

**Jämförelse av analysmetoder för clickstream-data**

**(HS-IKI-EA-04-303)**

**Fredrik Ekberg (a01freek@student.his.se)**

*Institutionen för kommunikation och information  
Högskolan i Skövde, Box 408  
S-54128 Skövde, SWEDEN*

Examensarbete på det systemvetenskapliga programmet under vårterminen 2004.

Handledare: Michael Andersson

**[Jämförelse av analysmetoder för clickstream-data]**

Examensrapport inlämnad av Fredrik Ekberg till Högskolan i Skövde, för Kandidatexamen (B.Sc.) vid Institutionen för kommunikation och information.

**[2004-06-08]**

Härmed intygas att allt material i denna rapport, vilket inte är mitt eget, har blivit tydligt identifierat och att inget material är inkluderat som tidigare använts för erhållande av annan examen.

Signerat: \_\_\_\_\_

## Jämförelse av analysmetoder för clickstream-data

Fredrik Ekberg (a01freek@student.his.se)

### Sammanfattning

Det här arbetet har som syfte att genom en jämförelse av olika analysmetoder för clickstream-data kunna fungera som en vägledning när en metod ska implementeras. Metoden som använts vid jämförelsen är litteraturstudie i och med att de analyseringsmetoder som ska undersökas redan är framtagna och kunskap om dem fås genom att studera litteratur i vilka de förekommer. Ett antal kriterier används sedan vid själva jämförelsen, anledningen till detta är att metoderna ska jämföras utifrån en gemensam grund.

De metoder som uppfyllde kraven för de olika kriterierna bäst var page events fact model och subsession fact model. Subsession fact model kan dock upplevas som det bästa valet i alla lägen men samtidigt är den kanske lite överdriven om clickstream-datan bara ska användas till att se hur besökarna använder varje individuell sida för att användas i designsupport syfte. Det går alltså att påvisa att syftet styr vilken metod som är mest lämpad.

**Nyckelord:** Clickstream, Analysmetoder, E-handel, Internet

# Innehållsförteckning

<b>1</b>	<b>Inledning</b> .....	<b>1</b>
<b>2</b>	<b>Bakgrund</b> .....	<b>2</b>
2.1	Clickstream-data.....	2
2.2	Vad kan clickstream-data användas till?.....	4
2.3	Analys av clickstream-data .....	5
<b>3</b>	<b>Problem</b> .....	<b>9</b>
3.1	Problemställning.....	9
3.2	Avgränsning .....	9
<b>4</b>	<b>Metod</b> .....	<b>10</b>
4.1	Kriterier .....	10
<b>5</b>	<b>Resultat</b> .....	<b>12</b>
5.1	Session fact model.....	12
5.1.1	Personifiering av webbinnehåll.....	12
5.1.2	Prefetching och Caching .....	13
5.1.3	Designsupport.....	13
5.1.4	E-handel .....	13
5.2	Page events fact model.....	13
5.2.1	Personifiering av webbinnehåll.....	14
5.2.2	Prefetching och Caching .....	14
5.2.3	Design support.....	15
5.2.4	E-handel .....	15
5.3	Subsession fact model.....	15
5.3.1	Personifiering av webbinnehåll.....	15
5.3.2	Prefetching och Caching .....	16
5.3.3	Design support.....	16
5.3.4	E-handel .....	16
<b>6</b>	<b>Analys</b> .....	<b>17</b>
6.1	Analys av resultatet.....	17
<b>7</b>	<b>Diskussion</b> .....	<b>19</b>
7.1	Metod och resultat .....	19
7.2	Förslag till fortsatt arbete .....	20
7.3	Utvärdering .....	20

<b>Referenser .....</b>	<b>20</b>
<b>Bilaga 1</b> Dimensionsattribut	

# 1 Inledning

I dagsläget har de flesta företag etablerat någon typ av närvaro on-line, antingen bara för att marknadsföra företagets image eller för att presentera sina produkter och sin service, men även för att sälja dessa produkter och servicetjänster (Theusinger & Huber, 2000).

Medan besökare navigerar runt på ett företags webbsida sparas deras interaktion i webblogger. Analyser av dessa webblogger ger värdefull information om vad besökarna är intresserade av för produkter eller tjänster (Theusinger & Huber, 2000). Enligt Theusinger och Huber (2000) handlar spåranalyser mestadels om en besökares navigationsbeteende. Det är av den anledningen som analyser av så kallad webbloggerdata har en så pass central roll för att hjälpa företagets strategi. Anledningen till att den här typen av analyser är speciellt viktig för webbaffärer är eftersom konkurrensen bara är ett musklick bort (Theusinger & Huber, 2000).

Navigeringsdesign handlar om att besvara tre fundamentala frågor (Nielsen, 2000, & Danielson, 2002): Var är jag? Var har jag varit? Vart kan jag gå? Enligt Danielson (2002) handlar webbdesign delvis om att förekomma rörelser med hjälp av beteendedata och att presentera navigeringshjälp utifrån de förväntade beteendena. För att sedan se hur användare navigerar runt på sidan kan något som kallas clickstream användas. Clickstream-data är de fotavtryck som en användare lämnar efter sig vid ett besök på en hemsida (Ulfelder, 2001). Ulfelder menar även på att trots användandet av sådan här data har ökat markant, så har inte företagets möjlighet att analysera den data som samlats in hängt med i utvecklingen.

Fördelarna med att kunna genomföra analyser av clickstream-data är något företag som är närvarande på webben fått upp ögonen för. Företagen ser det som en möjlighet att kunna förbättra sin webbplats, då de kan dra stora fördelar av att analysera information genom hur webbplatsen används (Andersen m fl. 2000). Enligt Andersen m fl. (2000) skulle ett spännande sätt att förbättra webbplatser vara genom att göra dem mer personliga. En webbplats som använder sig av personifiering skulle ge sina kunder en bättre upplevelse, både genom att användare skulle behöva göra färre klick för att få vad denne vill ha och för att sidan blir mer relevant för dem (Andersen m fl., 2000). Det här kräver dock att de har ett effektivt sätt att analysera den data som de samlar in i sina webbloggerfiler (Andersen m fl., 2000).

Andersen m fl. (2000) påpekar att det krävs en noggrann analys av den clickstream-data som samlas in för att kunna dra slutsatser angående en användares beteende. Men vilken metod ska användas för att göra detta? Problemet är att varje forskare som tar fram en metod att analysera clickstream-data förespråkar sin metod.

## 2 Bakgrund

Sedan Internet trädde fram som ett möjligt medium för kommersiella syften har besökare noga övervakats av analytiker för att kunna få ett mått på hur framgångsrik en handelssida är. Sättet att mäta hur framgångsrik en handelssida är genom med antalet besök per månad ger en dålig insikt i besökares individuella beteenden på sidan (Moe m fl., 2001). I miljöer som förändras som till exempel Internet utvecklas ett beteende över tiden. Förståelse för den här sortens utveckling kan ge viktiga kunskaper i hur en affärssida sköts och marknadsförs (Moe m fl., 2001).

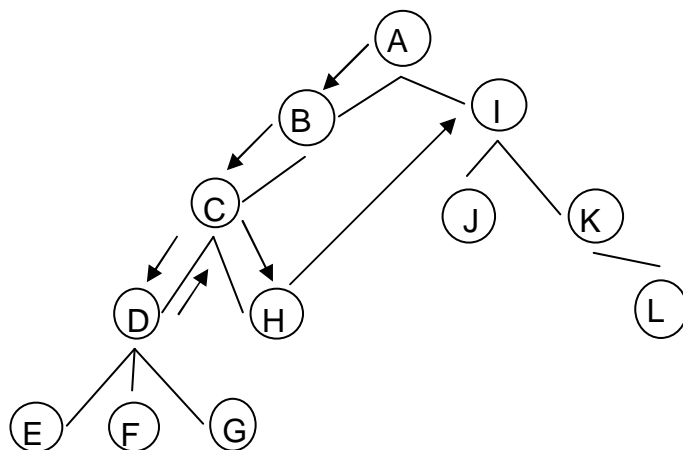
I detta avsnitt kommer clickstream att presenteras, vad det är och hur det kan analyseras. Först kommer clickstream-data att förklaras ingående, vad det består av och hur det samlas in. Därefter kommer ett avsnitt om vad clickstream-data kan användas till och till sist ett avsnitt om hur clickstream-data kan analyseras för att vara användbart.

