

# Webanalyse

Definitionen, Verfahren und Unterschiede

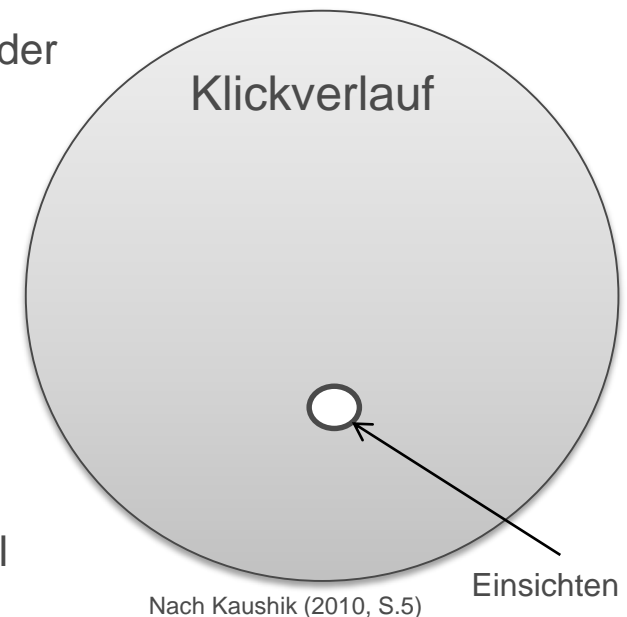
Stefan Keil, DIPF

# Agenda

1. Einleitung
  - a. Erhebungsmethoden
2. Definitionen verschiedener Metriken
3. Datenschutz – Eckpunkte
4. Zusammenfassung – Fazit
5. Diskussion

# 1. Einleitung

- Nutzungsverhalten und Webanalyse heutzutage zentrales Thema bei dem Betreiben einer Webseite
- Vielfalt an Tools und vergleichsweise einfache Erhebung der Daten
- „Paradox of data“ (vgl. Kaushik 2010, S. 2):
  - Steigende Datenmenge  $\neq$  steigende Erkenntnisanzahl



# 1. Einleitung - Begriffe

## Webmonitoring:

- Unter Webmonitoring wird im allgemeinen die zeitkritische und systematische Sammlung und Auswertung von Daten im Web verstanden“

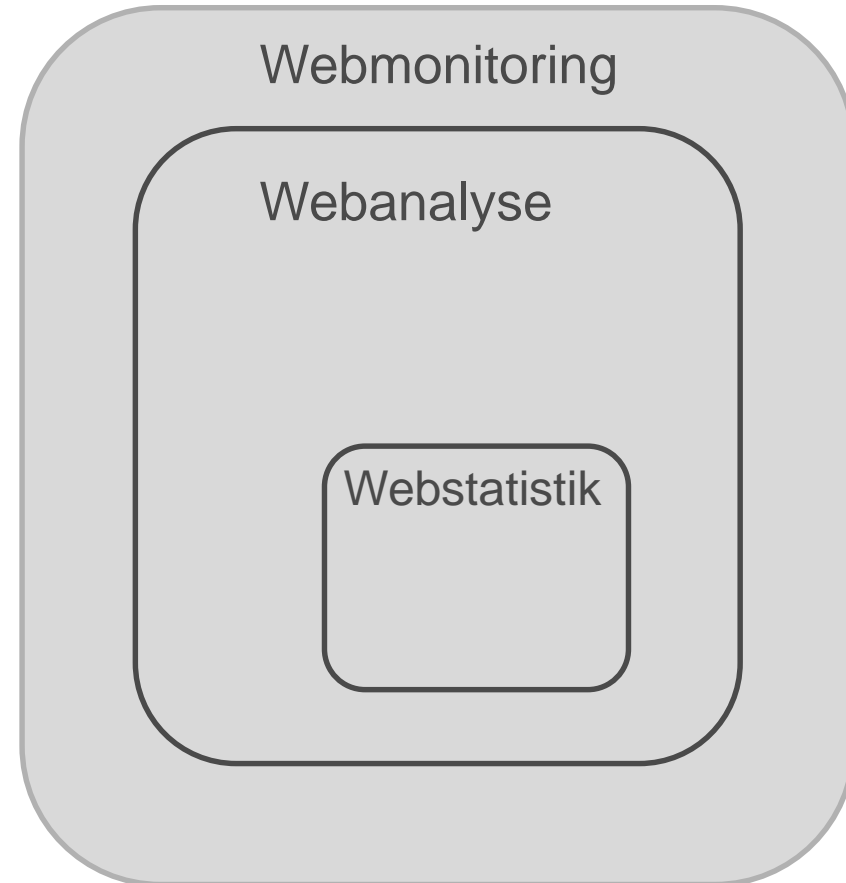
(Höchstötter & Lewandowski 2014, S. 28)

## Webanalyse:

- Webanalyse ist die Messung, Erfassung, Aufbereitung und Analyse von Daten, die bei der Nutzung einer Webseite entstehen (Wikipedia 2014)

## Webstatistik:

- Eher „einfache Betrachtung“ von aggregierten Zahlen



# 1. Einleitung - Definition Webanalyse

- **Webanalyse** ist die Messung, Erfassung, Aufbereitung und Analyse von **Daten**, die bei der **Nutzung einer Webseite** entstehen

## Onsite

- Seitenaufrufe
- Downloads
- Suchen
- ...

