

Version 2023a

Grundlagen der Web-Programmierung Search Engine Optimization (SEO)

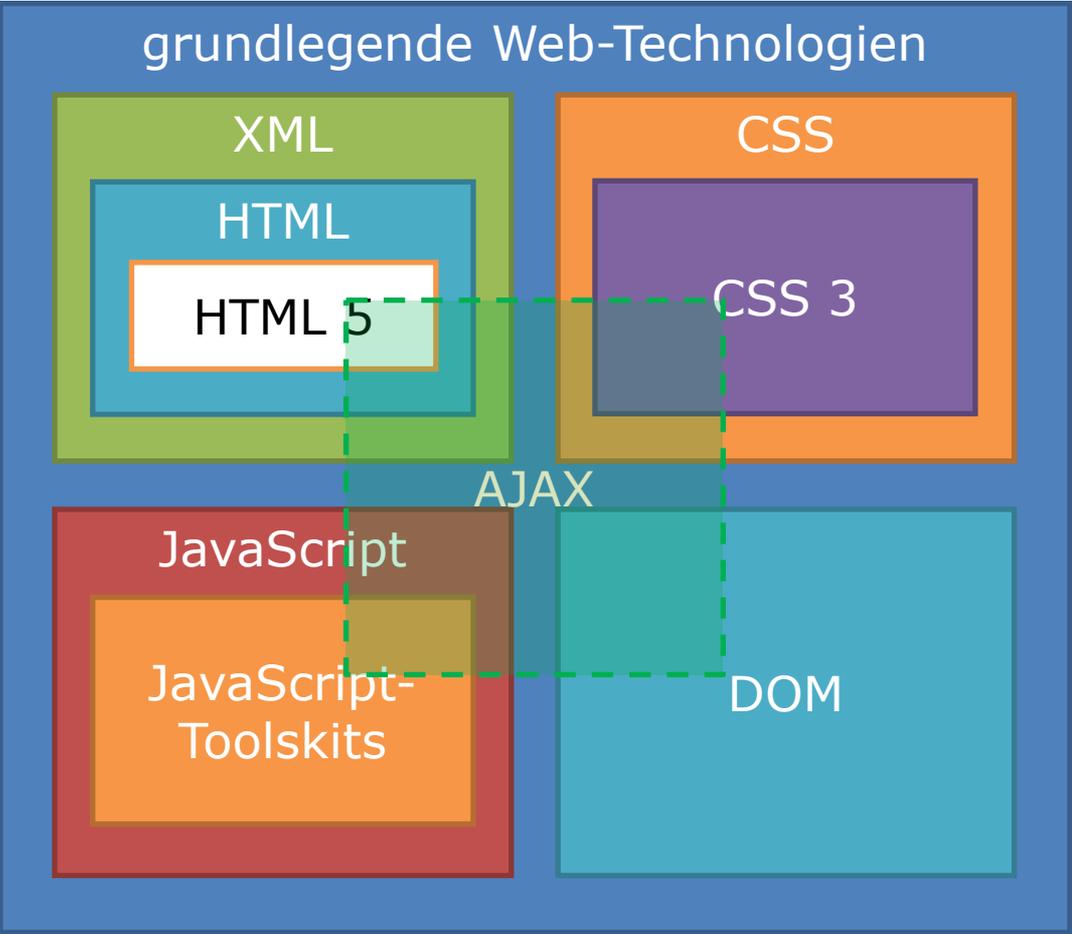
Prof. Dr.-Ing. Tenshi Hara
tenshi.hara@ba-sachsen.de



AUFBAU DER LEHRVERANSTALTUNG

Progressive Web Apps

Search Engine Optimization



Proprietäre Standards

SUCHERGEBNISSE

- Suchmaschinen führen einen Index
- Index ist manipulierbar
- Indexierungsalgorithmus ist Einstiegspunkt
- umso besser die Anpassung an den Algorithmus, umso höher die Wahrscheinlichkeit in Suchergebnisse hoch gelistet zu werden

WEBSITE-NUTZER

- mobil nutzbare und barrierefreie Webseiten werden belohnt
- von besserer mobiler Performance profitieren auch PC-Nutzer
- übersichtliche Inhalte sind leicht zu indexieren
- übersichtliche Inhalte begrenzen die Informationsflut

SEARCH ENGINE OPTIMIZATION

Search Engine Optimization (SEO) ist ein Prozess

- Anpassung der Seitenstruktur
 - Anpassung der Darstellung
 - Anpassung der Hyperlinks
 - Anpassung der Server-Performance
 - Anpassung der Seiteninhalte
 - Anpassung der Metadaten
- } technische SEO
- } semantische SEO

ASPEKTE DER SEO

- TF:IDF
- Pagerank
- Website-Gestaltung
- Server-Einstellungen
- weitere Google-Metriken

} technische SEO

semantische SEO { • Google Analytics

TF:IDF

CONTENT INDEX

- Informationsmenge basierend auf Dokumentsammlung
- reduziert Seiteninhalte auf das Notwendigste, z.B.
 - Entfernen aller Format-, Struktur- und Markup-Daten
 - Entfernen aller Satzzeichen; verbleibende Wörter sind
 - nur noch durch Leerzeichen voneinander getrennt
 - die Terme/Token des Index'
 - Entfernen unnötiger und nur Grammatik tragender Wörter (z.B. der, die, das, ich, du, ...)
 - Kanonisierung der Terme (z.B. „Programme“, „Programmieren“ und „programmierte“ zu „Programm“)
- Häufig: Erstellung eines invertierten Index' → jedem Term wird eine Liste von Dokumenten zugeordnet, in denen er vorkommt

BEISPIELINDEX

Term	Dokument
Zitat	D1, D2
Friedrich	D1, D3
Schiller	D1, D3
Literatur	D1
Nietzsche	D3
Handschrift	D2



Term	Dokument
Zitat	(D1,1), (D2,1)
Friedrich	(D1,3), (D3,6)
Schiller	(D1,4), (D3,20)
Literatur	(D1,31)
Nietzsche	(D3,7)
Handschrift	(D2,12)

Wegen ihrer Schnittmenge sind D1 und D3 in der Ergebnismenge des Suchbegriffs „Friedrich Schiller“.

Problem: Was, wenn Suchende nur Ergebnisse mit der exakten Phrase „Friedrich Schiller“ wollen?

→ **Mit dem simplen invertierten Index nicht möglich.**

Durch Ergänzen der Termposition im Index kann die exakte Phrase „Friedrich Schiller“ gefunden werden.

→ **Berücksichtigt der Algorithmus auch Strukturdaten, können gezielt Überschriften, etc. gefunden werden.**

„TERM FREQUENCY“ UND „INVERSE DOCUMENT FREQUENCY“

TF:IDF ist ein Gewicht für die Relevanz eines Terms in einem Dokument innerhalb einer Dokumentensammlung.

- **Term Frequency** beschreibt, wie relevant Begriff in einem einzelnen Dokument ist.
- **Inverse Document Frequency** beschreibt, wie relevant ein Begriff innerhalb der gesamten Dokumentensammlung ist.

Beispielmaß:

$$TF:IDF_t = TF_t \times IDF_t = \frac{n_{t,J}}{N_J} \times \frac{D}{D_t}$$

Bei großen Dokumenten steigt Wahrscheinlichkeit, dass der Term darin enthalten ist.
 → Solche Dokumente sind weniger relevant als die, die kurz sind und den Term enthalten.

Wenn ein Term selten in der Dokumentensammlung ist, ist ein den Term enthaltendes Dokument sehr wichtig.

TF:IDF-BEISPIEL

Zitat von Friedrich Schiller:

Das Gesetz ist der Freund des Schwachen.

D1

Zitat von Arthur Schopenhauer:

Die Freunde nennen sich aufrichtig. Die Feinde sind es.

D2

Zitat von Friedrich Nietzsche:

Wer ein Warum hat, dem ist kein Wie zu schwer.

D3

Für die Suchphrase „Zitate Friedrich“ ergeben sich:

$$TF_{Zitat,D1} = \frac{1}{11}$$

$$TF_{Zitat,D2} = \frac{1}{13}$$

$$TF_{Zitat,D3} = \frac{1}{14}$$

$$TF:IDF_{D1} = 1 \times \frac{1}{11} + 1 \frac{1}{2} \times \frac{1}{11} = \frac{5}{22} = 0,227$$

$$TF_{Friedrich,D1} = \frac{1}{11}$$

$$TF_{Friedrich,D2} = 0$$

$$TF_{Friedrich,D3} = \frac{1}{14}$$

$$TF:IDF_{D2} = 1 \times \frac{1}{13} + 1 \frac{1}{2} \times 0 = \frac{1}{13} = 0,077$$

$$IDF_{Zitat} = 1$$

$$IDF_{Friedrich} = 1 \frac{1}{2}$$

$$TF:IDF_{D3} = 1 \times \frac{1}{14} + 1 \frac{1}{2} \times \frac{1}{14} = \frac{5}{28} = 0,179$$

⇒ sortiert nach TF:IDF ist die Reihenfolge der Dokumente: D1, D3, D2

⇒ TF:IDF kann unabhängig von Anfragen vorab berechnet werden!

