its aftermath and new expenses to fight terrorism, officials said.

The total could be as low as \$60 billion because Pentagon budget planners don't know how long a military conflict will last, whether U.S. allies will contribute more than token sums to the effort and what damage Saddam Hussein might do

to his own country to retaliate against conquering forces.

Budget planners also are awaiting the outcome of an intense internal debate over whether to include \$13 billion in the requests to Congress that the Pentagon says it needs to fund the broader war on terrorism, as well as for stepped up homeland security. The White House Office of Management and Budget argues that the money might not be necessary. President Bush, Defense Secretary Donald Rumsfeld and budget director Mitchell Daniels Jr. met yesterday to discuss the matter but didn't reach a final agreement. Mr. Rumsfeld plans to continue pressing his

## Cat and Mouse

### As Google Becomes Web's Gatekeeper, Sites Fight to Get In

Search Engine Punishes Firms That Try to Game System; Outlawing the 'Link Farms'

Exotic leatherwear Gets Cut Off

By MICHAEL TOTTY  
And MYLENE MANGALINDAN

Joy Holman sells provocative leather clothing on the Web. She wants what nearly everyone doing business online wants: more exposure on Google.

So from the time she launched exoticleatherwear.com last May, she tried all sorts of tricks to get her site to show up among the first listings when a user of Google Inc.'s popular search engine typed in "women's leatherwear" or "leather apparel." She buried hidden words in her Web pages intended to fool Google's computers. She signed up with a service that promised to have hundreds of sites link to her online store—thereby boosting a crucial measure in Google's system of ranking sites.

The techniques worked for a





# Web Sites Fight for Prime Real Estate on Google

*Continued From First Page*  
advertising that tried to capitalize on Google's formula for ranking sites. In effect, SearchKing was offering its clients a chance to boost their own Google rankings by buying ads on more-popular sites. SearchKing filed suit against the search company in federal court in Oklahoma, claiming that Google "purposefully devalued" SearchKing and its customers, damaging its reputation and hurting its advertising sales.

Google won't comment on the case. In court filings, the company said SearchKing "engaged in behavior that would lower the quality of Google search results" and alter the company's ranking system.

Google, a closely held company founded by Stanford University graduate students Sergey Brin and Larry Page, says Web companies that want to rank high should concentrate on improving their Web pages rather than gaming its system. "When people try to take scoring into their own hands, that turns into a worse experience for users," says Matt Cutts, a Google software engineer.

## Coding Trickery

Efforts to outfox the search engines have been around since search engines first became popular in the early 1990s. Early tricks included stuffing thousands of widely used search terms in hidden coding, called "metatags." The coding fools a search engine into identifying a site with popular words and phrases that may not actually appear on the site.

Another gimmick was hiding words or terms against a same-color background. The hidden coding deceived search engines that relied heavily on the number of times a word or phrase appeared in ranking a site. But Google's system, based on links, wasn't fooled.

Mr. Brin, 29, one of Google's two founders and now its president of technology, boasted to a San Francisco search-engine conference in 2000 that Google wasn't worried about having its results clogged with irrelevant results because its search methods couldn't be manipulated.

That didn't stop search optimizers from finding other ways to outfox the system. Attempts to manipulate Google's results even became a sport, called Google-bombing. Bombsters would try to

creating Web sites that were nothing more than collections of links to the clients' site, called "link farms." Since Google ranks a site largely by how many links or "votes" it gets, the link farms could boost a site's popularity.

In a similar technique, called a link exchange, a group of unrelated sites would agree to all link to each other, thereby fooling Google into thinking the sites have a multitude of votes. Many sites also found they could buy links to themselves to boost their rankings.

Ms. Holman, the leatherwear retailer, discovered the consequences of trying to fool Google. The 42-year-old hospital laboratory technician, who learned computer skills by troubleshooting her hospital's

**'The big search engines determine the laws of how commerce runs,' says Mr. Massa.**

equipment, operates her online apparel store as a side business that she hopes can someday replace her day job.

When she launched her Exotic Leather Wear store from her home in Mesa, Ariz., she quickly learned the importance of appearing near the top of search-engine results, especially on Google. She boned up on search techniques, visiting online discussion groups dedicated to search engines and reading what material she could find on the Web.

