

STUDIERETNINGSPROJEKT 2016

Navn: Tore Vendelhaven	Klasse: 3Z	Fødselsdato: 21.06.1997
---------------------------	---------------	----------------------------

1.fag og niveau: Kommunikation og IT A	vejleder: Bo Alsworth Tvede
2.fag og niveau: Matematik A	vejleder: Stig Salomonsen

Emne: Søgemaskiner
<p>Redegør kort for brugen af matricer og lineær algebra i søgemaskiner og PageRank.</p> <p>Redegør kort for forskellige typer af søgemaskiner.</p> <p>Analyser udvalgte søgemaskinernes strategi, hvor du også inddrager PageRank og lignende funktioner, og vurder om de henvender sig til forskellige målgrupper.</p> <p>Vurder hvilke fordele og ulemper de forskellige søgemaskiner har, og diskuter hvilke etiske problemstillinger, der kan opstå i brugen af dem.</p>

Opgaveformulering udleveres:	13. december 2016 klokken 14.00
Opgaven afleveres:	20. december 2016 senest kl. 14.00

Ved at aflevere denne opgave i www.netproever.dk bekræfter jeg, at opgavebesvarelsen er udarbejdet af mig. Jeg har ikke anvendt tidligere bedømt arbejde uden henvisning hertil, og opgavebesvarelsen er udfærdiget uden anvendelse af uretmæssig hjælp og uden brug af hjælpemidler, der ikke har været tilladt under prøven.

Antal tegn: 34862

STUDIERETNINGSPROJEKT 2016

Indholdsfortegnelse

1. Abstract	3
2. Indledning	3
3. Generelt om søgemaskiner	4
4. Forklaring af PageRank-princippet	5
5. Lineær algebra og Matricer	5
5.1. Matrixaddition og -multiplikation	6
6. Matricer i PageRank og Googlematricen	6
7. Søgemaskiners markedsstrategier	9
7.1. Google og generelle søgemaskiner	9
7.2. Specifikke søgemaskiner	12
8. Brugen af PageRank og alternativer	13
9. Fordele og ulemper ved søgemaskiner og PageRank	14
10. Privatliv eller personaliseret indhold	16
11. Konklusion	17
Referencer	19

STUDIERETNINGSPROJEKT 2016

1. Abstract

The most used website in the world is google.com. All major websites at the web are either a search engine or have a search engine implemented. Google have a market share at 90%, which of the searches made which is a huge plus for them, because they get a lot of data, they can adapt their search results after.

Google use an algorithm called PageRank to rank websites in their search results. PageRank is based on the hyperlink structure that the internet is constructed after. The algorithm tries to estimate a websites relevance and trustworthiness based on inbound links. The hole principal is, that the algorithm is independent on where you start measuring from.

This was their first search criteria; however, Google now uses around 200 different criteria's, including the PageRank-algorithm and the Panda-algorithm.

All the major search engines use nearly the same marketing strategies and are actually going after the same audience.

Of minor search engine, it is relevant to mention DuckDuckGo, which is an anonymous search engine, that also can be visited from the tor-network.

There also exist specified search engines with specified audience like Google Scholar and bibliotek.kk.dk.

2. Indledning

I denne afhandling vil jeg gennemgå forskellige typer af søgemaskiner, hvor jeg bl.a. vil komme ind på generelle og mere specificerede typer af søgeværktøjer.

Jeg vil se på hvad der definerer en søgemaskine, hvilke ligheder og forskelle der findes i forskellige typer af søgemaskiners strategier samt fordele og ulemper ved forskellige strategier.

Princippet og matematikken bag Googles oprindelige indekseringsalgoritme til søgeresultater, PageRank, vil blive forklaret og anvendt på et opstillet scenario.

Herudover vil de etiske problemstillinger, der er opstået i forbindelse med udbredelsen af personaliseret indhold, forårsaget af den stigende overvågning på internettet, blive diskuteret, hvorefter jeg kort vil berøre, hvordan man kan minimere risikoen for at blive kigget over skulderen.

STUDIERETNINGSPROJEKT 2016

3. Generelt om søgemaskiner

Verdens mest besøgte hjemmeside er google.com. På Alexas¹ top 10 over de mest populære hjemmesider er hele fem af hjemmesiderne helt generelle søgemaskiner, mens de resterende fem har interne søgeværktøjer indbygget.²

Søgemaskiner er et af de mest brugte værktøjer på internettet, hvilket kan ses på at næsten alle større hjemmesider har en indbygget søgefunktion. Om det er nyhedsmedier som TV2, streamingsider som Netflix, shoppingsider som Amazon eller en helt fjerde, så har de enten en intern søgefunktion eller implementeret et eksternt søgeværktøj implementeret; eksempelvis Facebooks brug af Bing som søgeværktøj.

På verdensplan findes der utallige søgemaskiner af forskellige slags til forskellige formål. Dette kan være alt fra Google med en markedsandel på ca. 90% til bibliotek.kk.dk, der er en meget specificeret søgemaskine med minimal markedsandel.³

En søgemaskine som bibliotek.kk.dk er ikke nødvendigvis interesseret i en større markedsandel, da den henvender sig til en målgruppe, der leder efter specifikke data, i dette tilfælde bøger, Google ikke nødvendigvis har indekseret.

Derimod findes der andre søgemaskiner som Bing, Yahoo!, Ask Baidu, der er generelle søgemaskiner, som gerne vil udvide deres markedsandel, og er direkte konkurrenter til Google.

Generelle søgemaskiner er modsat bibliotek.kk.dk en søgemaskine, der søger på hele internettet og prøver at give de mest relevante resultater blandt alle former for sider, troværdige som scam-sider opbygget efter utallige forskellige principper.

Ligesom bibliotek.kk.dk er en specifik søgemaskine, findes der også andre typer af specifikke søgemaskiner som Google Scholar og Infomedia, der søger blandt videnskabelig litteratur eller WolframAlpha, der minder om et leksikon snarere end en traditionel søgemaskine.

Blandt de generelle søgemaskiner er det værd at nævne DuckDuckGo, der bryster sig med, at de ikke gemmer informationer om brugeren. Her har debatten om masseovervågning, personaliserede søgeresultater og reklamer mm. bidraget til at dette marked er vokset.⁴

En optimal generel søgemaskine har al information indekseret og har samtidig faciliteter til at ordne resultaterne efter relevans og troværdighed.

Google bruger et princip til denne vurdering kaldet PageRank, der vurderer en hjemmeside efter hvilke sider der linker hertil. Selvom PageRank kun er ét af 200 kriterier Google bruger til vurdering af sider, er det en af de mest kendte metoder, som bruges af andre søgemaskiner.⁵

¹ Alexa er et rangeringssystem til vurdering af en sides popularitet ud fra faktorer som daglige besøg, samt besøg inden for sidste måned. Alexa er ejet af Amazon.com.

² (Alexa Internet, 2016)

³ (Statista, 2016)

⁴ (Hern, 2014)

STUDIERETNINGSPROJEKT 2016

4. Forklaring af PageRank-princippet

For at forstå PageRank vælger vi at beskrive internetsiderne med en kvadratisk matrix, hvor det (i,j) 'te koordinat beskriver sandsynligheden for at gå fra hjemmeside j til hjemmeside i .