## Offsite

- Verlinkung
- Social Media Erwähnungen
- Sichtbarkeit in Suchmaschinen
- ...

Vgl. Wikipedia (2014)

# 1. Einleitung – Erhebungsmethoden

Logfiles	Java-Script-Tagging
Basiert auf Server Logfiles	„Markierung“ (Page-Tagging) zu erfassender Seiten
Filterung	Konzeption und Kontrolle der Implementierung
Keine externen Links	Ggf. externe Links
Caching & Robots	JavaScript-Aktivierung
<b>z.B. http-Analyze, Webalyzer, Urchin</b>	<b>Piwik, Google Analytics</b>

# 1. Einleitung – Filtern der Logfiles

- Welche Zugriffe werte ich aus?
  - Crawler? (UserAgent: crawl-66-249-78-221.googlebot.com)
  - Wie filtere ich? (Positivlisten (z.B. gängige Browser)? „Nur“ Ausschlüsse bestimmter Crawler/IP-Adressen?)
- Welche Seiten werte ich aus?
  - Nur „Webseiten? (HTML-Dokumente?)
  - Beispiel: „/get\_sw.php“ 15.838 Aufrufe in den Logfiles

```
{  
  query:'qual',  
  suggestions:['Qualität', 'Qualifikation', 'Qualitätssicherung', 'Qualifizierung',  
'Qualifikationsanforderung', 'Qualitätsmanagement', 'Qualitätsentwicklung']  
}
```

# 1. Einleitung – Vergleich Tools

- ..., obwohl gerne so getan wird, als seien es immer sehr präzise Zahlen, unterscheiden sich diese teils sehr stark zwischen den Tools/Erhebungsmethoden (Seitenaufrufe ein Tag)

URL	Urchin	Piwik	Google Analytics
Startseite	1555	757	702
URL1	214	304	304
URL2	91	155	153





## 2. Metriken

- (eindeutiger) Seitenaufruf



- Besucher

- Sitzung

- Absprungrate

- Ausstiege

- Verweildauer

- „Pageview“
- Aufruf einer bestimmten Seite
- Eindeutig: Mehrfachaufrufe in einer Sitzung werden nur einmal gezählt

## 2. Metriken

- (eindeutiger) Seitenaufruf

- Besucher



- „unique visitor“
- Anzahl verschiedener Browser (tlw. IPs), die die Seite in einem bestimmten Zeitraum besucht haben

- Sitzung

- Absprungrate

- Ausstiege

- Verweildauer

## 2. Metriken

- (eindeutiger) Seitenaufruf

- Besucher

- Sitzung



- „Besuch“, „visit“

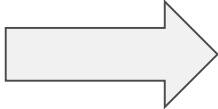
- „Zusammenhängende“  
Aktionen eines  
Besuchers in einem  
**bestimmten** Zeitraum

- Absprungrate

- Ausstiege

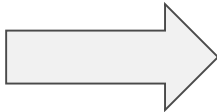
- Verweildauer

## 2. Metriken

- (eindeutiger) Seitenaufruf
- Besucher
- Sitzung
- Absprungrate 
- Ausstiege
- Verweildauer


- „bounce rate“
- Anteil der Sitzungen, die nur eine Aktion (z.B. einen Seitenaufruf) umfassen
- Wird auf Seitenebene bestimmt

## 2. Metriken

- (eindeutiger) Seitenaufruf
- Besucher
- Sitzung
- Absprungrate
- Ausstiege 
- Verweildauer

- „Ausstiegsrate“, „exit rate“
- Anteil der Sitzungen bei denen die Seite die letzte Aktion war

