Semantic Web y SEO
Historia de los buscadores
Archie (or "Archive" without the "v")
1990
• First search engine
• FTP site hosted an index of downloadable directory listings.
• Due to limited space, only the listings were available and not the contents for each site.
Infoseek
January 1994
• Webmasters could submit a page in realtime
December 95: Netscape began using them as their default search

E|Net Galaxy
January 1994
• Efficient in its utilization of different web search features

1994

Altavista
January 1994
• Unlimited bandwith (for the time)
• First to allow natural language queries
• Advanced searching techniques
• Add or delete your own URL within 24 hours
• Search tips and new features
• Inktomi and Google sent them into irrelevancy

February 2003: Overture intends to purchase Altavista for $80 million in stock, $60 million in cash
• Yahoo bought Overture in 2003
WebCrawler
April 1994
- First crawler that indexed entire pages
- Too popular to be used during daytime
- June 1995: AOL buys WebCrawler
- 1997: Excite bought out WebCrawler
- AOL begins using Excite

Lycos
July 1994
- Went public with catalog of 54,000 documents
- Ranked Relevance retrieval

Yahoo! Directory
April 1994
- Created by David Filo and Jerry Yang
- Began as a collection of favorable web pages
- Increasing size influenced them to become a searchable directory
- A man-made description with each URL
- Informational sites added for free, but they expanded to include commercial sites. This is still available for $300 a year.
- Long wait time to be included
- Yahoo! didn't work on their search engine until 2002 when they began acquiring other search directories. Until that time they had outsourced their search services
**Inktomi: Hotbot**

May 1996
- Search engine Hotbot
- Listed on Hotwire
- October 2001: Inktomi exposed for accidentally allowing public to access database of spam sites (over one million)
- Inktomi pioneered paid inclusion model
- Not as efficient as pay per click by Overture
- **December 2003:** Yahoo bought them out for $235 million

**Google**

January 1996
- Larry and Sergey began working on BackRub, a search engine which utilized backlinks for search
- It ranked pages using citation notation, meaning any mention of a website on another site would count it as a vote toward the mentioned site
- A website’s “authority” or reliability came from how many people linked to that site, and how trustworthy the linking sites were
Cómo funcionan los buscadores web y sus algoritmos
THE PERIODIC TABLE OF SEO RANKING FACTORS

Search engine optimization — SEO — seems like alchemy to the uninitiated. But there’s a science to it. Search engines reward pages with the right combination of ranking factors. Below, some major factors or “signals” you should seek to have.

ON THE PAGE SEO RANKING FACTORS
These elements are in the direct control of the publisher.

<table>
<thead>
<tr>
<th>CONTENT</th>
<th>HTML</th>
<th>ARCHITECTURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cq</td>
<td>Ht</td>
<td>Ac</td>
</tr>
<tr>
<td>Cq QUALITY</td>
<td>Ht Titles</td>
<td>Ac Crawl</td>
</tr>
<tr>
<td>Cr</td>
<td>Hd</td>
<td>As</td>
</tr>
<tr>
<td>Cr RESEARCH</td>
<td>Hd Description</td>
<td>As Speed</td>
</tr>
<tr>
<td>Cw</td>
<td>Hh</td>
<td>Au</td>
</tr>
<tr>
<td>Cw Words</td>
<td>Hh Headers</td>
<td>Au URLs</td>
</tr>
<tr>
<td>Cf</td>
<td>Ln</td>
<td>Pl</td>
</tr>
<tr>
<td>Cf FRESH</td>
<td>Ln Numbers</td>
<td>Pl Locality</td>
</tr>
</tbody>
</table>

FACTORS WORK TOGETHER
Some factors influence more than others, which we’ve weighted 1 (weakest) to 3 (strongest). No single factor guarantees top rankings. Several favorable factors increase odds of success. Negative “Violation” and “Blocking” factors harm your chances.

VIOLATIONS
Spam techniques may cause ranking penalty or ban.

<table>
<thead>
<tr>
<th>VIOLATIONS</th>
<th>BLOCKING</th>
</tr>
</thead>
<tbody>
<tr>
<td>VT THIN</td>
<td>VT</td>
</tr>
<tr>
<td>VS STUFFING</td>
<td>VT</td>
</tr>
<tr>
<td>VC CLOAKING</td>
<td>VP Paid Links</td>
</tr>
<tr>
<td>VH HIDDEN</td>
<td>VI Link Spam</td>
</tr>
</tbody>
</table>

VI

LINES
Are pages well written & have substantial quality content?

