

Google's Text Analytics War on Spam

*Sentiment Analysis Symposium
October 30, 2012*



Mike Moran
Chief Strategist, Converseon

Black Hat SEO experts fool everybody

- They fool Google
- They fool searchers
- They are extremely smart and hardworking
- In the last two years, Google has declared war on spammers



The search ranking algorithm decides who's first

- This is not about paid search results
- How does Google decide the #1 organic result?

Google search results for "digital cameras".

Search results include:

- Sony® Digital Cameras | store.sony.com**
store.sony.com/Cybershot
Digital Cameras for Beautiful Pictures. Free Shipping Order Now!
- Digital Cameras - Large Brand Selection in Stock Now.**
www.staples.com
staples.com is rated ★★★★★ 8,426 reviews
Free Shipping on All Orders \$45+
- 10 Top Digital Cameras | Amazon.com**
www.amazon.com/Cameras
amazon.com is rated ★★★★★ 7,043 reviews
Save on 10 top digital cameras Free Shipping Available with Amazon

Related searches for digital cameras:
Stores: Best Buy Amazon Ritz Camera Overstock Walmart
Brands: Olympus Canon Sony Nikon Kodak

Cameras: Digital Cameras & Accessories - Best Buy
www.bestbuy.com/site/Cameras.../Digital-Cameras/abcat0401000.c?i...
Shop digital cameras at Best Buy and capture picture-perfect memories with digital cameras and accessories designed to meet your needs.
Point & Shoot Cameras - Digital Camera Package Deals - Nikon logo

89% of clicks on page 1

Google won the first spam war with link analysis

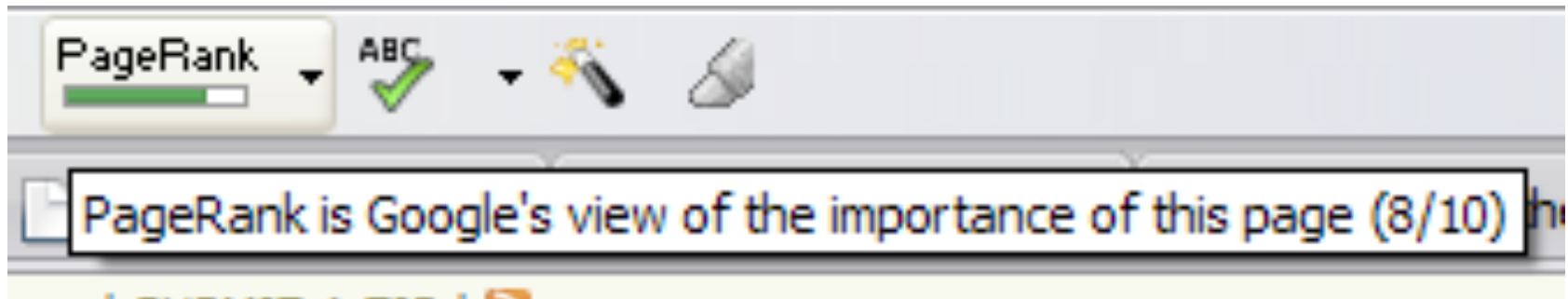
- Links are votes for quality
- They mimic citations in scientific research papers
- The best papers have the most citations



How do search engines value links?

- The most links
- From the best sites
- With the right anchor text

**PageRank is
Google's name for
its page factor
ranking**



Some people reverse engineer the algorithm

- Google wants the keywords on the page
- Google wants the keywords in the title
- Google wants links to the site



**What Google really wants
is high quality search results
for the searchers**

The kinds of links Google has declared war on

- Spam techniques
 - ▶ Blog spamming
 - ▶ Link farms
 - ▶ Hidden links
- Paying for links
- Negative links



There's no such thing as bad publicity?

- Some clever characters found that *bad* reviews actually improved their search rankings?
- Google was embarrassed by a New York Times story outlining the practice

HOME PAGE | TODAY'S PAPER | VIDEO | MOST POPULAR | TIMES TOPICS

The New York Times **Business Day**

WORLD | U.S. | N.Y. / REGION | BUSINESS | TECHNOLOGY | SCIENCE | HEALTH | S

Search Global DealBook Markets Economy

A Bully Finds a Pulpit on the Web

By DAVID SEGAL
Published: November 26, 2010

SHOPPING online in late July, Clarabelle Rodriguez typed the name of her favorite eyeglass brand into [Google's](#) search bar.