### 2.1 Clickstream-data

Clickstream-data är den detaljerade elektroniska informationen som är spårad och sparad om Internet-användande och e-handelstransaktioner. Datan ger empiriska forskare ett utmärkt tillfälle att förbättra förståelsen och förutsägbarheten hos konsumenters beteende som har med val att göra (Bucklin m fl., 2002). Bucklin m fl. (2002) påpekar också att eftersom clickstream-data spårar mycket av det som sker före själva köpet så kan den här typen av data ge mer detaljerad information angående konsumenters, än i de fallen då register sparas med hjälp av skannrar ute i dagligvarubutiker.

Uttrycket "clickstream" används för att beteckna den väg som en besökare nyttjar genom en eller flera webbsidor. Vägen som en besökare tagit visar på en serie av val som besökaren tagit ställning till både på en webbsida och över flera webbsidor (Bucklin m fl., 2002). En serie av webbsidor som efterfrågats av en besökare kallas för en session och clickstream-data på en webbsida är en samling av sessioner på den sidan (Lee m fl., 2001). Lee m fl. förklarar också att clickstream-data kan härledas från rena sidförfrågningar och den informationen som hör till. Exempel på sådan information är tid, IP adress, URL, status, antalet överförda bytes, användaragent och ibland cookie data vilka sparas i Webbserver-loggfiler. Analyser av clickstream-data visar hur en användare navigerar och använder en webbsida.

En webbplats består av ett antal sidor som är hyperlänkade. I Figur 1 nedan visar ett exempel på en webbplats där noderna är sidor och strecken är hyperlänkar (Theusinger & Huber, 2000).



Figur 1: En webbplats med ett spår avbildad från Theusinger & Huber (2000)

Under ett besök på en webbplats navigerar sig en besökare runt antingen genom hyperlänkar, interna sökningar eller genom eventuella bokmärken för att ta sig till önskad sida (Theusinger & Huber, 2000). Ett exempel på hur en besökare navigerar runt på en webbplats kan ses i bilden ovan, där pilarna markerar hur besökaren rört sig. Varje sida som besökaren besökt sparas i webbloggen som ett separat protokoll (Theusinger & Huber, 2000). Enligt Theusinger och Huber (2000) är det den här sekvensen som kallas "spår" eller ibland "clickstream" och detta sparas i form av en loggfil. Spåren kan sedan analyseras för att få reda på hur användare navigerar på sidan, den här informationen kallas även för "e-intelligens", vilket är en fördel för de organisationer som använder den menar Theusinger och Huber (2000).

Loggfiler från en webbserver och cookies är två sätt att samla in data på. Loggfilerna från en webbserver skapas automatiskt när en användare besöker en URL på en sida. I en loggfil registreras besökarens IP-adress, tiden då besökaren går in på sidan samt hur länge denne är där och vilken URL som besöks (Perner & Fiss, 2002). Med hjälp av den här informationen går det att få fram hur en användare rör sig på sidan. Loggfilerna är viktiga för att kunna upptäcka användares beteende på en sida, dock leder inte alltid den sparade IP-adressen till en specifik användare, då adressen kan ha ändrats av en proxyserver och att de heuristisk metoder som används för att identifiera en användarsession inte alltid håller (Perner & Fiss, 2002). Det här är anledningen till att cookies kan vara att föredra. Cookies är en kort textfil som skapas av servern på besökarens dator medan denne besöker sidan (Perner & Fiss, 2002). Perner och Fiss förklarar vidare att en cookie ger möjlighet att dela ut ett identifieringsnummer eller kod till en specifik användare, vilka sedan kan användas för att identifiera en användare varje gång denne besöker sidan. Enligt Perner och Fiss (2002) finns det dock en nackdel med cookies och det är att användaren måste ge sitt tillstånd vilket inte alltid är fallet, därför är en kombination av cookies och loggfiler enda möjligheten för att få en god grund för data mining.



## 2.2 Vad kan clickstream-data användas till?

Enligt Kimball och Merz (2000) så är clickstream-data en typ av data som är överväldigande. Detta beror på att det finns så mycket data och att delar av det är så intressant att det skulle gå att lägga all sin energi på att vrida och vända på datan utan att fokusera på ett mål. Kimball och Merz (2000) påpekar därför att det är viktigt att tänka på att datan i sig är praktiskt taget oanvändbar utan att först ha organiserat den för att kunna se spår, dessa spår kan kallas information. Men bara att ha fått fram den här informationen leder inte till att beslut kan fattas på det underlaget utan först måste informationen omvandlas till kunskap sedan kan beslut fattas (Kimball & Merz, 2000).

Den information som går att få fram från en clickstream-analys är enligt Danielson (2002) intressant för loggfilsanalytiker samt informationsdesigners. Detta i och med att typen av rörelsemönster och en användares navigeringsöverskikt går att få fram, vilka kan användas för att förutspå framgång och nöjda användare med en webbplats (Danielson, 2002).

Enligt Perner och Fiss (2002) så är det lika viktigt i en webbaffär som i en vanlig affär med god rådgivning utifrån en kunds förfrågningar samt hjälp efter försäljningen för att få kundförtroende. Perner och Fiss (2002) fortsätter med att påpeka att detta kräver att affären har exakt kunskap om vad kunden föredrar, som i sin tur inte gillar långa utfrågningar eller användandet av andra kommunikations sätt. En webbaffärs-ägare måste alltså samla in den här informationen genom att studera kundens interaktion på sidan och datan från säljprocessen, vilken kräver en omfattande dataanalys av den clickstream-data som samlats in. Det är viktigt att generera återkommande affärer då flera undersökningar visat att det är minst fem gånger så dyrt att skaffa en ny kund jämfört med att få en gammal kund att komma tillbaka för återköp (Falonius & Hedberg, 1995).

Det finns ett antal användarmodeller: den individuella användarmodellen, den kortsiktiga användarmodellen och den långsiktiga användarmodellen (Perner & Fiss, 2002). Den individuella användarmodellen kan erhållas genom att skapa elektroniska profiler. Genom att använda sig av frågeformulär kan en användares behov och önskemål samlas in. De elektroniska profilerna kan sedan användas varje gång användaren går in på sidan för att visa innehållet utifrån dennes behov och önskemål samt vilken typ av hård- och mjukvara som användaren har (Perner & Fiss, 2002). Men som Perner och Fiss (2002) påpekar att då det krävs ett visst engagemang från användarens sida genom att fylla i ett frågeformulär samt att många inte vill lämna ut den nödvändiga informationen om sig själva är det möjligt att dela upp användarna i flera grupper. Varje användargrupp representerar en utmärkande och tillräckligt stor grupp av användare som delar ett antal karaktärsdrag. Identifieringen av dessa grupper kan ske genom att se hur de navigerar runt på sidan (Perner & Fiss, 2002).

Enligt Andersen m fl. (2000) så är ett viktigt användningsområde av clickstream information att kunna göra webbplatser mer personliga utifrån en användares behov. Exempel på detta kan vara att leverera reklam och tjänster utifrån intresse vilket skulle leda till att kvaliteten av användarens interaktion förbättras samt till högre

kundlojalitet. För att lyckas med detta är det viktigt att identifiera användargrupper och då speciellt den användargrupp som butiken vänder sig mot för att på så sätt generera så många lönsamma affärer som möjligt (Andersen m fl., 2000).

En användares intresse kan förändras med tiden så borde användarmodeller som tar vara på detta skapas (Perner & Fiss, 2002). Detta har lett till skapandet av en kortsiktig och en långsiktig användarmodell. Där den kortsiktiga är baserad på mycket specifik information medan den långsiktiga är baserad på mer generell information (Perner & Fiss, 2002). Den kortsiktiga användarmodellen är skapad enbart utifrån de senaste observationerna. Den representerar användarmodeller som snabbt kan ändras på grund av ändrat intresse. Om den kortsiktiga användarmodellen inte kan klassificera en användare så skickas denne vidare till den långsiktiga användarmodellen som representerar mer stereotypa användargrupper (Perner & Fiss, 2002). Enligt Perner och Fiss (2002) så är syftet med den långsiktiga modellen att modellera användares generella förväntningar på speciella produkter som inte kunde bli klassificerade av den kortsiktiga modellen.

