

MAGISTERUPPSATS I BIBLIOTEKS- OCH INFORMATIONSVETENSKAP  
VID INSTITUTIONEN BIBLIOTEKS- OCH INFORMATIONSVETENSKAP/BIBLIOTEKSHÖGSKOLAN  
2009:76  
ISSN 1654-0247

Folksonomi på webben  
Indexeringskonsistens och konstruktion av taggar  
bland Delicious användare

ANNIKA SAMUELSSON



HÖGSKOLAN I BORÅS  
VETENSKAP FÖR PROFESSION

© **Annika Samuelsson**  
Mångfaldigande och spridande av innehållet i denna uppsats  
– helt eller delvis – är förbjudet utan medgivande.

Svensk titel:	Folksonomi på webben: indexeringskonsistens och konstruktion av taggar bland Delicious användare
Engelsk titel:	Folksonomy on the web: indexing consistency and construction of tags among Delicious' users
Författare:	Annika Samuelsson
Kollegium:	2
Färdigställt:	2009
Handledare:	Katriina Byström
Abstract:	<p>The aim for this master's thesis is to investigate the indexing consistency in a folksonomy Delicious. This is done by investigating how users construct tags. By applying a semiotic perspective and use analysis of content the occurrences of inflected forms, hyponymy, meronymy and users own tag creations were established.</p> <p>Lancaster's seven factors is the starting point for the discussion about indexing and the factors that affects it in either a negative or positive way. The indexing consistency is calculated for the first twenty users of the five objects that are investigated. For the calculation Rolling's method has been used but since it focuses on the indexing consistency between two users, it was adapted for the calculation of the average consistency of the first 20 users.</p> <p>The thesis results show that users prefer to use much the same words which mean that the average consistency for all objects is 28.34 %. Like professional indexers, the users of Delicious prefer to tag the objects with nouns and they often use the singular form. Problems arise when users create their own tags because many of these tags only occur one or two times for every object. This affects the indexing consistency in a negative way.</p>
Nyckelord:	folksonomi, delicious, indexeringskonsistens, indexering, social taggning

# INNEHÅLLSFÖRTECKNING

<b>1 Inledning</b> .....	<b>1</b>
1.1 Problemformulering .....	1
1.2 Syfte och frågeställning .....	2
1.3 Avgränsningar .....	3
1.4 Disposition.....	4
<b>2 Folksonomi</b> .....	<b>5</b>
2.1 Bred och smal folksonomi.....	6
2.2 Problematiken i folksonomier .....	6
2.3 Positiva tendenser i folksonomier .....	7
2.4 Förbättringar som kan göras .....	7
<b>3. Teoretisk anknytning</b> .....	<b>8</b>
3.1 Ämnesindexering .....	8
3.1.1 Problematik inom ämnesindexering .....	9
3.2 Indexeringskonsistens.....	10
3.2.1 Faktorer som påverkar indexeringskonsistensen .....	10
3.2.2 Selektiv och uttömmande indexering .....	11
3.2.3 Bra indexering .....	11
3.2.4 Taggning i en folksonomi.....	12
<b>4 Metod</b> .....	<b>13</b>
4.1 Ett semiotisk perspektiv .....	14
4.1.1 Språk, kultur och mening.....	14
4.1.2 Hyponymi och Meronymi .....	15
4.1.3 Egenkonstruerade ord i Delicious.....	15
4.2 Urval .....	16
4.3 Mätning av indexeringskonsistens .....	17
4.3.1 Rollings mått .....	17
4.4 Innehållsanalys .....	18
4.4.1 Utformning av kodningsschema.....	19
4.5 Felkällor.....	20
<b>5 Delicious</b> .....	<b>21</b>
<b>6 Resultat och analys</b> .....	<b>24</b>
6.1 Objekt A .....	24
6.1.1. Böjningsvarianter .....	24
6.1.2. Hyponymer och Meronymer .....	25
6.1.3 Egenkonstruerade taggar .....	25
6.1.4 Indexeringskonsistens.....	26
6.2 Objekt B.....	28
6.2.1 Böjningsvarianter .....	28
6.2.2 Hyponymer och Meronymer .....	29
6.2.3 Egenkonstruerade taggar .....	29
6.2.4 Indexeringskonsistens.....	30
6.3 Objekt C.....	32
6.3.1. Böjningsvarianter .....	32
6.3.2. Hyponymer och Meronymer .....	33
6.3.3 Egenkonstruerade taggar .....	33
6.3.4 Indexeringskonsistens.....	34