WAS BEDEUTET TF:IDF FÜR SEO?

- kurze, prägnante Texte
- Terme konzentrieren und nicht über alle Seiten der Website streuen
- Algorithmus durch gezielte Strukturelemente unterstützen
 - Überschriften mit H1 bis H6 kenntlich machen:
 - `<h1>Oberste Ebene</h1>`
 - ...
 - `<h6>Unterste Ebene</h6>`
 - zusammenhängende Wörter in SPAN zusammenfassen:
„Die `Cloud & Heat Communications GmbH hat ihre Webseiteninhalte einer Search Engine Optimization unterzogen.“`

Pagerank

REPUTATION: PRESTIGE UND GLAUBWÜRDIGKEIT (1/2)

- neben den Termen (siehe TF:IDF) ist die **Reputation** einer Webseite für die Platzierung in den Suchergebnissen besonders wichtig
- umso bedeutender die Inhalte einer Webseite sind, umso mehr traut der Indexierungsalgorithmus ihnen
- umso angesehener die von einer Webseite referenzierten Targets sind, umso mehr traut der Indexierungsalgorithmus der Quelle

REPUTATION: PRESTIGE UND GLAUBWÜRDIGKEIT (2/2)

Reputation einer Webseite ist das Produkt ihrer

- **Prestige:**
wie viele andere (relevante) Webseiten verlinken auf sie
- **Credibility** (Glaubwürdigkeit):
auf wie viele andere (relevante) Webseiten verlinkt sie

$$\textit{Reputation} = \textit{Prestige} \times \textit{Credibility}$$

PRESTIGE

Grob gesagt: umso mehr Webseiten eingehende Links geben, umso besser

Dabei muss das Verhältnis

$$\frac{\text{Anzahl relevanter eingehender Links}}{\text{Gesamtanzahl eingehender Links}}$$

gegen 1 gehen (keine relevanten Quellen: 0; nur relevante Quellen: 1)

Beispiel: 8 eingehende Links, davon 5 von relevanten Webseiten

$$\text{Prestige} = \frac{5}{8} = 0,625$$

Dementsprechend ist eine „Internetquelle“ sehr schlecht, da $\text{Prestige} = 0$.
(wegen Produkt wird dann auch $\text{Reputation} = 0 \times \text{Credibility} = 0$)

GLAUBWÜRDIGKEIT

Grob gesagt: umso relevanter die Ziele ausgehender Links, umso besser

Dabei muss das Verhältnis

$$\frac{\text{Anzahl relevanter ausgehender Links}}{\text{Gesamtanzahl ausgehender kanonischer Links}}$$

gegen 1 gehen (keine relevanten Ziele: 0; nur relevante Ziele: >1)

Beispiel: 9 ausgehende kanonische Links; 17 relevante Ziele

$$\text{Credibility} = \frac{17}{9} = 1,8\bar{8}$$

Daher ist eine „Internetsenke“ sehr schlecht, da $\text{Credibility} = 0$.

(wegen Produkt wird dann auch $\text{Reputation} = \text{Prestige} \times 0 = 0$)

REPUTATION DES BEISPIELS

Unser Beispiel hatte:

- 8 eingehende Links, davon 5 von relevanten Webseiten
- 9 ausgehende kanonische Links
- 17 relevante Ziele

Daraus ergibt sich:

$$\textit{Reputation} = \underbrace{\textit{Prestige}}_{\frac{5}{8}=0,625} \times \underbrace{\textit{Credibility}}_{\frac{17}{9}=1,8} = 1 \frac{13}{72} \cong 1,18$$

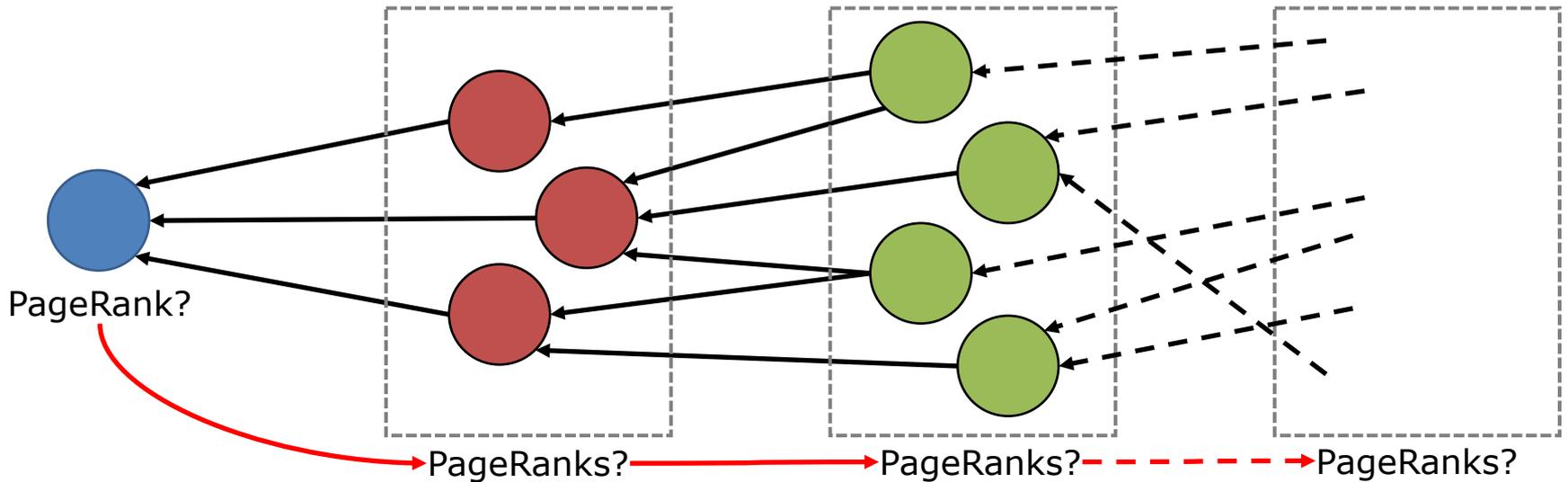
PAGERANK (1/2)

Was macht der Algorithmus nun mit der Reputation?

→ Er berechnet den PageRank!

Der PageRank berechnet sich rekursiv durch Reputationsweitergabe, d.h. die Qualität einer Webseite korreliert mit

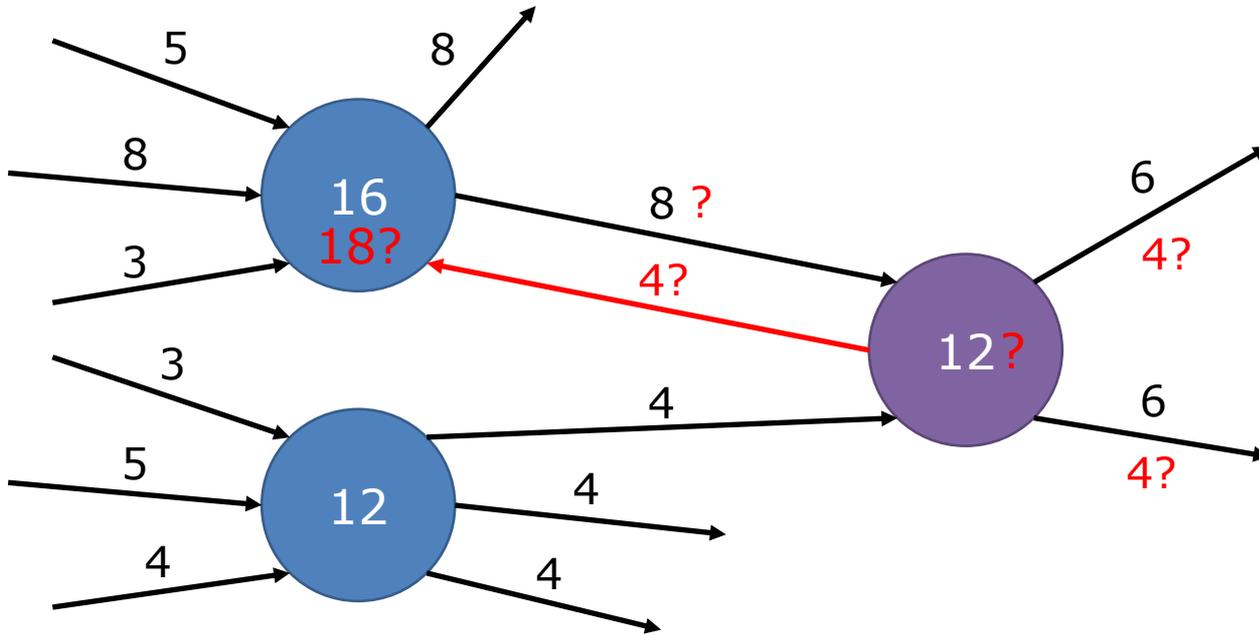
- ihrem Eingangsgrad und
- ihrer Zielqualität



PAGERANK (2/2)

Jede Webseite

- erhält Teile der Reputation ihrer Linksetzer
- gibt ihre vollständige Reputation anteilig an ihre Linkziele weiter



⇒ Verfahren wird hinreichend kompliziert sobald sich Webseiten gegenseitig verlinken... Verwendete Lösungen sind Trade-Secrets.

