

# Google Caffeine

Was ist es, was ändert sich,  
wie bereite ich mich vor?



# Wer ist das?

## Johannes Beus, SISTRIX

- Suchmaschinenoptimierung seit 5 Monaten
- Betrieb eigener Webprojekte unterschiedlichster Themengebiete und Zielgruppen
- SISTRIX Toolbox

# Google Caffeine

- Was ist Google Caffeine?
- Was ändert sich durch Google Caffeine?
- Wie kann ich mich auf Google Caffeine vorbereiten?

# 1 | Was ist Google Caffeine?

- im August 2009 der Öffentlichkeit vorgestellt
- monatelang in einer öffentlichen Sandbox testbar
- für Januar 2010 angekündigt
- ... und immer noch nicht richtig live.

Achtung: die folgenden Folien sind „Educated Guess“, nicht unbedingt Realität.

# 1 | Was ist Google Caffeine?

Vereinfachter Aufbau einer Suchmaschine

**Crawlen**



**Indexieren**



**Ranken**



# 1 | Was ist Google Caffeine?

Vereinfachter Aufbau einer Suchmaschine

**Crawlen**



**Indexieren**



**Ranken**



# 1 | Was ist Google Caffeine?

Unter dem Arbeitstitel **Caffeine** wird ein Teil der Suchinfrastruktur, also der eben gezeigt mittlere Teil der Suchmaschine, neu konzipiert und umgesetzt.

Zu der Infrastruktur gehört Software wie:

- MapReduce – paralleles Verarbeiten sehr großer Datenmengen
- BigTable – Speichern und Durchsuchen umfangreicher Tabellen
- GFS (Google File System) – Verteiltes Dateisystem als Grundlage der Software von Google

# 1 | Was ist Google Caffeine?

## GFS – Google File System

Als Google das GFS konzipiert hat, wurden Entscheidungen getroffen wie: „*Hoher Durchsatz ist wichtiger als kurze Wartezeit*“

Gut für monatliche oder wöchentliche Updates eines Suchindexes, schlecht für Anforderungen, die heute an Google gestellt werden:



**Realtime-Search**

# 1 | Was ist Google Caffeine?

## GFS2 – Google File System Version 2

Das Google File System wird neu konzipiert und umgesetzt. Dabei werden Fehler und Grenzen, die in den letzten Jahren aufgetreten sind beseitigt. Dadurch kann Google Informationen in Zukunft schneller verarbeiten und Software, die auf dem GFS basiert, kann besser arbeiten.

Nerd-Wissen:

- Kleinere Chunks (von 64MB auf 1MB, dadurch weniger Latenz)
- Multi-Master-Setup (Dadurch werden kleinere Chunks erst möglich, da Metadaten als B-Tree im RAM des Masters liegen)

# 1 | Was ist Google Caffeine?

## Zusammenfassung

- Google Caffeine ist **kein Algorithmus-Update**, sondern Google erneuert die Software-Infrastruktur, da sie den wachsenden Anforderungen nicht mehr Stand hält.
- Das Google-System wird neue Informationen in Zukunft schneller verarbeiten können. Google legt damit den Software-Grundstein für eine bessere und schnellere **Realtime-Suche**.
- Da die erneuerte Infrastruktur Google nicht mehr einschränkt, werden in Zukunft neue und **andere Signale** in den Algorithmus einfließen können.

## 2 | Was ändert sich durch Google Caffeine?

### Der Realtime-Index rückt näher

Google ist in den letzten Jahren zwar schneller geworden, ein Realtime-Index mit unterschiedlichen Datenquellen ist das aber noch nicht.

Google Caffeine legt die Grundlagen, damit sowas in Zukunft umgesetzt werden kann. Anzeichen dafür wie die Vorstellung von PuSH (Realtime Syndication Protokoll) sind ausreichend vorhanden.