6.4 Objekt D .....	36
6.4.1. Böjningsvarianter .....	36
6.4.2 Hyponymer och Meronymer .....	37
6.4.3 Egenkonstruerade taggar .....	38
6.4.4 Indexeringskonsistens.....	38
6.5 Objekt E.....	40
6.5.1 Böjningsvarianter .....	40
6.5.2 Hyponymer och meronymer.....	41
6.5.3 Egenkonstruerade taggar .....	41
6.5.4 Indexeringskonsistens.....	42
<b>7 Övergripande analys .....</b>	<b>44</b>
7.1 Böjningsvarianter .....	44
7.2 Hyponymi och meronymi .....	45
7.3 Egenkonstruerade taggar .....	46
7.4 Indexeringskonsistens.....	47
7.4.1 Lancasters faktorer .....	47
7.5 Jämförelse mellan objekt D och E.....	48
<b>8 Diskussion.....</b>	<b>50</b>
8.1 Sammanfattning av mitt resultat.....	50
8.2 Indexeringskonsistens ur ett folksonomiperspektiv .....	51
8.3 Metoden ur ett indexeringsperspektiv .....	53
8.4 Vidare studier .....	54
<b>Sammanfattning av uppsatsen .....</b>	<b>55</b>
<b>Källförteckning.....</b>	<b>57</b>
<b>Bilaga .....</b>	<b>60</b>
Objekt A .....	60

# 1 Inledning

Webben är ett ständigt växande medium som dagligen används av människor världen över. Försök har gjorts att göra den enorma mängd av information som finns tillgänglig på webben sökbar. Ett exempel är sökmaskinen Google vars index omfattar mer än en miljard webbsidor (Google 2009). På senare år har webbans användare börjat begagna sig av ett system vid namn folksonomi, vilket innebär en möjlighet att göra material sökbart.

Primärt innebär begreppet folksonomi att elektroniskt publicerat material lokaliseras och klassificeras av webbans användare. Användarna tillämpar vid klassifikationen en fri vokabulär och tillskriver objekt så kallade taggar, vilka är ekvivalenta med ämnesord. Thomas Vander Wal, som myntade detta begrepp, definierar folksonomi som ett resultat av att människor tillskriver objekt taggar, sk. taggning, för personligt bruk i en miljö öppen för andra och med målet att återvinna objekten vid ett senare tillfälle. (Vander Wal 2007)

En av dessa öppna miljöer är webbplatsen Delicious. Där erbjuds användare att publicera och klassificera sina bokmärken samt göra dem tillgängliga för andra användare, så kallad bokmärkeshantering. Delicious är en av de äldsta webbsajter som använder sig av systemet folksonomi och lanserades redan år 2003 av analytikern Joshua Schachter (Surowiecki 2006).

I denna uppsats är webbplatsen Delicious, dess användare och deras konstruerade taggar de tre huvudsakliga komponenterna. Jag vill med denna studie undersöka närmare hur de taggar vilka Delicious användare konstruerar är uppbyggda.

Vidare vill jag uppmärksamma läsaren på att denna uppsats är ett resultat av en delad uppsats mellan Marie Franzén och Annika Samuelsson och att frågeställningen därmed delades. Marie Franzén fokuserade på förekomsten av böjningsvarianter, synonymer och indexeringskonsistensen. Samtidigt har jag koncentrerat min uppsats kring böjningsvarianter, hyponymer, meronymer, egenkonstruerade taggar samt indexeringskonsistensen mellan användare. Följden har blivit att en del kapitel är omarbetade samt att en del kapitel är helt nya. Avgränsningar, Disposition, Delicious, Övergripande analys, Diskussion och Sammanfattning är de kapitel som tillkom efter delningen. Folksonomi, Indexeringskonsistens har genomgått omfattande omarbetning. Inledning, Syfte och frågeställning, Ämnesindexering, Metod samt Resultat och analys har genomgått mindre förändringar. Problemformuleringen är fortfarande densamma då problematiken inte har påverkats av vår delning.

## 1.1 Problemformulering

Även om fler och fler internetsajter använder sig av folksonomisystemet är det än idag ett relativt outforskat område. Fokus har främst lagts på att beskriva folksonomi och att undersöka för- och nackdelar samt skillnader mellan användarnas taggar och en indexerat ämnesord.

Tillämpandet av en fri vokabulär vid taggning av objekt har av flera skribenter, bland annat Noruzis (2007) och Golder och Huberman (2005) vars åsikter jag tagit del av, betraktats som en nackdel. Detta då det bland annat går att tillskriva objekt taggar i alla

olika böjningsvarianter. Enligt kritikerna innebär en fri vokabulär således att två objekt som handlar om böcker kan komma att bli tillskrivna två olika böjningsvarianter på samma ord, bok och böcker. Resultatet blir då att enbart det ena av de båda objekten återvinns vid sökning. Detta minskar möjligheten för användarna att hitta relevanta objekt. Samtidigt finns det inte något i systemet som talar om för användarna hur specifika taggar de bör begagna sig av. Detta innebär att användarna både använder sig av hyponymer, som visar på hierarkiska relationer, och meronymer, som innebär att orden har en 'del av' relation till varandra. Följden blir då att om användarna taggar med olika specifitet kan möjligheten att återvinna relevanta objekt minska ytterligare.