Om clickstream utnyttjas i en E-handels butik så går det att få ut information för förståelse om hur effektiv marknadsföringen varit, som till exempel hur kunden hittade butiken, vilka produkter de tittar på och vad de köpte (Lee m fl., 2001). En analys av information som ingår i clickstream-data är nödvändig för att kunna förbättra effektiviteten av webbmarknadsföring och exponering i E-handels butiker. På grund av detta intresse att analysera data om webbanvändandet i webbloggfiler har resulterat i en marknad för verktyg som kan genomföra analyser, summera och visualisera mönster i webbanvändandet ur webbloggfiler (Lee m fl., 2001).

### **2.3 Analys av clickstream-data**

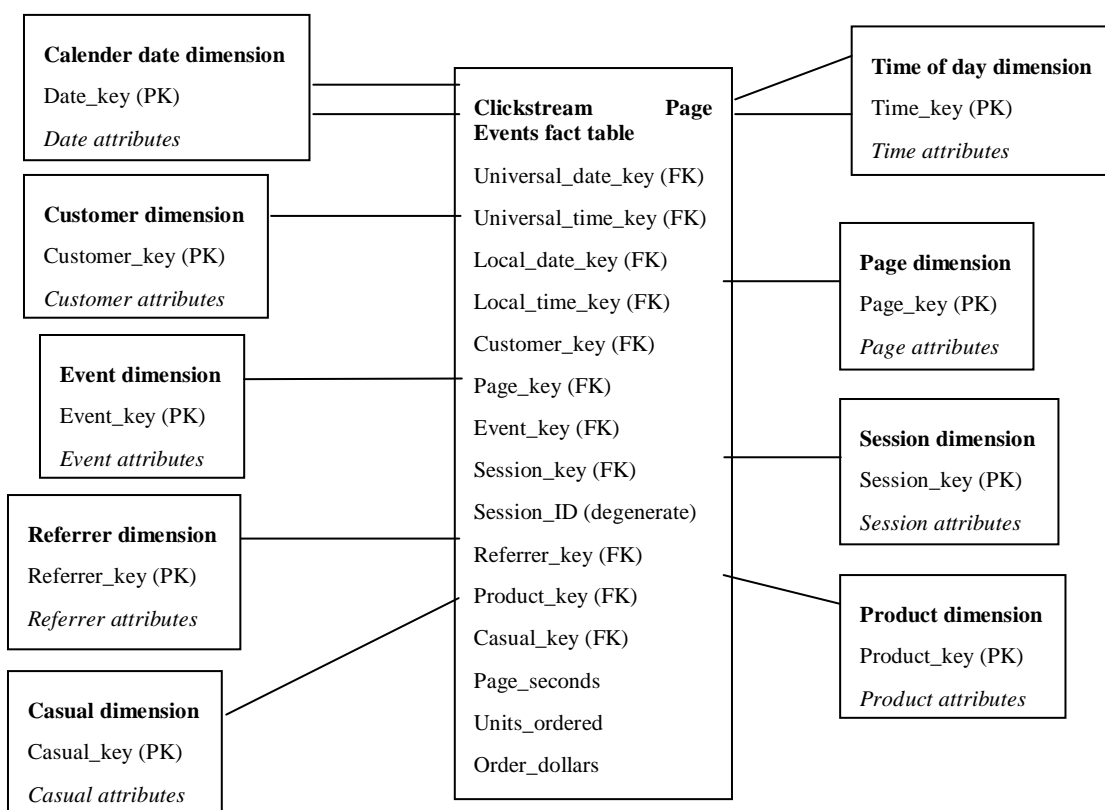
Enligt Theusinger och Huber (2000) så kan en datamodell för clickstream-data delas upp i två typer, transaktionsbaserad data och kundbaserad data. Det mesta av den transaktionsbaserade datan kan fås ut från en loggfil. I en sådan loggfil skapar varje klick vanligtvis en rad som innehåller IP adress, tidpunkt, referens adress, vilken sida och serveradress. Baserat på den här datan kan ett första steg vara att särskilja transaktioner utifrån syfte, nästa steg skulle sedan kunna vara att utföra associerings och sekvensanalys för att få en inblick i besökta webbplatser och kunders spår (Theusinger & Huber, 2000). Om en mer sofistikerad analys efterfrågas så behövs det mer information om transaktionerna så därför behöver nya variabler att tas med i beräkningen från loggfilen (Theusinger & Huber, 2000). Exempel på sådana variabler är enligt Theusinger och Huber (2000): antal klick, genomsnittlig tid per webbplats, webbläsar typ, vilka webbplatser som besöktes, i vilken ordning och vilken veckodag. Baserat på den här typen av data så kan det gå att hitta segment av kunder med liknande beteende (Theusinger & Huber, 2000).

Enligt Danielson (2002) kan analyser av clickstream-data i huvudsak genomföras på tre sätt, som till exempel användandet av "Tillbakaknappen" i webbnavigering: (1) medelfrekvens (2) medelproportioner av någon totalsumma, som till exempel en användares totala antal av navigeringsaktiviteter, eller (3) frekvens och proportion

över alla navigatörer. Alternativ tre används vanligtvis vid globala navigeringsanalyser, vilket ofta används när individuella externa platsanvändare inte kan skiljas från varandra (Danielson, 2002).

Enligt Andersen m fl. (2000) så är de två vanligaste modellerna för modellering av clickstream click fact table och sessions fact table. Dessa överensstämmer dock inte helt med de modeller som finns i Kimball och Merz (2000) vilket är den källa som Andersen m fl. (2000) har använt som referens. Click fact table som Andersen m fl. (2000) nämner heter i Kimball och Merz (2000) clickstream page events table. Det senare namnet är det som kommer att användas i fortsättningen.

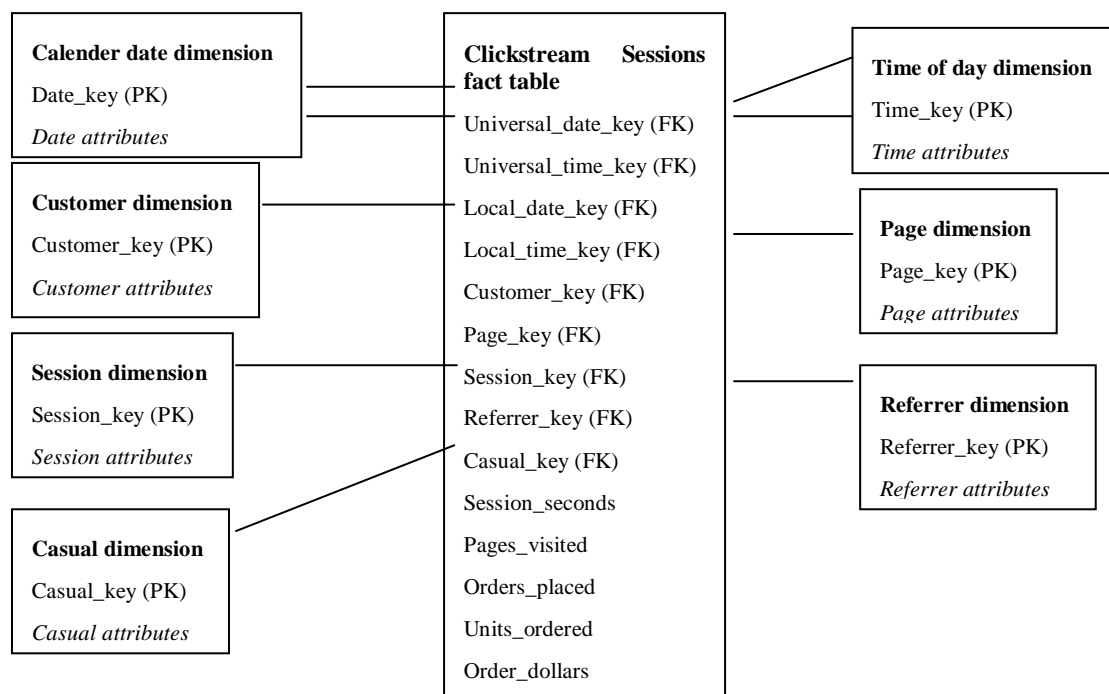
En page events fact table är ett stjärnschema som modellerar enskilda klick i faktatabellen (Andersen m fl., 2000). Figur 2 nedan visar en page events fact table. Se bilaga 1 för att få reda på vilken funktion en dimension har samt vilka attribut dimensionen består av. Fakta tabellen modellerar självmant informations som nummer i sessionen och klicksekunder, där nummer i sessionen illustrerar var i sessionen ett specifikt klick ägde rum (Andersen m fl., 2000).



