



1. Suchmaschinen und Information Retrieval
2. Architektur von Suchmaschinen
3. Evaluierung von Suchmaschinen
4. Retrieval-Modelle
5. Ranking mit Indexstrukturen
6. Textverarbeitung
7. Anfragen / Benutzerschnittstellen / Interaktion
8. Crawling und Texterfassung
9. Suchmaschinenoptimierung, Marketing, ...
10. Bilder und vertikale Suchlösungen

Grundlage der Vorlesung:

Croft, W Bruce; Metzler, Donald;  
Strohman, Trevor: Search Engines:  
Information Retrieval in Practice.  
International edition. Boston:  
Addison-Wesley / Pearson, 2009.

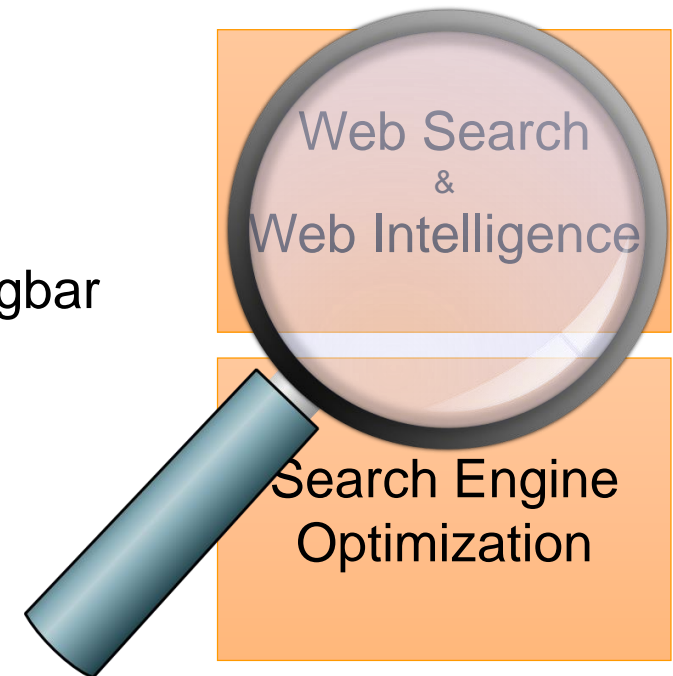
# 1 Web Intelligence



- Die Situation:
  - Milliarden von Dokumenten
  - Relativ wenig Dateiformate
  - Recht einfach zu analysieren
- Dann:
  - hervorragende Quelle zur Analyse
  - Informationen zum Teil sehr früh verfügbar
- Aber:
  - starke Qualitätsunterschiede
  - auch viele falsche Informationen



Suche im Web



## Diese Themen interessieren die Welt

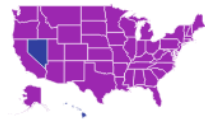
Suchbegriff oder Thema eingeben



Möchten Sie mit einem Beispiel loslegen?

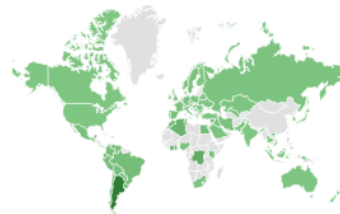
AUSBLENDEN

● Taylor Swift ● Kim Kardashian



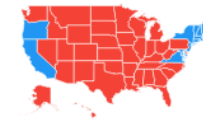
Interesse nach Unterregion, Letzte 7 Tage, Vereinigte Staaten