Dessutom försvårar det för återvinningen om användarna väljer att tillskriva objekt, begrepp eller flerdelade namn samt taggar utan information för andra användare, vilket jag ser som egenkonstruerade ord. Är de konstruerade på samma sätt? En större konsekvens av att användarna kan begagna sig av alla olika böjningsvarianter, hyponymer, meronymer och egenkonstruerade ord är att systemet kan komma att innehålla så många taggar att det blir svårt för användarna att navigera i och nyttja folksonomin.

Genererar då den fria vokabulären inom folksonomi verkligen så dålig indexeringskonsistens, vilket är ett mått på hur lika taggar/termer som tillskrivs, mellan de tillskrivna taggarna för ett objekt som kritikerna tycks förutsätta? Någon forskning kring detta har jag inte funnit. Ett antal studier kring folksonomi har bedrivits och användarnas konstruktion av taggar har varit föremål för ett par av dessa. Dock finns endast ett fåtal studier som undersökt hur indexeringskonsistensen, det vill säga överensstämmelsen bland tillskrivna ämnesord, ser ut i en folksonomi genom användandet av något mått som beräknar indexeringskonsistensen. Därför har jag framför allt begagnat mig av den forskning inom området indexeringskonsistens som bedrivits kring de traditionella systemen. Detta då det inte bör föreligga någon skillnad i vilka faktorer som inverkar på indexeringskonsistensen oavsett vilket system som används, folksonomi eller traditionella taxonomier, se 3.2.1.

Då jag inte har stött på någon studie som fokuserat på användarkonstruerade taggar och hur dessa överensstämmer med varandra tycker jag att det är av intresse att utvärdera om kritiken mot folksonomi är befogad eller ej. Även med bakgrund av att folksonomi är ett växande fenomen och att fler sajter använder sig av systemet anser jag det vara av intresse att undersöka hur återvinningsmöjligheterna faktiskt ser ut. Detta genom att fokusera på indexeringskonsistensen mellan användarna.

## 1.2 Syfte och frågeställning

Syftet med denna uppsats är att undersöka indexeringskonsistensen i en folksonomi. Grunden till detta görs genom att närmare studera förekomsten av böjningsvarianter, hyponymer och meronymer samt hur användarna begagnar sig av möjligheten att konstruera egna taggar. Detta kommer att studeras närmare genom följande frågeställningar:

1. Hur konstrueras taggar med avseende på böjningsvarianter och hur vanligt förekommande är böjningsvarianter?
2. Vilken utbredning har närliggande ord i form av hyponymer och meronymer bland användarna?

### 3. I vilken utsträckning begagnar användarna sig av egenkonstruerade taggar?

Dessa tre frågeställningar representerar de tre områden jag kommer undersöka för att kunna få en inblick i hur indexeringskonsistensen ser ut i en folksonomi.

## 1.3 Avgränsningar

Då jag läste in mig på ämnet var det framför allt två sajter som nämndes, Delicious och Flickr. Av dessa två valde jag den som påstod sig vara störst vad gäller antalet användare nämligen Delicious. Valet innebär att det resultat som jag kommit fram till endast är relevant för Delicious, resultatet kan följaktligen inte appliceras på någon annan sajt som använder sig av folksonomi. Även valet av att endast använda mig av listan 'popular bookmarks' när jag valde ut objekt avgränsar studien ytterligare. Resultatet gäller därmed endast på de poster jag valde ut, vilket blev de fem översta i listan. Det innebär också att resultatet på min frågeställning kunde ha blivit ett annat om jag valt att begagna mig av några andra än de fem mest populära bokmärkena i listan.

Vidare har jag inte fokuserat på någon speciell användargrupp. Detta hade varit omöjligt då användarna av Delicious finns över hela världen. Kontentan är att jag inte har tagit någon hänsyn till användaren och vilken expertis som denne besitter. Användarens expertis är dock en viktig del för att förstå varför användaren väljer att tillskriva de taggar till objektet som denne valt. Att försöka undersöka och få fram den informationen hade dock varit omöjligt i denna studie. Samtidigt har jag också valt att enbart nyttja engelska taggar då flertalet av taggarna är skrivna på engelska. Visserligen går det inte att se om det är användarnas modersmål och följden blir att de kanske inte skulle uttrycka sig på samma sätt som på sitt första språk. Möjligen skulle mer detaljerade ord kunna användas eller också finns det, till följd av detta, risk för stavfel vilket leder till en minskad konsistens då folksonomin ser det som en ny tagg.