EXTREMFÄLLE

Authority: Webseite mit sehr vielen eingehenden Links

- scheint wichtig zu sein
- algorithmische Reputation entspricht evtl. nicht realer Reputation
- Beispiele: Standards, Hersteller-Webseiten, Nachrichten, ...

Hub: Webseite mit sehr vielen ausgehenden Links

- könnte sein, dass keine eigenen Inhalte vorliegen
- könnte aber auch sein, dass Autor viele gute Webseiten kennt
- algorithmische Reputation entspricht evtl. nicht realer Reputation
- Beispiele: Testberichte, Blogs, Suchmaschinen, ...

WAS BEDEUTET PAGERANK FÜR SEO?

- regelmäßige (!) Qualitätskontrolle bei Links
 - nur auf relevante Ziele verlinken
 - nur auf vertraute Ziele verlinken
 - auf kanonisierbare Ziele achten und relevante URLs auswählen (statt auf die Startseite oder indifferente Artikel zu verlinken, lieber auf das möglichst exakte Ziele achten)
- regelmäßiges (!) Prüfen von Server-Logs auf eingehende Links
 - ggf. Take-Down-Notices an unerwünschte Quellen senden (z.B. Porno-Webseiten, Konkurrenz, ...)
 - kanonisierbare URLs sicherstellen
 - Permalinks für besonders wichtige Inhalte erstellen

Website-Gestaltung

BARRIEREFREIHEIT

Webseiten sollten vom Großteil der Bevölkerung genutzt werden können

- gute Kontraste und leicht erfassbare Schriftarten verwenden (häufig auch einfache Sprache und Erklärvideos für Behinderte)
- niemals feste Inhaltsgrößen verwenden
→ stattdessen nur Relativgrößen, insb. bezogen auf die minimal erlaubte Schriftgröße (Browser-Einstellung)
- Screenreader unterstützen
 - alle Elemente benennen (immer alt- und title-Tag verwenden!)
 - niemals Tabellen zum Layouten verwenden, stattdessen
 - div-Container oder
 - HTML5-Semantikelemente (<article>, <aside>, <details>, <figcaption>, <figure>, <footer>, <header>, <main>, <mark>, <nav>, <section>, <summary>, <time>)

MOBILE FIRST

Webseiten sollten vor allem anderen zuerst für mobile Endgeräte entwickelt werden

- mobile Endgeräte haben i.d.R.
 - weniger Performance als PCs
 - deutlich kleinere Bildschirme als PCs
- mobiles Internet hat schlechtere Durchsatz- und Latenzwerte
 - alle Inhalte sollten adressierbar sein
 - alle Anfrage sollten bevorzugt aus dem Cache bedient werden
- von optimierten mobilen Webseiten profitieren auch PCs
 - lediglich das Design sollte sich anpassen (CSS Media Queries verwenden!)

DYNAMISCHE INHALTE – ASYNCHRONE NUTZUNG DER WEBSEITE

dynamische Inhalte sind OK (insb. nachgeladene Inhalte; AJAX) und wegen Netzwerklast und Nutzerblockierung zu bevorzugen, aber

- immer statische Alternativen anbieten (kein JavaScript-Zwang!)
- Navigationsbewusstsein darf nicht verloren gehen
 - Breadcrumbs oder Positionshinweise
 - beim Nachladen von Inhalten URL anpassen
 - Inhalte der Vor- und Zurück-Schaltflächen des Browsers aktualisieren

Server-Einstellungen

SCHNELLER INHALTSAUFBAU

- besonders wichtig sind die Informationsinhalte einer Webseite
 - diese sollten als erstes auf dem Endgerät ausgewertet werden
 - Webseite kann dann bereits in Teilen angezeigt werden, während Inhalte noch nachladen
 - blockierende Skripte erst am Ende laden
 - möglichst Interlace-Bilder verwenden
 - externe Ressourcen erst spät laden (insb. WebFonts)

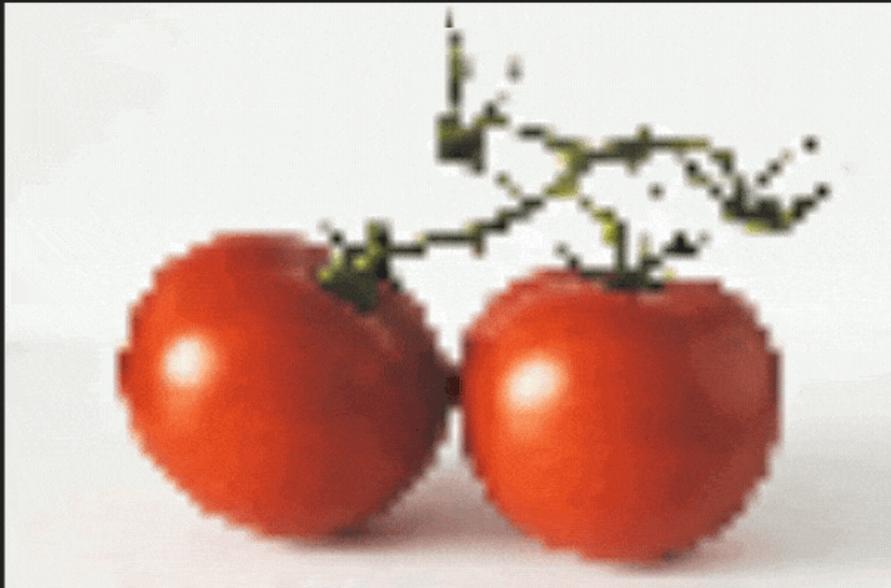
SICHERERE INHALTSAUSLIEFERUNG

neben den Inhalten an sich ist ihre sichere Übertragung wichtig

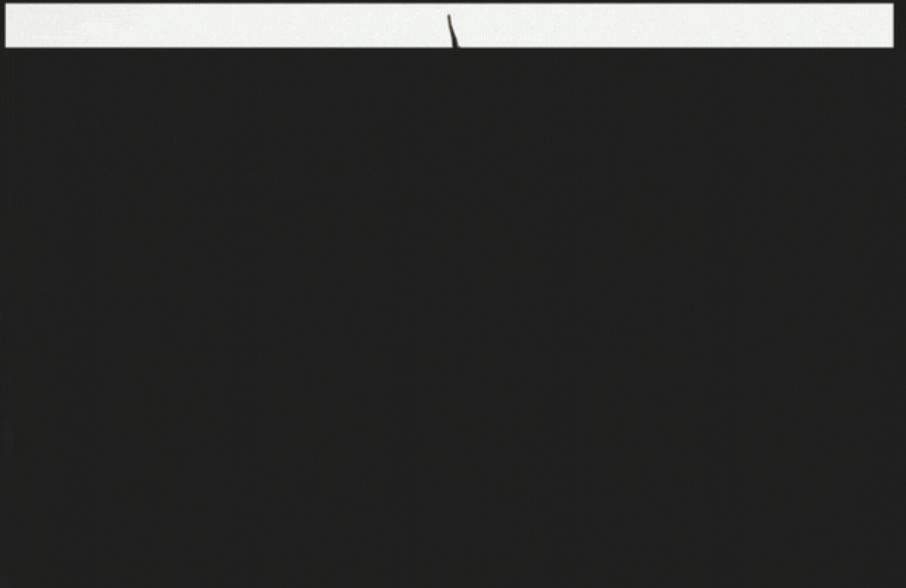
- generell HTTPs aktivieren (Google delistet reine HTTP-Seiten)
- aktuelle Cipher-Verfahren
- SSL-Labs-Empfehlungen befolgen (<https://www.ssllabs.com/ssltest/>)

INTERLACE-BILDER

Interlaced



NOT interlaced



LADEREIHENFOLGE VON INHALTEN



SERVER RICHTIG KONFIGURIEREN

- statische und dynamische Inhalte voneinander trennen
→ oft: nginx (statische Inhalte) + separater Server für Rest
 - Inhalte nach Möglichkeit komprimieren (ist nicht immer sinnvoll)
 - Inhalte in der richtigen Reihenfolge ausliefern
 - Pagespeed-Module verwenden
 - HTTP/2 mit Server-Push verwenden
- schnelle Speicher bevorzugen
 - Server-seitiges Cachen um Look-up-Time zu minimieren
 - wenig Loggen
- persistente Verbindungen erlauben
 - TCP Connection Keep Alive
 - WebSockets

weitere Google-Metriken

SINN DER GESTALTUNGS- UND SERVER-VORGABEN

Ziel: gefühlte Performance der Webseite maximieren
(nach Heise Zeitschriftenverlag: „Schwuppdizität“)

