

Web Search Basics

Introduction to Information Retrieval
Informatics 141 / CS 121
Donald J. Patterson

Content adapted from Hinrich Schütze
<http://www.informationretrieval.org>



Overview

- Introduction
- Classic Information Retrieval
- Web IR
- Sponsored Search
- Web Search Basics
 - Size of the Web
- Web Users
- Spam

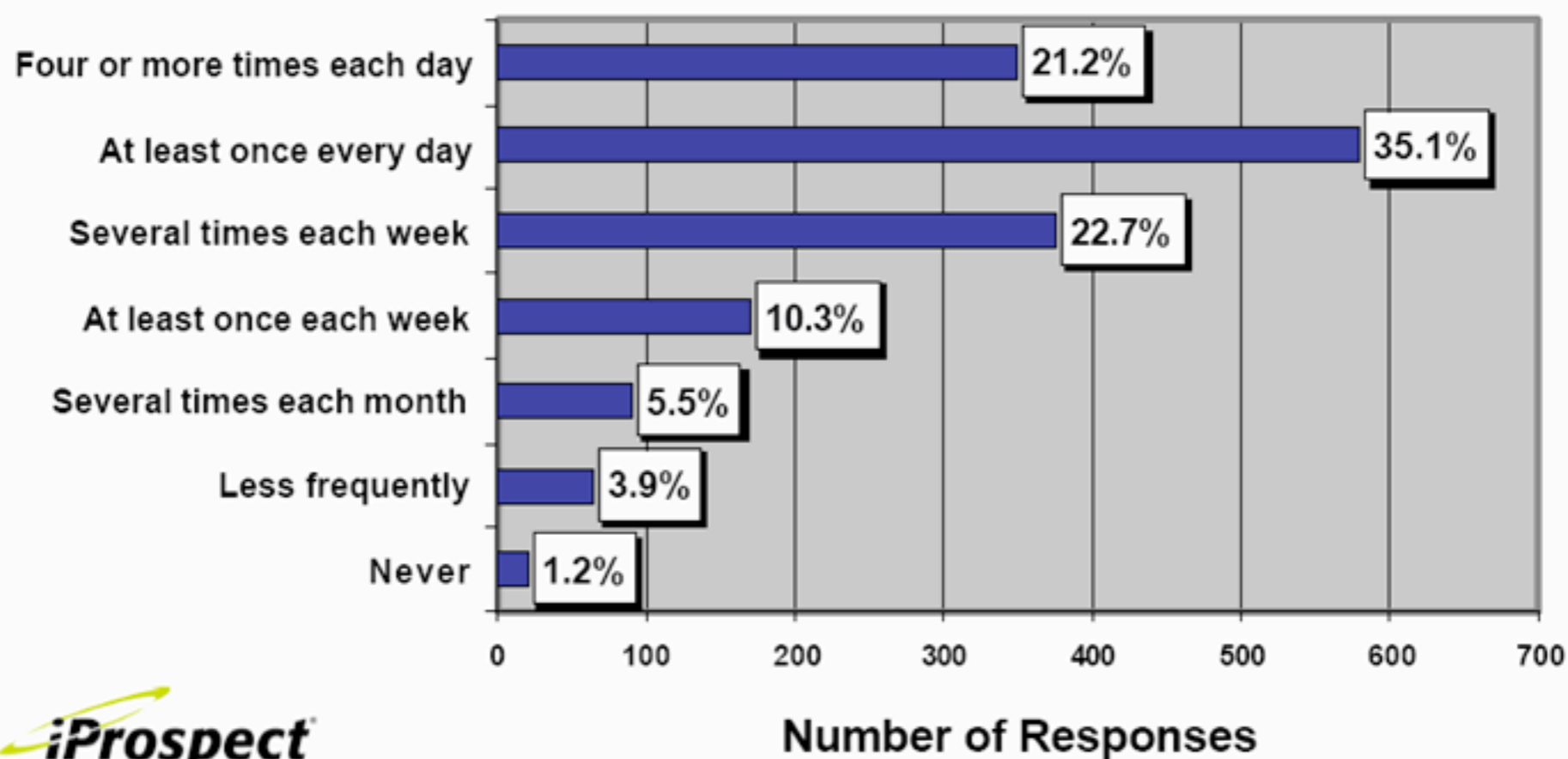


Search use ...

(iProspect Survey, 4/04,

<http://www.iprospect.com/premiumPDFs/iProspectSurveyComplete.pdf>)

How often do you use search engines on the Internet?