At first, Ms. Holman limited herself to modest changes, such as loading her page with hidden metatag coding that would help steer a search toward her site when a user entered words such as "haltertops" or "leather miniskirts." Since Google doesn't give much weight to metatags in determining its rankings, the efforts had little effect on her search results.

She then received an e-mail advertisement from AutomatedLinks.com, a Wirral, England, company that promised to send traffic "through the roof" by linking more than 2,000 Web sites to hers. Aside from attracting customers, the links were designed to improve her site's search engine rankings by taking

In theory, when Google encounters the AutomatedLinks code, it treats it as a legitimate referral to the other sites and counts them in totting up the sites' popularity.

Shortly after Ms. Holman signed up with AutomatedLinks in July, she read on an online discussion group that Google objected to such link arrangements. She says she immediately stripped the code from her Web pages. For a while her site gradually worked its way up in Google search results, and business steadily improved because links to her site still remained on the sites of other AutomatedLinks customers. Then, sometime in November, her site was suddenly no longer appearing among the top results. Her orders plunged as much as 80%.

Ms. Holman, who e-mailed Google and AutomatedLinks, says she has been unable to get answers. But in the last few months, other AutomatedLinks customers say they have seen their sites apparently penalized by Google. Graham McLeay, who runs a small chauffeur service north of London, saw revenue cut in half during the two months he believes his site was penalized by Google.

The high-stakes fight between Google and the optimizers can leave some Web-site owners confused. "I don't know how people are supposed to judge what is right and wrong," says Mr. McLeay.

AutomatedLinks didn't respond to requests for comment. Google declined to comment on the case. But Mr. Cutts, the Google engineer, warns that the rules are clear and that it's better to follow them rather than try to get a problem fixed after a site has been penalized. "We want to return the most relevant pages we can," Mr. Cutts says. "The best way for a site owner to do that is follow our guidelines."

## Crackdown

Google has been stepping up its enforcement since 2001. It warned Webmasters that using trickery could get their sites kicked out of the Google index and it provided a list of forbidden activities, including hiding text and "link schemes," such as the link farms. Google also warned against "cloaking"—showing a search engine a page that's designed to score well while giving visitors a different, more attractive page—or creating multiple Web addresses that take visitors to a single site.

To stay one step ahead of the Web

homa City-based SearchKing, an online directory for hundreds of small, specialty Web sites. SearchKing also sells advertising links designed both to deliver traffic to an advertiser and boost its rankings in Google and other search results.

Bob Massa, SearchKing's chief executive, last August launched the PR Ad Network as a way to capitalize on Google's page-ranking system, known as PageRank. PageRank rates Web sites on a scale of one to 10 based on their popularity, and the rankings can be viewed by Web users if they install special Google software. PR Ad Network sells ads that are priced according to a site's PageRank, with higher-ranked sites commanding higher prices. When a site buys an advertising link on a highly ranked site, the ad buyer could see its ratings improve because of the greater weight Google gives to that link.

Shortly after publicizing the ad network, Mr. Massa discovered that his site suddenly dropped in Google's rankings. What's more, sites that participated in the separate SearchKing directory also had their Google rankings lowered. He filed a lawsuit in Oklahoma City federal court, claiming Google was punishing him for trying to profit from the company's page-ranking system.

A Google spokesman won't comment on the case. In its court filings, Google said it demoted pages on the SearchKing site because of SearchKing's attempts to manipulate search results. The company has asked for the suit to be dismissed, arguing that the PageRank represents its opinion of the value of a Web site and as such is protected by the First Amendment.

"The big search engines determine the laws of how commerce runs," says Mr. Massa, who is persisting with the lawsuit even though the sites have had their page rankings partly restored. "Someone needs to demand accountability."

Google is taking steps that many say could satisfy businesses trying to boost their rankings. Google has long sold sponsored links that show up on the top of many search-results pages, separate from the main listings. Last year, the company expanded its paid-listings program, so that there are now more slots where sites can pay for a prominent place in the results. Many sites now are turning to advertising instead of tactics to optimize their rankings.

# Home Depot E Amid First Qu

By CHAD TERHUNE

ATLANTA—Home Depot Inc. reported fiscal fourth-quarter earnings down 3.4% on disappointing sales.

