

CS155b: E-Commerce

Lecture 16: April 10, 2001

WWW Searching and Google

WWW Digraph

- More than 1 Billion Nodes (Pages)
- Average Degree (links/Page) is 5-15.
(Hard to Compute!)
- Massive, Distributed, Explicit Digraph
(Not Like Call Graphs)

“Hot” Research Area

- Graph Representation
- Duplicate Elimination
- Clustering
- Ranking Query Results

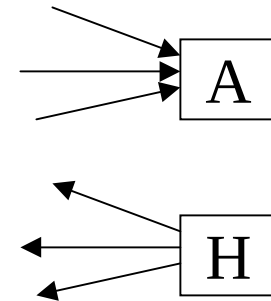
<http://theory.stanford.edu/~focs98/tutorials.html>

A. Broder & M. Henzinger

“Abundance” Problem

<http://simon.cs.cornell.edu/home/kleinber/kleinber.html>

- Given a query find:
 - Good Content (“Authorities”)
 - Good Sources of Links (“Hubs”)
- Mutually Reinforcing
- Simple (Core) Algorithm



$$T \triangleq \{\text{n Pages}\}, A \triangleq \{\text{Links}\}$$

$X_p \in \mathbb{R}^{\geq 0}, p \in T$ non-negative “Authority Weights”

$Y_p \in \mathbb{R}^{\geq 0}, p \in T$ non-negative “Hub Weights”

I operation Update Authority Weights

$$X_p \leftarrow \sum_{(q,p) \in A} Y_q$$

O operation Update Hub Weights

$$Y_p \leftarrow \sum_{(p,q) \in A} X_q$$

Normalize: $\sum_{p \in T} X_p^2 = \sum_{p \in T} Y_p^2 = 1$

Core Algorithm

$Z \leftarrow (1, 1, \dots, 1)$

$X \leftarrow Y \leftarrow Z$

Repeat until Convergence

 Apply I /* Update Authority weights */

 Apply O /* Update Hub Weights */

 Normalize

Return Limit (X^* , Y^*)

Convergence of $(X^i, Y^i) \triangleq (OI)^i(Z, Z)$

$A \triangleq n \times n$ “Adjacency Matrix”

Rewrite I and O:

$$\mathbf{X} \leftarrow A^T \mathbf{Y} \quad ; \quad \mathbf{Y} \leftarrow A \mathbf{X}$$

$$\mathbf{X}^i = (A^T A)^{i-1} A^T \mathbf{Z} \quad ; \quad \mathbf{Y}^i = (A A^T)^i \mathbf{Z}$$

AA^T Symm., Non-negative and $Z = (1, 1, \dots, 1)$

$$\mathbf{X}^* \triangleq \lim_{i \rightarrow \infty} \mathbf{X}^i = \omega_1(A^T A)$$

$$\mathbf{Y}^* \triangleq \lim_{i \rightarrow \infty} \mathbf{Y}^i = \omega_1(A A^T)$$

Whole Algorithm (k,d,c)

q Search Engine $|S| \leq k$

Base Set T:

(In S, $S \rightarrow , \rightarrow S$) and $\leq d$ links/page

Remove “Internal Links”

Run Core Algorithm on T

From Result (X, Y), Select

C pages with max X^* values

C pages with max Y^* values

Examples (k= 200, d=5)

q = censorship + net

www.EFF.org

www.EFF.org/BlueRib.html

www.CDT.org

www.VTW.org

www.ACLU.prg

q = Gates

www.roadahead.com

www.microsoft.com

www.ms.com/corpinfo/bill-g.html

[Compares well with Yahoo, Galaxy, etc.]

Approach to “Massiveness”: Throw Out Most of G!!

- Non-principal Eigenvectors correspond to “Non-principal Communities”

- Open (?):

Objective Performance Criteria

Dependence on Search Engine

Nondeterministic Choice of S and T

Google History

- **Founded in 1998 by Larry Page and Sergey Brin, two Stanford Ph.D. candidates.**
- **Privately held company, whose backers include Kleiner Perkins Caufield & Byers and Sequoia Capital.**
- **Continues to win top awards for Search Engines. Computer Scientists love it!!!**

Major Partners

- **Yahoo!**
- **Palm**
- **Nextel**
- **Netscape**
- **Cisco Systems**
- **Virgin Net**
- **Netease.com**
- **RedHat**
- **Virgilio**
- **Washingtonpost.com**