Dies bedeutet Minimierung von insb.:

- PLT: Page Load Time
- FCP: First Contentful Paint
- LCP: Largest Contentful Paint
- FMP: First Meaningful Paint
- TTI: Time to Interactive
- FCI: First CPU Idle
- FID: First Input Delay
- TTFB: Time to First Byte
- HRC: HTTP Request Count



https://www.cloudandheat.com/



Labdaten

▲ First Contentful Paint	8,8 s	▲ Time to Interactive	14,6 s
▲ Speed Index	18,6 s	▲ Total Blocking Time	950 ms
▲ Largest Contentful Paint	16,1 s	● Cumulative Layout Shift	0,003

Die Werte sind geschätzt und können variieren. Die [Leistungsbewertung](#) wird direkt aus diesen Messwerten berechnet. [See calculator.](#)



PLT – PAGE LOAD TIME

- Zeit, bis Webseite komplett geladen wurde
- meistens werden aber vorher schon Inhalte angezeigt:
 - Interlace-Bilder: erste Anzeige nach ein paar Millisekunden, aber einige Sekunden bis Bild in voller Auflösung angezeigt wird
 - WebFonts: textueller Inhalt wird schon in der Standardschriftart des Browsers angezeigt bevor die WebFont nachgeladen wird

Wie erreicht man PLT-Minimum?

- so viele Inhalte wie möglich in so wenig Dateien wie möglich
- Kompression aktivieren
- nur benötigte Inhalte senden
- Wartezeit auf Server minimieren (beeinflusst auch TTFB)

FCP – FIRST CONTENTFUL PAINT

- Zeit, bis erste Inhalte sichtbar sind
- meistens werden aber Struktur und Layout nachgebessert:
 - CSS-Dateien: erst durch die CSS-Angaben ergibt sich das Layout
 - Bilder: Platzhalter für die Bilder können bereits in der Struktur angezeigt werden

Wie erreicht man FCP-Minimum?

- Größenangaben für die meisten Bilder entweder im `img`-Tag, dem umgebenden Container oder in der CSS-Klasse
- Auslieferungsreihenfolge der Elemente optimieren; Texte bevorzugen
- Inhalte möglichst nicht durch JavaScript platzieren oder gar generieren lassen (jQuery, etc. nur in Maßen!)

LCP – LARGEST CONTENTFUL PAINT

- Zeit, bis das größte Inhaltselement sichtbar ist
- i.d.R. Zeit für das größte Bild

Wie erreicht man LCP-Minimum?

- Bilder gut komprimieren (gilt auch für andere Medien)
- prüfen, ob darzustellender Inhalt wirklich nur durch ein Bild repräsentiert werden kann → Aufteilen in mehrere Teilbilder
- Medien vorladen und cachen

FMP – FIRST MEANINGFUL PAINT

- Zeit, bis sinnvolle Inhalte sichtbar sind, insb. die, welche die Besuchenden sehen wollen
- meistens werden falsche Prioritäten gesetzt: Laden von Hintergrundbildern und Schriftarten bevor Inhalte da sind

Wie erreicht man FMP-Minimum?

- wichtige Inhalte in der Auslieferungsreihenfolge bevorzugen
- zwischen Seiten so viel Struktur und Layout recyceln wie möglich (Ausnutzen, dass Browser insb. CSS- und JS-Dateien cachen)

TTI – TIME TO INTERACTIVE

- Webseite zeigt gemäß FCP sinnvolle Inhalte an
- Event-Handler wurden für die meisten sichtbaren Elemente registriert
- Webseite reagiert binnen 50ms auf Nutzerinteraktion
(Nutzer können also mit den gebotenen Inhalten sinnvoll interagieren)

Vorsicht: TTI und FMP können gegeneinander arbeiten; während man versucht, ein Element zu bedienen, wird dieses durch nachgeladene Struktur oder Layout weg bewegt

Wie erreicht man TTI-Minimum?

- wichtige Inhalte mitsamt Struktur und Layout in der Auslieferungsreihenfolge bevorzugen
- Elemente im Quelltext so platzieren, dass sie in Nachbarschaft ihrer späteren FMP-Position liegen

FCI – FIRST CPU IDLE

- Zeit, bis Webseite minimal-interaktiv wird
- erste Elemente können bedient werden, auch wenn der (angezeigte) Rest noch nicht interaktiv ist
- FCI ist naturgemäß früher als TTI erreicht

Wie erreicht man FCI-Minimum?

- Befolgen der FCP- und TTI-Tipps
- es ist nicht sinnvoll, die Webseite primär auf FCI zu optimieren (es frustriert Nutzende nur, wenn manche Elemente benutzt werden können und andere nicht)

FID – FIRST INPUT DELAY

- Zeit zwischen erster Nutzerinteraktion und erster möglicher Reaktion des Browsers darauf
- Indiz für Lade-Flaschenhalse zwischen erstem Laden und letztem Interaktionselement

Wie erreicht man FID-Minimum?

- blockierende Skripte vermeiden
- parallele Skripte vermeiden
- Request-Cancellation ermöglichen
(Beispiel: Führt eine Interaktion zum Verlassen der Webseite, sollten alle verbleibenden Handler abgebrochen werden, damit der Browser sich sofort um das Laden der anderen Webseite kümmern kann)

TTFB – TIME TO FIRST BYTE (1/2)

- Zeit, bis das erste Byte der angeforderten Webseite ankommt
- TTFB entspricht im Grund der Round-Trip-Time (RTT) und besteht aus vielen Teilzeiten:
 - Request Time (ReqT): Zeit, bis die Anfrage beim Server ankommt
 - Processing Time (ProcT): Zeit, die der Server zum Verarbeiten der Anfrage und Bereitstellen der angeforderten Ressourcen benötigt
 - Response Time (ResT): Zeit, bis die Antwort beim Browser ankommt

TTFB – TIME TO FIRST BYTE (2/2)

Wie erreicht man TTFB-Minimum?

- Content Delivery Netze (CDN) verwenden
- Inhalte möglichst statisch ausliefern
- Medien (Videos und Audio) streamen
- Server richtig konfigurieren (siehe Abschnitt „Server-Einstellungen“)
- Load-Balancing: reduziert Auslastung der einzelnen Server
- Netzwerk-Latenzen vermeiden: Internetanschluss richtig skalieren
- DNS richtig konfigurieren: Verschachtelungen und DNS-Weiterleitungen vermeiden → Domainname muss möglichst zur IP-Adresse übersetzen
 - Negativbeispiel: cloudandheat.de → www.cloudandheat.de → cloudandheat.com → www.cloudandheat.com → 185.128.118.144
 - Besser: cloudandheat.de → 185.128.118.144

HRC – HTTP REQUEST COUNT

- Anzahl echter HTTP-Anfragen, die Browser an Server senden muss
- jede Anfrage braucht mindestens eine RTT zur Beantwortung
- jede einzelne Ressource (Text, Bild, CSS, Skript, ...) muss beim Server angefordert werden

Wie erreicht man HRC-Minimum?

- mindestens HTTP/1.1 mit TCP Connection Keep Alive verwenden
- möglichst viele Ressourcen konsolidieren
 - Image-Maps verwenden
 - JavaScript-Dateien zusammenfassen
 - CSS-Dateien zusammenfassen
- Parallelisierung von Anfragen ermöglichen (bspw. Medien von einer eigenen Subdomain aus ausliefern)
- Wo möglich, auf HTTP/2 mit Pagespeed/Oxygen und Server-Push setzen

Page Experience Ranking

PAGE EXPERIENCE RANKING

seit 20. Mai 2021 für Desktop-Suchergebnisse

- Barrierefreiheit der Webseite
- Sicherheit und Datenschutz
- Ladetempo
- **User Annoyance** (Nerv-Faktor)

USER ANNOYANCE

- Auswertung missbräuchlicher Einblendungen
- Bewegungsgrad von Seiteninhalten (**Cumulative Layout Shift**) (bspw. durch Nachladen von Werbung)
 - betrachtet Änderungen im sichtbaren Bereich der Website in den ersten 5 Sekunden nach dem Laden (nach LCP und FID):

$$\frac{\text{bewegte Elemente}}{\text{Gesamtelemente}} \times \frac{\sum(\text{verschobene Pixel}) / (\text{bewegte Elemente})}{\text{sichtbare Pixel}}$$
$$= \frac{\sum(\text{verschobene Pixel})}{(\text{Gesamtelemente}) \times (\text{sichtbare Pixel})} \leq \underbrace{0,1}$$

- **User Annoyance soll unter 0,1 liegen**
- vom User gewollte Seitenveränderungen werden nicht mit einbezogen (etwa nach Klick auf einen Link)
- nützliche/notwendige Interstitials werden nicht bestraft (Cookie-Banner, Event-Banner, Corona-Hinweise, ...)