Speaking to investors and industry analysts, the company's chairman and chief executive, Bob Nardelli, said Home Depot is prepared to win back dissatisfied customers and answer the competitive challenge from its chief rival with remodeled stores, increased inventory and improved customer service.

The nation's largest home-improvement retailer said net income for the quarter ended Feb. 2 decreased to \$686 million or 30 cents a share, from \$710 million or 30 cents a share, a year earlier. Sales fell 2% to \$13.21 billion from \$13.49 billion, the first quarterly sales decline in the company's 24-year history. Home Depot's net income in the latest quarter was a week shorter than a year earlier. Using comparable 13-week periods, the company said quarterly sales increased 5% and net income rose 8%.

Same-store sales, or sales at stores open at least a year, declined 6% in the quarter. Home Depot said stronger sales last month offset a disastrous December and helped the retailer avoid its earliest estimate that same-store sales could drop as much as 10%. In 4 p.m. New York Stock Exchange composite trading, Home Depot shares rose 66 cents to \$22.84.

# Fiat Patriarch Is Set to Becom

By ALESSANDRA GALLONI

ROME—Umberto Agnelli is due to take over as chairman of Fiat SpA on Friday, stepping into the driver's seat as the Italian conglomerate works on an 11th-hour refinancing of its unprofitable car unit.

Mr. Agnelli, the 68-year-old brother of Fiat patriarch Gianni Agnelli, who last month, was widely expected to take over from current chairman, Franco Bernabè, later this year. But Mr. Bernabè, who has served as chairman since





# PageRank Issues

## Spamming

- Link Farms
- Google Bombs

SEARCH  [Low Graphics version](#) | [Change edition](#)[Feedback](#) | [Help](#)**BBC NEWS** WORLD EDITION **WATCH/LISTEN TO BBC NEWS****News Front Page**[Africa](#)[Americas](#)[Asia-Pacific](#)[Europe](#)[Middle East](#)[South Asia](#)[UK](#)[Business](#)[Health](#)[Science/Nature](#)[Technology](#)[Entertainment](#)[Have Your Say](#)[Country Profiles](#)[In Depth](#)[Programmes](#)**RELATED SITES**[BBC SPORT](#)[BBC WEATHER](#)[BBC ON THIS DAY](#)**LANGUAGES**[ESPAÑOL](#)[BRASIL](#)[CARIBBEAN](#)

Last Updated: Sunday, 7 December, 2003, 15:04 GMT

[E-mail this to a friend](#) [Printable version](#)**'Miserable failure' links to Bush****George W Bush has been Google bombed.**

Web users entering the words "miserable failure" into the popular search engine are directed to the biography of the president on the White House website.

The trick is possible because Google searches more than just the contents of web pages - it also counts how often a site is linked to, and with what words.

Thus, members of an online community can affect the results of Google searches - called "Google bombing" - by linking their sites to a chosen one.

Weblogger Adam Mathes is credited with inventing the practice in 2001, when he used it to link the phrase "talentless hack" to a friend's website.

The search engine can be manipulated by a fairly small group of users, one report suggested.

Newsday newspaper says as few as 32 web pages with the words "miserable failure" link to the Bush biography.

The Bush administration has been on the receiving end of pointed Google bombs before.

In the run-up to the Iraq war, internet users manipulated Google so the phrase "weapons of mass destruction" led to a joke page saying "These Weapons of Mass Destruction cannot be displayed."

The site suggests "clicking the regime change button", or "If you are George Bush and typed the country's name in the address bar, make sure that it is spelled correctly (IRAQ)".



Bush has been the target of similar pranks before

**SEE ALSO:**[WMD spoof is internet hit](#)[04 Jul 03 | West Midlands](#)[Google hit by link bombers](#)[13 Mar 02 | Science/Nature](#)**RELATED INTERNET LINKS:**[White House](#)[Google bombing](#)

The BBC is not responsible for the content of external internet sites

**TOP AMERICAS STORIES NOW**[US army battles to keep soldiers](#)[Report backs US Catholic bishops](#)[Envoys bid to ease BSE fears](#)[Protests widen over sky marshals](#)

**“ If you are George Bush and typed the country's name in the address bar, make sure that it is spelled correctly (IRAQ) ”**