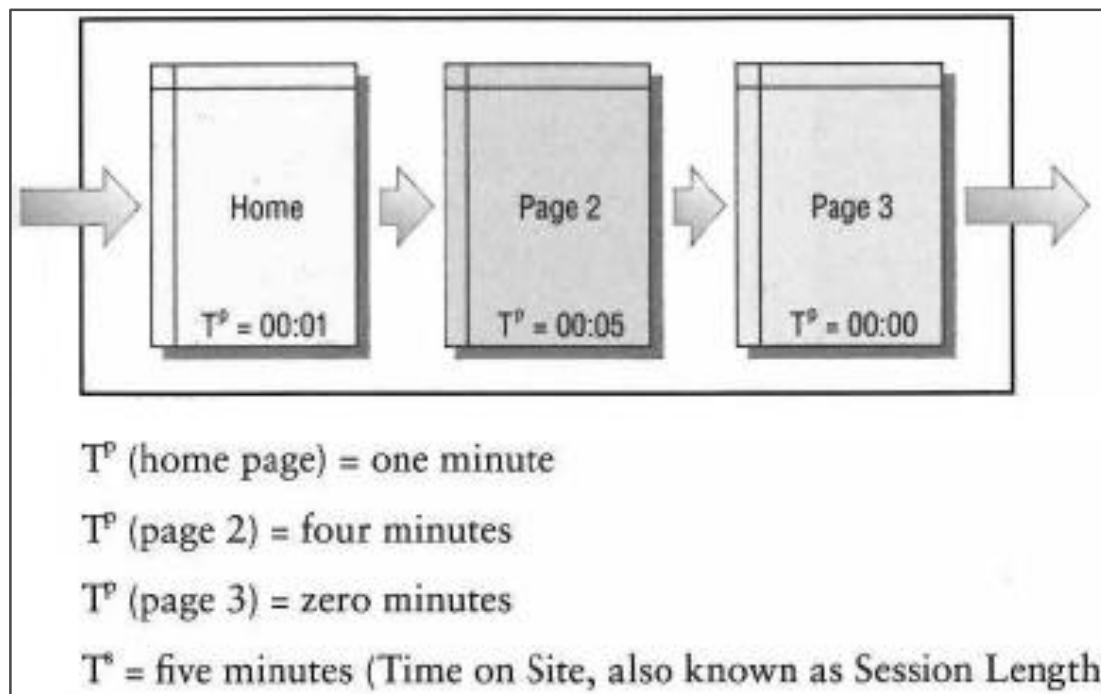
## 2. Metriken

- (eindeutiger) Seitenaufruf
- Besucher
- Sitzung
- Absprungrate
- Ausstiege
- Verweildauer 

- „**Schätzung**“ der Zeit, die auf einer Seite verbracht wird
- Berechnet sich aus dem zeitlichen Abstand **zwischen zwei** Aktionen

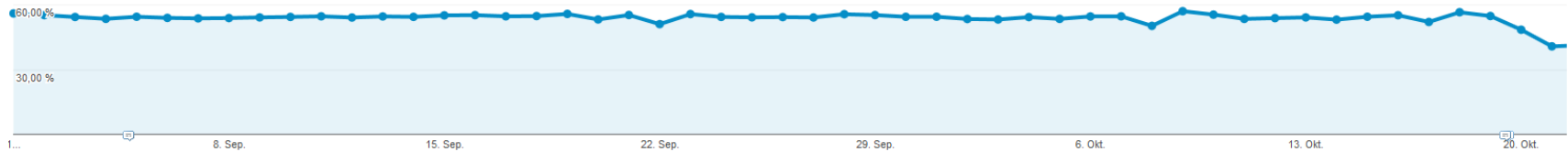
## 2. Metriken - Verweildauer

- ...berechnet sich aus dem zeitlichen Abstand **zwischen zwei** Aktionen

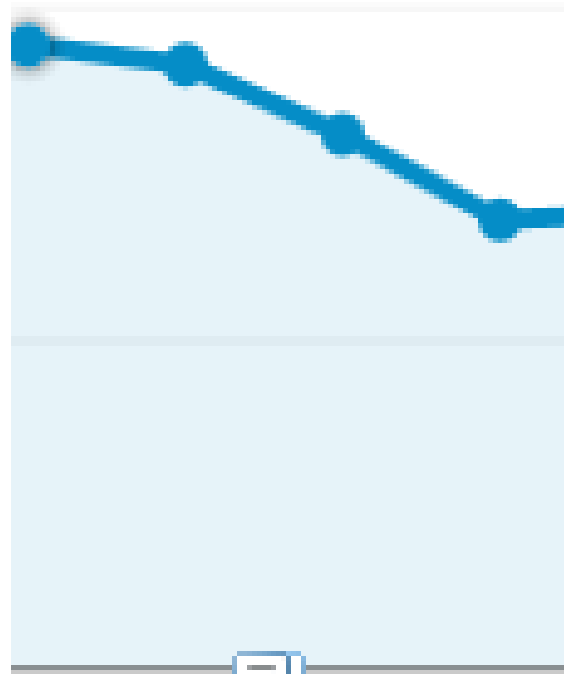


aus Kaushik (2010, S.47)

## 2. Metriken - Absprungrate



54,80%



40,78%

20. Okt.



## 2. Metriken - Sitzung

Aus der Praxis:

Sitzungen			
	Sep 13	%-Differenz	Sep 14
HTTP-Analyze	116.055	-23,89%	88.325
Piwik	29.517	15,64%	34.133

Was muss ich berücksichtigen?

