About me

• Chief Data Scientist at Searchmetrics since mid-2012

• PhD in Computer Science at Technische Universität Berlin in 2011

• Interested in Information Retrieval, Natural Language Processing & Machine Learning – both research & applications
Study details

- **Dataset:**
  - Google.com
  - Top 3 SERPs
  - 10,000 Keywords (by search volume, excluding "navigational")
  - Organic results (including video integrations)
  - Jan 2014

- Analyzed 95 factors (2013: 60), and more to come
- Metric: Spearman's Rank Correlation Coefficient
- Averaged Values: according to the respective position(s) in the SERPs
Key observations (preliminary)

- Good on-page technical implementation is a must

- Relevant content that satisfies users' information needs matters!

- Backlink quality & diversity remain important
Overview
On-page Technical Implementation

Factor correlation - 2014

<table>
<thead>
<tr>
<th>Factor</th>
<th>Correlation 2014</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sitespeed</td>
<td>0.11</td>
</tr>
<tr>
<td>Existence of H1</td>
<td>0.05</td>
</tr>
<tr>
<td>Existence of Description</td>
<td>0.05</td>
</tr>
<tr>
<td>Video Integration</td>
<td>0.05</td>
</tr>
<tr>
<td>Keyword in Title</td>
<td>0.02</td>
</tr>
<tr>
<td>Keyword in Description</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Factor correlation - vs 2013

<table>
<thead>
<tr>
<th>Factor</th>
<th>Correlation 2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sitespeed</td>
<td>0.11</td>
</tr>
<tr>
<td>Existence of H1</td>
<td>0.05</td>
</tr>
<tr>
<td>Existence of Description</td>
<td>0.05</td>
</tr>
<tr>
<td>Video Integration</td>
<td>0.05</td>
</tr>
<tr>
<td>Keyword in Title</td>
<td>0.02</td>
</tr>
<tr>
<td>Keyword in Description</td>
<td>0.01</td>
</tr>
</tbody>
</table>
• Consistently high averages -> implies this to be a must-have (but results in low correlation)
On-page Technical Implementation

H1 existing

- A large fraction of pages have at least one H1 tag – lower for brands

\[ \rho_{2014} = 0.05 \]

\[ \rho_{2013} = 0.03 \]
On-page Technical Implementation
Sitespeed (Pageloadtime, by domain mean)

- Longer load times—why? Possible explanation follows...
- Higher correlation—more consistent “increase” in pageload times by position

\[ \rho_{2014} = 0.11 \]
\[ \rho_{2013} = 0.04 \]
Overview

On-page Content

Factor correlation - 2014

- Relevant Keywords: 0.34
- Proof Keywords: 0.20
- Keywords in Body: 0.15
- Text Character Length: 0.13
- Word Count: 0.13
- Flesch Reading Ease: 0.00
On-page Content
Text length in characters

- 12.5% increase in text length!

$\rho_{2014} = 0.13$

$\mu_{2014} \approx 7500$

$\rho_{2013} = 0.11$

$\mu_{2013} \approx 6700$

BrandFactor
On-page Content

Wordcount

- ≈20% increase in words

![Graph showing word count vs Google position with equations: \( \rho_{2014} = 0.13 \), \( \mu_{2014} \approx 650 \), \( \rho_{2013} = 0.11 \), \( \mu_{2013} \approx 540 \).]
On-page Content
Proof & relevant keywords

- Keyterms/phrases extracted from ranking URLs' contents
- Proof – keyterms occurring in most documents
- Relevant – other keyterms
Significant correlation – comparable to "# of backlinks" or "GP +1s"

Spearman Correlation - Study of Google search results
On-page Content

Relevant terms

Top results contain significantly more "relevant" terms than lower-ranked results.

\[ \rho_{2014} = 0.34 \]
Top results contain more "proof" terms than lower-ranked results

$\rho_{2014} = 0.20$
On-page Content

Why are proof and relevant terms important?

“You shall know a keyword by the company it keeps” (Firth, 1957)
On-page Content

Why are proof and relevant terms important?

Key terms and overall term distribution help to make your topic/intent understandable for Google’s/Bing’s algorithms

Source: Google
Overview

Backlinks

Factor correlations - 2014

- Number of Backlinks: 0.31
- Ø SEO-Visibility of backlinking URL: 0.26
- Number of ref. Domains to Homepage: 0.24
- New Backlinks: 0.20
- URL brand anchors: 0.15
- Domain brand anchors: 0.11
Backlinks

Number of backlinks

More Backlinks!

$\rho_{2014} = 0.31$
Backlinks
SEO Visibility of backlinks

BetterLinks!

$\rho_{2014} = 0.26$
Interesting content that people link to ranks from P3 onward?
Backlinks
Brand anchor texts

Higher percentage of brand anchors for top-ranking results

\[ \rho_{2014} = 0.15 \]
Summary: Key Observations

• Relevant content that satisfies users' information needs matters!
  – Significant increase in text length/word counts observed
  – Key terms & overall term distribution help to make your topic/intent understandable for search engine algorithms

• Good on-page technical implementation is a must

• Backlink quality & diversity remain important
  – Fresh links, high-quality domains, diversity…
Still to come

New factors & Faceted analyses

- Facets: Mobile, Industry sectors, more countries, ...
THANK YOU
FOR YOUR TIME

The Power for Success.
SEO, Social and Content intelligence to boost your revenue!