● Fußball-Weltmeisterschaft



Interesse nach Region, Letzte 7 Tage, Weltweit

● Fußball ● American Football



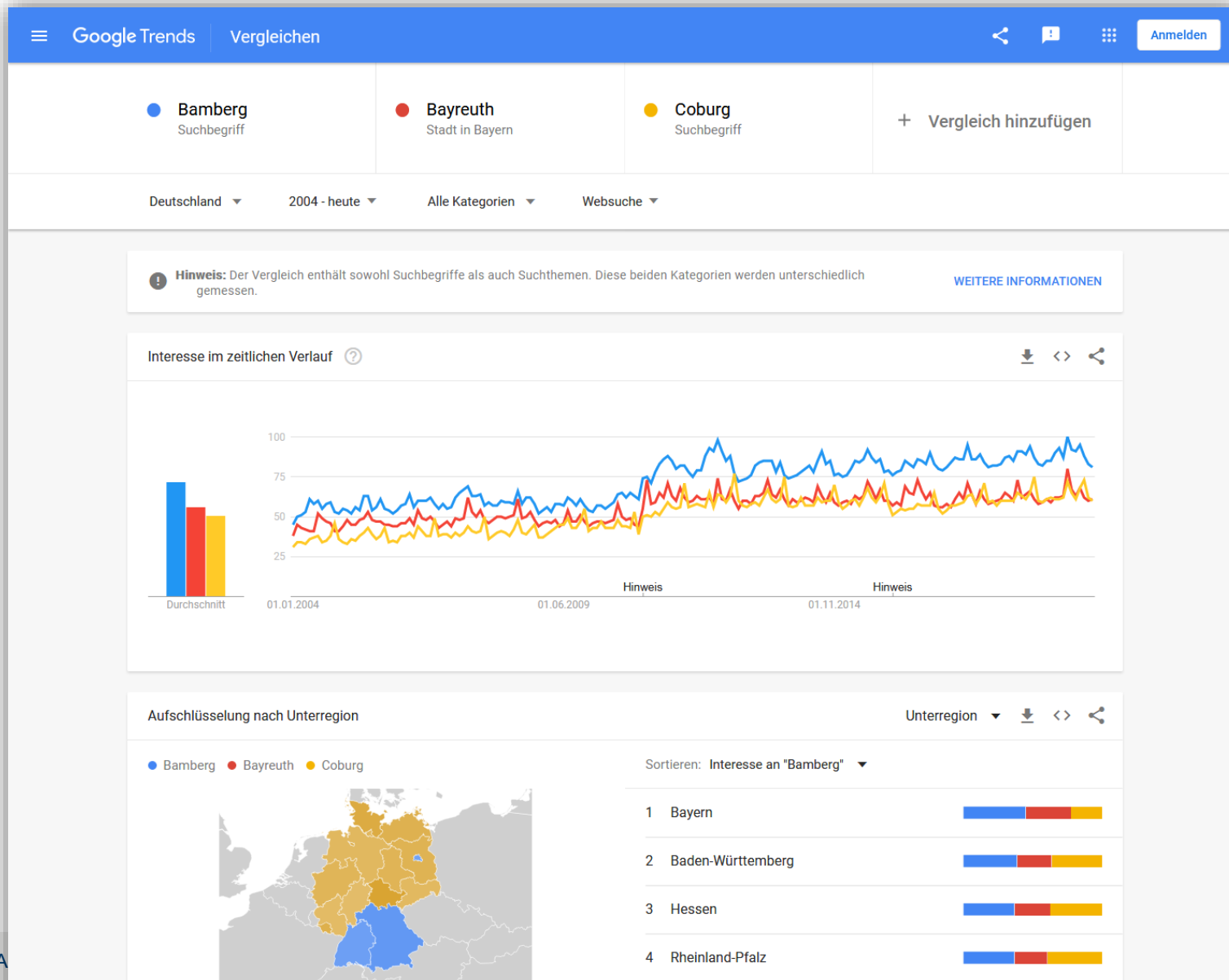
Interesse nach Unterregion, 2004 - heute, Vereinigte Staaten

< 1-3 von 6 Beispielen werden angezeigt >

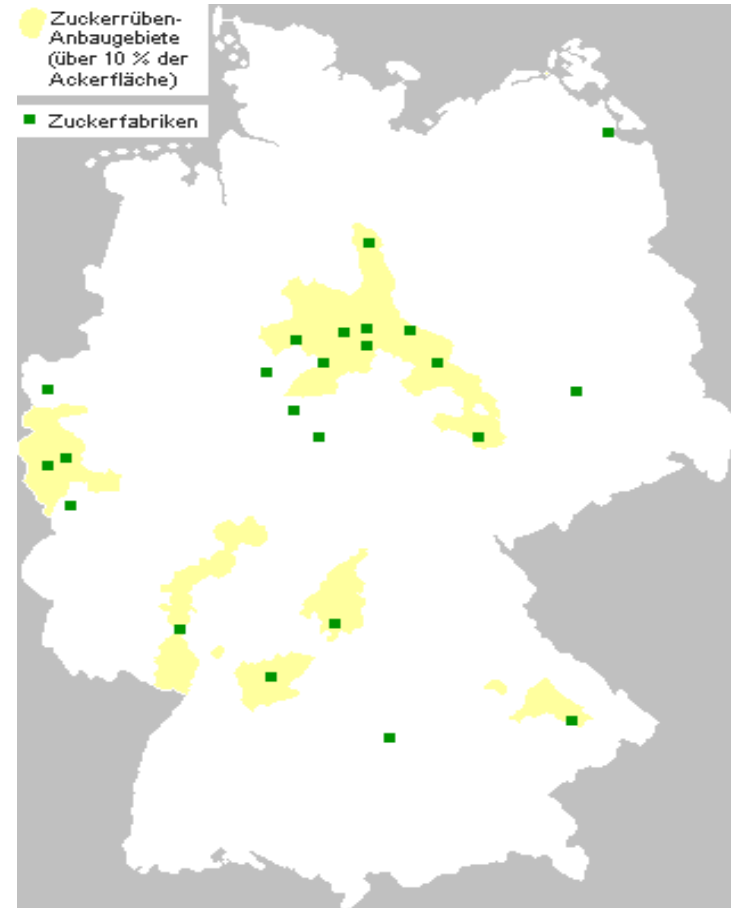
### Aktuelle Meldungen und Empfehlungen

Hier erfahren Sie, wie Daten von Google verwendet werden können, um eine Geschichte zu erzählen.