Vi definerer for side P en samlet værdi for relevans og troværdighed mellem 0-1 kaldet sidens PageRank $PR(P)$.

Når forfatteren af hjemmeside P_j vælger at linke til hjemmeside P_i , må man forvente at side P_i er relevant ifølge forfatteren af side P_j 's optik. Her overfører side P_j en del af sin troværdighed til side P_i lig $1/l_j$, hvor l_j er antallet af udgående links på side P_j .

Her er der dog ikke taget højde for antallet af indgående links til side P_i , da andre sider end P_j også kan linke til P_i . Mængden af sider med indgående links til P_i betegnes B_i .⁶

En side med tre udgående links vil altså sende 1/3-del af sidens egen PageRank videre til de tre sider.

Hver side med link til P_j overfører sin egen PageRank divideret med antallet af udgående links på den pågældende side. Det kan samlet skrives som summen:⁷

$$PR(P_i) = \sum_{P_j \in B_i} \frac{PR(P_j)}{l_j} \quad (1)$$

Denne beregning har dog ét væsentligt problem; sidernes PageRank afhænger af andre sideres PageRank, så hvis to sider linker til hinanden kan det have betydning, hvor man beregner PageRanken fra. Det er dog ikke umuligt at løse, man skal blot se på ovenstående som en matrix.

5. Lineær algebra og Matricer

En matrix er kort sagt et skema af tal. Matrixregning er en central del af lineær algebra. En matrix defineres ofte som en $m \times n$ -matrix, hvor m er antallet af rækker og n antallet af søjler.

Matricen $A^{m \times n}$ er defineret

$$A^{m \times n} = \begin{bmatrix} v_{1,1} & \cdots & v_{1,n} \\ \vdots & \ddots & \vdots \\ v_{m,1} & \cdots & v_{m,n} \end{bmatrix}$$

I sammenhæng med PageRank er det hovedsageligt matrixaddition og -multiplikation, der er relevant da Googlematricen tager brug heraf.

⁵ (Blaavand, 2010)

⁶ (Blaavand, 2010)

⁷ (Blaavand, 2010)

STUDIERETNINGSPROJEKT 2016

5.1. Matrixaddition og -multiplikation

Addition af to matricer $A^{m \times n}$ og $B^{m \times n}$ foregår ved at hvert koordinat lægges til det pågældende i koordinat i den anden matrix. Dette er kun muligt, fordi matricerne har samme dimensioner.

$$A^{m \times n} + B^{m \times n} = \begin{bmatrix} A_{1,1} + B_{1,1} & \cdots & A_{1,n} + B_{1,n} \\ \vdots & \ddots & \vdots \\ A_{m,1} + B_{m,1} & \cdots & A_{m,n} + B_{m,n} \end{bmatrix}$$

Inden for matrixmultiplikation snakkes der om matrix-vektorproduktet og matrix-matrixproduktet.

Matrixmultiplikation kan kun ske, hvis antallet af søjler i matrix A er lig med antal rækker i matrix B.

Lad os omtale matricerne $A^{m \times n}$ og $B^{n \times p}$.

Her vil man kunne danne matrix-matrixproduktet AB men ikke BA, da $m \neq p$. Matrixmultiplikation er altså ikke kommutativt.

Derfor er det også kun muligt at sætte en kvadratisk matrix i potens.

Matrix-vektorproduktet for matrix A og vektor \vec{x} kan beregnes ved:⁸

$$A \cdot \vec{x} = x_1 \cdot \vec{v}_1 + x_2 \cdot \vec{v}_2 + \cdots + x_m \cdot \vec{v}_m$$

Når matrix A og vektor \vec{x} er defineret ved:

$$A = [\vec{v}_1, \vec{v}_2, \dots, \vec{v}_m] \quad \vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$$

Matrix-matrixproduktet for matricerne $A^{m \times n}$ og $B^{n \times p}$ kan beregnes ved:⁹

$$A \cdot B = A \cdot [\vec{z}_1, \vec{z}_2, \dots, \vec{z}_p] = [A\vec{z}_1, A\vec{z}_2, \dots, A\vec{z}_p] = [A_{i1}B_{1j} + A_{i2}B_{2j} + \cdots + A_{in}B_{nj}]$$

Når matricerne A og B er defineret ved:

$$A = [\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n] \quad \text{og} \quad B = [\vec{z}_1, \vec{z}_2, \dots, \vec{z}_p]$$

En god huskeregel er at hvis du opskriver dine to matricer ved siden af hinanden, eksempelvis 3×4 , 4×5 går den op så længe de to understregede tal er de samme $3 \times \underline{4}$, $\underline{4} \times 5$.

6. Matricer i PageRank og Googlematricen

For at løse problemstillingen fra afsnit 4, vil vi indføre matrixregning til beregning af PageRank-vektoren.

Denne matrix vil altid være kvadratisk, da hver side på internettet er repræsenteret både som søjle og række, så $m=n$.

Hvis en side eksempelvis er repræsenteret af række i vil siden også være repræsenteret af søjle i ; denne matrix kaldes *hyperlinkmatricen*, H_{ij} .

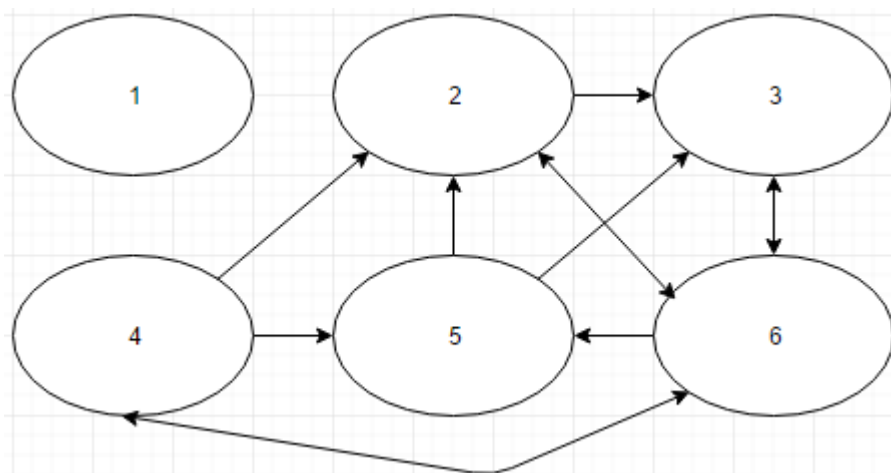
⁸ (Undervisningsministeriet, 2014)

⁹ (Undervisningsministeriet, 2014)

STUDIERETNINGSPROJEKT 2016

For hyperlinkmatricen gælder det, at alle værdier er lig nul eller større, samt at hver søjles sum er 1, med mindre siden ikke har nogen udgående links; dertil er summen 0.

Dette kan godt virke uoverskueligt, derfor vil jeg vise det med et lille lukket system af hjemmesider, for at give en ide om hvordan det fungerer.



Figur 1: Et isoleret netværk af hjemmesider, bestående af siderne 1, 2, 3, 4, 5 og 6.

side 1 har hverken indgående eller udgående links og PageRanken vil altså være 1, hvis du starter på side 1 og 0 hvis du starter på en af de andre sider, hvis vi bruger formel (1).