In moments, she found the perfect frames — made by a French company called Lafont — on a Web site that looked snazzy and stood at the top of the search results. Not the tippy-top, where the paid ads are found, but under those, on Google's version of t gold-medal podium, where the most relevant and popular site is displayed

Ms. Rodriguez placed an order for



Enlarge This Image

David G. Klein

Sentiment analysis to the rescue?

- Google has not publicly admitted this...
- ...but sentiment analysis seems to be in use
- Good reviews now appear to be votes for quality but bad reviews do not
- Google might be reticent to discuss this, because it becomes a new spam technique



For awhile, Google was showing reviews

- You could sort by reviews and sentiment analysis extracted interesting reviews

[Drooling Dog Bar B Q - Colfax, CA](#)

★★★★☆ 15 reviews - Price range: \$\$

Drooling Dog has some really good BBQ. I had the pulled pork sandwich, Drooling Dog BBQ is a great place to stop at on your way up the hill to Tahoe ...

www.yelp.com/biz/drooling-dog-bar-b-q-colfax - 75k - [Cached](#) - [Similar pages](#)

- But now they show Zagat ratings

[McCormick & Kuleto's](#)

www.mccormickandkuletos.com/

Zagat: **22** / 30 - 687 Google reviews

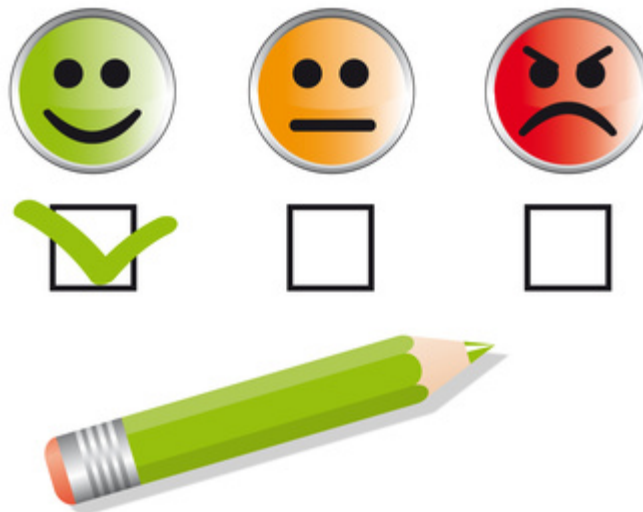
[Woodhouse Fish Company](#)

woodhousefish.com/

Zagat: **22** / 30 - 336 Google reviews

Google keeps writing papers and filing patents

- Summarizing local reviews with sentiment
- Classifying reviews with sentiment
- Analyzing sentiment differently by domain



They clearly have sentiment analysis technology


Google Developers

Google Prediction API X Search



Home Products Events Showcase Live Groups

Google Prediction API

- ▶ Getting Started
- ▶ Using Prediction
- ▶ Reference
- ▶ Tools and Resources
- ▶ Feedback and Discussion
- ▶ Related APIs
- FAQ