Prank website

[E-mail this to a friend](#) [Printable version](#)**LINKS TO MORE AMERICAS STORIES**[E-mail services](#) | [Desktop ticker](#) | [Mobiles/PDAs](#) |

© BBC MMIV

[Back to top ^^](#)[News Front Page](#) | [Africa](#) | [Americas](#) | [Asia-Pacific](#) | [Europe](#) | [Middle East](#) | [South Asia](#)[UK](#) | [Business](#) | [Entertainment](#) | [Science/Nature](#) | [Technology](#) | [Health](#)[Have Your Say](#) | [Country Profiles](#) | [In Depth](#) | [Programmes](#)[BBCi Homepage >>](#) | [BBC Sport >>](#) | [BBC Weather >>](#) | [BBC World Service >>](#)[ABOUT BBC NEWS](#) | [Help](#) | [Feedback](#) | [News sources](#) | [Privacy](#) | [About the BBC](#)

# BLAH3.COM

Dusty & Yellowing - [The Blah3 Archives](#)  
Complaints, compliments, arguments? [Email me](#)



[<< "Happy Ramadan, y'all..."](#) [\[Main Index\]](#) [>>"His heart just isn't in it...."](#)

## 10/27/2003 Archived Entry: "I'm taking part in a new web project..."

I'm taking part in a new web project...

From this day forth, I will refer to George W. Bush as a [Miserable Failure](#) at least once a day. Why, you ask? Well, someone came up with this great idea to link George W. Bush and [Miserable Failure](#) in popular search engines. **If you have a blog or web site, help raise the link between George W. Bush and the phrase 'miserable failure' by copying this link and placing somewhere on your site or blog.**

Thank you very much for your participation.

**Replies: 16 people speak up**

Great idea!

Posted by [rlr](#) @ 10/27/2003 10:06 PM NY

That is genius. I could add a few other keywords, like "pathetic". I will post it on my blog now...

Posted by [Political Pulpit](#) @ 10/28/2003 02:32 PM NY

Miserable Failure? I'm down with that....

Stay tuned...

Posted by [Drewcifer](#) @ 10/28/2003 02:35 PM NY

Done!

Posted by [Maru](#) @ 10/28/2003 08:46 PM NY

thats great, another thing I think  
might be good to use: tax cuts for the wealthy....welfare for the wealthy. just my 2 cents.

Posted by [doodaa](#) @ 10/29/2003 03:01 AM NY

Call me a liberal lemming, I guess. :) I'm in.

Posted by [BJ](#) @ 10/29/2003 09:28 AM NY

The key is stating it in connection with terms that will be widely searched. It does no good to simply say "George Bush is a miserable failure" because no one will ever search for that. It might be fun at a parties to show how often the two are in the same sentence in a Google search, but otherwise it does little to advance the theme.

What will work is connecting it to frequent search times, such as "Iraq policy". For instance "George Bush's Iraq Policy is a miserable failure."

The plan shouldn't be to link Miserable Failure to George Bush, but to link Miserable Failure to George Bush and two or three choice, frequently searched phrases.

Overture.com has a keyword suggestion tool that shows how many times certain terms are coming up in searches. Using that tool, I can determine that in September the search for "bush george iraq saddam" gets about 12 times more queries than "george bush iraq speech". "george bush biography" gets a huge amounts of hits compared to something like "george bush policy".

So someone needs to write about three complete sentences using these terms based on verifiable search results and including the "miserable failure" phrase and then advocate for that exact usage.

According to Overture, the phrases "george Bush miserable failure" were not queried even once in their sample during the month just passed.

Posted by [Joe Briefcase](#) @ 10/29/2003 10:51 AM NY

how about drunken, illiterate, mendacious, runt-like miserable failure?

Posted by [tim](#) @ 10/29/2003 11:58 AM NY

Hahaha, that's very productive. This is why everyone knows that liberals are stupid. They do stupid things.

Posted by [Reek Stankleberry](#) @ 10/29/2003 12:04 PM NY

how about, instead of calling it lies--anyone can lie--how about calling it HORSEFEATHERS AND CODSWALLOP! Pin that on him too.