Figur 2: Modellen av en Clickstream Page Events fact table avbildad från Kimball och Merz (2000). Alla kopplingar ska vara 1..N, där många är på fact table sidan.

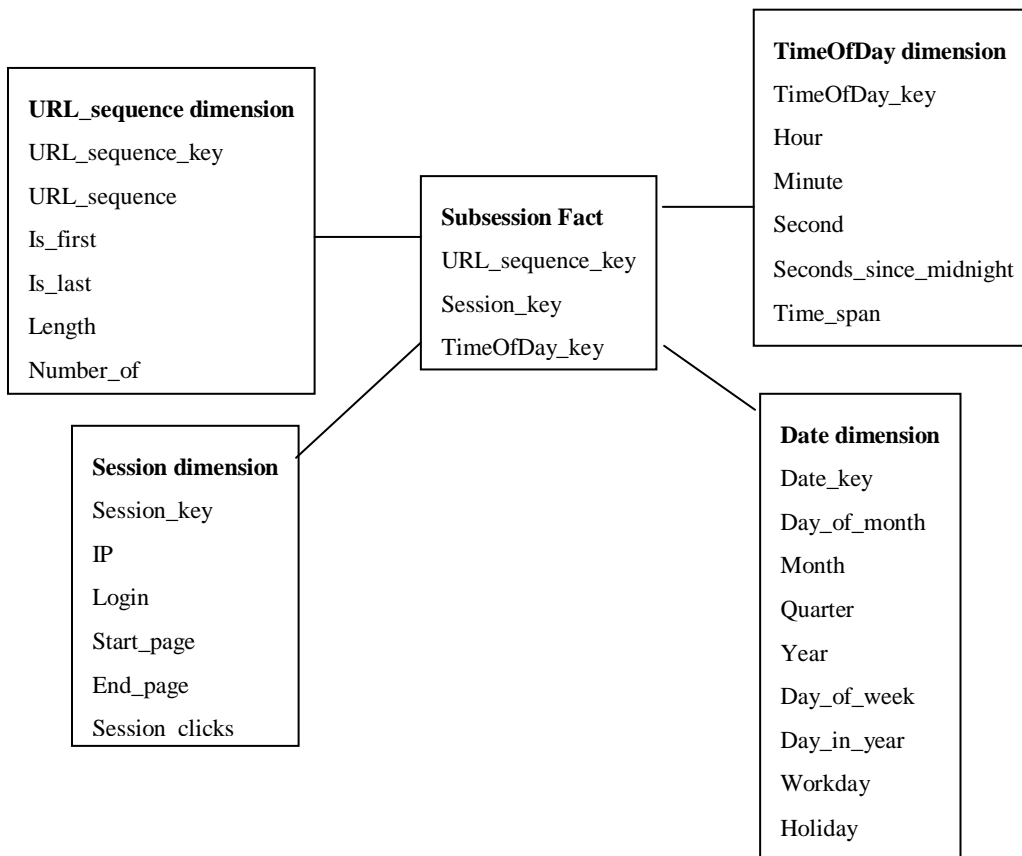
Enligt Andersen m fl. (2000) är en sessions fact table ett rekommenderat alternativ till page events fact table samt ett komplement till densamma. Sessions fact table modellerar kompletta sessioner istället för enskilda klick i fakta tabell (Andersen m

fl., 2000). I figur 3 syns en modell över sessions fact table. De attribut som ingår i en dimension finns listade i bilaga 1, där förklaras även en dimensions användningsområde kort. De attribut som finns i faktatabellen och som inte är främmande nycklar är uppmätta fakta (Kimball och Merz, 2000).

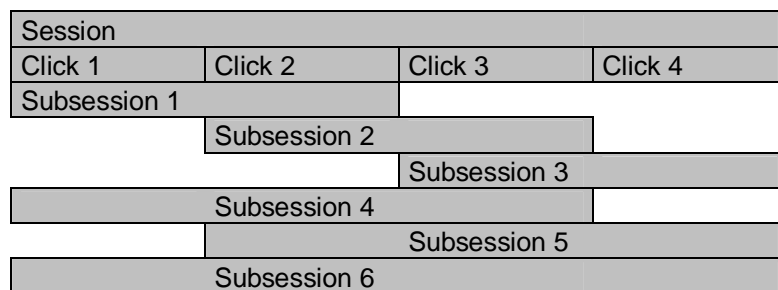


Figur 3: Modellen av en Clickstream Sessions fact table avbildad från Kimball och Merz (2000). Alla kopplingar ska vara 1..N, där många är på fact table sidan.

Andersen m fl. (2000) presenterar ett annat sätt att analysera clickstream-data på. Deras metod går ut på att använda sig av undersessioner för att på ett bättre sätt kunna analysera sekvenser. Enligt Andersen m fl. (2000) så är meningen med subsession fact model att modellera sekvenser av på varandra följande klickningar från varje session, en sådan sekvens kallar de undersession. Figur 4 visar hur en möjlig subsession fact table ser ut. De dimensioner som ingår i en subsession fact table är valfria så länge URL sequence dimensionen finns med. En genomsnittlig användarsession kommer att sträcka sig över flera undersessioner och undersessionerna kommer att överlappa varandra i många lägen. I figur 5 där en session med 4 klickningar är uppdelad i sex undersessioner. Undersessioner som bara har en längd av ett är inte med då de är singel klickningar och enklast modelleras med en page events fact table. Längden på en undersession är sådan att i en session med 10 klickningar så skulle det skapa en 10-klicks undersession, två 9-klicks undersessions, tre 8-klicks undersession och så vidare (Andersen m fl., 2000).



Figur 4: Modellen av en Subsessions fact table avbildad från Andersen m fl. (2000).



Figur 5: Undersessioner som sträcker sig över en 4-klicks session, återskapad från Andersen m fl. (2000).

## 3 Problem

I det här avsnittet kommer problemet som rapporten utgått ifrån att presenteras samt de aktuella avgränsningarna som används för att få ett mer hanterbart problem. Motivet till varför detta problem är av intresse kommer också att presenteras.

### 3.1 Problemställning

Enligt Andersen m fl. (2000) så håller analysering av clickstream-data på att bli en av de absolut viktigaste aktiviteterna för företag i alla branscher eftersom de flesta företagen blir e-företag. Då en analys av clickstream-data kan ge besked om användarnas mönster på företagets webbplats skulle detta ge en betydligt ökad förståelse för kundernas beteende. Den här kunskapen skulle sedan användas till att öka kundernas belåtenhet med webbplatsen och företaget i stort och generera ett stort konkurrensövertag (Andersen m fl., 2000).

Enligt Xing och Shen (2004) så är det ett brådskande problem för Internet E-handel att skapa det föredragna användarmönstret på ett korrekt sätt vilket skulle leda till optimering av webbsidor och förbättrad affärsstrategi.

För att kunna göra webbplatser personliga automatiskt så krävs det att webbsystemet genomför en noggrann analys av användares tidigare beteenden för att kunna fastställa vad de föredrar samt deras vanor (Andersen m fl., 2000). Andersen m fl. (2000) menar därför att det är väldigt viktigt att hitta ett effektivt och kraftfullt sätt att analysera stora mängder av clickstream-data och då speciellt sekvenser av klick.