## 2 | Was ändert sich durch Google Caffeine?

### Bessere und aktuellere Signale für den Google Algorithmus

Viele der aktuell genutzten Ranking-Signale sind langsam und entsprechen damit nicht mehr den Erfordernissen an einen Realtime-Index. Google versucht das durch domainweite Signale wie Trust zu kompensieren.

Durch Caffeine können vermutlich auch Signale wie die Anzahl der ReTweets oder Ähnliches in den Algorithmus einfließen und für sehr frischen Inhalte bessere Werte als die alten Signale liefern.



## 2 | Was ändert sich durch Google Caffeine?

### Google indexiert mehr Dokumente

Tests an der öffentlichen Sandbox haben gezeigt, dass Google Caffeine häufig mehr Ergebnisse findet als die bekannte Google-Suche. Crawling von Ajax-Seiten und Teilen des Deepwebs scheinen dazu beizutragen.

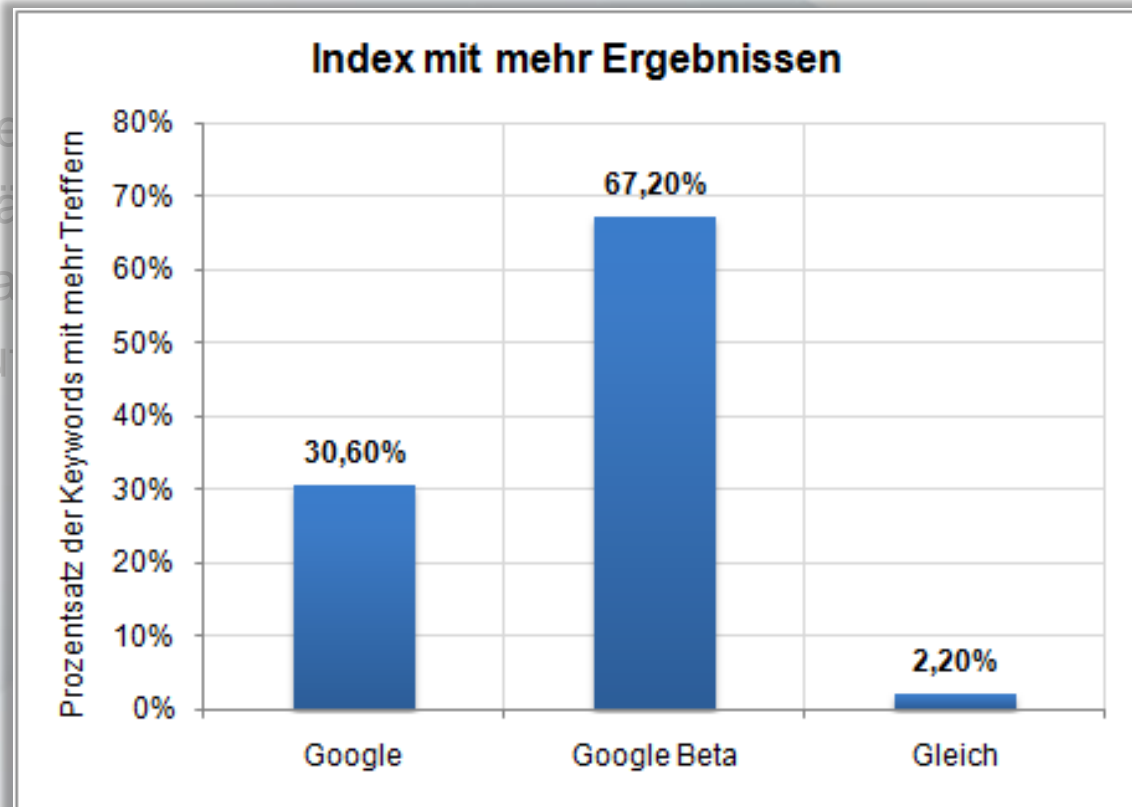


## 2 | Was ändert sich durch Google Caffeine?

Google indexiert mehr Dokumente

Tests an de  
Caffeine hä  
Suche. Cra  
dazu beizu

gle  
ogle-  
scheinen



## 2 | Was ändert sich durch Google Caffeine?

### Die Antwortzeiten der Googlesuche sinken

Viele Features und Umsetzungen der aktuellen Google-Suche mussten um die Beschränkungen der derzeitigen Infrastruktur „drumherum“ entwickelt werden. Auf Basis von Caffeine können effizientere Methoden gewählt werden und die Antwortzeit sinkt.

