



Web-statistik - og hvad så?

Løvschall, Kasper

Published in:
Biblioteksårbog 2003

Publication date:
2004

Document Version
Også kaldet Forlagets PDF

[Link to publication from Aalborg University](#)

Citation for published version (APA):

Løvschall, K. (2004). Web-statistik - og hvad så? I Rasch, Anna : Schneider, Annette : Shin, Lissa : Thorhauge, Jens (red.) (red.), *Biblioteksårbog 2003* (s. 140-143). Biblioteksstyrelsen.
<http://www.bs.dk/publikationer/aarbog/2003/>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Web-statistik – og hvad så?

Statistik er mere end at tælle – tælle på frekvenser. Det er mere end matematik, da der skal anvendes en god portion sund fornuft til at udlede rigtige konklusioner, og det kræver samtidig et godt kendskab til det vi tæller på; helt ned til baggrunden for hvordan datamateriale er opstået, og med hvilket forsæt brugeren har givet os det.

Og ja, det er brugeren der leverer materialet, og det er brugeren der skal have gavn af de resultater vi udleder. Brugeren er i centrum. Det er vi forhåbentlig alle enige om.

Hvad er web-statistik

Når vores bruger med sin browser besøger en af vores websider, gemmer webserveren en masse oplysninger om besøget. Nærmere bestemt én registrering pr. element på siden – et 'hit', et ord de fleste nok nikker genkendende til i sammenhængen.

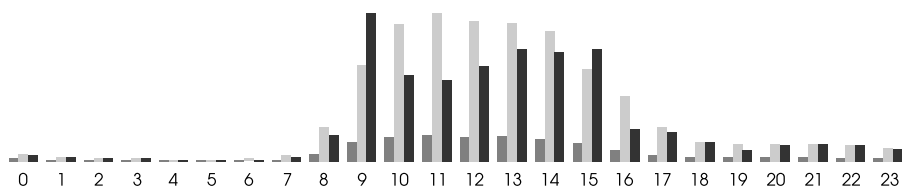
En webside består ud over selve siden ofte af mange yderligere elementer som altså hver især bidrager med et hit til det aktuelle besøg. Det vil sige at fx alle grafikelementer hver registreres af webserveren. Så der kan være rigtig mange hits pr. side.

Og hvad er web-statistik så? Jo, det er selvfølgelig en statistisk analyse af ovenstående datamateriale over tid.

Hvad kan det bruges til?

I 'gamle dage' gik vi næppe helt så meget op i de tilbud vi stillede til rådighed via vores websites. Succeskriteriet var det at vi var on-line og med på noderne. Der var statistikken et mere kuriøst bidrag; vi var imponerede over et besøg fra det store udland. Vi kunne se når folk holdt frokostpause, og edbfolkene havde et påskud til at konkludere at med den trafik, så var der da brug for en ny server.

I dag ser det anderledes ud. Vi har et behov for at kunne dokumentere benyttelsen og kvaliteten eller resultatet heraf. Vi stiller os selv spørgsmålene: Hvad benytter vores brugere, og hvad er de ude efter? Hvilken adfærd anvender de i deres søgen efter oplysninger? Hvem er brugeren, hvor ofte kommer de forbi og i hvor lang tid? Hvordan kom de hertil, og hvor forlod de os? Og det ultimative spørgsmål: Fandt min bruger det han/hun kom efter?



Timer (server tid)	Sider	Hits	Bites	Timer (server tid)	Sider	Hits	Bites
00	6631	25843	386,30 MB	12	87002	510393	6,41 GB
01	4045	13248	224,09 MB	13	90167	506133	7,42 GB
02	2808	9232	176,12 MB	14	83564	479577	7,23 GB
03	2279	6659	142,19 MB	15	60608	336388	7,51 GB
04	2225	4993	90,15 MB	16	39673	235588	2,16 GB
05	1847	4822	71,80 MB	17	23668	124182	1,92 GB
06	2499	7468	106,94 MB	18	15779	69200	1,26 GB
07	5226	22722	294,38 MB	19	13041	63435	773,13 MB
08	27018	123143	1,69 GB	20	13307	64829	1,06 GB
09	66349	352431	9,82 GB	21	13593	64857	1,19 GB
10	89157	501823	5,68 GB	22	11891	57183	1,06 GB
11	94164	540112	5,39 GB	23	10041	44725	883,68 MB

At vores brugere sover mellem midnat og kl. syv om morgenen, overrasker nok ikke de fleste, men finder de det de leder efter, når de er stået op?

I konflikt med fundamentet

Vi er desværre allerede kommet ikke så lidt på kollisionskurs med datagrundlaget. For vores behov for svar fra statistikken har udviklet sig voldsomt, mens datagrundlaget stadig er på "jeg har brug for en ny server"-stadiet.