'Den of Thieves'  
The ? Campaign, 2002  
'Fair & Balanced' Day



**Blahroll**

[A Level Gaze](#)  
[A Skeptical Blog](#)  
[Ain't No Bad Dude](#)  
[Angry Bear](#)  
[Ann Slanders](#)  
[Apathy, Inc.](#)  
[Army of Fun](#)  
[Atrios](#)  
[Attorney At Arms](#)  
[Avedon's Sideshow](#)  
[Bag Times](#)  
[BartCop E!](#)  
[BartCop!](#)  
[Bellum Americanum](#)  
[Big Picnic](#)  
[Bitter Obscurity](#)  
[Booknotes](#)  
[Bunsen](#)  
[Burgblog](#)  
[Bush Is A Moron](#)  
[BushFlash](#)  
[BushLiar](#)  
[BusyBusyBusy](#)  
[Byrd's Brain](#)  
[Certain Shade of Green](#)  
[Chimes at Midnight](#)  
[Chris Nelson](#)  
[Circumspect](#)  
[CNN Lies](#)  
[Connipion](#)  
[Counterspin](#)  
[Cursor](#)  
[Daily Brew](#)  
[Daily Cynic](#)  
[Daily Kos](#)  
[Daily Outrage](#)  
[Daily War News](#)  
[Damfacrats](#)  
[Deckie Holmes](#)  
[Democratic Veteran](#)  
[Dodona](#)  
[dratfink](#)  
[Duckwing](#)  
[E Pluribus Unum](#)  
[Estimated Prophet](#)  
[Ethel](#)  
[Federal Examiner](#)  
[Fengi](#)  
[For Freedom Century](#)  
[FrogN'Blog](#)  
[Ge. JC Christian](#)  
[GeekPol](#)  
[Genoan Sailor](#)  
[GeoDog](#)  
[Get Donkey!](#)


[Advanced Search](#) [Preferences](#) [Language Tools](#) [Search Tips](#)

miserable failure

Google Search

[Web](#) [Images](#) [Groups](#) [Directory](#) [News](#)
Searched the web for **miserable failure**. Results **1 - 10** of about **257,000**. Search took **0.08** seconds.

Tip: In most browsers you can just hit the return key instead of clicking on the search button.

### [Michael Moore.com](#)

Wednesday, January 14th, 2004 I'll Be Voting For Wesley Clark /

Good-Bye Mr. Bush — by Michael Moore. Many of you have written ...

Description: Official site of the gadfly of corporations, creator of the film Roger and Me and the television show...

Category: Arts &gt; Celebrities &gt; M &gt; Moore, Michael

[www.michaelmoore.com/](http://www.michaelmoore.com/) - 43k - [Cached](#) - [Similar pages](#)

### [Biography of President George W. Bush](#)

Home &gt; President &gt; Biography President George W. Bush En Español.

George W. Bush is the 43rd President of the United States. He ...

Description: Biography of the president from the official White House web site.

Category: Kids and Teens &gt; School Time &gt; ... &gt; Bush, George Walker

[www.whitehouse.gov/president/gwbbio.html](http://www.whitehouse.gov/president/gwbbio.html) - 29k - [Cached](#) - [Similar pages](#)

### [Biography of Jimmy Carter](#)

Home &gt; History &amp; Tours &gt; Past Presidents &gt; Jimmy Carter. Jimmy Carter.

Jimmy Carter aspired to make Government "competent and compassionate ...

Description: Short biography from the official White House site.

Category: Society &gt; History &gt; ... &gt; Presidents &gt; Carter, James Earl

[www.whitehouse.gov/history/presidents/jc39.html](http://www.whitehouse.gov/history/presidents/jc39.html) - 36k - [Cached](#) - [Similar pages](#)

### [Senator Hillary Rodham Clinton: Online Office Welcome Page](#)

Dear Friend,. Thank you for visiting my on-line office! I appreciate

your interest in the issues before the United States Senate. ...

Description: Official US Senate web site of Senator Hillary Rodham Clinton (D - NY).

Category: Society &gt; History &gt; ... &gt; First Ladies &gt; Clinton, Hillary

[clinton.senate.gov/](http://clinton.senate.gov/) - 9k - [Cached](#) - [Similar pages](#)

### [BBC NEWS | Americas | 'Miserable failure' links to Bush](#)

'Miserable failure' links to Bush. ... Prank website. Newsday newspaper says as few as 32 web pages with the words "miserable failure" link to the Bush biography. ...