Med udgangspunkt i side 3 ville beregningen af denne sides PageRank med (1) se således ud:

$$\sum_{P_j \in B_3} \frac{PR(P_j)}{l_j} = \frac{PR(2)}{2} + \frac{PR(5)}{2} + \frac{PR(6)}{4}$$

Beregning 1: (1) på et lukket kredsløb. $B_3 = \{P_2, P_5, P_6\}$, dvs. $P_2, P_5, P_6 \in B_3$

Problemet herved er at vores PageRank er afhængig af startside, og dermed er sandsynligheden for hver side defineret af I^0 .

Det skyldes at der ikke er taget højde for at brugeren kan vælge at tilgå en tilfældig side på nettet frem for kun at følge links.

Google har approksimeret sandsynligheden for at en bruger vælger en tilfældig side på internettet frem for at følge links til 15%. Sandsynligheden for man derimod følger et link på siden er altså ifølge Google 85%. Denne værdi omtales alpha (α) og er en approksimeret konstant fra Google.¹⁰

Foruden α opstilles en matrix A, der består af 0'er, bortset fra i de søjler, der repræsenterer sider uden udgående links, hvori der skrives $1/n$, hvor n er det samlede antal sider på nettet. Denne matrix lægges til hyperlinkmatricen for at skabe matrix S.

¹⁰ (Blaavand, 2010)

STUDIERETNINGSPROJEKT 2016

$$S = H + A \quad (2)$$

Nu har vi fastlagt en værdi for sandsynligheden for, at en bruger følger linkstrukturen frem for at tilgå en tilfældig hjemmeside.

Det er nu Googlematricen bliver interessant. Googlematricen er en formel, der er defineret af Google og er den hidtil mest præcise simulering af en brugers adfærd på internettet.¹¹

$$G = \alpha S + (1 - \alpha) \cdot \frac{1}{n} \mathbf{1} \quad (3)$$

Men for at I^k er uafhængig af I^0 kræver det at nedenstående gælder.

$$GI^k = G^k I^0 \quad (4)$$

Det kan være svært at tyde, men 1-tallet sidst i (4) symboliserer en matrix med samme dimensioner som H, men bestående af lutter 1-taller.

For at opfylde både (3) og (4) sættes beregningen op:

$$GI^k = \left(\alpha S + (1 - \alpha) \cdot \frac{1}{n} \mathbf{1} \right)^k * I^0 \quad (5)$$

Den approksimerede PageRank vil altid konvergere når k går mod uendelig. Vores endelige resultat for figur 1 vil være af dette format:

$$G^{6 \times 1} = \begin{bmatrix} PR(1) \\ PR(2) \\ \vdots \\ PR(6) \end{bmatrix} \quad (6)$$

Den beregnede PageRank-vektor til I^{50} er derfor

$$G^{6 \times 1} = \begin{bmatrix} 0,0291262135942341 \\ 0,181870097112797 \\ 0,230987247459736 \\ 0,0994504973982950 \\ 0,127628138324613 \\ 0,330937806135394 \end{bmatrix}$$

Resultat 1: PageRank-vektor (G) for figur 1 ved I^{50} . Jf. (6) for forståelse heraf.

og konvergerer jo større k bliver.

¹¹ (Blaavand, 2010)

STUDIERETNINGSPROJEKT 2016

Her kan det ses at sandsynligheden for jeg ved et tilfældigt besøg på netværket fra figur 1 vil besøge side 6 er 33% mens sandsynligheden for jeg vil se side 1 er 2,9%.

Resultat 1 er fundet ved nedenstående beregning:

$$G := \left(0.85 \cdot \begin{pmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{4} \\ 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{4} \\ 0 & \frac{1}{2} & 1 & \frac{1}{3} & 0 & 0 \end{bmatrix} + \begin{bmatrix} \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \frac{0.15}{6} \cdot \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \right) \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}^{50}$$

Beregning 2: Beregning af PageRank af figur 1.

Den fulde beregning af PageRank-vektoren til figur 1 kan ses i bilag 1.

Det skal hertil bemærkes, at vektoren yderst til højre repræsenterer I^0 , og så længe denne summerer til 1, er resten af beregning 2 uafhængig heraf.

7. Søgmaskiners markedsstrategier

Da der findes mange forskellige søgemaskiner, findes der også mange forskellige strategier for optimal udformning og funktion. Her har Google lagt baghjul til alle konkurrenter inden for det man kan kalde generelle søgemaskiner.

7.1. Google og generelle søgemaskiner

En af Googles strategier har fra starten været at vise et ry minimalistisk design, hvor brugeren ikke behøver et forkendskab til værktøjet for at bruge det.

Når man gør designet simpelt, risikerer man også at miste funktioner. Her har Google været gode til at skabe balance, og har bevist at designet er ikke det eneste, der har relevans for hvor populær og succesfuld en søgemaskine bliver.

Google har altid været gode til at sortere i deres resultater og vise de vigtigste resultater øverst. Dette har de netop formået at være førende i, bl.a. ved udvikling af en algoritme til sortering af resultaterne efter omkring 200 parametre, herunder PageRank.¹²

¹² (Blaavand, 2010)

STUDIERETNINGSPROJEKT 2016

Derudover har Google deres ikoniske "Prøv Lykken"-knap, der oprindeligt valgte det øverste resultat på listen. Efterhånden som siden er blevet mere interaktiv, er det faktisk ikke muligt at klikke herpå længere, da man får vist søgeresultater så snart man indtaster noget i søgefeltet.

Google har ligesom mange andre store søgemaskiner en lukket kildekode, da dette er en måde hvorpå man kan forhindre andre søgemaskiner i at kopiere deres metoder til sortering af resultater.

Dette er en af årsagerne til at Google fortsætter med at have førertrøjen på inden for søgninger.

Udover selve algoritmen handler det også om hvilke data man har tilgængelige.

Når Google sidder med ca. 90% af markedsandelen mens eksempelvis Bing har ca. 4%, har Google altså over 20 gange så meget data om brugeres adfærd på internettet at tage udgangspunkt i, når de skal vurdere siders placering i resultaterne.¹³

Da smartphones kom frem åbnede det et nyt marked for søgemaskiner, samt muligheder for indsamling af en ny form for data. Her har både Google og Microsoft mf. prøvet at vinde markedsandele, men man kan konstatere at Google er løbet med sejren. Dette skyldes bl.a. Googles store succes med styresystemet Android, der netop havde Google som standard-søgeenhed.¹⁴

Når man kigger på de forskellige markeder, hvor Google enten allerede eksisterer eller prøver at komme ind, kan man se at det er informationsrige markeder, og langt de fleste har relevante data i forhold til vurderingen af siders placering i deres søgeresultater.

Der er eksempelvis Youtube, der er ejet af Google, hvor de har mulighed for at indsamle data om hvilke videoer man ser, liker, disliker mm. som de så kan bruge til tilpasning af søgeresultater i deres søgemaskine.

Men Youtube er ikke det mest dybdeborende værktøj Google har til rådighed. Her hentyder jeg naturligvis til Google Analytics, der er et værktøj udviklet til hjemmesideadministratorer, der giver overblik over hvordan brugere opfører sig på ens side.

