

INFORMATION RETRIEVAL

Search Engine Optimization

Paid Ad Placement

- It costs money
- So...

Search Engine Optimization (SEO)

- Tuning of a web page/site to rank highly in search results for certain queries
- Alternative to paying for ad placement
 - ▣ Bonus: visibility through rank, people trust
- It's marketing: getting your content to your audience

SEO

□ Motives

- Commercial
- Political
- Religious
- Lobbying

□ Who does this?

- Internally: webmasters, writers
- Commercially: companies, consultants
- Hosting services
- Plugins to popular CMSs

SEO

□ How to do it:

- ▣ <https://support.google.com/webmasters/answer/40349?hl=en>
- ▣ <http://freetools.webmasterworld.com/category/seo-tools>

SEO



- Ethical and unethical ways of doing it
- Legitimate approach:
 - ▣ Indexed age of the pages (older is better)
 - ▣ Good incoming links
 - ▣ Good content, well written, well organized, up to date
 - ▣ Good use of web standards/practices

SEO

- Unethical approaches (aka spam):
 - ▣ fake pages
 - ▣ fake sites that point to your site
 - ▣ fake comments/engagement
 - ▣ in short: “alternative facts” aka lies
- Sometimes the line between legitimate and illegitimate practices is hard to find. There’s a large grey area

SEO

- Search results (also) depend on which data center receives the query
 - ▣ google.com vs. google.fr vs google.pt
 - ▣ e.g. “Paris”



Coercion Techniques

Keyword Stuffing

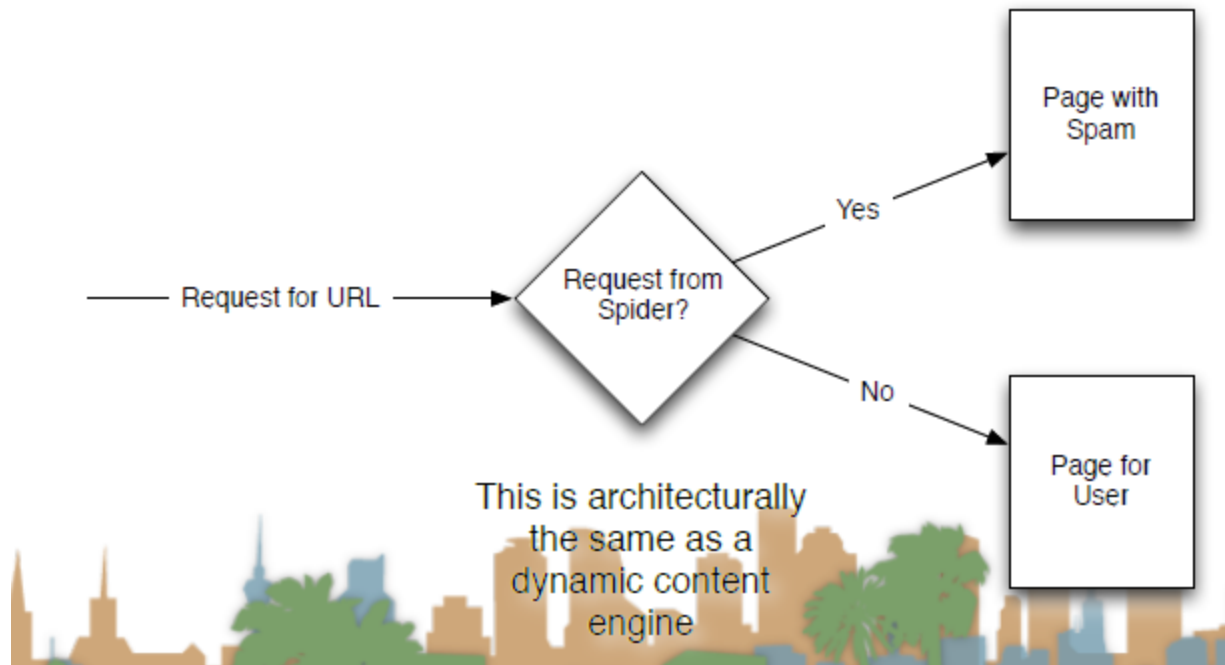
- 1st- gen search engines relied heavily on textual content and frequency of words
- SEO moved to play around with keywords
 - ▣ Misleading meta-tags

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.d
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
<COMMENT TITLE="MONITOR"></COMMENT>
<meta http-equiv="Content-Language" content="en-us" />
<meta http-equiv="Content-type" content="text/html; charset=iso-8859-1"/>
<META NAME="ROBOTS" CONTENT="NOODP"><meta name="verify-v1" content="aeVxP6zTHeQzT620ipj5+ikXd/VXcdlKoYUJ/C6vVdY=" />
<META NAME="keywords" content="Expedia, Travel, Cheap Airfare, Car, Hotels, Vacations, Airfare, Car Rental, Cruises,
<META NAME="description" content="Purchase airline tickets, make hotel reservations, find vacation packages, car rent
```

- ▣ Repeating words over and over
- ▣ Playing games with colors (white on white)
 - visible to spiders but not users

Cloaking

- Serving different content to a spider than to a browser



Doorway Pages

- Like cloaking but using a redirect (302)
 - ▣ Initial page is optimized for spider, then redirect takes user to actual content

Link exchanges

- ❑ I link to you, you link to me
- ❑ “Translations”
- ❑ Universities and professors are targets



Spam

Link spamming

- Bots that search for blogs and leave comments with links

Clicker bots

- Bots that issue queries and “click” on targeted query results

Spam industry

The First Page of Google and How I Get my Clients There

www.malleeblue.com/1st-page-google-optimization-tips/

★★★★★ Rating: 10/10 - 275 reviews

Nov 8, 2016 - Can I get your website or page to rank on the first page of Google? ... If you want to get 1st page of Google search engine results listing, make ...

First Page of Google GUARANTEED !! | How To Get Your Business At ...

https://www.youtube.com/watch?v=qjASS_NslL8

Dec 26, 2013 - Uploaded by Spotlight Ventures

Business on The First Page of Google GUARANTEED !! or Goto: ... video, blog, or Business Listing on the ...

How To Get On The First Page of Google In 24 hours - YouTube



https://www.youtube.com/watch?v=2O_pdMbjR3k

May 15, 2016 - Uploaded by Amazing Tricks World

how to get first page rank on google how to get first rank in google search how to rank first on google how ...

Spamming Web contests

- <https://www.wholewhale.com/tips/how-to-win-or-cheat-any-online-voting-contest/>

The war on spam

- Quality indicators
 - ▣ Statistical analysis of links
 - ▣ Votes
- Usage indicators
 - ▣ Analytics
- Anti-bot mechanisms
 - ▣ Captchas



The war on spam

- Limits on meta keywords
- Spam recognition by machine learning
- Family-friendly filters (humans and ML)
- Robust link analysis
 - ▣ Ignore statistically improbable links
 - ▣ Detect cycles
 - ▣ Use link analysis to detect spammers
 - Guilt by association
- Editorial intervention

Webmaster Guidelines

- Search engines have SEO policies
 - ▣ What is allowed and not allowed
- Must not be ignored
 - ▣ Once a site is blacklisted by a search engine, it will virtually disappear from the Web