*Syftet med arbetet är att jämföra metoder för analys av clickstream-data.*

Den här rapporten har för avsikt att ge en bild av ett antal metoder för analyser av clickstream-data samt deras fördelar och nackdelar. Detta för att någon sedan ska kunna använda rapporten till att avgöra vilken metod som kan passa dem bäst utifrån deras behov. Anledningen till att det finns ett behov av detta är att varje författare förespråkar den metod som denne själv kommit på.

### 3.2 Avgränsning

Rapporten kommer inte att gå in på några tekniska detaljer ifråga om vad som behövs för att implementera någon av analysmetoderna.

Vidare kommer bara Session fact model, Page events fact model samt Subsession fact model att användas vid jämförelserna. Anledningen till att den fjärde metoden som beskrivs i kapitel 2.3 inte är med är att den inte är så komplett som de övriga. Den har ett användningsområde och det är att se hur mycket olika knappar används på en sida.

## 4 Metod

Litteraturstudie är den grundläggande metoden som kommer att användas i genomförandet av det här arbetet. Anledningen till detta är att ett antal redan framtagna analysmetoder ska jämföras, så för att få så god kunskap om dessa som möjligt krävs det en hel del litteraturstudier.

För att sedan kunna göra en jämförelse av olika analysmetoder kommer ett antal kriterier att användas. Anledningen till att ett antal kriterier kommer att användas är för att få en grund utifrån vilka analysmetoderna kan jämföras. Hade inte kriterier använts så hade jämförelserna mer blivit en sammanfattning av vad författarna påstår om sina metoder. Kriterierna kommer att vara av en sådan karaktär som kan förväntas att en användare av clickstream-data vill kunna använda datan till. Dessa kriterier kommer sedan att appliceras på varje analysmetod för att kunna göra en utvärdering om eller hur väl den data som går att få fram kan bidra till att genomföra det som vardera kriteriet är anpassat för.

### 4.1 Kriterier

De kriterier som kommer att användas är hämtade från Facca och Lanzi (2003) där de presenterar ett antal användningsområden för information om navigeringsspår. Dessa användningsområden är:

- **Personifiering av webbinnehåll**, vilket innebär att en webbplats anpassas efter den användaren som använder den. Till exempel är det möjligt att i realtid förutse en användares beteende genom att jämföra det aktuella navigeringsspåret med typiska spår från tidigare webblogger. De dimensioner som bör finnas med här är customer-, event- och page dimensionen.
- **”Prefetching” och ”Caching”**, vilket innebär att informationen kan hjälpa till att utveckla användbara prefetching och caching strategier för att minska en servers svarstid. För att uppfylla det här kriteriet är de dimensioner som bör finnas med här är referrer-, time of day- samt calender date dimensionen.
- **Designsupport**, användbarhet är ett av de största ämnena när det gäller design och implementation av en webbplats. Ett användningsområde är att informationen används för att se hur sidan används och göra justeringar utifrån denna genom att ta fram riktlinjer för design av webbapplikationer. Dimensionerna som bör finnas med här är event- och page dimensionen.
- **E-handel**, att utvinna affärskunskap utifrån web usage data är av stor vikt för företag som bedriver E-handel. Anledningen till detta är att höja andelen nöjda kunder. Detta kan göras genom att titta på, kundlockelse, behållandet av kunder, att kunna sälja nya produktlinjer till befintliga kunder samt kunder som lämnar. I det här kriteriet är det bra pm följande dimensioner finns med customer-, product- och page dimensionen.

Anledningen till att dessa kriterier har valts är att dessa förekommer i de flesta artiklar som skrivs inom användning av clickstream och web usage mining. Enligt Facca och Lanzi (2003) är web usage mining den del av web mining som hanterat framtagandet av kunskap ur serverloggfiler. Den data som används i web usage mining kommer huvudsakligen från tre källor: (1) webbservrar, (2) proxyservrar och (3) webbklienter. Huvudmålet med web usage mining är att samla in intressant information om användares navigeringsspår för att sedan kunna förbättra webbplatsen utifrån användarnas synpunkter (Facca & Lanzi, 2003). Perner och Fiss (2002) samt Lee m fl. (2001) påpekar hur viktigt det är för ett E-handels företag dels för att skapa nöjda kunder samt för att kunna se hur effektiv marknadsföringen har varit. Personifiering av webbinnehåll är något som både Andersen m fl. (2000) och Perner och Fiss (2002) nämner som en viktig uppgift för användandet av clickstream-data. I och med att clickstream-datan kartlägger det spår som en användare lämnar efter sig så är den här informationen intressant för informationsdesigners enligt Danielson (2002).



## 5 Resultat

I följande stycke kommer det resultat att presenteras som framkommit vid jämförelserna. Varje metod kommer att presenteras för sig utifrån vardera av de fyra kriterierna från kapitel 4.1.

### 5.1 Session fact model

I följande stycke kommer Session fact model att undersökas hur den hanterar de ovan nämnda kriterierna.

#### 5.1.1 Personifiering av webbinnehåll

I och med att customer dimensionen är en del av modellen så innebär det att information finns om vem besökaren är. Detta är en förutsättning för att personifiering ska kunna genomföras. Beroende hur mycket vi vet om kunden, desto fler attribut vi har att basera en personifiering på, så som till exempel hälsningsfraser.

De attribut som session dimensionen bidrar med, se bilaga 1, gör det möjligt att se vilket syfte besökaren hade med sitt besök genom att de lagrar information om en sidas innehåll men även genom att spara information utifrån den bana som en användare tagit. Detta förstärks genom att information om varifrån besökaren kom och vilken sida som var besökarens startsida.

Det går även att ta reda på om det är en besökare med ett specifikt syfte eller om det bara är en besökare som kikar runt lite på webbplatsen. Den här informationen går att få fram genom att se på hur många sidor med hjälp av pages visited attributet i modellens fact table, som besökaren tittade på under sin session. Antagandet är att en besökare med ett specifikt syfte besöker färre sidor än en som kikar runt på webbplatsen.

Dock så går det inte att göra någon personifiering utifrån vilka produkter som besökaren är intresserad av då den här typen av information inte lagras på något sätt. Det går inte heller att genomföra några justeringar utifrån besökarens beteende på hemsidan eftersom inte heller denna information lagras.

Som sammanfattning angående personifiering av en webbplats med hjälp av session fact model så är det inte något att föredra då det endast går att göra personifiering utifrån den informationen som finns i kund dimensionen. Detta innebär att det går att göra personifieringar såsom att besökaren välkomnas med namn. Vad som inte går att göra är personifieringar utifrån produkter som brukar intressera kunden eller om kunden på något sätt har personligt sätt att använda webbplatsen.

### **5.1.2 Prefetching och Caching**

Genom time of day dimensionen så går det att ta reda på om det är en speciell tid på dygnet som trycket på servern är speciellt högt. Calendar date dimensionen bidrar med att det går att se om trycket på servern är speciellt högt någon speciell dag.

Trots att dessa två dimensioner gör att det går att se när servern får arbeta som allra mest så bidrar det inte till att en effektiv strategi för att minska serverns svarstider i och med att det inte finns något spår att följa för att se vart besökaren är på väg samt att det inte går att dra några slutsatser utifrån besökarens tidigare beteenden i och med att sådan information inte lagras.

Ett par undantag finns som till exempel om att det går att dra några slutsatser genom att göra en analys utifrån syftet med besöken som finns i session dimensionen ihop med calendar date dimensionen och time of day dimensionen. Men detta är endast om det går att se om syftet med besöken varierar utifrån dag och tid. Ett annat undantag är att se om vartifrån en besökare kommer varierar beroende på dag och tid. Dessa två undantag skulle kunna innebära en viss minskning i svarstider men inte någon större då det endast innebär att ett fåtal sidor kan påverkas.

### **5.1.3 Designsupport**

Det finns inget stöd för det här kriteriet i och med att det inte lagras någon information om hur en besökare använder sidan. Vad som eventuellt skulle gå att undersöka är om det finns någon annan sida än den tänkta startsidan som flera besökare får som första kontakt med webbplatsen.