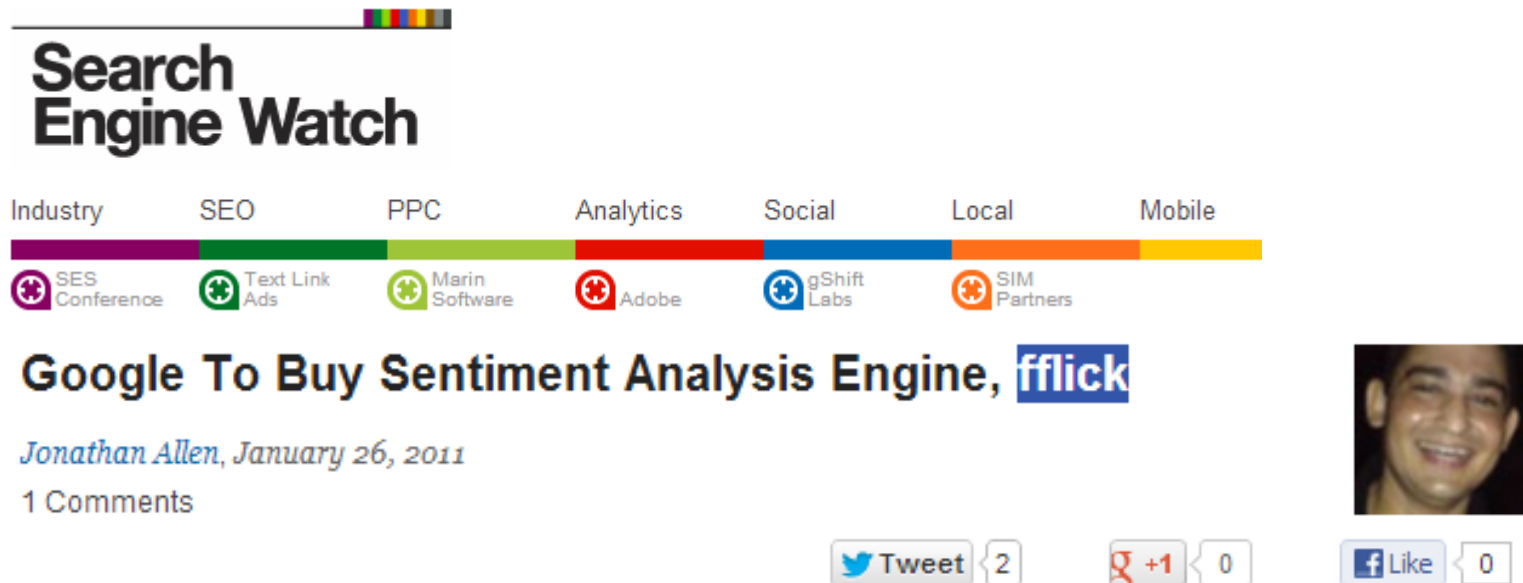
 What is the Google Predictic

Google's cloud-based machine learning tools can help an

-  Customer sentiment analysis
-  Message routing decisions

And they acquired an interesting company

- Movies are recommended based on user conversations
- But sentiment analysis can't fight all the spam links



The screenshot shows a Search Engine Watch article. At the top, the title "Search Engine Watch" is displayed in a large, bold font. Below the title, a horizontal bar categorizes the article into several areas: Industry, SEO, PPC, Analytics, Social, Local, and Mobile. Each category is represented by a colored segment and a corresponding icon. Below the bar, the article title "Google To Buy Sentiment Analysis Engine, fflick" is shown in a large, bold font. The author's name "Jonathan Allen" and the date "January 26, 2011" are listed below the title. The article has "1 Comments". At the bottom of the article, there are social media sharing buttons for Twitter (2 tweets), Google+ (+1), and Facebook (Like).

Search Engine Watch

Industry SEO PPC Analytics Social Local Mobile

SES Conference Text Link Ads Marin Software Adobe gShift Labs SIM Partners

Google To Buy Sentiment Analysis Engine, **fflick**

Jonathan Allen, January 26, 2011

1 Comments

Tweet 2 +1 0 Like 0

If social activity doesn't match link analysis...

- ...it seems more suspicious to Google as spam
- It's difficult for spammers to simulate both



Maybe we can use old-fashioned sentiment—ask!

- Human beings rate a small subset of search results:
 - ▶ Nice design?
 - ▶ Speedy response?
 - ▶ Quality content?
 - ▶ Would you return?
- Sites that people like get bumped higher in ranking
- The sites they don't like are shoved down



Machine learning scales the human ratings

- Even Google can't afford human ratings for every page for every search
- So, it looks for patterns—common features
- If your site looks like the low-rated sites, your site gets ranked lower



What does it mean to “look like” low quality?

- Machine learning detects patterns in what pages look like based on “features”
- What could the features be?
 - ▶ The length of the title tag
 - ▶ The ratio of words to pictures
 - ▶ The commonality of runs of words