Anledningen till att jag valt att begränsa mig till fem objekt beror på antalet taggar dessa fem genererade. Jag ansåg att det var viktigare att jag kunde analysera materialet mer ingående än att möjligtvis få ett resultat som har en högre reliabilitet. Vidare är detta också anledningen till att jag begränsade antalet användare till 20 stycken. Dessutom ville jag ha en möjlighet att kunna studera samtliga olika taggar när jag undersökte hur användare konstruerar taggar med avseende på böjningsvarianter, meronymer och hyponymer samt egenkonstruerade ord. Som en konsekvens valde jag därför aktivt objekt med mindre än 500 användare vilket jag ansåg gav mig en bra översikt. Denna del av undersökningen omfattar dock alla olika taggar från de fem olika objekten.

Jag vill understryka att denna uppsats endast kan ge en ögonblicksbild av hur det såg ut på Delicious den nionde juni 2007 då jag samlade in materialet, på de poster som jag valde att begagna mig av. Det går inte att dra några generella slutsatser av vare sig hur det ser ut på Delicious eller i någon av de övriga folksonomierna på nätet. Därför kan jag varken säga att min studie har hög validitet eller reliabilitet. Istället ser jag uppsatsen mer som ett första steg på vägen till att lyfta upp frågan om indexeringskonsistensen och därigenom möjligtvis kunna intressera forskare för denna inriktning inom området folksonomi.

## 1.4 Disposition

Då jag redogör för *tidigare forskning* inom ämnet har jag valt att kalla kapitlet för folksonomi. I detta kapitlet redogör jag också mer ingående för vad systemet folksonomi innebär samt den tidigare forskning som jag använder mig av.

Min teoretiska *anknytning* har jag valt att delar upp i två delar, ämnesindexering och indexeringskonsistens. Detta för att göra det mer lättöverskådligt för läsaren. Jag gör en genomgång av det viktigaste inom dessa ämnen samt det som är relevant för min uppsats.

I *metodkapitlet* beskriver jag hur jag har gått tillväga i undersökningen, var och när jag fått mitt material, urvalsmetod samt vilken beräkningsmetod jag nyttjat och hur jag utformar mina kodningsscheman. Då det semiotiska synsättet och även semantiken ligger till grund för hur beräkningarna har genomförts och hur kodningsscheman har utformats anser jag att dessa är mer relevanta för att tydliggöra min metod än för att förklara den teoretiska anknytningen vilken är indexeringskonsistens.

*Resultat- och analyskapitlet* är uppdelat efter de objekt som undersökningen utgår ifrån. Detta för att tydliggöra för läsaren de resultat som framkommer för de enskilda objekten. För att kunna se fenomen som gäller för alla objekt och tydliggöra dessa använder jag mig av kapitlet *övergripande analys* där jag kommer att sammanställa de generella resultat som jag har sett hos de enskilda objekten. Vidare tar jag upp saker från de enskilda objekten som jag anser vara av intresse för uppsatsen. För att underlätta för läsaren är den *övergripande analysen* uppdelad efter samma mönster som objekten i *resultat- och analys kapitlet*.

I *diskussionskapitlet* börjar jag med att sammanfatta och tydliggöra mitt resultat. Detta genom att återupprepa min frågeställning och redogör under varje frågeställning vad jag kommit fram till. Vidare relaterar jag mina resultat till de resultat som andra kommit fram till tidigare, samt redogöra för egna tankar om hur indexeringskonsistensen skulle kunna påverkas positivt. Slutligen kommer jag att diskutera hur den valda metoden kan ha påverkat resultatet samt ge några förslag till fortsatta studier inom ämnet.

Den *bilaga* som jag har valt att ha med är ett exempel på hur mitt obearbetade material såg ut.



## 2 Folksonomi

En folksonomi är ett socialt nätverk där användare interagerar med systemet och varandra. En folksonomi erbjuder sina användare att spara ett objekt för senare återvinning. Det som krävs är dock att objektet har en URL, Uniform Resource Locator (URL 2009), vilket är en förkortning av Uniform Resource Locator och innebär att objektet har en adress. Förutom detta karaktäriseras folksonomier av att det publicerade materialet beskrivs genom fria ämnesord av folksonomins användare, vilket betyder att användarna kan tillskriva det aktuella objektet specifika ord som de associerar till objektet (Vander Wal 2007). Det är dessa ämnesord eller taggar som kommer att undersökas i uppsatsen. Viktigt att betänka är att det är själva taggningen som skapar folksonomin och att folksonomi är ett system för att kunna organisera upp information.