Google Analytics

TUTORIAL

Ein sehr umfangreiches Tutorial (4h) gibt es auf Youtube:

<https://youtu.be/e6ntvZDErQ4>

etwas kürzer als Blogeintrag:

<https://moz.com/blog/absolute-beginners-guide-to-google-analytics>

oder

<https://blog.hubspot.com/marketing/google-analytics>

DATENBASIS

Google Analytics sammelt mit jedem Seitenzugriff Metriken, mindestens

- Browser (Typ und Version)
- Gerätetyp und Betriebssystem
- IP-Adresse
- Zugriffszeit und -dauer
- falls gesetzt: Referrer und Exit

Darüber hinaus kann Analytics Nutzer über mehrere Seiten verfolgen

- z.B. weil sie Social Media verwenden oder
- in ihr Google Konto eingeloggt sind
- erfasst dann Zugriffsart (Privat, geschäftlich), Geschlecht, Rasse, Alter, Suchinteressen, sexuelle Orientierung, Zielgruppenklammer, ...

AUSWERTUNG

- viele voreingestellte Auswertungen möglich
- sinnvoll ist jedoch Definition eigener KPI
 - Wer sind die Besucher?
 - Was interessiert die Besucher?
 - Was machen die Besucher auf der Webseite?
 - ...

ZIELE DEFINIEREN

The screenshot displays the Google Analytics interface with three main panels: Account, Property, and View. The left sidebar shows the 'REPORTS' menu with 'Admin' highlighted. A red dashed arrow labeled '1' points to the 'Admin' button. The 'Account' panel shows 'physik.aneiki.de' with options like 'Account Settings' and 'Account User Management'. The 'Property' panel shows 'http://physik.aneiki.de (UA-2635381-1)' with options like 'GA4 Setup Assistant' and 'Property Settings'. The 'View' panel shows 'physik.aneiki.de' with options like 'View Settings' and 'View User Management'. A red dashed arrow labeled '2' points to the 'Goals' option in the 'View' panel.

Ziele erzeugen unterschiedliche Ansichten (Views). Jede Webseite soll min. 1 View haben.

ZIELE DEFINIEREN (2/5)

The screenshot shows the Google Analytics interface. On the left is a navigation sidebar with various settings and tools. The main content area is titled 'Goals' and contains a table with columns for 'Goal', 'ID', 'Type', 'Past 7 day conversions', and 'Recording'. The table is currently empty, displaying the message 'There is no data for this view.' A red arrow points from a red circle containing the number '1' to the '+ NEW GOAL' button in the top left of the table area.

Goal	ID	Type	Past 7 day conversions	Recording
There is no data for this view.				

ZIELE DEFINIEREN (3/5)

View [+ Create View](#)

- View Settings
- View User Management
- Goals**
- Content Grouping
- Filters
- Channel Settings
- E-commerce Settings
- Calculated metrics BETA
- PERSONAL TOOLS & ASSETS
- Segments
- Annotations
- Multi-Channel Funnels Settings
- Custom Channel Grouping BETA
- Custom Alerts
- Scheduled Emails
- Saved Reports
- Share Assets

1 Goal description

Name

Pages seen

Goal slot ID

Goal ID 1 / Goal Set 1

Type

- Destination e.g. thanks.html
- Duration e.g. 5 minutes or more
- Pages/Screens per session e.g. 3 pages
- Event e.g. played a video
- Smart Goal Smart Goal not available.

Measure the most engaged visits to your website and automatically turn those visits into Goals. Then use those Goals to improve your Google Ads bidding. [Learn more](#)

[Continue](#) [Cancel](#)

2 Goal details

[Cancel](#)

Ziele können für verschiedene Metriken definiert werden.

ZIELE DEFINIEREN (4/5)

- View [+ Create View](#)
- View Settings
- View User Management
- Goals**
- Content Grouping
- Filters
- Channel Settings
- E-commerce Settings
- Calculated metrics BETA
- PERSONAL TOOLS & ASSETS
- Segments
- Annotations
- Multi-Channel Funnels Settings
- Custom Channel Grouping BETA
- Custom Alerts
- Scheduled Emails
- Saved Reports
- Share Assets

Goal description [Edit](#)
Name: *Pages seen*
Goal type: *Pages/Screens per session*

2 Goal details

Pages/Screens per session **greater than**

Value optional
 OFF Assign a monetary value to the conversion.

[Verify this Goal](#) See how often this Goal would have converted based on your data from the past 7 days.

Basierend auf dem Typ der Metrik kann dann das Ziel angegeben werden.

Den meisten Zielen kann ein monetärer Wert zugewiesen werden.

ZIELE DEFINIEREN (5/5)

View [+ Create View](#)

- View Settings
- View User Management
- Goals**
- Content Grouping
- Filters
- Channel Settings
- E-commerce Settings
- Calculated metrics BETA
- PERSONAL TOOLS & ASSETS
 - Segments
 - Annotations
 - Multi-Channel Funnels Settings
 - Custom Channel Grouping BETA
 - Custom Alerts
 - Scheduled Emails
 - Saved Reports
 - Share Assets

+ NEW GOAL	Import from Gallery	Search			
<input type="checkbox"/>	Goal	ID	Type	Past 7 day conversions	Recording
<input type="checkbox"/>	Pages seen	Goal ID 1 / Goal Set 1	Pages/Screens per session	0	<input checked="" type="checkbox"/> ON

Show rows 10 1 - 1 of 1 < >

Bis zu 20 Ziele sind möglich.

Es müssen nicht alle Ziele aktiv sein.



BESUCHERVERHALTEN PRÜFEN: FUNNEL

für bestimmte Inhalt ist es gut zu wissen, ob Besucher sie in bestimmten Reihenfolgen aufsuchen

- Definieren eines Trichters (Funnel) zur Ziel-Seite, z.B.
 1. Einstieg: Startseite
 2. Software
 3. Customers
 4. Ziel: Kontakt
- den Funneln können wieder Geldwerte zugewiesen werden
- Überprüfen der Funnel via [Conversions](#) → [Goals](#) → [Funnel Visualization](#)

BESUCHERVERHALTEN PRÜFEN: PAGES PER SESSION (1/2)

Wie schnell finden Besucher die gewünschten Inhalte?

- Definieren eines Zielwertes für besuchte Seiten pro Session
 - individuelle URLs werden als Seiten gezählt
(d.h. /software und /software/#collapse1 sind *eine* Seite)
 - jede Session dauert mindestens 30 Sekunden
(d.h., finden zwei sehr kurze Besuche binnen 30 Sekunden statt, zählt das als *eine* Session)
 - jede Session dauert höchstens 3 Minuten
(d.h., wenn Besucher längere Artikel lesen und dann die nächste Seite aufrufen, kann es sein, dass das als neue Session zählt)
→ Individuelle Sessions können länger als 3 Minuten dauern

BESUCHERVERHALTEN PRÜFEN: PAGES PER SESSION (2/2)

Wie schnell finden Besucher die gewünschten Inhalte?

- **Vorsicht: niedrige Pages-per-Session-Werte bedeuten nicht zwingend, dass Inhalte schnell gefunden werden**
 - kann auch bedeuten, dass Besucher schnell abspringen, weil sie meinen, den gewünschten Inhalt nicht finden zu werden...
- Was ist dann ein gutes Ziel?
 - Etwas mehr als der Durchschnitt des relevanten Betrachtungszeitraum?
 - Ein (mit Absicht) höher gesetztes Ziel!
 - Erlaubt Analyse nach Nutzertyp.
 - Was hat gut geklappt? Was eher nicht?
 - Wo kann verbessert werden?

BESUCHERVERHALTEN PRÜFEN: EVENTS

Was, wenn Besucher Ereignisse auslösen, die Analytics nicht nativ erfassen kann?