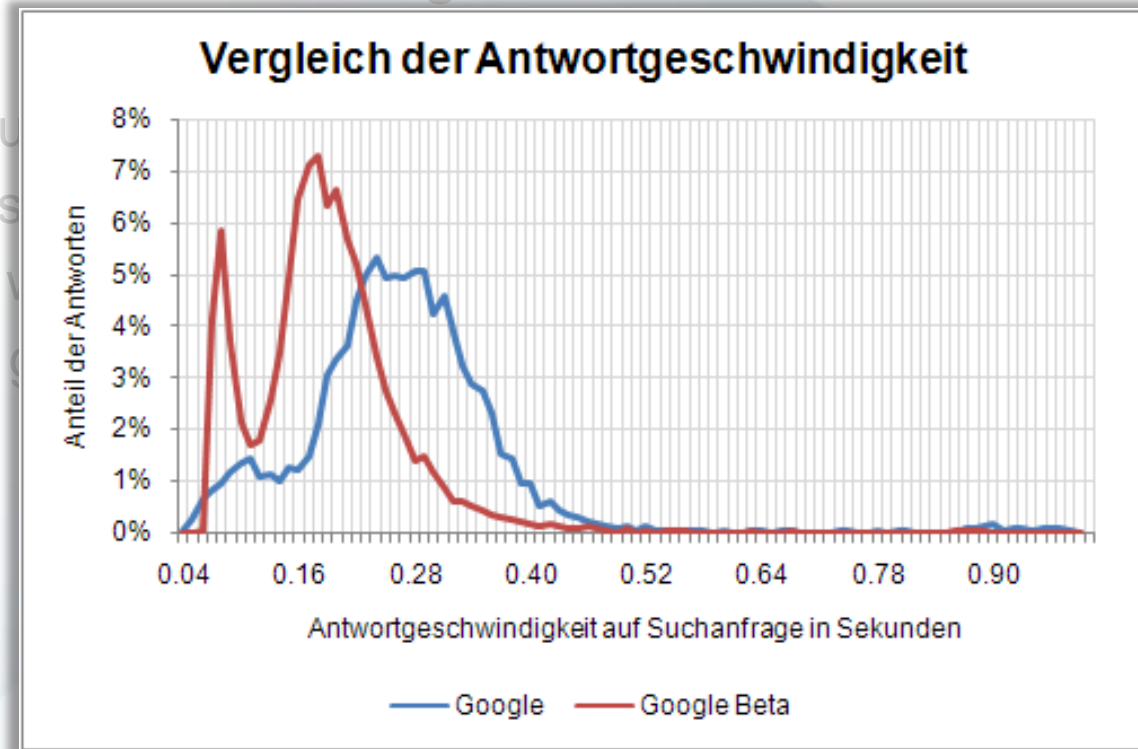


## 2 | Was ändert sich durch Google Caffeine?

Die Antwortzeiten der Google Suche sinken

Viele Features  
um die Benutzer  
entwickelt  
Methoden

... mussten  
"herum"  
re



## 2 | Was ändert sich durch Google Caffeine?

### Zusammenfassung

- Google entwickelt sich in Richtung **Realtime-Suchmaschine**. Der Index wird bei vielen Themen aktueller und schnelllebiger.
- **Neue Signale** werden Einzug in den Algorithmus erhalten. Diese Signale werden häufig eine Beziehung zu Erkenntnissen aus dem Realtime-Index haben.
- Der Google-Index wird (deutlich) **umfangreicher** und trotzdem wird Google in der Lage sein, Ergebnisse **schneller** als bisher zu liefern.