Den typ av fenomen som begreppet folksonomi beskriver är inte nytt. Indexering genom beskrivande ord har förekommit länge, bland annat inom biblioteksvärlden i form av klassificering och indexering. I jämförelse med den professionella indexeringen är indexeringen gjord av vem som helst, genom att tillskriva taggar, fortfarande relativt nytt. Vander Wal var den första att använda sig av begreppet folksonomi för att beskriva denna typ av system. Det gjorde han 2004 och sedan dess har begreppet vunnit mer och mer mark. (Vander Wal, 2007)

På engelska betyder ordet folk 'vanliga människor' medan ordet nomi eller nomia betyder 'metod' vilket innebär att folksonomi ungefär betyder 'vanliga människors metod' (Taxonomy 2001) (Folk 2001). Då det betänkts att vanliga människor taggar objekten i en folksonomi efter sina egna intressen och behov med hjälp av sina egna metoder och ord blir det tydligt att begreppet folksonomi beskriver systemets syfte. Syftet är att låta vanliga människor utan krav på utbildning inom indexering tagga bokmärken som är av intresse för dem.

En folksonomi har inga fasta relationer mellan orden vilket betyder att den, i motsats till en taxonomi, är icke hierarkisk vilket gör att den är vad som brukar kallas platt. Taggarna har ingen inbördes rangordning utan har samma värde, de kan sägas vara linjära till varandra. Följden är att de nya taggar som tillkommer har samma värde som en gammal tagg. Det gör att uttryck och begrepp kan appliceras relativt snabbt i en folksonomi om det blir prefererat av användarna, vilket innebär att uttrycket eller begreppet är relevant för många av de objekt som sparas. Resultatet blir att förändringar i språket snabbt appliceras på folksonomi utan några extra kostnader, vilket inte är fallet med en taxonomi. Värdet på taggar går dock att mäta genom begagnandet av vad som brukar kallas för 'power curve' och 'long tail'. (Stock 2007)

Genom att begagna en 'power curve' går det att fastställa vilka taggar som prefereras av användarna. En 'power curve' är i grunden en graf som visar hur frekvent taggarna förekommer, vilket gör det möjligt att se olika mönster vad gäller användandet av taggar. Det som kan sägas som en generell regel är att det oftast endast är några få taggar som står för merparten av det totala antalet taggar. Många av taggarna är dock av sådan karaktär att de bara nyttjas ett fåtal gånger, några används bara en gång. Grafen blir då väldigt utdragen åt höger och det ger upphov till en svansliknande figur vilken då kallas för 'long tail' vilket Morville förklarar i sin bok (Morville 2005, s. 51).

## 2.1 Bred och smal folksonomi

Vander Wal (2005) har också definierat två olika typer av folksonomier, så kallad bred och smal folksonomi. Bred folksonomi betyder att flera personer kan tillskriva taggar till samma objekt helt oberoende av hur de tidigare personerna har taggat, vilket medför att samma tagg kan tillskrivas ett obegränsat antal gånger. Alla användare kan alltså använda samma tagg. I denna typ av folksonomi går det att tillämpa 'power curve' och 'long tail'. I motsats till en bred folksonomi kan samma tagg i den smala folksonomin endast tillskrivas en gång. Det innebär att två användare inte kan begagna sig av samma tagg för samma objekt. I denna typ är det alltså omöjligt att försöka nyttja power curve. Följden blir också att taggarna här inte kan få något värde eftersom att de endast kan förekomma en gång för varje objekt. En slutsats som kan dras är att en bred folksonomi får ett annat djup. Samtidigt blir den automatiskt större då alla taggar sparas vilket gör att den också blir svårare att navigera i eftersom mer kopplingar mellan olika taggar görs. (Vander Wal 2005)

## 2.2 Problematiken i folksonomier

Det som orsakar den enskilt största problematiken är fri vokabulär vilket en folksonomi använder sig av. Det går till exempel att tagga med både singular- och pluralformen av ett ord. Om en sökning då görs på singularformen hittas inte de objekt som blivit tillskrivna pluralformen av ordet. Det gör att återvinningen av relevanta objekt minskar.

