# Digitala metoder - inledning

Hur många gånger har du inte läst ”en sökning på Google gav tvåhundrafemtio tusen träffar”? Meningen påträffas – i olika variationer – i uppsatser, tidningskrönikor och de förekommer till och med i vetenskapliga texter. Google används ofta som en termometer för att få en indikation på hur populärt ett specifikt begrepp är. Ju fler träffar som sökningen ger, desto vanligare tänker man sig att förekomsten av ett visst fenomen är. Det är helt förståeligt att man använder Google på detta sätt, sökmotorn är ju nuförtiden långt mera använd än bibliotekens kataloger eller de dammiga uppslagsverken i bokhyllorna. Men vad är ”tvåhundrafemtio tusen” egentligen för typ av mått? Spelar det egentligen någon större roll om det skulle stått ”femhundra tusen” eller ”femhundra miljoner”?

Att räkna träffar från Google skulle man kunna betrakta som en mycket enkel digital metod. Men det dyker genast upp ett stort problem. Hur har Google kommit fram till siffran ”tvåhundrafemtio tusen”? För att vi ska kunna använda en metod på ett vetenskapligt sätt, oavsett om den är digital eller analog, måste vi kunna redovisa steg för steg vad metoden gjort med det empiriska materialet vi har valt att undersöka. Vi måste kunna svara på vilket urval som gjorts, hur detta sedan har bearbetats och lagrats, och inte minst hur vi har räknat fram resultatet. Men inget av detta blir synligt när vi använder oss av Googles fantastiskt effektiva gratistjänst. Vi skriver bara in vad vi vill söka efter, trycker på en knapp och en halv sekund senare har vi fått resultatet serverat. Men vad hände egentligen i Googles serverhallar någonstans i Kalifornien? För att ta reda på det måste vi förstå hur algoritmen\* PageRank (Brin and Page 1998) fungerar, alltså på vilket sätt den formaliserade proceduren som räknar ut hur miljontals hemsidor ska rankas sinsemellan egentligen är inställd. Vi måste öppna en ”svart låda” och se vad som är gömt inuti Googleapparaten som vi använder dagligen utan att tänka på hur den fungerar.

PageRank har förändrats över åren, men en grundläggande princip är att hemsidor får en högre rankning ju oftare de länkas av andra sidor. För att kunna räkna på detta sätt förutsätter algoritmen att det finns en ”random surfer”, en slags genomsnittlig internetanvändare, som helt slumpmässigt klickar sig fram från hemsida till hemsida. Ju större chansen är att denna användare klickar på sidan som ska rankas, desto högre rankning får den enligt algoritmen, enkelt sammanfattat. Exakt en sådan användare finns inte i verkligheten, ingen människa kan fungera helt slumpmässigt. Men Sergey Brin och Lawrence Page, som kring millennieskiftet skulle komma att grunda ett av världens största teknikföretag, antog att på en aggregerad nivå\*\*, alltså när vi lägger samman miljontals surfbeteenden, så skulle denna slumpmässighet ändå utgöra ett realistiskt sätt att räkna på.

Men även om vi i princip kan läsa oss till hur PageRank och liknande sökalgoritmer fungerar så är det i princip omöjligt att använda oss av dessa sökmotorer på ett systematiskt sätt. Det finns helt enkelt för många variabler som inte är redovisade fullt ut gällande vilka hemsidor som inkluderas och exkluderas. Tjänsterna är dessutom kommersiella och därmed utformade för andra ändamål än att generera transparent kunskap.

Denna bok kommer att introducera ett antal metoder som till skillnad från de färdiga tjänsterna och paketlösningarna kan användas *systematiskt* och därmed vetenskapligt. För detta måste man betala ett visst pris. Man måste göra sig bekant med varje steg i den metod man tillämpar och därtill kunna redovisa alla de beslut som man har fattat på väg till sitt forskningsresultat, oavsett om det handlar om att beskriva tekniska lösningar hos olika mjukvaror eller mera konventionella metodöverväganden. Grunden för en vetenskaplig inställning till metod ligger nämligen i möjligheten för någon annan att kunna komma fram till samma resultat som du. Man brukar inom naturvetenskaperna kalla detta för *replikerbarhet*, men detta begrepp känns kanske lite fyrkantigt när man nämner det i samband med samhällsvetenskap och humaniora eftersom det är intimt förknippat med experiment och laborationer. Men på en allmän nivå kan replikerbarhetskriteriet ändå utgöra en första metodprincip att bära med sig när man forskar. Applicerar vi det på en Googlesökning så faller metoden ganska fort. Antalet sökträffar beror på hur Googles index ser ut en viss dag, från vilket land du genomför sökningen och vilka tidigare sökningar du har gjort. Google gör vad de kan för att individualisera sökresultaten snarare än att leverera en neutral sökfunktion, som till exempel en bibliotekskatalog. Vidare så blir det svårt att i detalj redovisa alla inställningar för PageRank-algoritmen, eftersom dessa enbart är synliga för Googles tekniker. Med andra ord, den Googlesökning som du gör kan inte återskapas av vilken annan forskare som helst, eftersom det inte går att beskriva i detalj vad som gjorts.