# Google Trends



# Ortsanalyse: Beispiel „Zuckerrübenanbau“

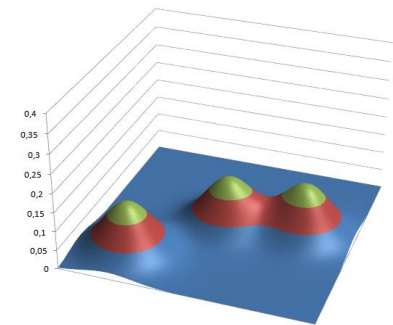
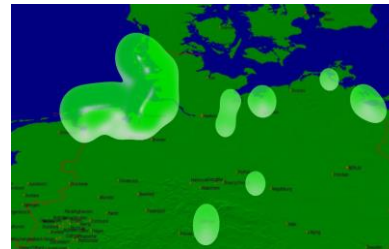
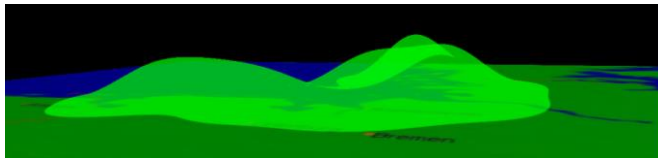


Bildquelle: <https://de.wikipedia.org/wiki/Zuckerr%C3%BCbe#/media/File:Zuckerr%C3%BCbengebiete-D.PNG>  
Datenquelle: Quelle: Wirtschaftliche Vereinigung Zucker - Verein der Zuckerindustrie

# Geo-Fußabdruck für beliebige Begriffe



1. Suche mit dem „Analyseziel“ als Anfrage nach Dokumenten
2. Nutze die  $\alpha$  Top Dokumente
3. Suche in diesen Dokumenten nach Toponymen
4. Addiere für alle Toponyme Gauß-Glocken mit Breite  $\beta$  und Höhe  $\gamma$
5. Bilde so eine Dichtefunktion über der Karte



6. Wähle einen Schwellenwert  $\tau$
7. Betrachte alles was über  $\tau$  liegt als Fußabdruck des „Anfrageortes“

# Beispiel: „Weisswurstäquator“



# ② Search Engine Optimization



- Die Situation:
  - Wer etwas ins Web stellt will, dass es gefunden wird!
  - Für viele Unternehmen ist es essentiell gefunden zu werden!
- Dann:
  - Unternehmen „optimieren“ ihre Webseiten für Suchmaschinen
- Aber:
  - Manche Maßnahmen sind erwünscht, andere (Spamming) werden abgestraft



Suche im Web

Web Search  
&  
Web Intelligence

Search Engine  
Optimization

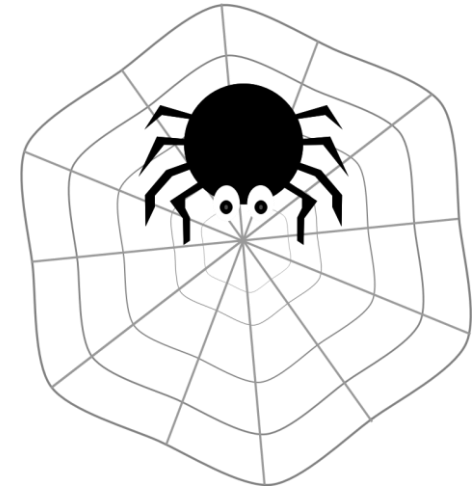




# Is Your Site Crawler Friendly?



## Google Tipps 1/3



- Make your site **crawlable** by engines
- Allow search engines to crawl in **robots.txt**
- Each page can be reached by clicking **links**
- Use **standard HTML**: less Flash, AJAX, ActiceX

Quelle: Matt Cutts, Google, SEO-Search Inside, 13.01.2014, [https://youtu.be/LQI\\_0L\\_ggno](https://youtu.be/LQI_0L_ggno)

## Google Tipps 2/3



- Pages should have clear **titles** and **descriptions**
- Use **text**, not just images of text
- Think about **what users will type**
- Example – **Search terms** for:



Quelle: Matt Cutts, Google, SEO-Search Inside, 13.01.2014, [https://youtu.be/LQI\\_0L\\_ggno](https://youtu.be/LQI_0L_ggno)