[news.bbc.co.uk/2/hi/americas/3298443.stm](http://news.bbc.co.uk/2/hi/americas/3298443.stm) - 31k - [Cached](#) - [Similar pages](#)

### [Atlantic Unbound | Politics & Prose | 2003.09.24](#)

... Atlantic Unbound | September 24, 2003 Politics &amp; Prose | by Jack Beatty

"A Miserable Failure" Will Bush be re-elected? Only if voters ...

[www.theatlantic.com/unbound/polipro/pp2003-09-24.htm](http://www.theatlantic.com/unbound/polipro/pp2003-09-24.htm) - 22k - [Cached](#) - [Similar pages](#)

### [miserable failure | Hillary Clinton | Hildebeest](#)

... Miserable Failure. Quotes for the History Books. ... You may also want to check out the Miserable Failure Project. and the cuckolded dyke Project. and the ...

[miserable-failure.blogspot.com/](http://miserable-failure.blogspot.com/) - 60k - [Cached](#) - [Similar pages](#)

### [Dick Gephardt for President - Welcome](#)

... to preserve some large part of the Bush tax cut. I think retaining





# PageRank Issues

## Spamming

- Link Farms
- Google Bombs

## Updating

- The Google Dance



[What's Google Dance?](#) - [Google Web API](#) - [Google Humor](#) -  
[Google Crawler](#)  
[Keyword Importance](#) - [Webmaster Tools](#) - [About Us](#) -  
[Contact Us](#)

Below you will find the Google Dance results for the search keyword **pagerank**. If you notice that there are any differences in results between the different Google data centers then Google is in the middle of spidering the internet. It's that simple!

www.google.com

www2.google.com

www3.google.com



Enter your search:

pagerank

Google Search

I'm Feeling Lucky

1. [Google Technology](#)
2. [Google Web Directory Help](#)
3. [Pagerank Explained. Google's Pag](#)
4. [PageRank Calculator. WebWorksh](#)
5. [Pagerank Explained Correctly with](#)
6. [The Anatomy of a Search Engine](#)
7. [LinkAdage Auctions Link Exchange](#)
8. [PageRank is Dead \(Jeremy Zawodr](#)
9. [Google PageRank](#)
10. [The PageRank Citation Ranking:](#)

[Open Results in New Window](#)

Enter your search:

pagerank

Google Search

I'm Feeling Lucky

1. [Google Technology](#)
2. [Google Web Directory Help](#)
3. [Pagerank Explained. Google's Pag](#)
4. [PageRank Calculator. WebWorksh](#)
5. [Pagerank Explained Correctly with](#)
6. [The Anatomy of a Search Engine](#)
7. [PageRank is Dead \(Jeremy Zawodr](#)
8. [LinkAdage Auctions Link Exchange](#)
9. [Google PageRank](#)
10. [Google PageRank Calculator Val](#)

[Open Results in New Window](#)

Enter your search:

pagerank

Google Search

I'm Feeling Lucky

1. [Google Technology](#)
2. [Google Web Directory Help](#)
3. [Pagerank Explained. Google's Pag](#)
4. [PageRank Calculator. WebWorksh](#)
5. [Pagerank Explained Correctly with](#)
6. [The Anatomy of a Search Engine](#)
7. [PageRank is Dead \(Jeremy Zawodr](#)
8. [LinkAdage Auctions Link Exchange](#)
9. [Google PageRank](#)
10. [Google PageRank Calculator Val](#)

[Open Results in New Window](#)[Next Page>>](#)

pagerank

10

Lets Dance!