En av de mest citerade och använda studierna är den av Golder och Huberman (2005). Författarna tar upp några av de problem som berör olika ordklasser och typer av ord så som hyponymer och böjningsvarianter. De diskuterar också betydelsen av specifitet i taggarna, det vill säga hur precisa taggar som användarna tillskrivit objektet. Det betyder att två användare kan ha taggat samma objekt med två olika taggar, exempelvis *cats* och *birma* om objektet handlar om kattrasen Birma. (Golder och Huberman 2005)

De är inte ensamma om att ta upp denna typ av problematik. Noruzis (2007) tittar framför allt på fyra större problem: böjning av substantiv, synonymer, specifitet och ordtypen polysemi. Studien går ut på att visa hur ett kontrollerat vokabulär skulle kunna implementeras i en folksonomi. Detta för att kunna öka återvinningen av relevanta objekt vid en sökning. (Noruzis 2007) Ytterligare två som har fokuserat på denna fråga är Nauman och Khan (2007) vilka koncentrerat sig på ordtypen polysemi och hur problemen kring den typen av ord kan lösas. I sin studie har de försökt lösa problemet genom att använda ett söksätt som tar hänsyn till de tidigare sammanhang där användaren valt att begagna sig av den aktuella taggen och användarens intressekategorier. Genom detta försöker de klargöra vad polysemi avser. (Nauman och Kahn 2007)

Golder och Huberman (2005) diskuterar även betydelsen av olika ord. Om en användare har valt att begagna sig av taggen *lampa* avser denne kanske hela lampan, skärm och fot samt glödlampan medan en annan användare som använt sig av samma tagg endast avser glödlampan (Golder och Huberman 2005). Det är även denna diskussion som det semiotiska perspektivet avser sig.

## 2.3 Positiva tendenser i folksonomier

Det finns även studier som uppmärksammar några positiva aspekter. Bruce (2008) belyser i sin studie problemet med vilka taggar som användare av folksonomier tillskriver ett objekt och vilka ämnesord som en professionell indexerare väljer att tillskriva samma objekt. Han upptäckte att mindre än tio procent av taggarna/termerna överensstämmer vilket ofta leder till att användarna inte hittar de objekt som indexeraren har indexerat. I en folksonomi är det användarna som tillskriver taggarna. Det leder till att objekten blir återfunna eftersom det råder en överensstämmelse mellan användare och de som indexerar då dessa är samma personer. (Bruce 2008)

Golder och Huberman (2005) ser i sin studie en annan aspekt, nämligen att det finns olika mönster i hur användarna väljer att tillskriva taggar. De kunde bland annat påvisa att användarna tenderar att begagna sig av samma taggar för att beskriva samma objekt, vilket motsäger de negativa effekter som fri vokabulär annars sägs medföra. (Golder och Huberman 2005)

## 2.4 Förbättringar som kan göras

Det finns även studier som fokuserar på hur folksonomier skulle kunna förändras och göras bättre. Centralt här är hur det skulle vara möjligt att undvika en del av de problem som finns inom folksonomier, oftast är det problem som är länkade till det fria vokabulär som folksonomi är uppbyggd av.

Två studier har tittat på olika sätt att underlätta sökning och navigering i en folksonomi. Wus (2006) studie är framför allt inriktad mot folksonomier ur ett organisationsperspektiv, då det är av största vikt att systemet är både lätt att navigera och söka i. Författarna anser att så inte är fallet i en vanlig folksonomi och därför skapar de en taxonomi med utgång i de taggar som folksonomin har genererat med hjälp av olika klustertekniker. (Wu 2006)

Den andra studien har som mål att underlätta för sökningar i folksonomibaserade system. Den tar också upp problemen kring bland annat specifitet i taggarna men även användandet av synonymer och liknande. Författarna använder sig också här av kluster. Genom att skapa kluster av existerande taggar och senare undersöka relationerna mellan taggarna inom de enskilda klustren, lyckas de skapa kluster där taggarnas innebörd är relaterade till varandra. Förhoppningen är att det underlättar för sökningar då det på sikt finns en möjlighet att kunna erbjuda användaren en möjlighet att även söka på de andra taggarna som finns i klustret (Specia och Motta 2007). Användaren får alltså en möjlighet att både kunna utöka sökningen till andra relaterade taggar och även specificera sökningen om det återvunna materialet visar sig vara för brett.

Andra typer av förbättringar som föreslås är exempelvis att öka användarnas medvetenhet angående taggarna och taggning och hur dessa påverkar indexeringskonsistensen. Fokus ligger främst på Delicious rekommenderingsverktyg och hur detta skulle kunna förbättras. Författarna använder sig av användarens tidigare taggningar samt det aktuella objektets innehåll för att kunna rekommendera taggar till användaren. Genom att öka användarnas medvetenhet om hur taggarna inverkar på sökningarnas utfall hoppas de i förlängningen att rekommenderingsverktyget ska bidra till en ökad indexeringskonsistens. (Recommending smart tags in a social bookmarking system 2007)

### **3. Teoretisk anknytning**

Några begrepp som behöver presenteras samt relationen mellan dessa är: bokmärk, webbsida och objekt. Ett bokmärke på Delicious är en URL som har blivit sparad och taggad av en eller flera användare. Detta bokmärke guidar de som klickar på det till den webbsida som innehar den aktuella URLen. Det som jag refererar till som objekt i min uppsats är egentligen ett bokmärke. Anledningen till att jag valt att begagna mig av objekt istället för bokmärke är att jag menar att objekt är mer neutralt och att texten flyter bättre än om jag hade använt mig av begreppet bokmärke.