# Promote your site!

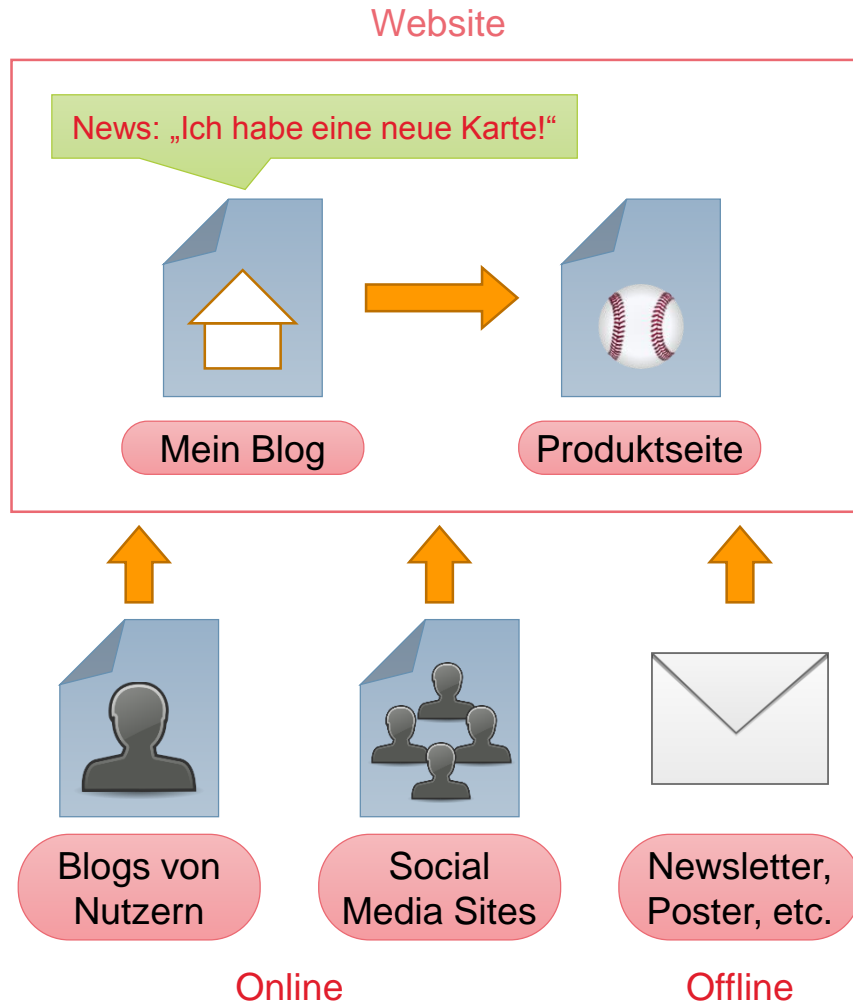


## Google Tipps 3/3



- Links matter, but don't spam to get links
- Good ways to get links: compelling content, blogging, newspaper stories.
- Social media can amplify your message.
- Free tools (z.B. <https://www.google.com/webmasters/>)

Quelle: Matt Cutts, Google, SEO-Search Inside, 13.01.2014, [https://youtu.be/LQI\\_0L\\_ggno](https://youtu.be/LQI_0L_ggno)



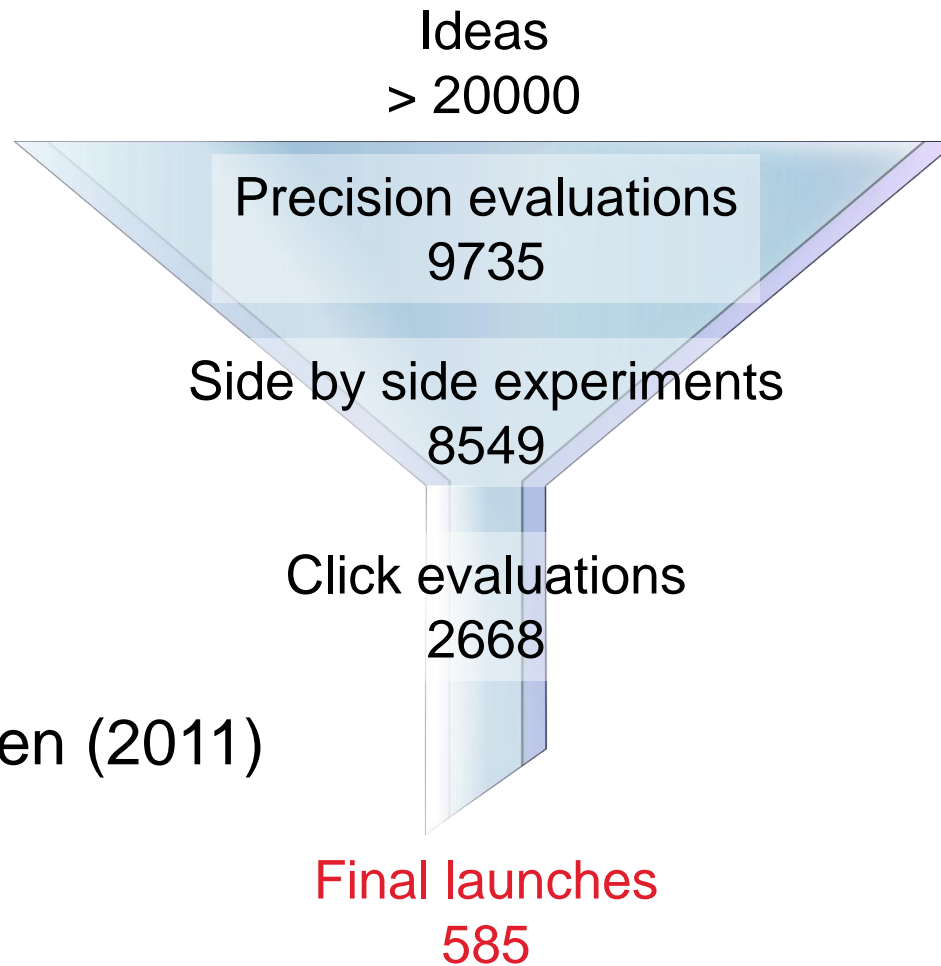
Google:

