

Webseiten Architektur

Status Codes

Ein HTTP-Statuscode wird von einem Server auf jede HTTP-Anfrage als Antwort geliefert.

Gängige Statuscodes:

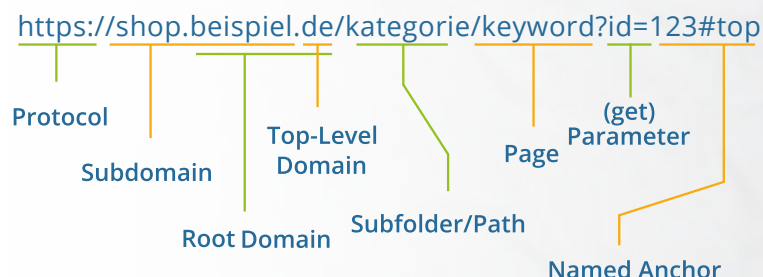
200 - OK ✓
301 - Dauerhafte Weiterleitung →
302 - Temporäre Weiterleitung
404 - Client Fehlermeldung ✗
500 - Server Fehler

NOT: Verwende intern möglichst keine Weiterleitungen nur wenn nötig dann 301.

TIPP: Beseitige 404-Fehler schnellstmöglich.

URL Elemente

Eine perfekt gestaltete URL enthält folgende 9 Elemente:



URL Struktur

Wichtige Merkmale für eine gute URL Struktur:

- Eine flache Hierarchie mit max. 4 Ebenen
- Benutze sprechende URLs (`www.domain.de/beispiel/`) und vermeide nicht-sprechende (`www.domain.de/cat47/pg69`)
- Verwende keine Sonderzeichen in der URL

Interne Verlinkung

Eine gleichmäßige Verlinkung der Seite sorgt für eine optimale Link-Juice Verteilung.

`verlinkter Text`

TIPP: Klickpfadlänge mit max. 4 Klicks.

Meta Daten Analyse

Gestalte einen aussagekräftigen Title und eine Description für ein klickstarkes Snippet in den Suchergebnissen.

Title Tag

Der Meta Title ist eine kurze und prägnante Beschreibung des HTML-Dokumentes. (maximal 512 Pixel, ca. 55- 60 Zeichen)

`<title>Aussagekräftiger Seitentitel</title>`

Description Tag

Meta Descriptions sind kurze Beschreibungstexte, die den Inhalt eines HTML-Dokumentes erläutern. (maximal 156 Zeichen)

`<meta name="description" content="Beschreibung">`

Alt-Attribute bei jedem Image hinterlegen

Hinterlege möglichst bei jedem Image ein beschreibendes Keyword als Alt-Attribut.

``

Canonical

Der Canonical Tag verweist auf eine kanonische URL – bei Websites mit gleichen oder fast gleichen Inhalten. Die URL auf die der Canonical verweist, ist die die in den Suchergebnissen erscheint

`<link rel="canonical" href="http://www.domain.de/beispiel/">`

robots-Attribut

Meta-Robots ist ein HTML Meta-Tag, das einem Suchmaschinenbot eine bestimmte Anweisung gibt, ob die Seite in den Suchmaschinenindex aufgenommen werden soll (index) und den Links auf der Seite gefolgt werden soll (follow).

So kann es deklariert werden:

```
<meta name="robots" content="noindex, nofollow" />
<meta name="robots" content="noindex, follow" />
<meta name="robots" content="index, nofollow" />
<meta name="robots" content="index, follow" />
```

Vermeide Duplicate Content

Doppelte Inhalte können leicht entstehen und sollten nach Möglichkeit vermieden werden.

`http://www.domain.de`
`http://domain.de`
`https://www.domain.de`
`http://www.domain.de/`

TIPP: Lösung über Canonical Tag:

`<link rel="canonical" href="https://www.domain.de/">`

robots.txt

Die robots.txt-Datei ist eine einfache Text-Datei. Seiten, die nicht gecrawled werden sollen, werden hier ausgeschlossen. Lege die robots.txt-Datei im root Verzeichnis ab.

Möchtest Du bestimmte Verzeichnisse ausschließen, dann sollte die robots.txt so aussehen:

User-agent: Googlebot
Disallow: /bestimmtes-Verzeichnis/

Sitemap.xml

Mit Hilfe der sitemap.xml bekommt die Suchmaschine einen besseren Überblick über die Seitenstruktur. Lege die sitemap.xml im root Verzeichnis ab.

Es können mehrere sitemaps existieren (Bilder, pdf, etc.).

Vorteile:

- Neue Seiten können schneller gecrawlt werden
- Versteckte oder schlecht erkennbare Inhalte können besser erkannt werden. (AJAX, Rich Media)
- Verwaiste oder schlecht verlinkte Seiten können besser ermittelt werden.

```
<?xml version="1.0" encoding="UTF-8"?>
<urlset xmlns="http://www.sitemaps.org/schemas/sitemap/0.9">
  <url>
    <loc>http://www.beispielseite.de/unterseite1.html</loc>
  </url>
</urlset>
```