Thought of a great name for your site?



yourdomainname

.com

REGISTER IT NOW

register.com

[What's Google Dance?](#) - [Google Web API](#) - [Google Humor](#) - [Google Crawler](#)  
[Keyword Importance](#) - [Webmaster Tools](#) - [About Us](#) - [Contact Us](#)

This site is in no way affiliated or associated with [Google](#) and/or or it's respective companies  
 Copyright © 2003 [Google Dance Tool](#) [Privacy Policy](#)



# PageRank Issues

## Spamming

- Link Farms
- Google Bombs

## Updating

- The Google Dance

## Speed Improvements

- Enhancing Power Method



[Congressional Affairs](#) | [Newsroom](#) | [Speeches](#) | [Priority Areas](#) | [What's Cool](#) | [Publications](#) | [Partners](#) | [History](#) | [About Us](#)

You are in: [NSF Home](#) > [OLPA Home](#) > [Newsroom](#) > [Press Releases](#) > NSF PR 03-56

## NSF Press Release

**NSF PR 03-56 - May 13, 2003**

**Media contact:** David Hart (703) 292-7737 [dhart@nsf.gov](mailto:dhart@nsf.gov)  
**Program contacts:** Maria Zemankova (703) 292-8930 [mzemanko@nsf.gov](mailto:mzemanko@nsf.gov)  
John Staudhammer (703) 292-8918 [jstaudham@nsf.gov](mailto:jstaudham@nsf.gov)

### **Researchers Develop Techniques for Computing Google-Style Web Rankings Up to Five Times Faster**

*Speed-up may make "topic-sensitive" page rankings feasible*

ARLINGTON, Va. — Computer science researchers at Stanford University have developed several new techniques that together may make it possible to calculate Web page rankings as used in the Google search engine up to five times faster. The speed-ups to Google's method may make it realistic to calculate page rankings personalized for an individual's interests or customized to a particular topic.

The Stanford team includes graduate students Sepandar Kamvar and Taher Haveliwala, noted numerical analyst Gene Golub and computer science professor Christopher Manning. They will present their first paper at the Twelfth Annual World Wide Web Conference (WWW2003) in Budapest, Hungary, May 20-24, 2003. The work was supported by the National Science Foundation, an independent federal agency that supports fundamental research and education in all fields of science and engineering.

Computing PageRank, the ranking algorithm behind the Google search engine, for a billion Web pages can take several days. Google currently ranks and searches 3 billion Web pages. Each personalized or topic-sensitive ranking would require a separate multi-day computation, but the payoff would be less time spent wading through irrelevant search results. For example, searching a sports-specific Google site for "Giants" would give more importance to pages about the New York or San Francisco Giants and less importance to pages about Jack and the Beanstalk.

"This work is a wonderful example of how NSF support for basic computer science research, including applied mathematics and algorithm research, has impacts in daily life," said NSF program officer Maria Zemankova. In the mid-1990s, an NSF digital library project and an NSF graduate fellowship also supported Stanford graduate students Larry Page and Sergey Brin while they developed what would become the Google search engine.

To speed up PageRank, the Stanford team developed a trio of techniques in numerical linear algebra. First, in the WWW2003 paper, they describe so-called "extrapolation" methods, which make some assumptions about the Web's link structure that aren't true, but permit a quick and easy computation of PageRank. Because the assumptions aren't true, the PageRank isn't exactly correct, but it's close and can be refined using the



# PageRank Issues

## Spamming

- Link Farms
- Google Bombs

## Updating

- The Google Dance

## Speed Improvements

- Enhancing Power Method
- Personalized PageRank

Microsoft extends life of its Java Virtual Machine 10:49AM

Hardware | [Software](#) | Security | Commentary | [Headline Archives](#) | Briefs**News Software**

## Searching for the personal touch

By [Stefanie Olsen](#)

CNET News.com

August 11, 2003, 4:00 AM PT

**TALK BACK!** [Add your opinion](#)Forward in [E-MAIL](#) Format for [PRINTER](#)

**A stealth start-up out of Stanford University is hoping to raise the heat on one of the toughest problems in Web search--and possibly out-Google Google in the process.**

**Kaltix** was formed in recent months by three members of Stanford's PageRank team--a research group created to advance the mathematical algorithm developed by Google co-founder and Stanford alum Larry Page that cemented Google's fame.

PageRank has helped steer people to Web sites like no other search technology before it, harnessing the link structure of the Web to determine the most popular pages. Now, Kaltix hopes to improve upon PageRank, with an attempt to speed up the underlying PageRank computations.

That, in turn, could lay the groundwork for a breakthrough in a cutting-edge area of Web search development known as "personalization," which aims to sort search results based on the specific needs and interests of individuals, instead of the consensus approach pioneered by Google.