### **5.1.4 E-handel**

I och med att det inte går att göra några mer avancerade personifieringar av webbinnehållet så är detta en brist då det gäller att höja kundnöjdheten hos kända kunder.

Session fact model är dock bra till att undersöka hur kunderna tar sig till webbplatsen. Med hjälp av den här informationen så går det att ta reda på hur eventuella marknadsföringsåtgärder på nätet fungerar. Exempel på sådana kan vara banners eller länkar som finns på externa webbplatser. Utifrån den här informationen kan innehavaren av webbplatsen se vilken typ av marknadsföring som fungerar bäst samt om det finns platser som är bättre än andra att bli länkade ifrån.

## **5.2 Page events fact model**

I följande stycke kommer Page events fact model att undersökas hur den hanterar de ovan nämnda kriterierna.

### **5.2.1 Personifiering av webbinnehåll**

En bra början för att kunna genomföra personifiering finns genom att customer dimensionen finns med, detta gör att vi kan identifiera besökaren. Genom customer dimensionen så har vi mer eller mindre information om den möjliga kunden detta beroende på om hur mycket information denne delat med sig av.

Information för att kunna genomföra en personifiering kan även hämtas från tre andra dimensioner som finns med, dessa är event dimensionen, page dimensionen och product dimensionen. Event dimensionen bidrar med information om vad som sker på en specifik sida. Page dimensionen innehåller information om hur den sidan som besöks innehåller i form av bilder och animationer men även vilken produkt som finns där. Den produkt som finns på sidan är sedan mer förklarad i product dimensionen. Genom att jämföra dessa dimensioner utifrån en besökares tidigare besök på webbplatsen om sådana finns så finns en god grund för att kunna genomföra en personifiering. Tack vare att calender date dimensionen och time of day dimensionen så går det att kontrollera att inte alla besök inte är från samma session.

Något som går att genomföra är att se om det till exempel är en vara eller flera varor som kunden ofta tittar på. När kunden genomför ett köp, har denne då några specifika rutiner innan köpet godkänns som kan göras på ett smidigare sätt för kunden.

Då varje rad som sparas bara innehåller information om vad som skett på en sida innebär detta att det inte går att få någon bra bild över besökaren beteende under hela sessionen. Detta borde inte vara något problem för en personifiering dock eftersom det går att se alla sidor som besökts, hur besökaren tog sig dit genom referrer dimensionen samt vilka händelser som sker på sidan.

### **5.2.2 Prefetching och Caching**

I och med att det går att se vilka sidor som besöks så går det att kontrollera vilka sidor som är speciellt populära vilket innebär att en server kan placera dessa i sitt cache-minne för att underlätta åtkomsten. Då även calender date dimensionen och time of day dimensionen finns med så går det att genomföra analyser om en viss sida är populärare under vissa dagar eller tider på dygnet, detta gör att de sidor som ligger i cache-minnet kan justeras utifrån dessa parametrar.

Med hjälp av referrer dimensionen går det att titta på hur besökare på en specifik sida har tagit sig dit. Om det sedan går att hitta ett mönster genom att det görs en jämförelse mellan flera besökare på samma sida och hur de har tagit sig dit, så går det att underlätta serverns arbete genom prefetching. Detta på grund av att det då går att anta att om en besökare besöker en viss sida så är det troligt att denne går vidare till den sidan som undersökts.

### **5.2.3 Design support**

Event dimensionen ger en god grund till att undersöka hur webbplatsen används. Anledningen till det är att den dimensionen lagrar information om vilka händelser som sker på en sida. Detta är också den informationen som behövs för att kunna dra slutsatser utifrån vilka eventuella förbättringar som skulle höja användbarheten på sidan.

Genom page dimensionen går det att göra undersökningar om det är någon typ av webbsidor som lockar fler besökare än andra i och med att den dimensionen har information om bilder med mera som finns på en specifik sida. Även detta är mycket nyttigt information för att kunna höja användbarheten på en webbplats.

### **5.2.4 E-handel**

I och med att den här modellen passar bra till att göra personifieringar så är den också ett bra alternativ till att förbättra relationen mellan kunden webbplatsen på grund av att den kan anpassas efter deras behov.

Då den här modellen även sparar information om sidornas design samt vilken produkt som beskrivs på sidan, går det att se om kunder föredrar att handla från en sida med en speciell design. Den här typen av information är viktig då innehavaren antagligen är någon typ av E-handlare som behöver sälja sina varor. Men den är inte bara viktig för att innehavaren ska kunna tjäna så mycket pengar som möjligt utan den är viktig för att se till så att kunderna får den information som de kan behöva för att göra ett köp.

Med hjälp av den här modellen har vi även möjlighet att se vilken typ av webbannonser som fungerar i och med att det går att se var ifrån en besökare på en sida kommer. Med den informationen så går det att gå vidare och se hur den sidan som besökaren kom ifrån såg ut, om det är en intern sida, och vad besökaren gjorde där.

## **5.3 Subsession fact model**

I följande stycke kommer Subsession fact model att undersökas hur den hanterar de ovan nämnda kriterierna.

### **5.3.1 Personifiering av webbinnehåll**

På grund av att den här modellen tillåter att dimensioner läggs till så länge URL sequence dimensionen finns kvar så ger den möjligheten att användas för att kunna genomföra personifieringar beroende på vilka dimensioner som läggs till. Det innebär att för det här ändamålet viktiga dimensioner som customer dimensionen och page dimensionen kan adderas till modellen.

Med hjälp av URL sequence dimensionen går det att undersöka om en specifik kund, om customer dimensionen finns med vill säga, har ett rörelsemönster som upprepar sig gång på gång. Det här innebär att utifrån de undersessioner som finns så går det att titta på om det på något sätt går att underlätta navigeringen för den aktuella kunden.

### **5.3.2 Prefetching och Caching**

Det samma gäller det här kriteriet som det ovan, att i och med att dimensioner kan adderas så är det valet av dimensioner som styr hur väl den här modellen kommer att hantera det här kriteriet.

Här kan dimensionen URL sequence dock spela en avgörande roll i och med att det går att undersöka om en specifik sekvens av sidor ofta används. Det skulle innebära en bra möjlighet att kunna ta fram prefetching strategier genom att en påbörjad sekvens jämförs med vanligt förekommande sekvenser.

### **5.3.3 Design support**

Valet av dimensioner är av största vikt här för att kunna studera hur besökare använder webbplatsen. De som skulle behövas är page dimensionen och event dimensionen eftersom de beskriver hur sidan ser ut och hur den används.

URL sequence dimensionen ger inte något stöd när det gäller det här kriteriet i och med att det är varje sida för sig som är viktig för det här kriteriet.

### **5.3.4 E-handel**

Som i alla de andra kriterierna ovan så beror det helt och hållet på vilka dimensioner som innehavaren av webbplatsen väljer att använda sig av. Dock så är det möjligt att med URL sequence dimensionen att kontrollera i fall en viss sekvens av sidbesök ofta leder till att en kund lämnar webbplatsen.

## 6 Analys

I det här kapitlet redovisas en analys av det som framkommit i resultatkapitlet och den problemställning som ligger till grund för det här arbetet. Problemställningen har varit att genomföra en jämförelse av analysmetoder för clickstream-data.

### 6.1 Analys av resultatet

Trots att alla tre metoder är väldigt lika varandra i design så har vissa skillnader kommit fram om hur väl de uppfyller de fyra kriterier som varje metod har testats utifrån.

Session fact model var inte direkt stark på något av kriterierna. Detta beror mycket på att den här metoden inte sparar något information om de sidor som besöks eller vad som sker på dessa sidor. Det som session fact model lämpar sig bäst för är enkla personifieringar, såsom att hälsa kända besökare välkomna med namn.

När det gäller page events fact model så sparar den information som lämpar sig väl för de kriterier som metoden har testats på. Detta är på grund av att den här metoden sparar mycket information om varje sida som besöks. Trots att Page events fact model fick bra resultat på samtliga kriterier går det inte att påvisa att metoden är bättre för ett kriterium än ett annat.