RESEARCH
Have you researched the keywords people may use to find your content?

WORDS
Do pages use words & phrases you hope they'll be found for?

ENGAGE
Do visitors spend time reading or “bounce” away quickly?

FRESH
Are pages fresh & about “hot” topics?

TITLES
Do HTML title tags contain keywords relevant to page topics?

DESCRIPTION
Do meta description tags describe what pages are about?

HEADERS
Do headlines and subheads use header tags with relevant keywords?

CRAWL
Can search engines easily “crawl” pages on site?

SPEED
Does site load quickly?

URLS
Are URLs short & contain meaningful keywords to page topics?

QUALITY
Lq
Are links from trusted, quality or respected web sites?

TEXT
Lt
Do links pointing at pages use words you hope they'll be found for?

NUMBER
Ln
Do many links point at your web pages?

REPUTATION
Sr
Do those respected on social networks share your content?

SHARES
Ss
Do many share your content on social networks?

AUTHORITY
Ta
Do links, shares & other factors make site a trusted authority?

HISTORY
Th
Has site or its domain been around a long time, operating in same way?

COUNTRY
Pc
What country is someone located in?

LOCALITY
Pl
What city or local area is someone located in?

HISTORY
Ph
Does someone regularly visit the site? Or “liked” it?

SOCIAL
Ps
What do your friends think of the site?

BLOCKING
If searchers “block” site, hurts both trust & personalization.

BLOCKING
Bt
Have many people blocked your site from search results?

Bp
Has someone blocked your site from their search results?

PAID LINKS
Vp
Have you purchased links in hopes of better rankings?

LINK SPAM
VI
Have you created many links by spamming blogs, forums or other places?
Tácticas de SEO
No técnicas

- escribir buen contenido
- escribir buenos title tags
- escribir meta descriptions
- links con buen anchor text relevante
Técnicas

- XML sitemaps
- Eliminar errores y 404
- Velocidad de carga
- Buen Markup HTML (semántico)
Web semántica
Microformatos
Schema.org
<p>Calle 4 Col. Independencia</p>

vs

<address>Calle 4 Col. Independencia</address>
Tipos de web semántica

RDFa
Microformatos
Microdata
RDFa
Microformatos
- Sin Microformato:
  <ul>
    <li>Joe Doe</li>
    <li>The Example Company</li>
    <li>604-555-1234</li>
    <li><a href="http://example.com/">http://example.com/</a></li>
  </ul>

- Con Microformato:
  <ul class="vcard">
    <li class="fn">Joe Doe</li>
    <li class="org">The Example Company</li>
    <li class="tel">604-555-1234</li>
    <li><a class="url" href="http://example.com/">http://example.com/</a></li>
  </ul>
Microdata
Hello, my name is John Doe, I am a graduate research assistant at the University of Dreams. My friends call me Johnny. You can visit my homepage at www.JohnnyD.com. I live at 1234 Peach Drive Warner Robins, Georgia.
Hello, my name is John Doe, I am a graduate research assistant at the University of Dreams. My friends call me Johnny. You can visit my homepage at www.JohnnyD.com.

I live at 1234 Peach Drive, Warner Robins, Georgia.
Schema.org
<div itemscope itemtype="http://schema.org/Movie">
  <h1 itemprop="name">Avatar</h1>
  <div itemprop="director" itemscope itemtype="http://schema.org/Person">
    Director: <span itemprop="name">James Cameron</span> (born <time itemprop="birthDate" datetime="1954-08-16">August 16, 1954</time>)
  </div>
  <span itemprop="genre">Science fiction</span>
  <a href="../movies/avatar-theatrical-trailer.html" itemprop="trailer">Trailer</a>
</div>
Ventajas de la web semántica

- poder generar un vCard de contacto para la agenda de tus clientes

- poder indicar navegacion en tu sitio para los buscadores

- poder indicar la ubicacion de tu negocio o evento

- que los buscadores encuentren tus recetas facilmente
Ventajas de la web semántica
Ventajas de la web semántica

- Location based search
- Quick links to popular pages
- Google Maps
- Street Address, taken from Google Places
- Reviews
Ventajas de la web semántica
Gracias

Contacto:

Gabriel Saldaña

gabriel@gabrielsaldana.org

Twitter/Identi.ca: @gabrielsaldana