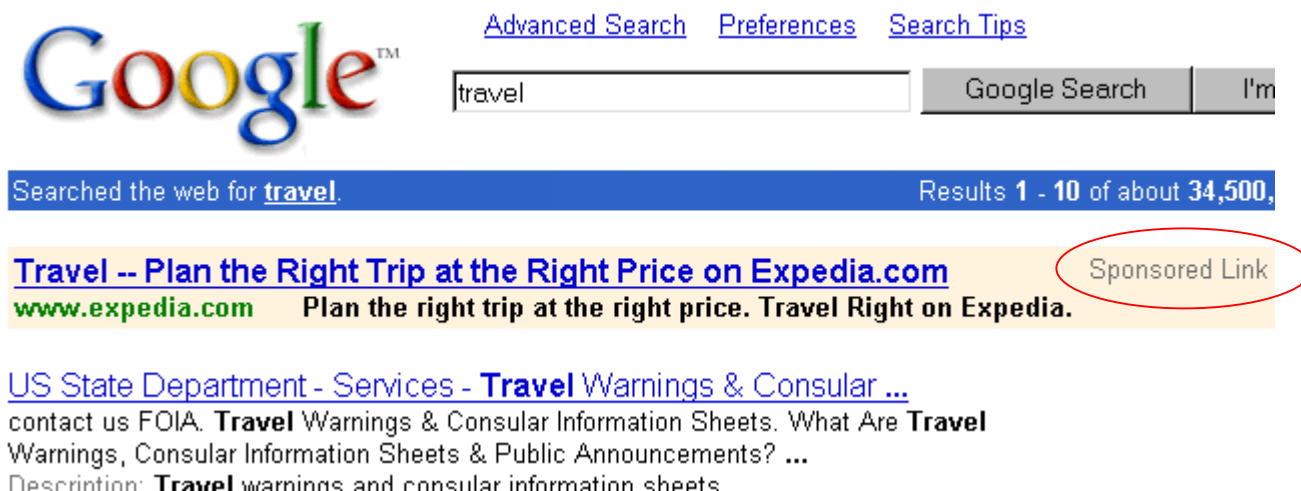
Business Model

- The company delivers services through its own web site at www.google.com and by licensing its search technology to commercial sites
- Advertising:
 - Premium Sponsorship – Purchase a keyword
 - AdWords – Manage your Ad text

I'd like to buy a Keyword

The advertiser's text-based ad will appear at the top of a Google results page whenever the keyword they have purchased is included in a user's search.

The ads appear adjacent to, but are distinguished from, the results listings.



The screenshot shows a Google search interface. At the top, the Google logo is on the left, and links for 'Advanced Search', 'Preferences', and 'Search Tips' are on the right. Below the logo is a search bar containing the text 'travel'. To the right of the search bar is a 'Google Search' button and a partially visible 'I'm' button. Below the search bar, a blue banner indicates 'Searched the web for travel' and 'Results 1 - 10 of about 34,500,'. The first search result is a sponsored link from Expedia.com, titled 'Travel -- Plan the Right Trip at the Right Price on Expedia.com' with the URL 'www.expedia.com' and the text 'Plan the right trip at the right price. Travel Right on Expedia.' The words 'Sponsored Link' are circled in red to the right of the ad. Below the ad is a search result from the US State Department, titled 'US State Department - Services - Travel Warnings & Consular ...' with a description about FOIA, travel warnings, and consular information sheets.

Google™ [Advanced Search](#) [Preferences](#) [Search Tips](#)

travel Google Search I'm

Searched the web for travel Results 1 - 10 of about 34,500,

Travel -- Plan the Right Trip at the Right Price on Expedia.com
www.expedia.com Plan the right trip at the right price. Travel Right on Expedia. Sponsored Link

[US State Department - Services - Travel Warnings & Consular ...](#)
contact us FOIA. **Travel** Warnings & Consular Information Sheets. What Are **Travel** Warnings, Consular Information Sheets & Public Announcements? ...
Description: **Travel** warnings and consular information sheets

Category purchase

- **Google uses a classification system to create an ongoing "Virtual Directory" of categories an advertiser can purchase.**
- **Advertisers can select the categories most appropriate to their business and Google will match the most relevant category ads to each user's search.**
- **The advantage of this approach is that it covers a broader audience that might be missed through the purchase of keywords alone.**