Without search engines the web wouldn't scale

- No incentive in creating content unless it can be found.
- Taxonomies, bookmarks can't keep up
- Or can they? (de.licio.us)
- The web is both a technology artifact and a social environment
- “The Web has become the ‘new normal’ in the America way of life; those who don't go online constitute an every-shrinking minority” [Pew Foundation report, January 2005]



Without search engines the web wouldn't scale

- Search engines make aggregation of interest possible:
 - Create incentives for very niche players
 - Economical - specialized stores, providers, etc.
 - Social - narrow interests, specialized communities
- The acceptance of search interaction makes “unlimited selection” stores possible
 - Amazon, Netflix, etc.



Without search engines the web wouldn't scale

- Search turned out to be the best mechanism for advertising on the web, a \$15 billion plus industry.
- Growing very fast (entire US advertising industry is \$250 billion though)
- Sponsored search marketing is about \$10 billion



Overview

- Introduction
- Classic Information Retrieval
- Web IR
- Sponsored Search
- Web Search Basics
 - Size of the Web
- Web Users
- Spam



Classic IR assumptions

- Corpus: Fixed document collection
- Goal: Retrieve information content relevant to information need



Classic IR Goal

- Classic “Relevance”
 - For each query, Q , and stored document, D , in a corpus there exists a relevance score: $R(Q,D)$
 - $R(Q,D)$ is averaged over users, U , and contexts, C
 - Maximize $R(Q,D)$ instead of $R(Q,D,U,C)$
 - Context is ignored
 - Individuals are ignored
 - Corpus is static



Overview

- Introduction
- Classic Information Retrieval
- Web IR
- Sponsored Search
- Web Search Basics
 - Size of the Web
- Web Users
- Spam



Web IR: Differences from traditional IR

- On the web, search and ads are intricately connected
- The web is huge
- The web is a rapidly changing collection.
- There is spam on the web
 - Adversarial IR
 - Huge difference from traditional IR
- One interface for hugely divergent needs
 - Queries, Maps, Stocks, Weather, Calculations



History

- Early keyword-based engines
 - (1995-1997) Altavista, Excite, Infoseek, Inktomi
- Paid placement ranking
 - Goto.com -> Overture.com -> Yahoo!
 - Results based on auction for keyword placement



Wilmington real estate.

Access 75% of all users now!
Premium Listings reach 75% of all
Internet users. [Sign up](#) for Premium
Listings today!

ib of
ow!
stings
of all
ers.
stings

1. [Wilmington Real Estate - Buddy Blake](#)
Wilmington's information and real estate guide. This is your on
anything to do with Wilmington.
www.buddyblake.com (Cost to advertiser: [10.38](#))
2. [Coldwell Banker Sea Coast Realty](#)
Wilmington's number one real estate company.
www.cbseacoast.com (Cost to advertiser: [10.37](#))
3. [Wilmington, NC Real Estate Becky Bullard](#)
Everything you need to know about buying or selling a home c
on my Web site!
www.iwwc.net (Cost to advertiser: [10.35](#))

History

- (1998+) Link-based ranking pioneered by Google
 - Links added the idea of “authoritativeness” to “relevance”
 - Blew away all early engines save Inktomi
 - Great user experience looking for a business model
 - Meanwhile Goto/Overture’s annual revenues were nearing \$1 billion



History

- Result
 - Google:
 - Added paid placement ads on the side
 - Differentiated from search results
 - Yahoo! built a similar architecture
 - Buys Overture for paid placement
 - Buys Inktomi for search



Overview

- Introduction
- Classic Information Retrieval
- Web IR
- Sponsored Search
- Web Search Basics
 - Size of the Web
- Web Users
- Spam



Sponsored Search



search engine optimization

Search

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

Web [Blogs](#) [News](#)

Personalized Results 1 - 1000,000 for [search engine optimization](#) (seconds)

[Search Engine Optimize](#)

[SEOP.com](#) Guaranteed Top Ranking w/ Warranty. Free Site Analysis! 877-231-155

[Guaranteed Page 1 Ranking](#)

[www.berankednumber1.com](#) Guaranteed Page 1 Rankings \$49.95 No Charge Until You are on Page 1

[Search engine optimization - Wikipedia, the free encyclopedia](#)

Search engine optimization (SEO) is the process of improving the volume and quality of traffic to a web site from **search** engines via "natural" ("organic" or ...

[en.wikipedia.org/wiki/Search_engine_optimization](#) - 87k - [Cached](#) - [Similar pages](#) - [Note this](#)

[Search Engine Optimization, Google Optimization - SEO Chat](#)

Search Engine Optimization, Google Optimization - SEO Chat.