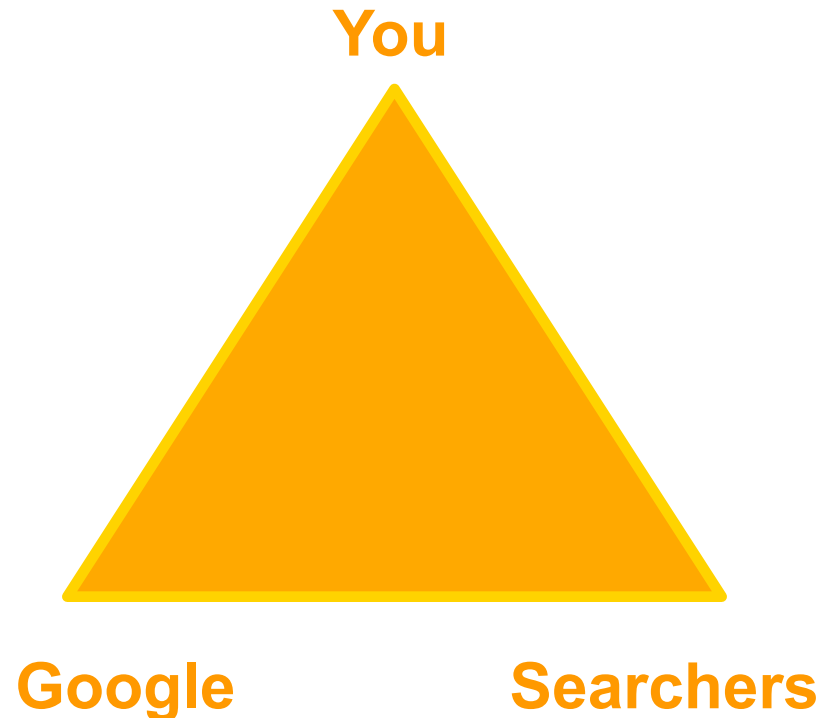
What is the practical effect of Panda?

- Sites that ranked highly with the old algorithm have been affected
- If your site was great for search engines, but not for actual people, time to up your game
- Who seemed to get hit?
 - ▶ “Content farms” and screen-scrapers
 - ▶ Older content
 - ▶ Sites loaded with ads
 - ▶ Vertical search sites—but not Google sites!



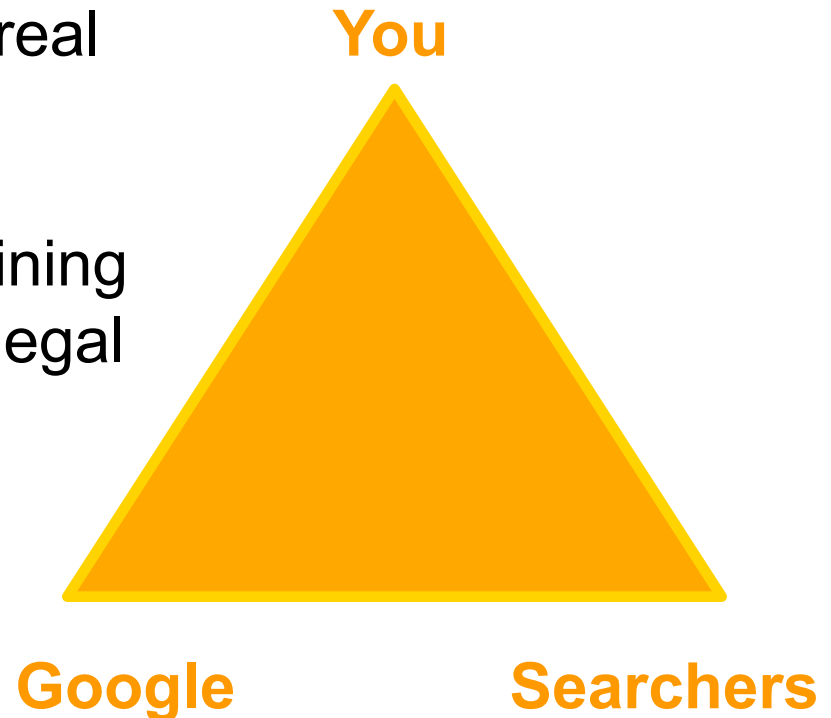
That wasn't just bad luck

- Tactics that help everyone get rewarded
- Tactics that help only you don't
- Maybe you should stick to what will work long-term



It's not just about spamming

- Most of you don't practice real black hat SEO
- But you don't win by examining the terms of service like a legal document
- You see what works for all three parties



Is Panda over? The world has now stabilized?