Problemet bliver i særdeleshed tydeligt når det forsøges at sammenligne bibliotekerne med hinanden. Det vi giver søger, er ofte sammenligning på fx kvalitet, men grundlaget er kvantitativt bestemt. Lad os tage et eksempel:

To lige store biblioteker sammenligner sig med hinanden. Bibliotek A har 100 sidevisninger dagligt, mens bibliotek B har 1.000 dagligt.

Har bibliotek B mere succes og en bedre hjemmeside end bibliotek A? Måske, men det er ikke nødvendigvis sikkert. For det kunne tænkes at brugerne af bibliotek A i gennemsnit fandt det de ledte efter på de første 1-2 sider, mens brugerne af bibliotek B måske først fandt det efter 4-5 sider. Altså vi kan ikke fortælle særligt meget om kvalitet ud fra tallene endsige om brugeren var i fokus.

Ses på aktivitet i stedet for kvalitet, har vi faktisk samme problem, for aktiviteten i sig selv siger ikke specielt meget. Den må sættes i relation til noget andet. Vi kan ikke direkte sammenligne udlånet af en bog med besøget på en webside. Et udlån kræver alt andet lige en større interaktion eller forsæt fra brugeren end et klik med en mus. Ved at holde brugeren indenfor vores websider øges aktiviteten, men vi bør lede brugeren ud og videre, når brugeren har behovet. Vi skal ikke dokumentere vores aktivitet med brugeren holdt som gidsel.

Mange andre kilder til fejlforklaringer

Desværre er der mange andre faktorer som må tages i betragtning, førend man går videre med analysen af sit datamateriale.

At identificere den unikke bruger var noget af det første der rokkede ved web-statistikens troværdighed. For at gøre en længere historie kort, så betyder det at vi i bund og grund ikke ved hvem vi har besøg af. Det vi ser som en bruger, kunne være alt fra en til flere tusinde. Der findes muligheder for at minimere problemet, men det ligger ofte udenfor web-statistikens område.

Valg af program til den statistiske analyse er også afgørende. De fleste programmer er såkaldte 'black-boxes', hvor vi føder data ind i den ene ende og får resultater ud af den anden. Men hvad foregår der inde i programmet? Hvilke antagelser ligger til grund for analysen? Resultater fra forskellige programmer kan bestemt ikke sammenlignes.

Alt i alt kan det siges at web-statistikken ikke nødvendigvis er forkert, man skal blot huske på at de kvantificerbare faktorer er flertydige og kender man ikke sin metode, kan man ikke drage holdbare konklusioner.

Står ikke alene

Derfor må vi anvende web-statistik på baggrund af det kendskab vi får på anden vis. Fx gennem fokusgrubeanalyser eller holde det op imod nye, mere direkte feedback-funktioner. Det kunne være at give brugeren mulighed for at tildele siderne karakter (fra 'oplysningerne på denne side kunne ikke anvendes' til 'oplysningerne på denne side var lige det, jeg ledte efter').

Men web-statistik har bestemt også sine styrker. Bruger vi materialet til at sammenligne de enkelte biblioteker med sig selv, kan man få en masse nyttig viden ud af fx ændringer over tid. Laves der om i sidestrukturen, vil det helt sikkert kunne spores i statistikken. Det er muligt at se om brugeren faktisk kommer på de sider man ønsker dem ledt hen til.

Web-statistik kan også anvendes til isolerede opgaver. Eksempelvis download af de digitaliserede publikationer mange biblioteker har gjort tilgængelige. Men stadigvæk kræver det at man holder tungen lige i munden. Som eksempel på dette viste en gennemgang af et kvartals download af Ph.d. rapporter fra Aalborg Universitetsbiblioteks hjemmeside at ud af 25.000 registrerede hits på disse var ca. 4.100 reelle visninger.

Tag ikke livet af web-statistikken

Det kan måske komme til at lyde som om jeg ikke tillægger web-statistik særlig stor værdi, eller ønsker at det ikke bliver anvendt i forbindelse med indberetning til Biblioteksstatistikken hvor vores indberetninger kommer til at stå i forlængelse af hinanden.

Det er ikke tilfældet, men anvendelsen forudsætter at vi i fællesskab er enige om de præmisser vi stiller op herfor. Vi skal kunne forsvare validiteten af de tal vi lægger på bordet. Både over for os selv, bevillingsgiverne og vores brugere. Gøre opmærksom på hvor vi kan tillade sammenligninger, og hvor vi ikke kan. Vi skal ud over black-box begreberne og måske omdefinere web-statistik, så det ikke udelukkende baserer sig på det materiale der opsamles i forvejen. Måske skal vi berige det med nogle kvalitative indikatorer. For igen: Brugeren er i centrum.