Den stora fördelen med subsession fact model är att vilka dimensioner som helst kan läggas till eller tas bort så länge URL sequence dimensionen finns kvar. Detta gör att den här metoden får ett bra resultat på alla kriterier i och med att den kan anpassas utifrån användningsområdet. Det som utmärker den här metoden är URL sequence dimensionen. Den här dimensionen gör att det är enklare att undersöka upprepade rörelsemönster som är bra information vid personifiering av webbinnehåll samt vid skapandet av prefetchingstrategier och cachingstrategier.

För att tydligare redogöra vilka kriterier en metod uppfyllde så har en tabell skapats (Figur 6). Utifrån det som presenteras ovan och tabellen är det lätt att få intrycket att Subsession fact model alltid är att föredra, vad som dock ska tas i beaktning är vilken typ av domän den ska appliceras på. Anledningen till det är att den information som URL sequence dimensionen sparar kan vara onödig om den insamlade clickstream-datan bara ska användas som designsupport.

	Session fact model	Page events fact model	Subsession fact model
Personifiering av webbinnehåll	X	X	X
Prefetching & Caching		X	X
Designsupport		X	X
E-handel		X	X

Figur 6: En sammanställning av resultatet. X markerar vilka kriterier som uppfylls av metoden.

Är det en informativ webbplats som ska undersökas är kanske inte behovet av att kartlägga alla undersessioner är så stort. Då är det antagligen tillräckligt att se hur en besökare använder webbplatsens sidor och om det är några sidor som besöks mer än andra för att kunna använda den här informationen vid förbättringar av webbplatsen, det vill säga designsupport. Det innebär att Page events fact model skulle fungera alldeles utmärkt. Är dock webbplatsen skapad för försäljning är även andra kriterier intressanta såsom personifiering av webbinnehåll och E-handel.

## 7 Diskussion

I det här kapitlet kommer en diskussion att göras utifrån den metod som valts samt det resultat som framkommit i och med jämförelsen. Ett avsnitt med förslag till fortsatta arbeten kommer att presenteras. Kapitlet avslutas med ett avsnitt tar upp de erfarenheter som det här arbetet har bidragit med, men även funderingar som funnits under arbetets gång.

### 7.1 Metod och resultat

I och med att den metod som använts är litteraturstudie ställs det stora krav på den litteratur som används. Vid genomförandet så har två källor använts var av den ena är en bok och den andra är en artikel. Nackdelen med att använda böcker är att de tar lång tid att skriva så det går inte att ta för givet att det är helt aktuellt det som står i en bok. Dessutom genomgår inte en bok samma granskning som artiklar som publiceras i vetenskapliga tidskrifter eller används som konferensmaterial. Fördelen med artikeln är alltså att den har blivit granskad i och med att den har använts som konferensmaterial. I den artikeln som använts är dock den använda boken en av källorna. Vad gäller övrig litteratur så är det främst artiklar som använts, det beror på att finns få böcker skrivna inom det här området. Alla artiklar är vetenskapliga artiklar förutom Ulfelder (2001) som dock bara har använts i inledningen av arbetet. I och med att detta är ett relativt nytt område har det varit svårt att hitta artiklar relevanta för det här arbetet.

Lite skeptisk till resultatet är jag dock eftersom det antagligen finns ytterligare material som inte använts vilket hade kunnat ge en större inblick i metoderna eller ämnet. För att få en högre reliabilitet i ett sådant här arbete krävs en annan metod, vilket i så fall skulle vara att implementera metoderna och genomföra tester utifrån detta, men detta kräver en hel del kunskap, tid och inte minst kapital, därför valdes en litteraturstudie.

I och med att hela resultatet bygger på en litteraturstudie så är den använda litteraturen av största betydelse. Vad är det som säger att de artiklar som använts faktiskt är trovärdiga, visst de har gått igenom en granskning i och med att de publicerades, men hur mycket kan jag lita på den granskningen. Förhoppningsvis får jag en viss reliabilitet i mitt resultat i och med att de artiklar som använts har granskats av personer som har kunskap nog att fria eller fälla en artikel inom mitt område. Därför görs antagandet att materialet som använts kan anses vara trovärdigt.

Resultatet på jämförelsen är helt beroende på de kriterier som valts, hade andra kriterier använts hade antagligen resultatet sett annorlunda ut. Men de kriterier som använts är som tidigare nämnts de som oftast förekommer som användningsområden för clickstream-data.



## 7.2 Förslag till fortsatt arbete

Som nämnts innan och som framgår av de referenser som använts så är det här ett relativt nytt ämne vilket innebär att det finns många olika aspekter att utreda.

Ett eventuellt framtida arbete kan vara ett praktiskt test av metoder för analys av clickstream-data. Det skulle vara intressant för att se om det faktiskt går att använda dem på de sätt som möjliga användare vill använda dem. Ett annat arbete skulle vara att göra en jämförelse mellan de kostnader som det innebär att införa ett clickstream-system för att göra användare analyser och den nytta som företaget får ut av det. Det skulle vara intressant för att se om det faktiskt ger så mycket fördelar som vissa artiklar kan påstå. Ett tredje och sista förslag på framtida arbeten är att göra en undersökning om hur företag faktiskt använder clickstream-data. Anledningen till att detta skulle vara intressant är för att se hur det används och jämföra det med hur diverse författare anser att det ska användas. Hur stämmer verkligheten ihop med forskarnas idéer?

## 7.3 Utvärdering

Vissa funderingar har funnits angående om det här problemet har tacklats på rätt sätt, finns det något annat sätt att genomföra jämförelsen är kriterierna som använts de mest lämpade. En annan fundering har varit, är verkligen den litteratur som använts den bästa, skulle mer litteratur ha sökts. Problemet är att någon gång måste sökandet av litteratur avslutas för att kunna gå vidare med arbetet.

En erfarenhet som det här arbetet har bidragit med är att alltid läsa texter på ett granskande sätt och inte bara ta det som står för givet. Det är dock lätt att göra genom antagandet att de som skrivit artiklarna vet bättre än en själv.

## Referenslista

Andersen, J., Larsen, R.S., Giversen, A., Pedersen, T.B., Jensen, A.H., Skyt, J. (2000) Analyzing clickstreams using subsessions. *Third ACM International Workshop on Data Warehousing and OLAP (DOLAP 2000)*, 25-32.

Bucklin, R.E., Lattin, J.M., Ansari, A., Gupta, S., Bell, D., Little, J.D.C., Mela, C., Montgomery, A., Steckel, J. (2002) Choice and the Internet: From clickstream to research stream. *Marketing Letters* 13-3, 245-258.

Danielson, D.R. (2002) Web navigation and the behavioral effects of constantly visible site maps. *Interacting with computers* 14 ,601-618.

Facca, F.M. & Lanzi, P.L. (2003) Recent developments in web usage mining research. *Lecture notes in computer science* 2737, 140-150.

Faloni, T. & Hedberg, T. (1995) *Spelet på marknaden*, IHM Förlag AB, Göteborg.

Kimball, R., Merz, R. (2000) *The data webhouse toolkit*, Wiley

Lee, J., Podlaseck, M., Schonberg, E., Hoch, R. (2001) Visualization and analysis of clickstream data of online stores for understanding web merchandising. *Data mining and knowledge discovery* 5, 59-84.

Moe, W. & Fader, P. (2001) *Capturing evolving visit behavior in clickstream data*, Working paper eller Marketing science institute, Cambridge.

Perner, P. & Fiss, G. (2002) Intelligent E-marketing with web mining, personalization, and user-adpated interfaces. *Lecture notes in computer science*, 37-52.

Theusinger, C. & Huber, K (2000) Analyzing the footsteps of your customers - A case study by ASK|net and SAS Institute GmbH. *WEBKDD'2000*.

Ulfelder, S. (2001) Plumb your click-stream data. *Computerworld*, 35, 16-18.

Xing, D. & Shen, J. (2004) Efficient data mining for web navigation patterns. *Information and software technology*, 46, 55-63.

## Bilaga 1

### Dimensions attribut

Beskrivning av dimensioner som kan användas vid dimension modellering. De dimension-strukturer som presenteras är de som Kimball och Merz (2000) rekommenderar.