- Downloads (z.B. PDF-Broschüre)
 - Öffnen und Schließen von Reitern
(/software → /software#collapse1)
 - Swipen von Bildanzeige
 - ...
-
- benötigt definierte Kategorie (Category) und Schlüsselwort (Tag)
 - Google Tag Manager
 - <https://analytics.google.com/analytics/academy/course/5>
 - <https://youtu.be/lohO5eFAK60>
 - <https://youtu.be/7zoJR6uq2g>
 - Ereignisse (Events) werden durch Interaktionen mit Tags ausgelöst

Google Tag Manager

EVENTS TRACKEN

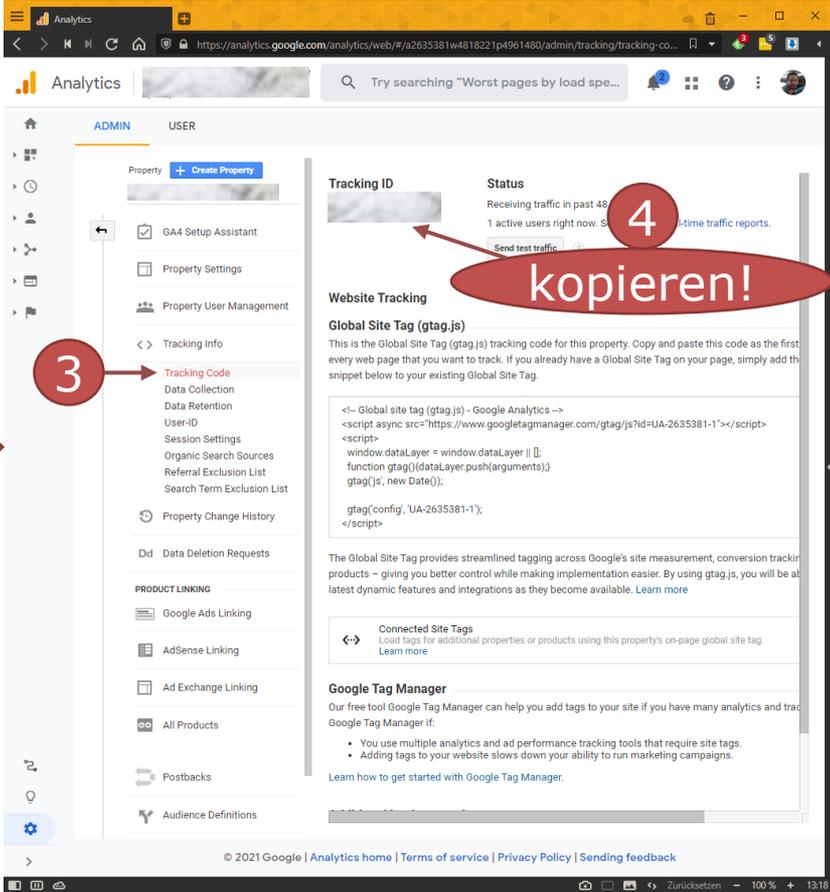
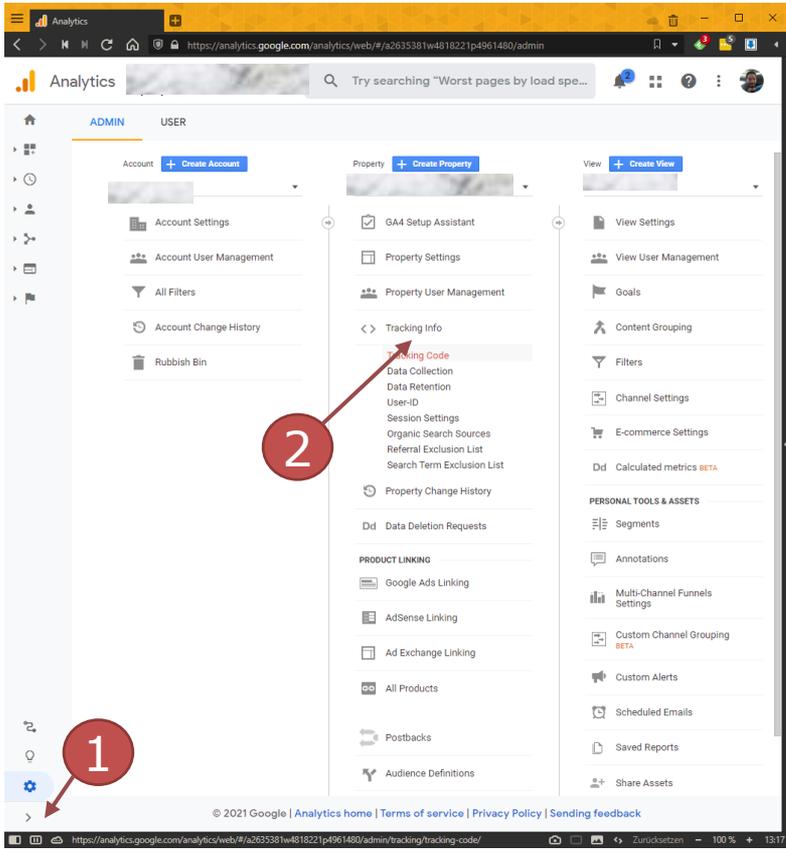
Events können nur mit definierten Tags getrackt werden!!

The screenshot shows the Google Tag Manager interface. At the top, there is a navigation bar with 'Tag Manager' and a search bar. Below this, there are tabs for 'Workspace', 'Versions', and 'Admin'. The main content area is titled 'CURRENT WORKSPACE' and shows a list of tags. The left sidebar contains navigation options: Overview, Tags (selected), Triggers, Variables, Folders, and Templates. The tag list has columns for Name, Type, Firing Triggers, and Last Edited.

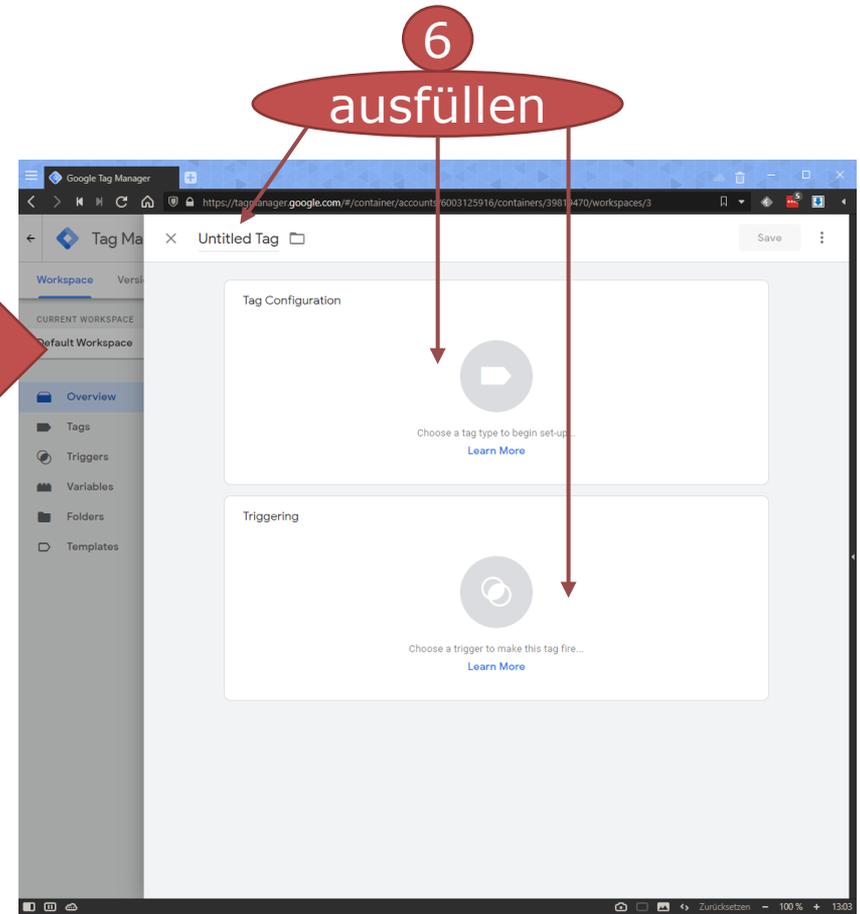
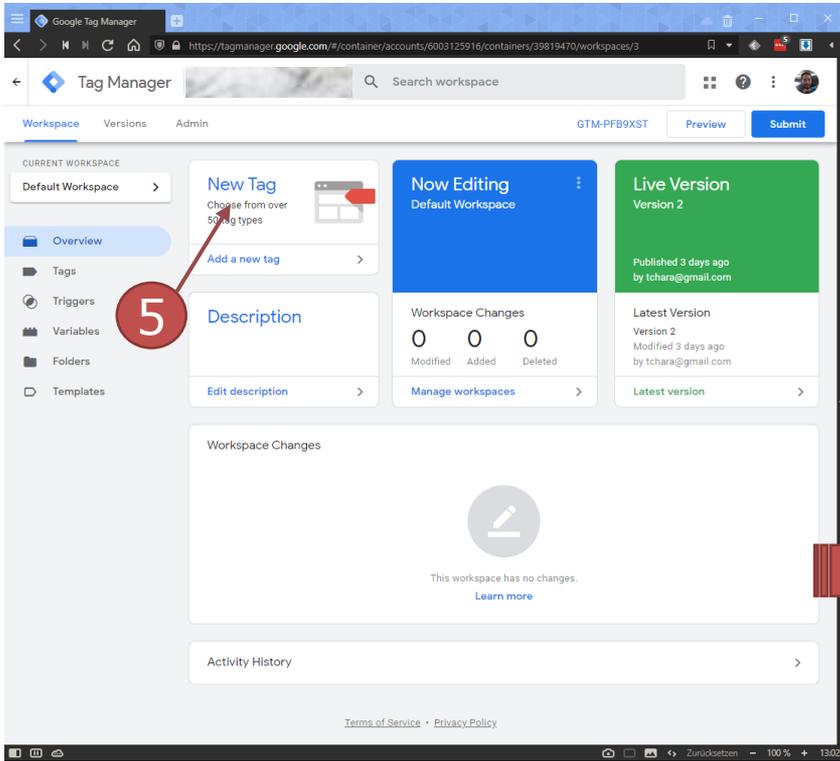
<input type="checkbox"/>	Name ↑	Type	Firing Triggers	Last Edited
<input type="checkbox"/>	Event Donations Amazon	Google Analytics: Universal Analytics	Amazon-Wunschliste	11 minutes ago
<input type="checkbox"/>	Event Donations PayPal	Google Analytics: Universal Analytics	PayPal Button	6 minutes ago
<input type="checkbox"/>	Google Analytics Universal Analytics	Google Analytics: Universal Analytics	All Pages	31 minutes ago
<input type="checkbox"/>	View of ETP	Google Analytics: Universal Analytics	ETP	24 minutes ago
<input type="checkbox"/>	View of phys. GP	Google Analytics: Universal Analytics	phys. GP	25 minutes ago