Men når man inviterer et værktøj som Google Analytics ind på sin side til at monitorere besøgende for en, får Google som tredjepart også en del informationer om dine besøgende, hvad de klikker på, hvor længe de bliver forskellige steder mm. Dette kan de sammenligne med data fra andre kilder og derved få et billede af individuelle brugeres færd på internettet. Dette bakkes yderligere op af at Google Analytics har en markedsandel på omkring 54,8%.¹⁵

Google er altså udformet for at være så nem og ligetil som muligt, hvilket også afspejles i deres målgruppe, der mere eller mindre er alle, som ikke nødvendigvis ser et problem i den store mængde information Google indsamler om brugerne.

En anden stor søgemaskine er Bing, som er ejet af Microsoft. Den bruger en del af de samme strategier som Google, bl.a. ved at det er standard-søgemaskinen i Windows og Windows Phone, men herudover bruger Facebook Bing som kort-tjeneste, hvilket er en stor fordel, da det er et lukket marked, hvor Bing er eneste udbyder.

¹³ (Blaavand, 2010)

¹⁴ (Statista, 2016)

¹⁵ (W3Techs.com, 2016)

STUDIERETNINGSPROJEKT 2016

En anden markedsstrategi er product placement. Eksempelvis siger en person 10 minutter og 43 sekunder inde i TV-serien *Hawaii Five-0, sæson 1, episode 6*: "I binged it, dude."

Dette er ikke nødvendigvis product placement, men ligner utrolig meget, da det ellers ville være en ret underlig replik, eftersom Bings markedsandel er på omkring 4%.

Men i Kina er det et helt andet billede man får frem når man sammenligner markedsandele. Her har Google kun en markedsandel på omkring 9%, mens den kinesiske søgegigant Baidu sidder på omkring 80% af markedet.¹⁶

Det er dog stadig imponerende at Google har en så stor markedsandel, da Google har været blokeret i Kina siden 2010, efter Google stoppede med at overholde de kinesiske regler for censur.¹⁷

Googles manglende tilstedeværelse på det kinesiske marked, har gjort Kina mere attraktivt for andre internationale søgemaskiner. Blandt andet er Yahoo! en søgemaskine som fungerer i fin fortræffelighed i Kina. Yahoo! har dog også et samarbejde med den kinesiske regering om sporing af systemkritikere.¹⁸

En helt anden type marketingstrategi kan findes hos Ask, der anderledes end de andre store søgeværktøjer bruger mere lyssky metoder end eksempelvis Google.

Deres strategi er at mange mindre-kendte programmer automatisk installerer Ask på din enhed, når du installerer dem. For selv om man i nogle af disse programmer kan fravælge at installere søgemaskinen, er det oftest pakket langt væk i installationen. Denne form for strategi, hvor man presser sin søgemaskine ind hos brugere ubemærket kan give et anspændt forhold til brugeren, men samtidig skabe nye brugere.

Mange af de store søgemaskiner som Google, Yahoo!, Ask, Baidu eller Bing har alle sammen helt generelle målgrupper, der ikke nødvendigvis er afhængige af køn, alder, politisk eller religiøs overbevisning, interesser o.l.

Baidu er dog den eneste af de ovenstående søgemaskiner, der har en geografisk forankret målgruppe, da den er henvendt udelukkende til kinesisk-talende brugere, modsat de andre, der findes på flere sprog.

Der findes dog også andre typer af generelle søgemaskiner, der ikke konkurrerer, eller nødvendigvis har interesse heri, med store søgegiganter som Google og Yahoo!.

Der er eksempelvis DuckDuckGo, som er en generel søgemaskine, men modsat de større søgemaskiner, der indsamler data om deres brugere til forbedring af søgeresultater, reklamerer DuckDuckGo med ikke at gemme nogen personlige data.¹⁹

Samtidig er størstedelen af DuckDuckGos kildekode open source og en del af videreudviklingen heraf sker på frivilligbasis, ved at opgaver bliver offentligt tilgængelige og man har mulighed for anonymt såvel som navngivet at komme med løsninger hertil. Dette har DuckDuckGo til dels succes med, fordi de gerne vil være den bedste søgemaskine for programmører.²⁰

DuckDuckGo er en af de foretrukne søgemaskiner i linux-baserede styresystemer, hvilket også hænger godt sammen med at den er open source. Den henvender sig mere til programmører og brugere der vil undgå at

¹⁶ (Incitez China, 2016)

¹⁷ (Waddell, 2016)

¹⁸ (Klein, 2009)

¹⁹ (DuckDuckGo, 2016)

²⁰ (DuckDuckGo, 2016)

STUDIERETNINGSPROJEKT 2016

blive set over skulderen end almindelige brugere på internettet. Dertil har DuckDuckGo også et *mirror*²¹ på Tor-netværket²² til brugere heraf.

Det er dog stadig relevant at nævne, at DuckDuckGo er baseret på Yahoo!'s søgeresultater og det derfor ikke kun er DuckDuckGo, der definerer sortering af resultaterne.²³

Hvis man skal definere en målgruppe for en søgemaskine som DuckDuckGo vil det være programmører eller andre teknisk interesserede, samt brugere der vil undgå at blive sporet.

7.2. Specifikke søgemaskiner

Udover store søgegiganter som Google eller DuckDuckGo med specificerede målgrupper, findes der også søgemaskiner, der giver specifikke former for data.

Søgemaskiner kan være specificerede inden for utallige områder.

Eksempelvis har Google udviklet et søgeværktøj inden for videnskabelig litteratur, der hedder Google Scholar. Dette værktøj giver muligheden for at søge efter forfattere, citater, afhandlinger og andet litteratur, der er videnskabeligt eller fagligt forankret. Denne type søgning giver mulighed for at finde forskellige vinkler og holdninger til emner, der ikke nødvendigvis er nemt tilgængelige.

Formålet med Google Scholar er, som de selv formulerer det, at tilbyde et værktøj, der giver overblik over faglig litteratur fra hele verdenen på et sted, nemlig Google Scholar.

Dertil er deres slogan "stå på skuldrene af giganter", hvilket hentyder til at man ikke nødvendigvis behøver vide alt, men ved hjælp af deres værktøj man kan få overblik over den relevante litteratur andre har skrevet, og derved kan man skabe sig et nuanceret billede af det pågældende område.²⁴

Derudover giver Google Scholar forfattere større mulighed for at se hvor, hvornår og hvem der citerer dem eller deres værker.

Google Scholars målgruppe er altså meget specifik i forhold til Googles generelle målgruppe, da Google Scholar er henvendt til forfattere af faglitteratur, forskere, eksperter og studerende, samt andre der har brug for litteratur der er mere videnskabeligt end hvad der nødvendigvis dukker op på Googles generelle søgemaskine.

WolframAlpha er en ret speciel søgemaskine, både i forhold til generelle søgemaskiner, men også i forhold til Google Scholar, da den fungerer fundamentalt anderledes end andre.

Når du søger på WolframAlpha får du ikke en række links til relevante sider, som på klassiske søgemaskiner. Derimod får du en masse informationer om det du søger på. WolframAlpha har dog visse begrænsninger: Hvis jeg eksempelvis hvis jeg søger på "yahoo china censorship" får jeg en masse informationer om domænet yahoo.com.cn, men ikke noget omkring censur.

²¹ En spejling af det oprindelige indhold på en side overført til en anden.