"Kaltix is a 'stealth mode' start-up...(leveraging) research done at Stanford University as well as several new technologies developed at Kaltix to provide large-scale personalized and context-sensitive search," a Kaltix representative said, declining to comment further.

Kaltix has disclosed few specifics about its plans or technology. But the company's general statements appear to place it in a sweet spot for innovation that's being pursued by all of the major search providers. Now that Web search has become a moneymaker for portals such as Yahoo and Microsoft's MSN, technologists from all the industry players are back in the labs developing formulas to personalize search.

Web companies outside the search industry have long made attempts to create personalization features, but most of these attempts have fallen short of expectations. Amazon.com, for example, regularly [serves up](#) book titles related to a visitor's previous purchases, which may no longer be relevant. A personalization feature offered through TiVo, a maker of video recording devices, was criticized when reports circulated that the device would recommend gay-themed television programs to viewers based on just a few program selections.

Despite these flawed attempts, developers continue to have faith that personalization technology can be created that will ultimately unleash marketing and revenue opportunities.

If search developers are successful in building such technology, they could help millions of people better

### Vendor Priorities from our sponsors

[Easy and Affordable  
Network Management: New  
HP ProCurve Tools for  
Network Administrators](#)

**Key findings:**

Easy-to-use VLAN management interface; an integrated, low-overhead traffic monitor shows detailed information on traffic

**Who/What:** White paper

[Intel® Xeon™ Processor MP  
Quick Reference Guide](#)

**Key findings:**

Designed specifically for 4-way, multi-processor (MP)-based servers, available up to 2.80 GHz and 2 MB iL3 cache

**Who/What:** Reference Guide (pdf)

[Free Office Pro 2003  
Evaluation Kit](#)

**Key findings:**

Test-drive of all the elements of the latest Microsoft Office System; evaluate before you upgrade

**Who/What:** 30-day trial copy, plus a resource guide CD

[ZDNet Power  
Centers](#)

[Webcasts & Video](#)

▼ advertisement

**The IBM ThinkPad X40**  
Our smallest, lightest, wireless notebook ever.

Order online and standard shipping included.

**\$1,499\***  
▶ #23861CU

[Learn more](#)



## Google Press Release

[Home](#)

[All About Google](#)

[Press Center](#)

### Press Kit

[Press Releases](#)

[In the News](#)

[Image Gallery](#)

[User Testimonials](#)

[Google Zeitgeist](#)

### Company

[Overview](#)

[Fact Sheet](#)

[Management](#)

[Milestones](#)

[Fun Facts](#)

[Investor Info](#)

[Awards](#)

### Technology

[Overview](#)

[Life of a Query](#)

### Business

[Overview](#)

[Case Studies](#)

## Google Acquires Kaltix Corp.

### New Technologies and Engineering Team Complement Google Search Engine

MOUNTAIN VIEW, Calif. - **Sept. 30, 2003** - Google Inc. today announced it acquired Kaltix Corp., a Palo Alto, Calif.-based search technology start-up. Financial terms of the deal were not disclosed.

"Google and Kaltix share a common commitment to developing innovative search technologies that make finding information faster, easier and more relevant," said Larry Page, co-founder and president of Products at Google. "Kaltix is working on a number of compelling search technologies, and Google is the ideal vehicle for the continued development of these advancements."

Kaltix Corp. was formed in June 2003 and focuses on developing personalized and context-sensitive search technologies that make it faster and easier for people to find information on the web.

### About Google

Google's innovative search technologies connect millions of people around the world with information every day. Founded in 1998 by Stanford Ph.D. students Larry Page and Sergey Brin, Google today is a top web property in all major global markets. Google's targeted advertising program, which is the largest and fastest growing in the industry, provides businesses of all sizes with measurable results, while enhancing the overall web experience for users. Google is headquartered in Silicon Valley with offices throughout North America, Europe, and Asia. For more information, visit [www.google.com](http://www.google.com).

###

*Google is a trademark of Google Inc. All other company and product names may be trademarks of the respective companies with which they are associated.*

### For further information:

Nathan Tyler

Google Inc.

+1 650-623-4311

[nate@google.com](mailto:nate@google.com)



©2003 Google - [Home](#) - [All About Google](#) - [We're Hiring](#) - [Site Map](#)