## Referenser

Brin, Sergey, and Lawrence Page. 1998. “The Anatomy of a Large-Scale Hypertextual Web Search Engine.” *Computer Networks and ISDN Systems*, Proceedings of the Seventh International World Wide Web Conference, 30 (1–7): 107–17. doi:10.1016/S0169-7552(98)00110-X.

# Synopsis

## Kort beskrivning av boken

*Digitala metoder för samhällsvetenskap och humaniora* (arbetstitel) vänder sig till studenter, forskare och yrkesverksamma som vill analysera olika typer av digitalt material. Tonvikten ligger på att använda de nya datamängderna som görs tillgängliga i varierande grad på internet och i databaser på ett vetenskapligt sätt, så att metoderna kan användas i uppsatser, artiklar och andra typer av vetenskapligt orienterade texter.

Boken består av XX kapitel som var för sig introducerar en metod på ett sätt som göra att läsaren kan applicera den helt eller delvis i sitt arbete. Kapitlen är framskrivna så att de både ger e dubbel förståelse av varje metod som presenteras. Dels ger de vägledning om generella aspekter av metoderna, såsom urval, lagring, hantering, bearbetning, presentation och tolkning av material. Men de ger även handfasta tips och exempel på hur man kan dra nytta av en digital bearbetning av materialet, hur specifika forskningsresultat kan presenteras och analyseras. I så stor utsträckning som möjligt ska kapitlen inte vara knutna till specifika mjukvaror, utan hellre presentera generella strategier som är giltiga för flera typer av mjukvara. Varje kapitel disponeras (med variationer) enligt en slags mall, med följande innehåll:

* Introduktion och beskrivning av metodens möjligheter att analysera ett samhälls-/humanvetenskapligt problem.
* Beskrivning av vad som undersöks (ex. interaktioner, nätverk, texter..)
* Metodproceduren.
* Urval
* Bortsortering/felkällor
* Analys/bearbetning (algoritmer)
* Resultatredovisning/Visualisering
* Rekommenderad läsning/relaterad forskning
* Länkar till ev. mjukvara
* Referenser

Utöver de rena metodkapitlena kommer boken även att innehålla ett/två inledande kapitel om vad det innebär att bedriva forskning på internet med digitala metoder. Denna reflektion består dels av en slags beskrivning av vad internet är för typ av social interaktionsyta, vad hypertext och hypermedialitet innebär. Men det kommer även att finnas en vetenskapsteoretisk del som diskuterar förhållandet mellan kvalitativ och kvantitativ metod, mikronivå/aggregerad nivå, algoritmen som forskningsprocedur / ”det algoritmiska tänkandet”, reproducerbarhet/replikerbarhet.

## Titel (arbetstitel)

Digitala Metoder för samhällsvetenskap och humaniora

## Vilka är de huvudsakliga konkurrenterna

Inga böcker på svenska existerar. På engelska finns däremot:

1. Richard Rogers ”Digital Methods”

<https://books.google.se/books/about/Digital_Methods.html?id=mLtEAgAAQBAJ&hl=en>

2. ”Digital Anthropology” (Heather A. Horst & Daniel Miller)

## Målgrupp/Målgrupper

1. Studenter i samhällsvetenskap och humaniora, speciellt på de avancerade nivåernas metodkurser samt för uppsatskurserna. Framförallt borde sociologi, statsvetenskap, medie- och kommunikationsvetenskap ligga nära till hands.

2. Forskare inom HumSam. Med tanke på trenden inom ”digital sociology” och ”digital humanities”.

3. Eventuellt: Journalister som vill närma sig datajournalistik.

## Syfte (kort)

Syftet med boken är att presentera ett flertal användbara digitala metoder som är applicerbara inom human- och samhällsvetenskaperna, samt att introducera metodkritiska reflektioner som är användbara för uppsats- och artikelskrivande. Metoderna skall även kunna vara användbara för framtida yrken som arbetar med att analysera interaktion i internetbaserade medier, exempelvis omvärldsanalys, datajournalistik, kommunikation etc.

## Innehåll/upplägg (kommentera bokens delar/kapitel och proportionerna dem emellan)

Introduktionskapitel:

1. Introduktion – Vad är digitala metoder och varför behövs de inom samhällsvetenskap och humaniora. Här ges en översikt över hur ett helt nytt landskap för forskning har öppnats upp men ännu inte blivit utforskat. Utgångspunkten är att vi aldrig har haft så mycket data om människors liv, deras interaktioner, vad de säger och vad de gör som vi har idag. Vi kan bokstavligen studera vad människor äter till frukost, vad de tycker om, vad de kommenterar och vad de delar med andra i realtid och långt i efterhand. Men denna ”dataflod” skapar ett överflöd av information som blir omöjligt att sortera eller göra begripbart utan att vi använder oss av systematiska metoder för att behandla dessa uppgifter. Det är här som digitala metoder blir helt nödvändiga. För att säga något relevant om det vi studerar måste vi både kunna göra urval och begränsningar samtidigt som vi kan redovisa vad vi har gjort på ett vetenskapligt sätt.