TRACKING-ID VON ANALYTICS HOLEN

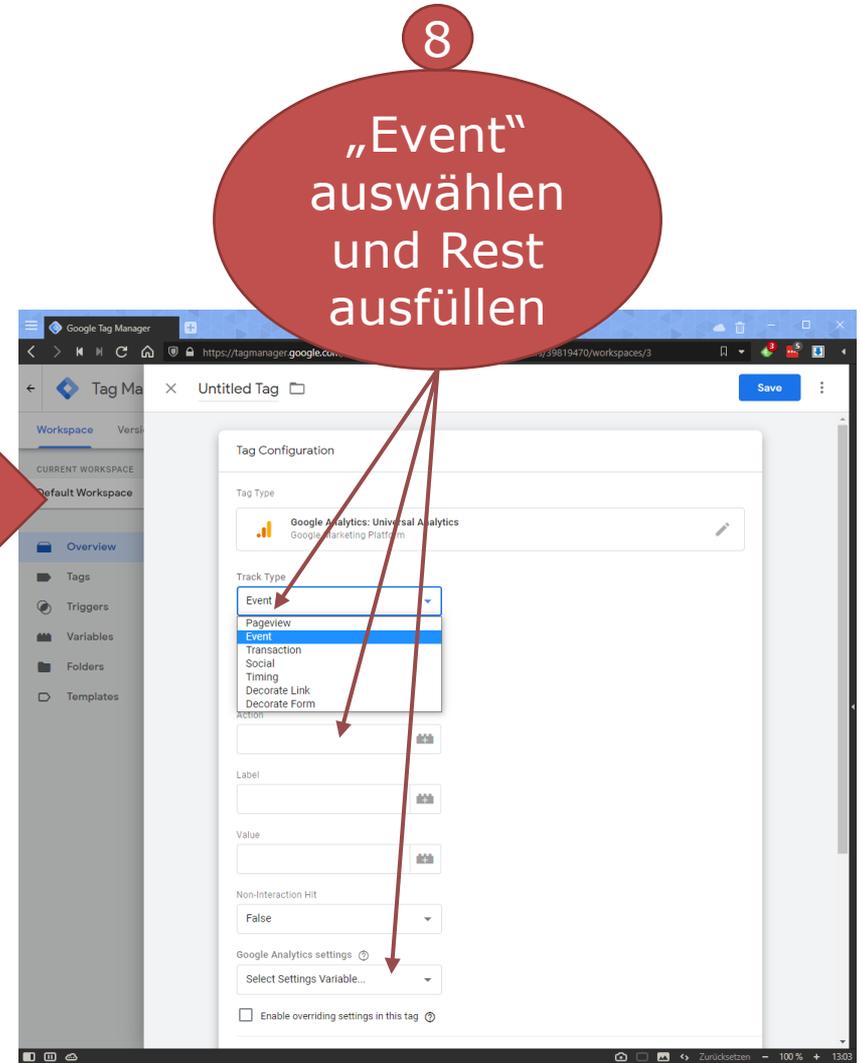
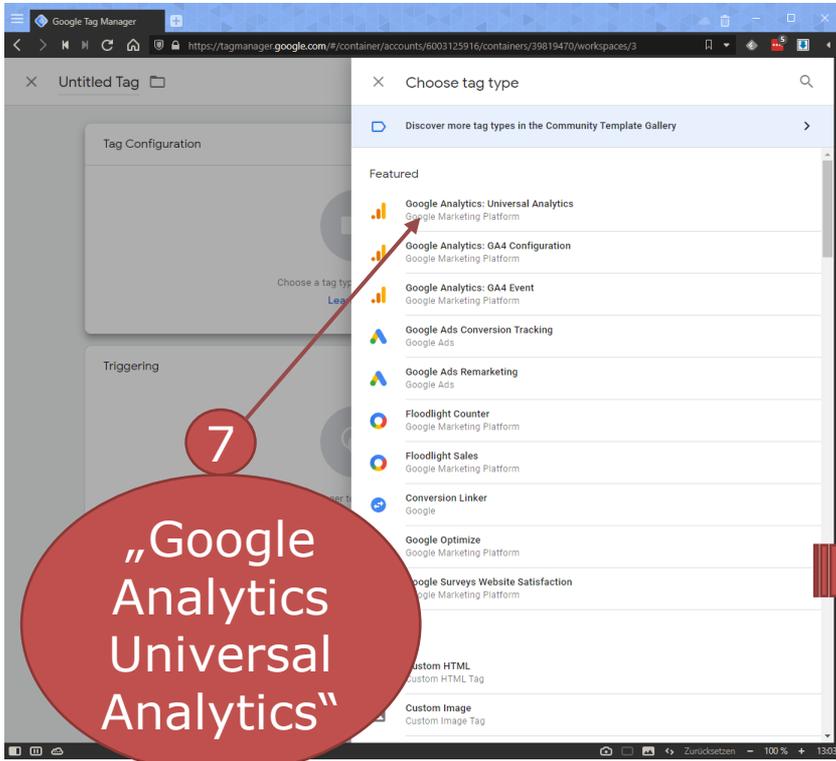
Aus Google Analytics wird die Tracking-ID benötigt, um Events im Tag-Manager anzulegen