„Das Promoten eurer Website und qualitativ hochwertige Links können die Reputation eurer Website verbessern.“

Quelle: Google Inc., Einführung in Suchmaschinenoptimierung, 2015,

<http://www.google.de/webmasters/docs/einfuehrung-in-suchmaschinenoptimierung.pdf>

# Wie oft ändert Google seine Kriterien?



Jährliche Zahlen (2011)

Quelle: Matt Cutts, Google, SEO-Search Inside, 13.01.2014, [https://youtu.be/LQI\\_0L\\_ggno](https://youtu.be/LQI_0L_ggno)

# On-The-Page SEO

# Off-The-Page SEO

## PERIODIC TABLE OF SEO FACTORS 2019

<http://searchengineland.com/guide/seo>

**Qu** +3

**QUALITY**  
Pages must be well-written and highly valuable.

CONTENT

**Rs** +3

**RESEARCH**  
Investigate keywords people may use to find your content.

CONTENT

**Cr** +3

**CRAWL**  
Search engines must easily crawl your pages.

ARCHITECTURE

**Kw** +2

**KEYWORDS**  
Build target keywords into your pages.

CONTENT

**Mo** +3

**MOBILE**  
Optimize for today's smartphones and tablets.

ARCHITECTURE

**Tt** +3

**TITLES**  
Build keywords into your titles.

HTML

**Co** +2

**COUNTRY**  
Consider the country of your searcher.

USER

**Fr** +2

**FRESHNESS**  
Create timely content; refresh or retire stagnant pages.

CONTENT

**Dd** +2

**DUPLICATE**  
Be smart. Use canonicals, redirects.

ARCHITECTURE

**Ds** +2

**DESCRIPTIONS**  
Meta tags should describe what pages are about.

HTML

**Lo** +2

**LOCALITY**  
Consider the region of your searcher.

USER

**Mm** +2

**MULTIMEDIA**  
Images, video and audio can set your content apart.

CONTENT

**Sp** +2

**SPEED**  
Your site should load quickly on any device.

ARCHITECTURE

**St** +2

**STRUCTURE**  
Use Schema and more to turn data into enhanced listings.

HTML

**Au** +3

**AUTHORITY**  
Authority is everything. Covet links, shares and other signals.

TRUST

**Va** +3

**VALUE**  
Use Schema from trusted, quality websites.

LINKS

**Ux** +2

**USER EXPERIENCE**  
Experience matters more every day.

USER

**An** +2

**ANSWERS**  
Create content that can be turned into answers in the SERP.

CONTENT

**Ps** +2

**HTTPS**  
Provide a secure connection for visitors.

ARCHITECTURE

**Hd** +2

**HEADINGS**  
Build keywords into your H1 - H2 tags, especially H1.

HTML

**En** +2

**ENGAGEMENT**  
Visitors should spend time with your pages, especially H1, not bounce.

TRUST

**Ac** +2

**ANCHORS**  
Seek links from pages that use similar keywords.

LINKS

**Hs** +1

**HISTORY**  
Consider how regularly someone has visited your site.

USER

**Dt** +2

**DEPTH**  
Shallow content fails. Aim for substance.

CONTENT

**Ur** +1

**URLs**  
Build keywords into your page addresses.

ARCHITECTURE

**Am** +1

**AMP**  
Google's mobile page format is built for speed.

HTML

**Re** +1

**REPUTATION**  
Sites operating the same way for years carry weight.

TRUST

**Ba** +1

**BACKLINKS**  
The more high-quality links, the better.

LINKS

**In** +1

**INTENT**  
Consider why someone is conducting a specific search.

USER

## TOXINS

**Cl** -3

**CLOAKING**  
Don't show the engines different pages than your searchers.

TOXINS

**Sc** -3

**SCHEMES**  
Buying links, spamming blogs, all terrible tactics.

TOXINS

**Sf** -2

**STUFFING**  
Don't be excessive with packing keywords into your copy.

TOXINS

**Hi** -2

**HIDDING**  
Making keywords invisible is a dirty trick.

TOXINS

**Ar** -1

**PIRACY**  
Hosting stolen content can get you flagged.

TOXINS

**Iv** -1

**INTRUSIVE**  
Ad-heavy content, intrusive interstitials are a bad idea.

TOXINS

## EMERGING VERTICALS

**Voi**

**VOICE**  
Say hi to Alexa, Google, Siri.

EMERGING VERTICALS

**Loc**

**LOCAL**  
Listings, reviews, maps and more

EMERGING VERTICALS

**Img**

**IMAGE**  
Photos, illustrations are rising in the SERP

EMERGING VERTICALS

**Vid**

**VIDEO**  
Searchers are looking for live action.

EMERGING VERTICALS

# On-The-Page SEO

# Off-The-Page SEO

## PERIODIC TABLE OF SEO FACTORS 2019

<http://searchengineland.com/guide/seo>

Wie kommen Suchmaschinen an Öffnungszeiten ...?