* Tarde's ”glashus”
* Varför digitala metoder?
* Data / Metadata
* Den ”kvalikvantitativa” analysnivån

1. Introduktion 2 – Centrala begrepp inom digitala metoder: Nätverk, noder, relationer, interaktioner, imitationer, kluster, algoritmer [fyll på i dialog med kapitelförfattare]. För att kunna dra nytta av digitala metoder måste vi delvis uppdatera den begreppsapparat som vi normalt använder oss av i Humsam-forskning. När vi studerar interaktioner på internet gör vi ofta det i termer av ”noder i nätverk” snarare än statistiska ”individer i massor”. Men sådana analyser är beroende av olika former av mjukvaror som baserar sig på algoritmer (stegvisa operationer, formaliserade tillvägagångssätt) som forskaren måste behärska och kunna redovisa. Fördelen med algoritmer är att de sedan kan appliceras på stora mängder data, men nackdelen är att de ibland kan te sig som väldigt tekniska. Detta kapitel förklarar hur detta bör hanteras av samhällsvetare och humanister på ett grundläggande sätt och hur man kan redovisa algoritmen som en central del av forskningsprocessen.

* Noder och nätverk
* Individ/massa → divid/bank
* Forskning eller övervakning? Vad är skillnaden?
* Hypotesprövning eller ny-induktivism
* Hur blir det vetenskapligt?

Metodspecifika kapitel:

1. Sociala medier: Facebook

I detta kapitel beskrivs hur man kan samla data från Facebook för att studera olika interaktonsmönster. Fokus ligger på att inhämta data och analysera/visualisera den.

Förslag på författare: Anders Olof Larsson - http://www.andersoloflarsson.se/publications/

1. Sociala medier: Twitter. Detta kapitel kommer att handla om hur man kan studera politisk opinionsbildning på Twitter. Fokus på hur man samlar och analyserar Twitter-data som sociala nätverk. Författare (konfirmerad) Jonas Andersson-Schwartz, Medie- och kommunikationsvetenskap, Södertörns Högskola.
2. Sociala medier: Bloggar (ännu inte hittat författare)
3. Sociala medier: Onlineforum (Förslag: Kania-Lundholm och Lindgren, se http://www.tandfonline.com/doi/pdf/10.1080/14608944.2015.1108958)
4. Sociala medier: Spotify och musikdelning: Författare Pelle Snickars, professor i Medie- och Kommunikationsvetenskap med inriktning mot digital humaniora. Skriver utifrån det pågående forskningsprojektet ”Strömmande kulturarv”.
5. Sociala medier: Onlineforum för poesi. Förslag på författare, Julia Pennlert, doktorand i litteraturvetenskap, Umeå universitet. Forskar om poesiforumet poeter.se
6. Stora textkorpus: Topic modelling (Letar författare)
7. Stora textkorpus: Sentiment analysis (Letar författare)
8. Stora textkorpus: Receptionsstudier <http://lir.gu.se/english/research/swedish-women-writers-on-export>
9. Forskningetik: Studera grupper som delar känsliga erfarenheter. Författare, Ylva Hård af Segerstad, Tillämpad IT, Göteborgs universitet/Chalmers
10. Forskningsetik: Att kartlägga människor utan att fråga om lov? (CK, utifrån ”övervakning som vetenskap”).
11. Bibliometri: Web of Science, Scopus. Fokus på att analysera citeringar och citeringsnätverk. Författare: Gustaf Nelhans, Vetenskapsteori, Högskolan i Borås.

Ytterligare intressanta personer att kontakta:

* Jussi Karlgren, KTH/Gavagai <http://spraktidningen.se/artiklar/2014/08/spraket-avslojar-hur-vi-rostar>
* Mattias Östmar, friforskare, <https://mattiasostmar.wordpress.com/2015/09/26/presentation-at-sodertorn-university-twitter-database-for-research/>
* Cecilia Lindhé, GU.
* Jessica Parland von Essen, http://digihist.se/

Begreppsordlista:

Sist men inte minst kan en bok av detta slag dra fördel av en ordlista. Varje kapitelförfattare bidrar med eventuella nyckelbegrepp som kan

* Nätverk
* Interaktion
* Algoritm
* Aggregerad data
* …

## Tidsplan

- Jan-Feb: Skriva och skicka ut en inledning till boken, ett första utkast till en några nyckelbegrepp som kan vara gemensamma hållpunkter för de kommande kapitlen.

- Feb-Mar: Hjälpa kapitelförfattare att komma igång, bolla idéer.

- Maj: Deadline för första kapitelutkasten.

- Juni-Aug: Granskning och redigering av kapitel, återsänds till författarna under Aug.

- Sep-Okt: Redigering av kapitel, sammanställning av ”ordlista” och konkluderande kapitel.

- Nov-Dec: Färdigställande redigering.

## Övrigt som du tycker är relevant eller viktigt att känna till kring projektet

Internationella kurser med relevant innehåll:

1. <http://studiegids.uva.nl/xmlpages/page/2015-2016-en/search-course/course/1132380>