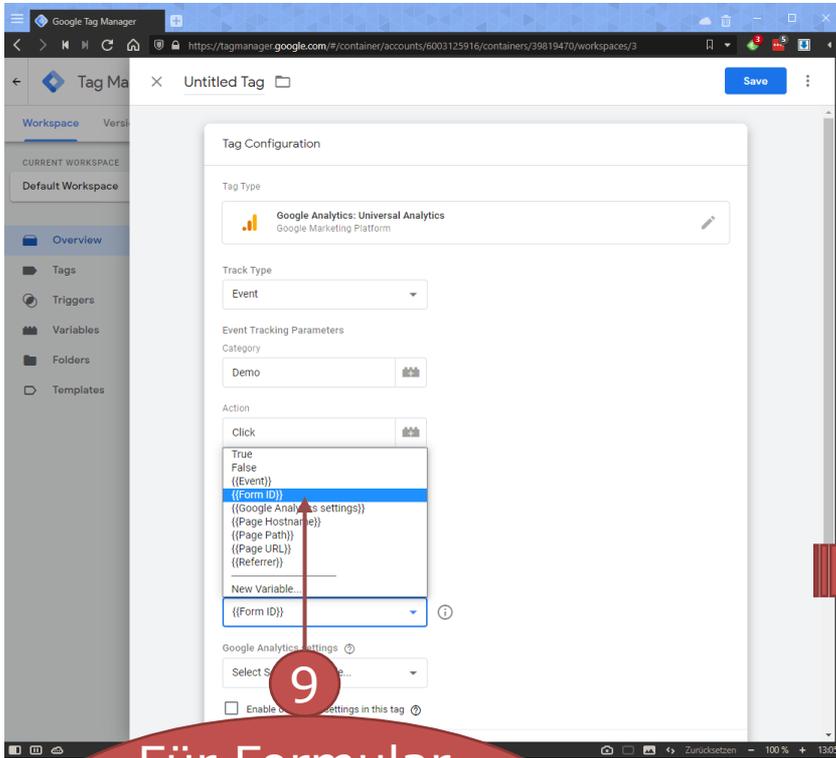
NEUEN TAG ANLEGEN



VERKNÜPFUNG ZU GOOGLE ANALYTICS ERZEUGEN



BINDEN AN FORMULAR IN WEBSEITE



Für Formular-basierte Events „{{Form ID}}“ auswählen

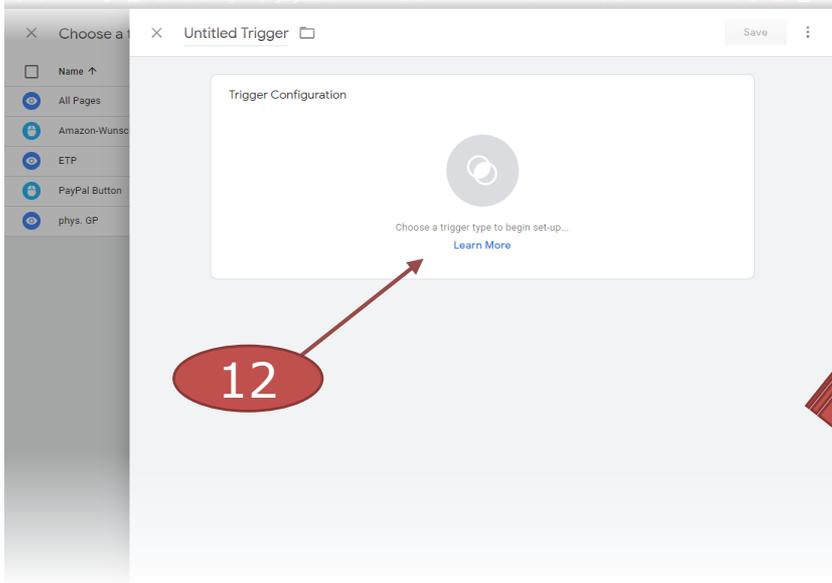


TRACKING ID EINFÜGEN

11
Die aus Analytics
kopierte
Tracking-ID
einfügen

The image shows a screenshot of the Google Tag Manager interface. The top panel, titled 'Variable Configuration', is for a 'Google Analytics settings' variable. It has a 'Tracking ID' field with a red arrow pointing to it from the callout box. Below it is a 'Cookie Domain' field set to 'auto'. A 'More Settings' link is at the bottom. The bottom panel, titled 'Untitled Tag', shows the configuration for a tag. The 'Value' field contains '({{Form ID}})'. Below it, the 'Google Analytics settings' dropdown is set to '({{Google Analytics settings}})'. A 'Triggerring' section is visible at the bottom.

TRIGGER DEFINIEREN

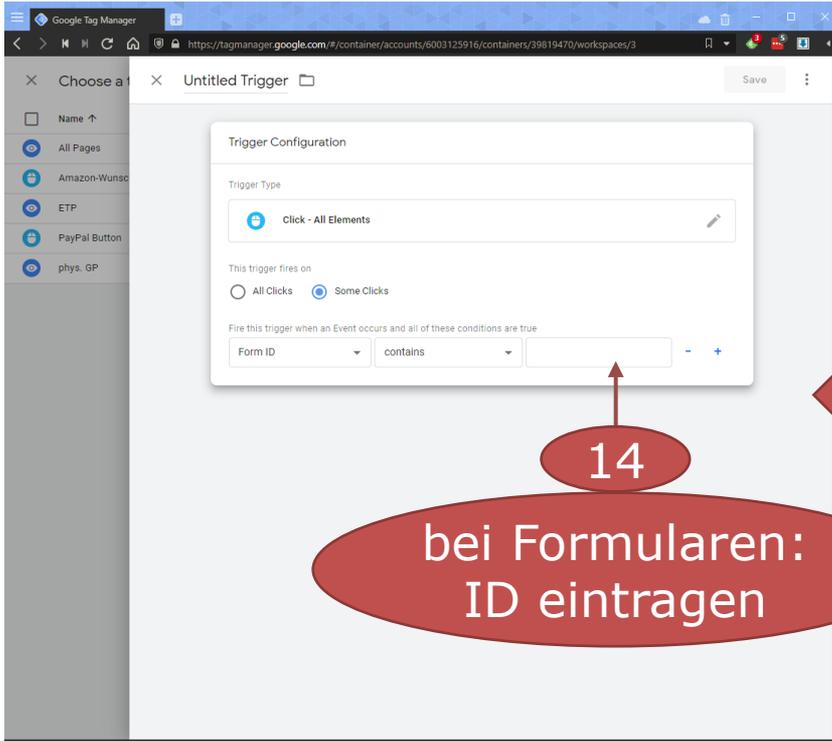


12



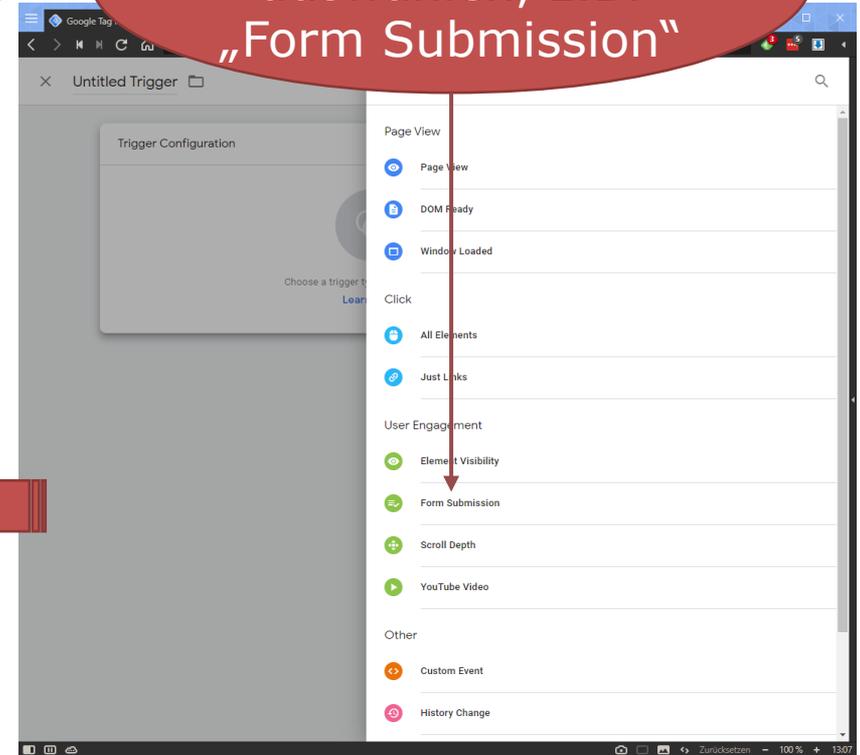
13

Trigger-Typ auswählen, z.B. „Form Submission“



14

bei Formularen:
ID eintragen



EVENT-TYPEN

- allgemein kann alles ein Event triggern, was Browser-Reaktionen hervorrufen kann (Element-Click, Formularfeld ausfüllen, Mouseover, ...)
- i.d.R. fein granular möglich, wenn Website-Quellcode es erlaubt
→ IDs bei wichtigen Elementen, Objekten, Formularen, ...
- umso mehr getrackt wird, umso langsamer die Webseite
→ **Tags spärlich verwenden!** (nur da, wo wirklich Analytics passiert)
- **gute Abstimmung zwischen Redakteuren (Content) und Administratoren (Backend) notwendig**

Konkurrenz zu Google

MATOMO

<https://matomo.org>

- Open Source und selbst gehostet (Matomo On-Premise)
- wenn richtig konfiguriert mit DS-GVO vereinbar
- ähnlich Google Analytics
- Datenimport von Google Analytics möglich (Bewahren historischer Berichte)
- Tracken von Funneln und Events ohne separaten Tag Manager

AWSTATS

<https://awstats.sourceforge.io>

- Open Source und selbst gehostet
- wenn richtig konfiguriert mit DS-GVO vereinbar
- Auswertung basiert auf Server-Logs
- Tracken von Funneln und Events nicht möglich
→ lediglich Erkennung von Unique Users und Sessions

GoACCESS

<https://goaccess.io>

- Open Source und selbst gehostet
- wenn richtig konfiguriert mit DS-GVO vereinbar
- Auswertung basiert auf Server-Logs
(entsprechend Log-Granularität oder in Echtzeit)
- Tracken von Funneln und Events nicht möglich
→ Metriken orientieren sich eher an technischer SEO

UMAMI

<https://github.com/mikecao/umami>

- Open Source und selbst gehostet
- wenn richtig konfiguriert mit DS-GVO vereinbar
- Auswertung basiert auf Tracking-JS
- Server-Logs sind mit etwas Aufwand einbindbar
- Tracken von Funneln und Events möglich
→ Pflegen der Datenbank zwingend notwendig!

Zusammenfassung

EINFLUSSFAKTOREN FÜR DIE TECHNISCHE SEO

- TF:IDF
- Pagerank
- Website-Gestaltung
- Server-Einstellungen
- weitere Google-Metriken
- Google Analytics

SEO-WERKZEUGE

- Pagespeed: <https://developers.google.com/speed/pagespeed/>
- Lighthouse: <https://web.dev/measure/>
- Barrierefreiheit:
 - <https://www.w3.org/WAI/ER/tools/>
(z.B. <https://www.alumnionlineservices.com/accessibility/scanner/>)
 - in Deutschland wichtig: BITV-2.0-Test und WCAG-Test
- Sicherheit: <https://www.ssllabs.com/ssltest/>

Am Beispiel www.ba-dresden.de:

- <https://developers.google.com/speed/pagespeed/insights/?url=www.ba-dresden.de>
- <https://lighthouse-dot-webdotdevsite.appspot.com//lh/html?url=https://www.ba-dresden.de>
- <https://www.webaccessibility.com/results/?url=https://www.ba-dresden.de>
- <https://www.ssllabs.com/ssltest/analyze.html?d=www.ba-dresden.de&hideResults=on>