Två andra centrala begrepp som jag vill redogöra för är ämnesindexering och indexeringskonsistens. I följande kapitel fokuserar jag på de viktigaste inom ämnesindexering och problematiken inom den samma. Vidare har jag uteslutande försökt att fokusera på de faktorer som påverkar indexeringskonsistensen. Det som framkommer här kommer jag senare att använda mig av i de olika analyserna som jag kommer att göra.

#### **3.1 Ämnesindexering**

Indexering innebär att ett objekt tillskrivs termer för att representera objektet i en databas. Indexeringen syftar till att de representerade dokumenten ska kunna återvinnas vid sökning i en databas.

F. W. Lancaster skriver i *Indexing and abstracting in theory and practice* att indexeringsprocessen kan delas upp i två steg: konceptuell analys och översättning. Konceptuell analys innebär att en indexerare analyserar ett objekt för att utröna objektets innehåll och vad som bör indexeras. Översättning däremot är processen att finna en term, eller snarare termer, vilka lämpligen representerar innehållet. Lancaster sammanfattar genom att påstå att konceptuell analys är att identifiera ämnena *X*, *Y* och *Z* i ett dokument medan översättning är att tillskriva termer som representerar *X*, *Y* och *Z*. (Lancaster 1998, s. 8-11)

Både konceptuell analys och översättning kräver kännedom om målgruppen för den databas de indexerade objekten ska inkluderas i. Målgruppen påverkar vilka objekt som blir föremål för indexering och vilka aspekter av ett objekts innehåll som bör fokuseras på vid indexeringen. (Lancaster 1998, s. 8-10) En artikel angående alternativa drivmedel för fordon är av intresse såväl för Volvo och deras produktion som för privatpersoner intresserade av ny teknik och klimat.

Den indexering som utförs på Volvos bibliotek kommer troligen att se annorlunda ut än den som utförs på ett folkbibliotek. Poneera att ett folkbibliotek nöjer sig med att konstatera att artikeln rör teknik, fordon och klimat. På Volvos bibliotek däremot lär artikeln indexeras med större exakthet, då artikelns innehåll är av stort intresse för fordonsbranschen och då Volvos databas troligtvis redan rymmer ett antal artiklar och böcker angående alternativa drivmedel. För att relevanta dokument ska kunna återvinnas och tillgodose målgruppens förfrågningar krävs alltså att indexering sker med målgruppen i åtanke.

Indexeringen görs för att representera dokument och för att dokument genom de representerande termerna ska kunna återvinnas. För att återvinningen ska kunna sägas

vara effektiv krävs dock att två böcker om samma ämne indexeras med samma termer. Alltså: om det är möjligt att finna bok *A* med termerna *X*, *Y* och *Z* ska samma termer generera en träff på bok *B*. Återvinning är således intimt förknippat med indexeringskonsistens.

Indexeringskonsistens kan definieras som i vilken utsträckning det råder enighet kring vilka termer som ska tillskrivas ett objekt. Inter-indexeringskonsistens berör i vilken utsträckning samförstånd råder mellan indexerare i valet av termer medan intra-indexeringskonsistens berör i vilken utsträckning en indexerare är konsekvent i valet av termer för ett objekt över tid. (Lancaster 1998, s. 62)

### **3.1.1 Problematik inom ämnesindexering**

Då indexeringsprocessen består av dessa två steg är det av största vikt att klargöra vad som försvårar dessa steg. Vad gäller konceptuell analys är det inte alltid lätt att avgöra vilket ämne som objektet behandlar, ofta är det flera olika ämnen som tas upp. Då är det indexerarens uppgift att bestämma vilket/vilka av dessa ämnen som är av intresse för återvinning. Här kan det uppstå en klyfta mellan indexeraren och användarna, de anser inte att objektet handlar om samma sak. Utifrån användarnas perspektiv har indexeraren identifierat fel ämne/ämnen.