# 3 | Wie bereite ich mich auf Google Caffeine vor?



# 3 | Wie bereite ich mich auf Google Caffeine vor?

## Vorbereitung auf neue Signale

Welche Signale Google in Zukunft zusätzlich in den Algorithmus aufnehmen wird, ist unklar. Gute Chancen dürften Realtime-Signale haben: Twitter, Facebook-Stream, sonstiger Social-Kram.

Also wenn noch nicht geschehen: Twitter- & Facebook-Account anlegen und sinnvolle (!) Kontakte knüpfen. Trickserien in sozialen Netzwerken sind leichter zu enttarnen als in Linknetzwerken ...



# 3 | Wie bereite ich mich auf Google Caffeine vor?

## Vorbereitung auf andere Darstellungs- und Auslieferungswege

Das Crawlen von Webseiten ist zum Datensammeln gut geeignet, wenn man es nicht eilig hat. Google hat es eilig und wird alternative Bereitstellungsformen fordern. PuSH (PubSubHubbub), eine Art erweiterter Atom/RSS-Feed wurde bereits vorgestellt, weitere werden vermutlich folgen.

Wer zeitkritischen Content hat, sollte seine Technik so aufstellen, dass sie die Anforderungen von Google flexibel umsetzen kann.



# 3 | Wie bereite ich mich auf Google Caffeine vor?

## Indexierung gezielt steuern

Der Google-Index wird mit Caffeine wachsen, Google muss also mehr Seiten crawlen. Deswegen fängt der Googlebot an, Ajax zu parsen, Formulare auszufüllen und ob Verbote in der robots.txt auf Dauer eingehalten werden, ist auch nicht so sicher.

Es ist wichtig, diese Aktivitäten von Google im Auge zu behalten und gegebenenfalls einzugreifen. Ihr wisst besser, welche Inhalte in den Index dürfen und welche nicht.

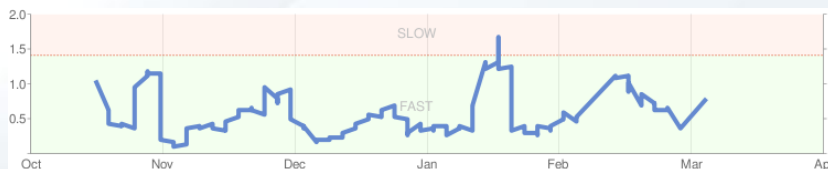


# 3 | Wie bereite ich mich auf Google Caffeine vor?

## Ladegeschwindigkeit optimieren

Auch wenn es keinen direkten Bezug zu Caffeine hat: achtet darauf, dass die Ladezeiten eurer Seiten in einem akzeptablen Rahmen bleiben. Das gilt sowohl für die Auslieferungszeit des reinen HTML als auch für die Ladezeit der kompletten Seite beim User.

Die Webmaster-Tools von Google zeigen euch unter „Labs“, welche Ladezeit-Daten Google derzeit von euch hat:



# 3 | Wie bereite ich mich auf Google Caffeine vor?

## Zusammenfassung

- Mögliche Realtime-Signale, die Google in Zukunft nutzen könnte für das eigene Umfeld ausmachen und sich darauf vorbereiten.
- Die Technik so flexibel aufstellen, dass verschiedene Distributionsmethoden genutzt werden können, wenn Google das will.
- Aufpassen, welche Inhalte Google indexiert und darauf achten, dass die ausgelieferten Inhalte schnell bei Google und beim Nutzer sind.

# Google Caffeine

Download und Zusammenfassung:

**[www.sistrix.de/news & seo.at](http://www.sistrix.de/news&seo.at)**

Twitter nicht vergessen, könnte wichtig werden:

**[twitter.com/sistrix](https://twitter.com/sistrix)**

Wenn sonst noch was ist:

**[j.beus@sistrix.com](mailto:j.beus@sistrix.com)**