[www.seochat.com/](#) - 111k - [Cached](#) - [Similar pages](#) - [Note this](#)

[Search Engine Optimization \(SEO\) Marketing Firm & Placement Company](#)

Offers **search engine optimization (SEO)** marketing services & placement since 1998.

Submit your website URL to 40 major **search** engines for FREE!

[www.submitexpress.com/](#) - 42k - [Cached](#) - [Similar pages](#) - [Note this](#)

[News results for search engine optimization](#)



[CIBER Selected as E-Commerce Vendor by Elite Island Resorts](#) - Jan 3, 2008

Their **search engine** marketing program will help us lower acquisition costs ... CIBER's advanced **search engine** marketing services will help Elite direct more ...

[FOX News](#) - [10 related articles](#) »

[bruceclay.com - Search Engine Optimization - SEO Training, Tools ...](#)

Search Engine Optimization, ranking, placement, and submission tutorial. Free step-by-step **SEO** tools and advice. **SEO** training and services offered. ...

[www.bruceclay.com/web_rank.htm](#) - 87k - [Cached](#) - [Similar pages](#) - [Note this](#)

[Inteliture™ Search Engine Optimization, Internet Marketing, and ...](#)

Inteliture™ a professional **search engine optimization** and internet marketing company.

Offers internet marketing solutions, **search engine optimization** ...

[www.inteliture.com/](#) - 12k - [Cached](#) - [Similar pages](#) - [Note this](#)

Ads

Ads

Algorithmic Results

[Search engine optimization](#)

Use Network Solutions online tools to drive business to your web site.

[marketing.networksolutions.com](#)

[Search Optimization Firm](#)

Looking for top rankings? Get real results. Receive a free analysis.

[www.customermagnetism.com](#)

[SEO Company](#)

Search Engine Optimization services since 1998 with proven results.

[www.iClimber.com](#)

[Get Optimization Help Now](#)

Top SEO Firms Want Your Business. Fast, Free Competitive Quotes!

[www.TopSeos.com/SEO](#)

[Check your SEO for Free](#)

PPC vs Natural **search** Keyword ranks costs & robot stats: 15 days free

[www.ClickTracks.com/15_Days_Free](#)

[Search Engine Marketing](#)

Boost Online Traffic and Sales! Free Site **Optimization** Analysis.

[www.corporatesearchoptimization.com](#)

[Free Website Visitors](#)

Free Visitors Plus Top 10 Positions In 8 Hours! FREE Trial Offer.

[www.EngineSeeker.com](#)

Ads vs. Search Results

- Google has maintained that ads (based on vendors bidding for search queries) do not affect vendors ranking in search results

Sponsored Links

[Search engine optimizer](#)

Use Network Solutions online tools to drive business to your web site.
marketing.networksolutions.com

[Search Optimization Firm](#)

Looking for top rankings? Get real results. Receive a free analysis.
www.customermagnetism.com

[SEO Company](#)

Search Engine Optimization services since 1998 with proven results.
www.iClimber.com

[Search engine optimization - Wikipedia, the free encyclopedia](#)

Search engine optimization (SEO) is the process of improving the volume and quality of traffic to a web site from search engines via "natural" ("organic" or ...

en.wikipedia.org/wiki/Search_engine_optimization - 87k - [Cached](#) - [Similar pages](#) - [Note this](#)

[Search Engine Optimization, Google Optimization - SEO Chat](#)

Search Engine Optimization, Google Optimization - SEO Chat.

www.seochat.com/ - 111k - [Cached](#) - [Similar pages](#) - [Note this](#)

[Search Engine Optimization \(SEO\) Marketing Firm & Placement Company](#)

Offers **search engine optimization (SEO)** marketing services & placement since 1998.

Submit your website URL to 40 major **search** engines for FREE!

www.submitexpress.com/ - 42k - [Cached](#) - [Similar pages](#) - [Note this](#)

[News results for search engine optimization](#)



[CIBER Selected as E-Commerce Vendor by Elite Island Resorts](#) - Jan 3, 2008

Their **search engine** marketing program will help us lower acquisition costs ... CIBER's advanced **search engine** marketing services will help Elite direct more ...

[FOX News](#) - [10 related articles »](#)

Ranking of ads

- Other search engines (Yahoo!, MSN) have made similar statements on occasion
 - Any of them can change at any time
 - Facebook is currently testing the waters in their “Newsfeeds”
- We will ignore the possibility of paid placement ads being interspersed in search results.



Ranking of ads

- Goto model:
 - Rank according to how much advertiser pays
- Current model:
 - Balance auction price and relevance
 - Irrelevant ads (few click-throughs)
 - Decrease opportunities for relevant ads
 - Harm the user experience
 - Idea: Well-targeted advertising is good for everyone