- <http://searchengineland.com/guide/seo/html-code-search-engine-ranking>
- <http://searchengineland.com/library/search-engine-optimization/schema-dot-org>
- <http://schema.org/docs/gs.html>
- <https://developers.google.com/structured-data/?hl=en&rd=1>
- <https://developers.google.com/structured-data/local-businesses/>

**Qu** +3  
QUALITY  
Pages must be well-written and highly valuable.  
CONTENT

**Rs** +3  
RESEARCH  
Investigate keywords people may use in your content.  
CONTENT

**Cr** +3

**Kw**  
KEYWORDS  
Build target keywords into your page.  
CONTENT

**Fr**  
FRESHNESS  
Create timely content, refresh or retire stagnant pages.  
CONTENT

canonicals, redirects.  
ARCHITECTURE

**Mm** +2  
MULTIMEDIA  
Images, video and audio can set your content apart.  
CONTENT

**Sp** +2  
SPEED  
Your site should load quickly on any device.  
ARCHITECTURE

**St** +2  
STRUCTURE  
Use Schema and more to turn data into enhanced listings.  
HTML

**Au** +3  
AUTHORITY  
Authority is everything. Covet links, shares and other signals.  
TRUST

**Va** +3  
VALUE  
Seek links from trusted, quality websites.  
LINKS

**Ux** +2  
USER EXPERIENCE  
Experience matters more every day.  
USER

your searcher.  
USER

can get you flagged.  
TOXINS

intrusive interstitials are a bad idea.  
TOXINS

EMERGING VERTICALS

**An** +2  
ANSWERS  
Create content that can be turned into answers in the SERP.  
CONTENT

**Ps** +2  
HTTPS  
Provide a secure connection for visitors.  
ARCHITECTURE

**Hd** +2  
HEADINGS  
Build keywords into your H1 - H2 tags, especially H1.  
HTML

**En** +2  
ENGAGEMENT  
Visitors should spend time with your pages, not bounce.  
TRUST

**Ac** +2  
ANCHORS  
Seek keywords from pages that use similar keywords.  
LINKS

**Hs** +1  
HISTORY  
Consider how regularly someone has visited your site.  
USER

**Voi**  
VOICE  
Say hi to Alexa, Google, Siri.  
EMERGING VERTICALS

**Loc**  
LOCAL  
Listings, reviews, maps and more.  
EMERGING VERTICALS

**Dt** +2  
DEPTH  
Shallow content fails. Aim for substance.  
CONTENT

**Ur** +1  
URLs  
Build keywords into your page addresses.  
ARCHITECTURE

**Am** +1  
AMP  
Google's mobile page format is built for speed.  
HTML

**Re** +1  
REPUTATION  
Sites operating the same way for years carry weight.  
TRUST

**Ba** +1  
BACKLINKS  
The more high-quality links, the better.  
LINKS

**In** +1  
INTENT  
Consider why someone is conducting a specific search.  
USER

**Img**  
IMAGE  
Photos, illustrations are rising in the SERP.  
EMERGING VERTICALS

**Vid**  
VIDEO  
Searchers are looking for live action.  
EMERGING VERTICALS

# SEO: Important Signals (Ranking Factors)



- Page Link Authority Features
- Page Level Social
- Page Level Anchor Text
- Page Keyword Usage
- Page Keyword Agnostic
- Domain Link Authority Features
- Domain Level Anchor Text
- Domain Keyword Usage
- Domain Keyword Agnostic
- Brand Metrics

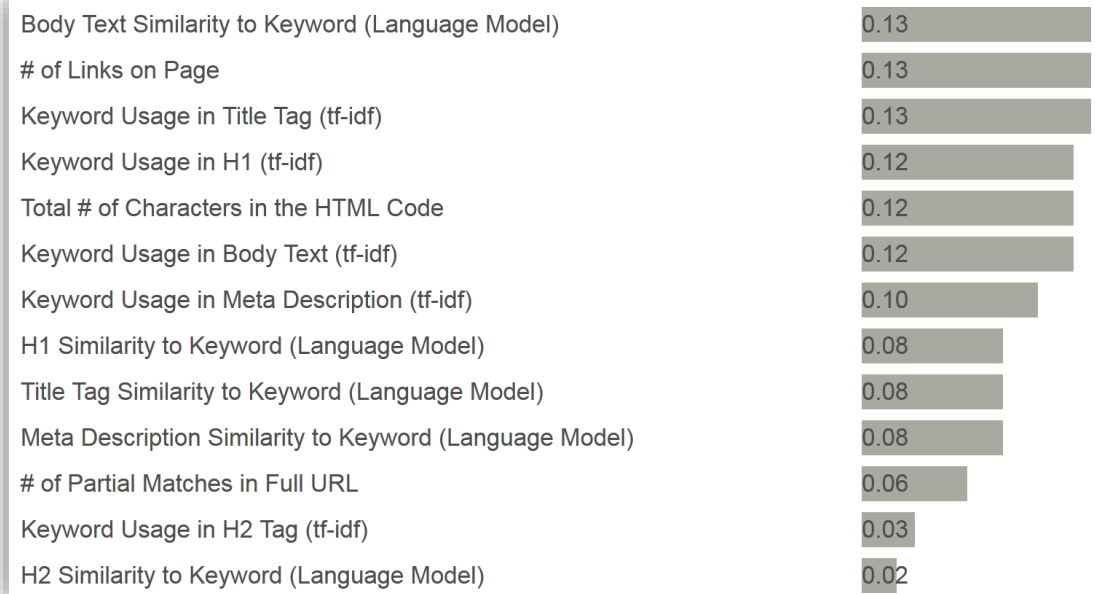
Quelle: SEOmoz, Inc., 2015, <https://moz.com/search-ranking-factors>