²² TOR-netværket er et anonymt overlay-netværk på internettet tilgængeligt for alle gennem værktøjer som TOR-browser, Freenet og lignende.

²³ (DuckDuckGo, 2016)

²⁴ (Google, 2016)

STUDIERETNINGSPROJEKT 2016

Her får man et langt mere brugbart resultat ud fra eksempelvis Bing, hvor man får en række links til hovedsageligt Wikipedia, der diskuterer dette emne eller Google Scholar, der giver en liste over *.edu-domæner der omtaler emnet.

Men hvis man foretager mere specifikke søgninger på WolframAlpha, får man en type resultater, der langt overgår både Google eller Bings linkstruktur.

Hvis man eksempelvis foretager en søgning på ordet "dog" bliver man spurgt om at specificere racen. Derudover får man en liste over informationer om selve ordet. Dette er alt fra det videnskabelige navn til gennemsnitsdiameteren på et hundeøje.²⁵

WolframAlpha fungerer i praksis rigtig meget som et leksikon, hvor målgruppen er brugere, der skal have informationer om helt specifikke områder.

8. Brugen af PageRank og alternativer

Langt de fleste søgemaskiner har lukkede kildekoder. Derfor ved man meget lidt om hvad der ligger til grund for resultaternes placeringer i søgealgoritmen. Nogle af de offentligt kendte algoritmer er PageRank-princippet, der bliver brugt af Google, og sandsynligvis også andre søgemaskiner.

Herudover kan nævnes HITS-algoritmen, der bygger på hubs og autoriteter. Mere specifikt er en god hub en linkdatabase med mange udgående links, mens en god autoritet er en side med mange indgående links fra hubs.²⁶

HITS-algoritmen er et lignende redskab til PageRank og bruges af eksempelvis Asks søgemaskine.²⁷

Herudover kan nævnes algoritmer som Edgerank som Facebook brugte til organisering af ens nyheds-feed indtil 2010, hvor de begyndte at bruge en algoritme baseret på maskinlæring.²⁸

Oprindeligt byggede Google udelukkende på PageRank-princippet, men efterhånden som markedet udviklede sig, udviklede Googles søgealgoritme sig også.

PageRank er i dag blot én af omkring 200 faktorer, som Google rangerer deres søgeresultater efter. En anden faktor hos Google er eksempelvis Panda-algoritmen, der er udviklet med det formål at vurdere hvor dårlig en sides kvalitet er og derigennem frasortere lavkvalitets-indhold, der formår at komme ind på Googles topresultater. Langt størstedelen af Googles kriterier er markedshemmeligheder, men nogle af disse har fagpersoner inden for SEO-industrien (Search Engine Optimization) prøvet at approksimere sig til ud fra empiriske data.^{29,30}

Beregning af PageRank-vektoren, jf. afsnit 4, for hele internettet, tager ifølge Google flere dage, og har den ulempe, at den er forældet sekundet efter beregningen er begyndt, fordi internettet er interaktivt og

²⁵ (Wolfram Language, 2016)

²⁶ (Cambridge University Press, 2009)

²⁷ (Cornell University, 2016)

²⁸ (McGee, 2013)

²⁹ (MOZ, 2016)

³⁰ (Search Engine Land, 2016)

STUDIERETNINGSPROJEKT 2016

linkstrukturen flydende. Googles præcise PageRank er ikke offentlig. Derfor beregner Google også kun PageRank en gang om måneden.³¹

Generelt vides der dog ikke særlig meget om algoritmerne, der ligger bag Google, Ask og andre søgemaskiner. Derfor kan det også være svært at se på, hvorvidt algoritmerne henvender sig til bestemte målgrupper, da den tilgængelige kildekode er meget begrænset.

9. Fordele og ulemper ved søgemaskiner og PageRank

I dette afsnit har jeg valgt at tage udgangspunkt i Google, da langt størstedelen af deres fordele og ulemper gælder for langt de fleste søgemaskiner.

Når man kigger på Googles opbygning er en af deres store fordele, at den er simpel og ligetil at bruge. Dertil kommer at resultaterne er af høj kvalitet.

Google giver mulighed for at specificere sin søgning, ved hjælp af deres avancerede søgeværktøj, men derudover kan man også gøre dette direkte i søgefeltet, hvis man da ved hvordan.

Eksempelvis kan man ved at skrive "mobning site:dr.dk" søge på alle sider, Google har indekseret, på domænet dr.dk, hvor ordet mobning står. Herudover kan man også søge på præcise sætninger hvis man indtaster sin søgning i anførselstegn.

Funktionen hvor man kan søge inden for bestemte domæner er dog blevet kritiseret for at fjerne hjemmesideadministratorens mulighed, for selv at vælge hvilke sider, der skal være tilgængelige. Dertil kan det udnyttes til at finde sikkerhedsbrister.

Eksempelvis udkom der en sikkerhedsopdatering til CMS-systemet Joomla, som resulterede i at mange sider store blev hacket, da udgivelsen af denne opdatering skete på et tidspunkt hvor få hjemmesideadministratorer var på arbejde. Her kan Googles værktøj eksempelvis bruges til at søge efter hjemmesider med forældede versioner af CMS-systemer, hvor der er kendte sikkerhedshuller.³²

Samtidig med at disse funktioner er praktiske, fordi de giver brugere mulighed for mere specificerede søgninger, kan de også være problematiske. For selv om det ikke er den eneste måde, hvorpå man kan søge specifikt blandt domæner, er den nemt tilgængeligt og fjerner meget af hjemmesideadministratorens kontrol af brugerens adfærd.

Googles store markedsandel bidrager bl.a. til at mange hjemmesider bliver udviklet specielt til Google, for at få en høj rangering, hvilket er en ulempe for deres konkurrenter, da dette både gør Googles arbejde nemmere, samtidig med konkurrenterne får sværere ved at vinde markedsandele, når sider udvikles efter SEO-principper, der oftest er baseret på Google.

En anden stor fordel for Google er at de allerede har et stort eksisterende marked, hvor de indsamler data. Dette gør deres mulighed for at forbedre søgemaskinen langt større end andre, fordi de er den

³¹ (Blaavand, 2010)

³² (Stensdal, 2015)

STUDIERETNINGSPROJEKT 2016

søgemaskine, der ved mest om brugeres adfærd på internettet. Dette kan de yderligere kombinere med de data, de indsamler fra deres styresystem Android.

Den indsamlede data fra Android indeholder mange andre typer af informationer end deres Google-søgninger. Eksempelvis bruges telefonernes konstante forbindelse til telefonmaster, til at forudsige trafikpropper o.l. som Google derefter kan bruge til at give de bedste ruteforslag på deres kort-tjeneste.

Med den store mængde information Google har om dig gennem dine søgninger, besøg på sider med Google Analytics og, hvis du har en Android-telefon, bevægelser, kan Google lave personaliserede søgeresultater, for at gøre det nemmere at finde frem til det man leder efter. Derudover giver det Google mulighed for at tilpasse reklamer efter, hvad de forventer du er interesseret i og mangler.

Selve den måde Googles PageRank-algoritme er opbygget på, giver mulighed for at koncentrere din PageRank på en bestemt side.

Eksempelvis hvis man på alle undersider på ens hjemmeside har et backlink til forsiden overfører man jo noget af PageRanken fra alle ens undersider til forsiden.