- Google will continue to roll out algorithm changes based on:
 - ▶ Panda feature changes and human rating changes
 - ▶ Google +1 button ratings
 - ▶ Other content quality assessments, including social media activity
- If Panda is working (and Google seems to be committed to it), expect Microsoft and other engines to follow



Google's newest update targets over-optimization

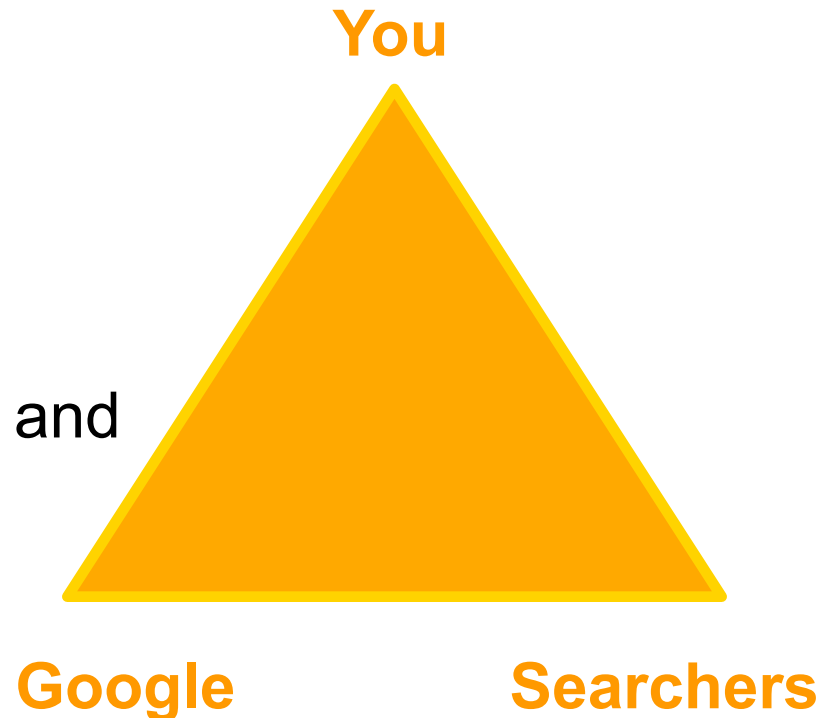
- Unusual link patterns
- Keyword stuffing
- Cloaking
- Content spinning

**If people told you this was
SEO, it's not anymore**



Don't be a good scout. Be good at business.

- You're not trying out for sainthood
- Do it because it works
- It's less risky, easier to do, and has fewer ups and downs
- Focus on a quality searcher experience, and success will follow

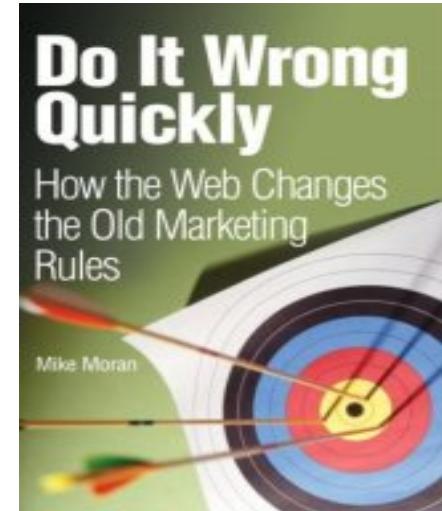
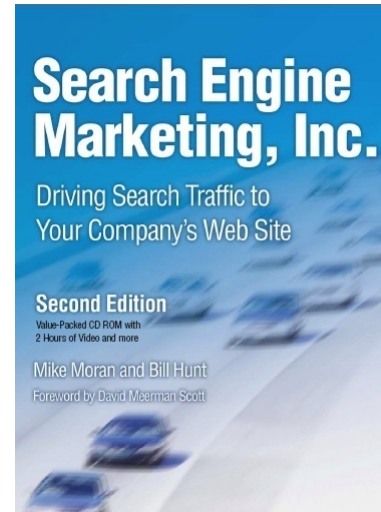


Thank you!

www.mikemoran.com



www.converseon.com



WOMMIE Award
Best Word of Mouth Program



OMMA Award
Best Use of Virtual Worlds



SAMMY Awards
Best Social Media Agency



Forrester Groundswell
Category: B2C Talking



Award of Excellence
Communications Research



PR News Award
Best Use of Virtual Worlds



Silver Quill Award of Merit
Heritage Region



Platinum Award
Blog Category



Webby Award
Official Honoree



Platinum Award
Podcast/Videocast



Online Publishing
Division



2009 Golden Quill Award
Social Synchronous
Technology Launch