# SEO: Important Signals (Ranking Factors)



- Page Link Authority Features
- Page Level Social
- Page Level Anchor Text
- ▶ Page Keyword Usage
- Page Keyword Agnostic
- Domain Link Authority Features
- Domain Level Anchor Text
- Domain Keyword Usage
- Domain Keyword Agnostic
- Brand Metrics

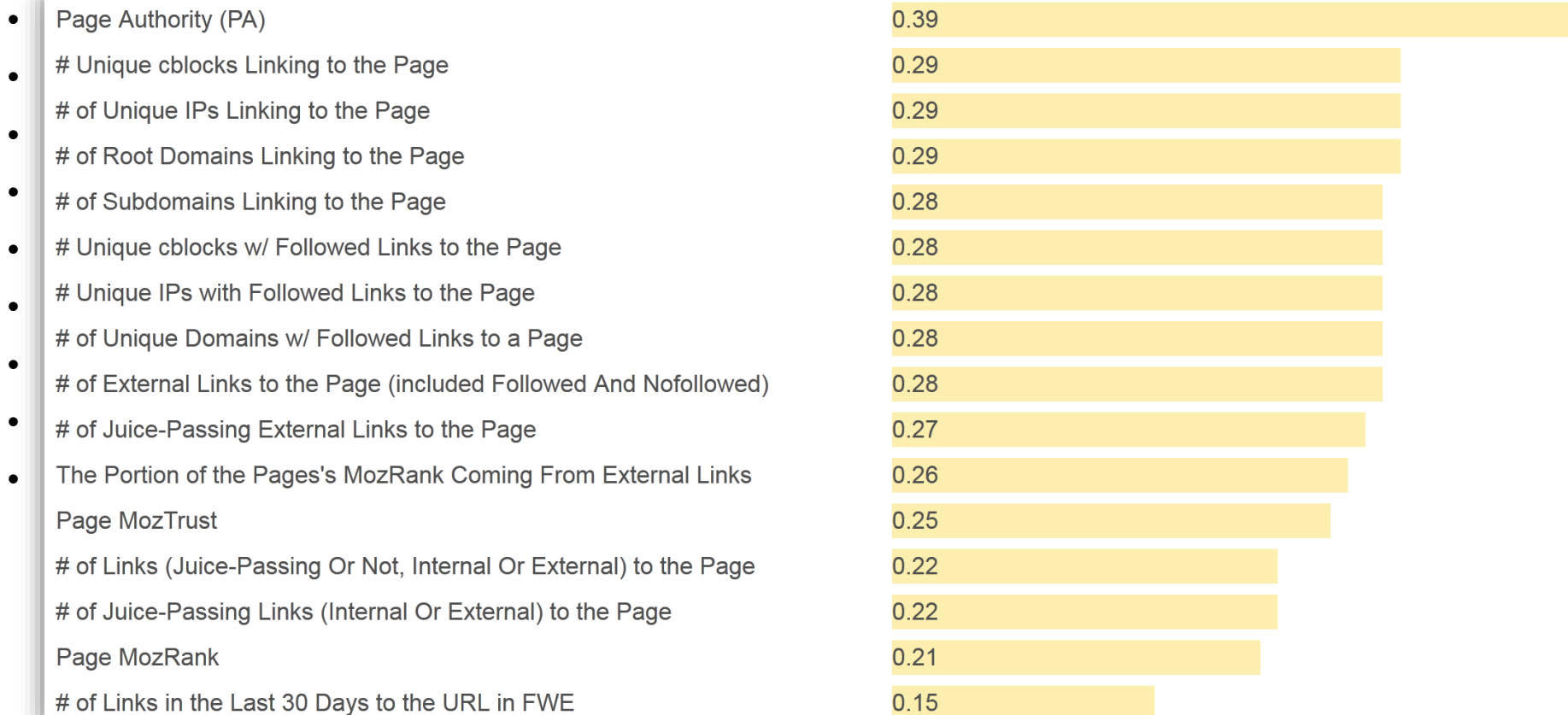


Quelle: SEOmoz, Inc., 2015, <https://moz.com/search-ranking-factors>

# SEO: Important Signals (Ranking Factors)



## ► Page Link Authority Features



Quelle: SEOMoz, Inc., 2015, <https://moz.com/search-ranking-factors>

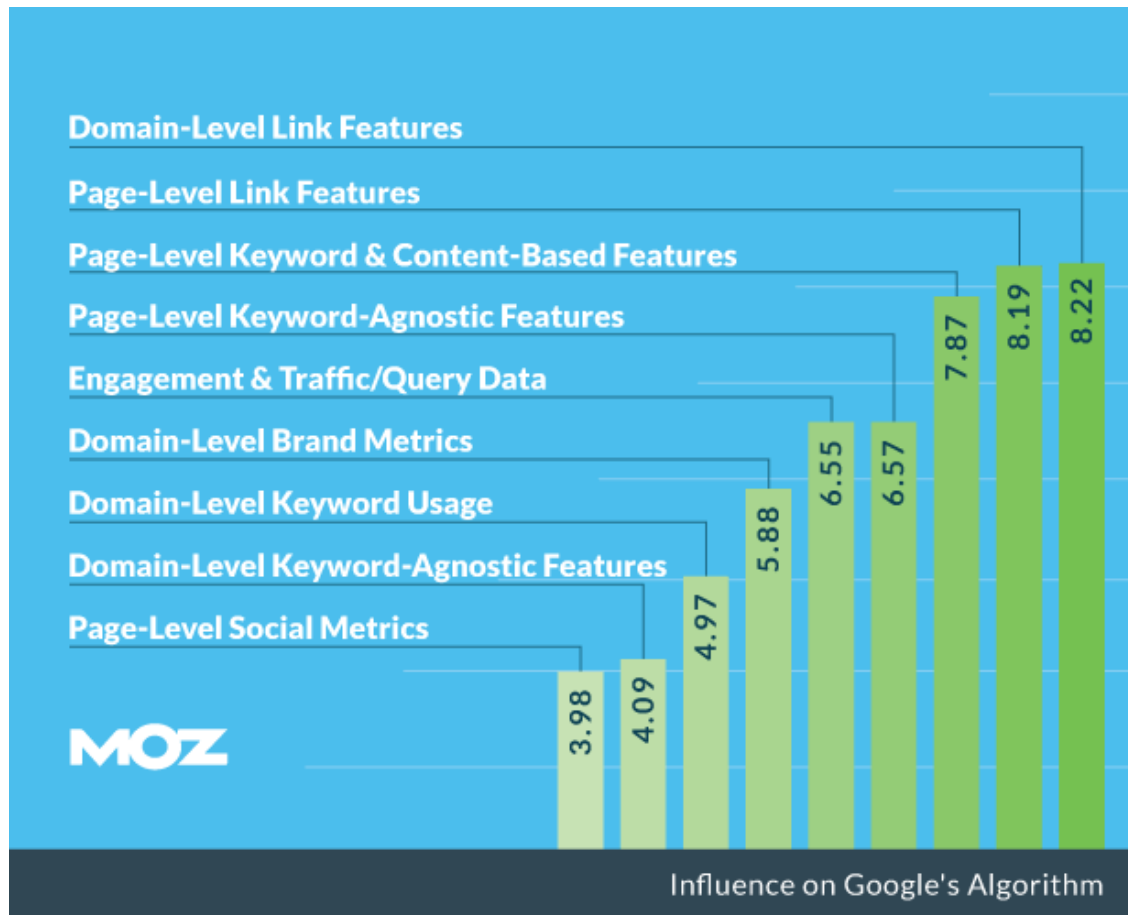
- Domain-Level, Keyword-Agnostic Features:
  - Domain name length, TLD extension, SSL certificate, etc.
- Domain-Level, Link Authority Features:
  - Based on link/citation metrics such as quantity of links, trust, domain-level PageRank, etc.
- Domain-Level Keyword Usage:
  - Exact-match keyword domains, partial-keyword matches, etc.
- Domain-Level Brand Metrics:
  - Offline usage of brand/domain name, mentions of brand/domain in news/media/press, toolbar/browser data of usage about the site, entity association, etc.

Quelle: <https://moz.com/search-ranking-factors>, 06.02.2017

- Page-Level Social Metrics:
  - Quantity/quality of tweeted links, Facebook shares, Google +1s, etc. to the page
- Page-Level Link Metrics:
  - PageRank, Trust metrics, quantity of linking root domains, links, anchor text distribution, quality/spamminess of linking sources, etc.
- Page-Level Keyword & Content-Based Metrics:
  - Content relevance scoring, on-page optimization of keyword usage, topic-modeling algorithm scores on content, content quantity/quality/relevance, etc.
- User Usage & Traffic/Query:
  - Data SERP engagement metrics, clickstream data, Visitor traffic/usage signals, quantity/diversity/CTR of queries, both on the domain and page level
- Page-Level, Keyword-Agnostic Features:
  - Content length, readability, Open Graph markup, uniqueness, load speed, structured data markup, HTTPS, etc.

Quelle: <https://moz.com/search-ranking-factors>, 06.02.2017

# Influence on Google's Algorithm



Quelle: <https://moz.com/search-ranking-factors>, 06.02.2017

# Aspekte der Suchmaschinenoptimierung