En anden metode, som nogle sider bruger til at hæve deres PageRank er ved at give et produkt væk. Dette kan eksempelvis være CMS-systemet Joomla, da der i bunden af sider, som bruger Joomla er et backlink til joomla.org. Således får joomla.org overført PageRank fra alle sider, der bruger deres CMS-system.

Når man kigger på de andre store søgemaskiner som Bing, Ask og Yahoo! kan man da også se, at de bruger samme strategier som Google ved at indsamle så meget data som muligt. Men modsat andre markeder, har det digitale det med at skabe monopoler. Google har ikke monopol på søgninger, men når så stor en markedsandel som de har, findes der ikke nogen reelle konkurrenter.

Prøv at se på brugen af styresystemer. Størstedelen af alle verdens computere bruger Windows og der ikke findes et reelt alternativ hertil, der er lige så allround. Mac er hovedsageligt henvendt til den kreative sektor, mens Linux bruges til servere og af it-folk.

Monopolerne har en tendens til at blive selvforstærkende, da nye it-produkter typisk udvikles til den mest udbredte platform. Når man udvikler et nyt produkt, f.eks. software eller en hjemmeside, er det utrolig krævende hvis det skal tilpasses alle platforme, og derfor udvikler man det oftest til den platform med størst markedsandel.

Mange computerspil udkommer til Windows men ikke nødvendigvis til Mac og Linux. Det samme sker med hjemmesider, hvor størstedelen optimeres til Google frem for Bing eller Yahoo!, da Google har den største markedsandel.

Markedet lider altså under at når én søgemaskine dominerer, har andre svært ved at finde fodfæste samt udvide deres territorie.

Der er dog plads til niche-søgemaskiner som WolframAlpha, da denne er fundamentalt anderledes. Den fungerer mere som et leksikon og hvis ens søgning ikke er præcis, får man sandsynligvis ikke svar på hvad man gerne ville.

Dertil kommer, at når de store søgegiganter baserer mere og mere af deres indhold på personaliserede resultater, udvider det markedet for søgemaskiner som DuckDuckGo og andre anonyme søgeværktøjer.

STUDIERETNINGSPROJEKT 2016

10. Privatliv eller personaliseret indhold

Uanset, om man tager udgangspunkt i Google, Ask, Yahoo! eller Bing, så bruger de alle personaliserede data til både reklamer og indhold i deres søgemaskiner. Det er dog forskelligt hvorvidt denne indblanding i søgeresultaterne kan kontrolleres og graden heraf.

Nogle ser det som et kæmpe problem at søgeresultater bliver baseret på ens tidligere færden på internettet, fordi det fjerner det nuancerede billede, da man ikke får et reelt billede af det man leder efter. Hvis man altid får et persontilpasset svar på en given problemstilling, kan man risikere at danne en holdning man ikke ville have haft, hvis man havde fået bredere svar i sine søgninger.

Her kan man frygte, at søgemaskinernes intension om at tilpasse søgesvar til brugernes præferencer, kan bidrage til øget radikaliserings mellem klasser i samfundet. Hvis man kun får søgeresultater, der underbygger ens politiske holdning, får man ikke muligheden for at sætte sig i modpartens situation.

Dette kan eksempelvis være, hvis man synes Danmark skal trække sig ud af Afghanistan-krigen og dette fremgår tydeligt af ens internet-færden, vil ens søgeresultater hovedsageligt have samme vinkel som en selv. Man hører altså ikke nødvendigvis om modpartens argumenter og kan derfor se sig sur på herpå, fordi man ikke har det fulde overblik over situationen. Når man først ser sig sur på en anden part, kan det bidrage til flere fordomme herom. Det er altså vigtigt at blive præsenteret for alle siderne af en aktuell situation for man kan danne sig et reflekteret overblik og undgå at bidrage til radikalismen.

Når Google har så stor en markedsandel, som de har har de også mulighed for at påvirke brugeres færden på internettet og forhindre informationer de ikke selv kan lide, i at nå ud til almindelige brugere. Hvis Google vil, kan de fjerne vinkler på problemstillinger de er uenige i fra deres søgeresultater hvilket sandsynligvis aldrig vil blive opdaget, men stadig bidrage til mindre refleksivitet hos den almene befolkning.

Det er altså et etisk dilemma når en privat virksomhed har så stor mulighed for direkte at påvirke ens hverdag.

Samtidig har virksomheder som Google, så meget information om én, at de har muligheden for at forudsige hvad man mangler.

Et eksempel er en sag, hvor Google gav reklamer om graviditetsprodukter til en kvinde, hvilket hendes far bemærkede, så kvinden måtte fortælle at hun var gravid. Pga. hendes søgeadfærd vidste Google således mere om hende, end hendes far.

Her kan man se, hvordan disse reklamer har påvirket en familie og rent faktisk fået en kvinde til at fortælle sin far om noget, hun måske ikke synes, at han skulle vide.

Den etiske konflikt heri ligger i, hvem der ejer ens data. Er det virksomheden der indsamler dem, eller dig, eller regeringen eller har digitale fodspor ingen ejer?

Et resultat af denne uklarhed om ejerskab til data har eksempelvis ført til at Yahoo! har tilladt USA's regering at installere sporingsværktøjer, der registrerer alle indkommende mails.³³

Men det behøver ikke være et etisk problem at skabe personaliserede reklamer og indhold, hvis det gøres med måde. Hvis man eksempelvis leder efter et klædeskab på diverse shoppingsider, kan Google hurtigt

³³ (Startpage Blog, 2016)

STUDIERETNINGSPROJEKT 2016

give dig relevante reklamer om lignende klædeskabe på deres egen søgemaskine eller de sider, der bruger annoncer fra Google.

Herudover kan informationer om hvor brugernes mobiler er, hjælpe politiet med at finde frem til kriminelle.

Om dette er en fordel eller en ulempe, er meget omdiskuteret.

Der findes utallige værktøjer til at undgå denne form for overvågning. Et af dem er DuckDuckGo, som ikke sporer brugernes færden. Herudover kan man yderligere bruge værktøjer som proxy-servere, virtuelle private servere og virtuelle private netværk samt tor-netværket.

Hvis jeg selv ville være anonym på internettet, ville jeg opsætte min internetforbindelse til at blive videresendt gennem først en VPS, derefter en VPN, efterfulgt af en SOCKS5-proxy for til sidst at tilgå TOR-netværket gennem TOR-browseren.



Figur 2: En forholdsvis simpel og sikker måde at tilgå TOR-netværket.

Det etiske dilemma handler om hvem, der har adgang til hvilke data om os og hvad disse data bliver brugt til samt hvem disse data tilhører.

11. Konklusion

Google har en markedsandel på ca. 90% af søgemaskinemarkedet. Denne store markedsandel bidrager til at de bevarer pladsen som den dominerende søgemaskine, sammen med deres styresystem Android, samt Youtube og Google Analytics, der alt sammen bidrager til deres database, søgemaskinen bruger til at vurdere, hvilke data, vi hver især får vist i søgeresultater og reklamer.

Googles målgruppe vil kunne defineres som alle, der ikke har et problem med den øgede overvågning på internettet.