### Calander Date Dimension

Beskriver vilken dag som ett besök ägde rum.

Date key	(ersättnings värde, 1..N)
Date type	(t ex., regular, unknown, corrupted)
Calander type	(t ex., Gregorian, Japanese, Islamic)
Full SQL date	(får ett värde om date type=regular, null annars)
Day of week	
Day number in week	
Day number in month	
Day number in year	
Day number in epoch	
Workday	
Holiday	
Weekday	
Last day in month	
Week number in year	
Week number in epoch	
Month	(januari, februari, m.m.)
Mon	(jan, feb, m.m.)
Month number in year	
Month number in epoch	
Quarter in year	(4Q1999, 1Q2000, m.m.)
Year	
Fiscal period	
Season	
Event	(speciella händelser t ex, Orkanen Hugo)

### Time of Day Dimension

Beskriver vid vilken tidpunkt ett besök ägde rum.

Time key	(ersättnings värde, 1..N)
Time type	(t ex., regular, unknown, corrupted)
Seconds since midnight	
Minutes since midnight	
SQL time stamp	
Hour	
Minute	
Second	
Time span	(t ex., lunch)

### **Customer Dimension**

Beskriver en besökare eller kund.

*Grupp 1 fält finns alltid vid en webbplats träff:*

Customer key	(ersättnings värde, 1..N)
Customer type	(t ex., regular, unknown, corrupted, eller olika typer av IP adresser eller cookies)
ISP address	(kan vara flera värden eftersom kunden kan koppla upp sig från flera olika ställen)
Cookie ID	
Last change datestamp	
Last change reason	

*Grupp 2 fält om vi har ett användbart kundnamn och har gett denna ett kundid:*

Customer ID	(ID utgett av webbplats ägaren om denne vet något mer än cookie ID:n)
Identifier	(gärna ett helt namn och riktigt namn)
Name type	(pseudonym, unverified real name, verified real name)
Salutation	(Mr., Mrs., Ms.)
First name	
Middle name	
Last name	
Cultural greeting style	(American, European, Latin, m.m.)
Gender	
IP city	
IP state	
IP country	

*Grupp 3 fält om kunden har delat med sig om sin grundadress och demografi:*

Customer type

Degrees

Ethnicity

Company name (Detta fält plus följande fem gäller bara kommersiella kunder, annars null)

Department name

Job title

Primary telephone

Primary Fax

E-mail

Website

Building

Floor

Mail stop

Address type (t ex., headquarters office, parcel delivery adress)

Intended use (t ex., domestic mail, foreign mail)

Street number

Street name

Street direction

Post office box

Locality

City

State

Postal code

Region

Country

Assembled address block (en korrekt sammansatt adress)

Currency (vilken valuta som används på den här platsen)

Marital state

Age range

Profession

Interest

Written greeting

Verbal greeting

Personal individual ID

Commercial entity ID

*Grupp 4 fält om kunden någonsin gjort ett köp:*

Recency	(datum på senast gjorda köp)
Frequency	(antal gjorda köp, livstid)
Intensity	(totalt värde på de gjorda inköpen, livstid)
Projected lifetime value	
Cluster	(ett eller flera fält som rubricerar kundens sammanlagda demografiska samling)
Purchase profile	(ett eller flera fält som beskriver kundens inköps profil)
Credit profile	(ett eller flera fält som beskriver kundens kredit profil)
Return profile	(ett eller flera fält som beskriver kundens benägenhet att returnera varor)
On line support profile	(ett eller flera fält som beskriver kundens användande av on-line support)
Phone support profile	(ett eller flera fält som beskriver kundens användande av telefon support)

**Page Dimension**

Beskriver en sidas innehåll för en webbsidas händelse.

Page key	(ersättnings värde, 1..N)
Page source	(t ex., static, dynamic, unknown)
Page function	(t ex., portal, search, product description)
Page template	(sparse, dense and others)
Item type	(product SKU, book ISBN number, telco rate type)
Graphics type	(t ex., GIF, JPG)
Animation type	(liknar Graphics type)
Sound type	(liknar Graphics type)
Page file name	

**Event Dimension**

Beskriver vad som har hänt på en specifik sida vid ett specifikt tillfälle.

Event key	(ersättnings värde, 1..N)
-----------	---------------------------

Event type	(t ex., Open Page, Refresh Page, Click link)
Event content	(applikations beroende fält vars innehåll kommer från XML taggar)

### **Session Dimension**

Bidrar med en eller flera steg av diagnoser på en användares session som helhet.

Session key	(ersättnings värde, 1..N)
Session type	(t ex., classified, unclassified, corrupted)
Local context	(huvudsakligen sid härlett sammanhang så som Requesting Product Information)
Overall session context	(huvudsakligen spår härlett sammanhang så som Ordering a Product)
Action sequence	(summerande rubrik för en total sekvens av aktiviteter som observerats under sessionen)
Success status	(om sessionens mål nåddes)
Customer status	(rubricerar sånt som High Value Reliable eller In Default)

### **Referral Dimension**

Förklarar hur kunden kom till den aktuella sidan.

Referral key	(ersättnings värde, 1..N)
Referral type	(t ex., intra site, remote site, search engine)
Referring URL	
Referring site	
Referring domain	
Search type	(t ex., simple text match, complex logical match)
Specification	(det faktiska sökord som använts, användbart om simple text, annars tveksamt)
Target	(var sökningen hittas sin matchning, t ex., i meta taggarna eller i texten)

### **Product (eller Service) Dimension**

Beskriver den produkt eller service som ämnet på en sida eller målet för en händelse.

Product key	(ersättnings värde, 1..N)
Product type	(t ex., regular, not measured, corrupted)
SKU code	(streckkoden)
Description	(möjlig unik beskrivning på en individuell produkt med version och realesedatum)
Brand	

Manufacturer	
Category	(återförsäljarens)
Departement	(återförsäljarens)
System type	(component, software, hardware, complete system)
Packaging	(piece, plastic bag, shelf box, floor box, other)
Package width	
Package depth	
Package height	
Package stacking height	
Package weight	
Standard cost	(här endast om det är konstant under en längre period, annars i en annan fakta tabell)
Standard price	(här endast om det är konstant under en längre period, annars i en annan fakta tabell)
Buyer	(här endast om det är konstant under en längre period, annars i en Buyer Dimension)
<i>Alternativt:</i>	
Service key	(ersättnings värde, 1..N)
Service type	(t ex., regular, not measured, corrupted)
Service code	(den interna kod som används av faktura och spårnings system, om synligt för chefer)
Description	(t ex., requirements gathering, end user training, extract system development)
Catagory	(t ex., warehouse development, management consulting)
Sector	(t ex., telco, banking, helth care, manufacturing)

### **Casual Dimension**

Beskriver handelsplatsen tillstånd vid det tillfälle som mätningar görs i fakta tabellen.

Casual key	(ersättnings värde, 1..N)
Casual type	(t ex., specific casual factor, no casual factor)
Price treatment	(t ex., regular, 2-for-1, 10% reduction, \$2 off)
Newspaper ad type	(t ex., large daily, small daily, large weekend)
Web ad type	(t ex., search engine constant, news portal time slice)
Radio ad type	(t ex., 24xdaily news station)
Store display type	(t ex., end aisle display, shelf tag)



Mfgr promo type (t ex., discount coupon, boundled extra product)  
Other casual event (t ex., war, hurricane, competitor discount)

### **Business Entity Dimension**

Beskriver en affärsverksamhets verklighet som är associerat med ett fakta protokoll.

Entity key (ersättnings värde, 1..N)  
Entity type (t ex., commercial, government, individual)  
Entity name  
Industry catagory  
Primary contact name  
Primary telephone  
Primary Fax  
E-mail  
Website  
Building  
Floor  
Mail stop  
Address type (t ex., headquarters office, parcel delivery adress)  
Intended use (t ex., domestic mail, foreign mail)  
Street number  
Street name  
Street direction  
Post office box  
Locality  
City  
State  
Postal code  
Region  
Country  
Assembled address block (en korrekt sammansatt adress)  
Currency (vilken valuta som används på den här platsen)