- Erhebungsmethoden → Logfiles vs. JavaScript
- Was bedeutet der Sitzungsbegriff in den Tools? → 65 Minuten HTTP-Analyze vs. 30 Minuten Piwik

... hilft jedoch auch nicht bei der Erklärung der Differenz der Zahlen...

HTTP-Analyze beinhaltet Robots! → Erklärt das generell wesentlich höhere Niveau der Zahlen

HTTP-Analyze	Sep 13	Sep 14	Differenz
Sitzungen Googlebot	49.544	5.410	-44.134

### 3. Datenschutz – Eckpunkte

#### Nur einige Punkte – keine vollständige Aufstellung!

- Klären Sie den Nutzer generell auf, dass seine Daten erfasst und ausgewertet werden!
  - Hinweis auf Piwik und Google Analytics
- Die Datenschutzerklärung muss eine eigene Seite, entsprechend verlinkt und von überall im Webangebot mit einem Klick aufrufbar sein

- IP-Anonymisierung;
  - Google Analytics: Änderung Trackingcode
  - Piwik: Im Backend mit „Plugin“
- Piwik + Google Analytics: „manuelle“ Deaktivierung
  - Opt-Out-Cookies für mobile Geräte

```
var _gaq = _gaq || [];
_gaq.push(['_setAccount', 'UA-12345678-1']);
_gaq.push(['_gat._anonymizeIp']);
_gaq.push(['_setDomainName', 'www.example.com']);
_gaq.push(['_trackPageview']);
```

**Widerspruch zur Erfassung mit Piwik**

Deaktivierung durchgeführt! Ihre Besuche auf dieser Webseite werden von der Webanalyse nicht mehr erfasst.  
Bitte beachten Sie, dass auch der Piwik-Deaktivierungs-Cookie dieser Webseite gelöscht wird, wenn Sie die in Ihrem Browser abgelegten Cookies entfernen. Außerdem müssen Sie, wenn Sie einen anderen Computer oder einen anderen Webbrowser verwenden, die Deaktivierungsprozedur nochmals absolvieren.

**Ihr Besuch dieser Webseite wird aktuell von der Piwik Webanalyse nicht erfasst. Klicken Sie hier, um Ihren Besuch wieder erfassen zu lassen.**

- Lese-Tipp (Google Analytics): Stellungnahme Hamburger Datenschutzbeauftragter

## 4. Zusammenfassung

- Zahlen immer kritisch beurteilen
  - Unterschiede in der Erhebung
  - Unterschiede in den Metriken
  - Vergleichbarkeit (verschiedener Angebote) „schwierig“
- Problem der Aussagekraft: eine sinnvolle/gute Nutzung des Angebots lässt sich nicht auf einfache Seitenaufrufe reduzieren
  - Sinnvolle „Schlüsselperformanzindikatoren“ bzw. „key performance indicators“ nötig, die das Webseitenziel berücksichtigen.
- Problemstellungen bei der Messung bzw. der Vergleichbarkeit.

# Danke für Ihre Aufmerksamkeit!

- Fragen und Diskussion....

# Quellen

- Höchstötter, Nadine; Lewandowski, Dirk (2014): Websuche und Webmonitoring. In: Höchstötter, Nadine (Hrsg.): Handbuch Webmonitoring 1. Social Media und Websitemonitoring. Heidelberg: Akademische Verlagsgesellschaft AKA, S. 23 - 46
- Kaushik, Avinash (2010): Web analytics 2.0. The art of online accountability & science of customer centricity. Indianapolis: Wiley Publishing, Inc.
- Wikipedia (2014): Web analytics.  
[http://en.wikipedia.org/w/index.php?title=Web\\_analytics&oldid=607653025](http://en.wikipedia.org/w/index.php?title=Web_analytics&oldid=607653025), zuletzt geprüft am 18.05.2014.

## Links Datenschutz:

- [https://www.datenschutz-hamburg.de/uploads/media/GoogleAnalytics\\_Hinweise\\_fuer\\_Webseitenbetreiber\\_in\\_Hamburg\\_01.pdf](https://www.datenschutz-hamburg.de/uploads/media/GoogleAnalytics_Hinweise_fuer_Webseitenbetreiber_in_Hamburg_01.pdf)
- <https://www.datenschutzbeauftragter-info.de/fachbeitraege/piwik-datenschutzkonform-einsetzen/>