Bing har ligesom Yahoo! og Ask en lille markedsandel, men bruger mange af de samme strategier som Google og har samme målgruppe.

Bing er brugt af Facebook som kort-tjeneste, hvilket er en stor fordel for dem, da det er et lukket marked, de har monopol på.

Disse store søgemaskiner bruger dog alle personaliserede reklamer og søgeresultater, hvilket kan skabe det etiske dilemma om hvem ens digitale fodspor tilhører samt hvilke data, det kan tillades at indsamle om brugerne.

Udover store søgegiganter som ovennævnte findes der også mindre generelle søgemaskiner som DuckDuckGo, der ikke bruger personaliseret indhold og ikke sporer en. DuckDuckGo kan derudover tilgås via TOR-netværket for øget anonymitet.

STUDIERETNINGSPROJEKT 2016

Der findes også søgeværktøjer, der er henvendt til specifikke målgrupper. Google Scholar et søgeværktøj er henvendt til forskere, studerende samt andre med interesse i faglitteratur.

WolframAlpha er anderledes fra andre søgemaskiner, da den fungerer som et leksikon og giver en information om søgefrasen, frem for en række relevante links.

PageRank-princippet er baseret på matrixregning inden for feltet lineær algebra. Princippet bag PageRank er at definere en numerisk værdi mellem 0-1, til vurdering af en sides troværdighed og relevans ud fra hvilke sider, linker hertil ved hjælp af internettets egen linkstruktur, hvor hver sides værdi er uafhængig af hvilken side beregningen starter fra. PageRank opstilles som en vektor med summen 1.

Et alternativ til PageRank-algoritmen er HITS-algoritmen, der bruges af Ask-søgemaskinen.

STUDIERETNINGSPROJEKT 2016

Referencer

- Alexa Internet. (2016, 12 18). *The top 500 sites on the web*. Retrieved from Alexa:
<http://www.alexa.com/topsites>
- Blaavand, J. L. (2010). *Filosofien og Matematikken bag Google*. Institut for Matematiske Fag. Aarhus Universitet. Retrieved 12 01, 2016, from Aarhus U: <http://home.math.au.dk/blaavand/google.pdf>
- Cambridge University Press. (2009, 04 07). *Hubs and Authorities*. Retrieved 12 19, 2016, from Stanford University: <http://nlp.stanford.edu/IR-book/html/htmledition/hubs-and-authorities-1.html>
- Cornell University. (2016, 12 19). *Lecture #3: PageRank Algorithm - The Mathematics of Google Search*. Retrieved from Cornell University:
<http://www.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture3/lecture3.html>
- Cornell University. (2016, 12 19). *Lecture #4: HITS Algorithm - Hubs and Authorities on the Internet*. Retrieved from Cornell University:
<http://www.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture4/lecture4.html>
- DuckDuckGo. (2016, 12 16). *DuckDuckGo*. Retrieved from GitHub: <https://github.com/duckduckgo>
- DuckDuckGo. (2016, 12 16). *Help us make DuckDuckGo the best search engine for programmers*. Retrieved from DuckDuckHack: <https://duckduckhack.com/>
- DuckDuckGo. (2016, 06 02). *New Features from a Stronger Yahoo Partnership*. Retrieved 12 18, 2016, from DuckDuckGo: <https://duck.co/blog/post/311/yahoo-partnership>
- DuckDuckGo. (2016, 12 16). *Privacy*. Retrieved from DuckDuckGo: <https://duckduckgo.com/privacy>
- Google. (2016, 12 19). *About*. Retrieved from Google Scholar:
<https://scholar.google.com/intl/en/scholar/about.html>
- Hern, A. (2014, 01 09). *Anonymous search tool DuckDuckGo answered 1bn queries in 2013*. Retrieved 12 18, 2016, from The Guardian: <https://www.theguardian.com/technology/2014/jan/09/anonymous-search-tool-duckduckgo-1bn-queries-2013-google>
- Incitez China. (2016, 04 18). *China Search Engine Market Overview 2015*. Retrieved 12 18, 2016, from China Internet Watch: <https://www.chinainternetwatch.com/17415/search-engine-2012-2018e/>
- Klein, N. (2009). Chokterapien i USA: Homeland Security-boblen. In N. Klein, *Chokdoktrinen* (p. 363). Århus: Klim.
- McGee, M. (2013, 09 16). *EdgeRank Is Dead: Facebook's News Feed Algorithm Now Has Close To 100K Weight Factors*. Retrieved 12 19, 2016, from Marketing Land: <http://marketingland.com/edgerank-is-dead-facebooks-news-feed-algorithm-now-has-close-to-100k-weight-factors-55908>
- MOZ. (2016, 12 19). *Title Tag*. Retrieved from MOZ: <https://moz.com/learn/seo/title-tag>
- Rogers, I. (2016, 12 15). *The Google Pagerank Algorithm and How it Works*. Retrieved from Page Rank Explained: <http://www.cs.princeton.edu/~chazelle/courses/BIB/pagerank.htm>
- Search Engine Land. (2016, 12 19). *Google: Panda Update*. Retrieved from Search Engine Land:
<http://searchengineland.com/library/google/google-panda-update>

STUDIERETNINGSPROJEKT 2016

Startpage Blog. (2016, 12 18). *How the Yahoo hack impacts you, even if you don't have a Yahoo account (and how to get safe again)*. Retrieved from Startpage Blog:

<https://www.startpage.com/blog/ixquick/how-the-yahoo-hack-impacts-you-even-if-you-dont-have-a-yahoo-account-and-how-to-get-safe-again/>

Statista. (2016, 12 18). *Global mobile OS market share in sales to end users from 1st quarter 2009 to 1st quarter 2016*. Retrieved from Statista: <https://www.statista.com/statistics/266136/global-market-share-held-by-smartphone-operating-systems/>

Statista. (2016, 12 16). *Worldwide desktop market share of leading search engines from January 2010 to October 2016*. Retrieved from Statista: <https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/>

Stensdal, K. (2015, 09 28). *Kendt CMS under voldsomt angreb kort efter udsendt patch - webmasterne var gået hjem*. Retrieved 12 19, 2016, from Computerworld: <http://www.computerworld.dk/art/235356/kendt-cms-under-voldsomt-angreb-kort-efter-udsendt-patch-webmasterne-var-gaaet-hjem>

Undervisningsministeriet. (2014, 05 26). *Matematik A - Højere teknisk eksamen - forberedelsesmateriale*. Danmark. doi:htx141.MAT/A-26052014

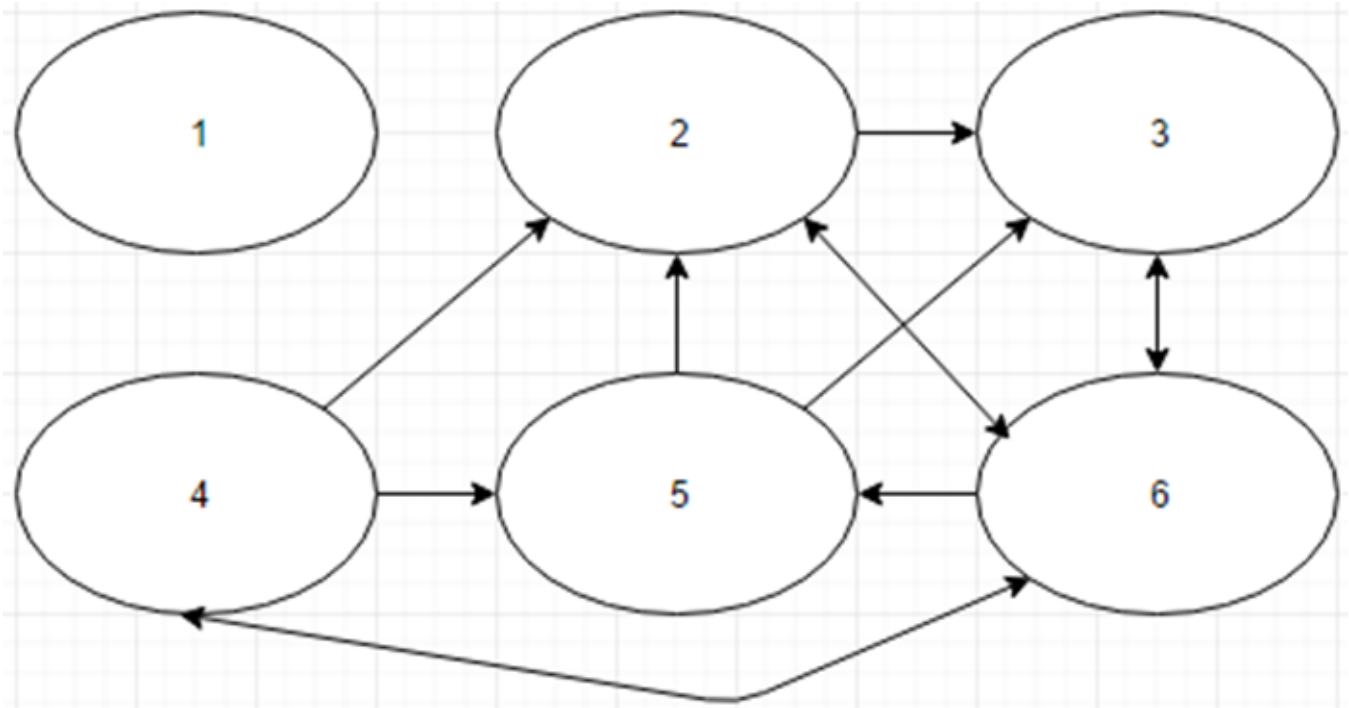
W3Techs.com. (2016, 12 18). *Usage of traffic analysis tools for websites*. Retrieved from W3Techs: https://w3techs.com/technologies/overview/traffic_analysis/all

Waddell, K. (2016, 01 19). *Why Google Quit China - and Why It's Heading Back*. Retrieved 12 18, 2016, from The Atlantic: <http://www.theatlantic.com/technology/archive/2016/01/why-google-quit-china-and-why-its-heading-back/424482/>

Wolfram Language. (2016, 12 19). *Dog*. Retrieved from WolframAlpha: <https://www.wolframalpha.com/input/?i=Dog>

Bilag 1

PageRank-beregning for et lukket netværk



Et isoleret netværk af hjemmesider, bestående af siderne 1, 2, 3, 4, 5 og 6.

Ovenstående netværk opstilt som hyperlinkmatrix.

$$\begin{bmatrix} \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

(2)

Matrix S:
 $S := H + A$

$$\begin{bmatrix} \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & \frac{1}{3} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{6} & \frac{1}{2} & 0 & 0 & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{6} & 0 & 0 & 0 & 0 & \frac{1}{4} \\ \frac{1}{6} & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{4} \\ \frac{1}{6} & \frac{1}{2} & 1 & \frac{1}{3} & 0 & 0 \end{bmatrix}$$

(3)

Startvektor (I^0):

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

(4)

Samlet beregning.

$$G := \left(0.85 \cdot \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{4} \\ 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{4} \\ 0 & \frac{1}{2} & 1 & \frac{1}{3} & 0 & 0 \end{bmatrix} + \begin{bmatrix} \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \frac{0.15}{6} \cdot \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \right)^{50}$$

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 0.0291262135942341 \\ 0.181870097112797 \\ 0.230987247459736 \\ 0.0994504973982950 \\ 0.127628138324613 \\ 0.330937806135394 \end{bmatrix}$$

(5)

$$G := \begin{pmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \cdot \left(0.85 \cdot \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{3} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{4} \\ 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{4} \\ 0 & \frac{1}{2} & 1 & \frac{1}{3} & 0 & 0 \end{bmatrix} + \begin{bmatrix} \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & 0 & 0 & 0 \end{bmatrix} + \frac{0.15}{6} \cdot \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \right)^{100}$$

$$\begin{bmatrix} 0.0291262135882344 \\ 0.181870097075333 \\ 0.230987247412155 \\ 0.0994504973778091 \\ 0.127628138298323 \\ 0.330937806067224 \end{bmatrix}$$

(6)

STUDIERETNINGSPROJEKT 2016

Navn: Tore Vendelhaven	Klasse: 3Z	Fødselsdato: 21.06.1997
---------------------------	---------------	----------------------------

1.fag og niveau: Kommunikation og IT A	vejleder: Bo Alsworth Tvede
2.fag og niveau: Matematik A	vejleder: Stig Salomonsen

Emne: Søgmaskiner
<p>Redegør kort for brugen af matricer og lineær algebra i søgemaskiner og PageRank.</p> <p>Redegør kort for forskellige typer af søgemaskiner.</p> <p>Analysér udvalgte søgemaskinernes strategi, hvor du også inddrager PageRank og lignende funktioner, og vurder om de henvender sig til forskellige målgrupper.</p> <p>Vurder hvilke fordele og ulemper de forskellige søgemaskiner har, og diskuter hvilke etiske problemstillinger, der kan opstå i brugen af dem.</p>

Opgaveformulering udleveres:	13. december 2016 klokken 14.00
Opgaven afleveres:	20. december 2016 senest kl. 14.00

Ved at aflevere denne opgave i www.netproever.dk bekræfter jeg, at opgavebesvarelsen er udarbejdet af mig. Jeg har ikke anvendt tidligere bedømt arbejde uden henvisning hertil, og opgavebesvarelsen er udfærdiget uden anvendelse af uretmæssig hjælp og uden brug af hjælpemidler, der ikke har været tilladt under prøven.

STUDIERETNINGSPROJEKT 2016

Vigtigt!

Afleveringen:

Opgaven skal afleveres elektronisk i Netprøver – læs vejledninger på www.netproever.dk

Du skal bruge Unilogin eller NemId til at logge på netproever.dk, så test i god tid, at dit login virker og test dit it-udstyr – se vejledning på www.netproever.dk

Afleveringen lukker præcis kl. 14 i Netprøver.

Alle opgaver kontrolleres for plagiat gennem Netprøver.

Derudover skal I aflevere opgaven elektronisk på Lectio. Der er oprettet en opgavemappe til hver klasse.

Opgaveformuleringen skal uploades som en pdf-fil i Netprøver som ekstramateriale ved afleveringen.

Vejledning:

Husk I har mulighed for at få vejledning i løbet af skriveugen. Det inkluderer dog ikke gennemlæsning af afsnit. Stil konkrete spørgsmål, så får I mest ud af vejledningen.