Detta kan bero på två saker. Indexeraren har inte tillräcklig kunskap inom det område som objektet berör och därav väljer fel ämnen att fokusera på. Det kan det också bero på att indexeraren misstolkar objektets text och följderna blir då att indexeraren feltolkar ämnet som objektet behandlar. Vidare påverkar detta också översättningen. Om indexerarens syn på ämnet skiljer sig från användarnas kan det medföra att de objekt som är relevanta för användaren inte blir återfunna eftersom de termer som användaren valt inte är desamma som de som indexeraren tillskrev objektet. Även här kan det bero på att indexeraren inte är tillräckligt insatt i ämnet. Följden kan bli att den term som tillskrivits inte är den term som bäst kan beskriva objektets ämne. Har en misstolkning av objektets ämne skett innebär det att en direkt felaktig term kan komma att tillskrivas objektet, då objektet inte handlar om det som indexeraren tror det handlar om. (Lancaster 1998, s. 85)

Ämnesindexering handlar mycket om hur indexeraren uppfattar och bedömer objektet och dess innehåll. En taxonomi utgörs av de termer som indexeraren tillskrivit men det finns inget som säger att de är de absolut rätta termerna. Vilka termer som tillskrivs beror på vad taxonomin ska användas till. Det medför att ämnesindexering omfattar en professionell bedömning i frågan om vilka termer som ska tillskrivas och vilka termer som inte ska användas. Dock görs bedömningarna med hjälp av olika verktyg som finns att tillgå men det går alltid att ifrågasätta det riktiga i bedömningen. (Morville & Rosenfeld 2006, s 205)

## 3.2 Indexeringskonsistens

I detta kapitel kommer enbart inter-indexeringskonsistens och faktorer som påverkar denna att diskuteras, då det är inter-indexeringskonsistens, och inte intra-indexeringskonsistens, som är föremål för denna uppsats. Intra-indexeringskonsistens innebär att påvisa hur lika samma indexerare indexerar samma dokument över tid. Inter-indexeringskonsistens visar i stället på hur lika olika indexerare indexerar samma dokument.

Inter-indexeringskonsistens kommer härifrån och framåt att refereras till som enbart indexeringskonsistens eller konsistens såvida inget annat anges i texten.

### 3.2.1 Faktorer som påverkar indexeringskonsistensen

Ett flertal studier kring indexeringskonsistens har bedrivits och dessa har visat att en hög konsistens är svåruppnåelig. De studier som har bedrivits har även gett mycket olika resultat. Lancaster hänvisar till en forskningsöversikt genomförd 1965 av R. S. Hooper. Hoopers översikt fann bland annat att indexeringskonsistensen bland sex studier, där samma mått på indexeringskonsistens användes, varierade mellan 24 % och 80 %. (Lancaster 1998, s. 69) Lancaster identifierar sju faktorer som kan tänkas påverka utfallet av indexeringskonsistensen (Lancaster 1998, s. 70-77). De troliga faktorerna är följande:

1. **Antalet tillskrivna termer**

Desto fler termer desto lägre indexeringskonsistens. Detta beror sannolikt på att indexerare generellt sett kan enas om de viktigaste aspekterna av ett objekts innehåll, men därutöver råder oenighet om vilka aspekter som är värda att tillkännages och uttryckas genom termer.

2. **Kontrollerad alternativt fri vokabulär**

Generellt sett genererar en kontrollerad vokabulär högre konsistens än en fri vokabulär, då en kontrollerad vokabulär innehåller färre termer att välja bland och tydliga direktiv om vilka termer som ska användas framför andra.

3. **Vokabulärens omfång och specifitet**

En vokabulär med större omfång tenderar att ha en mer specifik vokabulär. Indexerare tenderar att bli oense när de ska tillskriva termer med specifitet till ett objekt. Medan en indexerare hävdar att ett dokument behandlar telefoni anser en annan att dokumentet behandlar 3G.

4. **Ämnets utmärkande egenskaper och terminologi**

Är ämnet och dess terminologi av en konkret karaktär tenderar konsistensen att bli högre än om ämnet och dess terminologi är av en mer abstrakt karaktär. Det är svårare att hitta precisa termer om ett ämne som inte är precis i sin natur.

5. **Indexeraren**

Då indexerare har samma typ av bakgrund tenderar de att indexerar mer lika än indexerare som har vitt skilda bakgrunder. Även hur länge indexeraren har indexerat spelar in, kunskap fås över tid och om indexeraren har indexerat länge innehar han/hon en viss kunskap om indexering.

## **6. Tillgängliga redskap vid indexeringen**

Om samma verktyg är tillgängliga för indexerarna ökar även indexeringskonsistensen då de har samma utomstående referensram att luta sig tillbaka på.

## **7. Det indexerade objektets omfång**

Ett långt dokument har fler möjliga indexeringstermer att välja mellan än ett kort dokument vilket medför att indexeringskonsistensen tenderar att bli högre i ett kort dokument än i ett långt. Ett långt dokument kan också kräva att fler termer tillskrivs vilket då sänker indexeringskonsistensen (Lancaster 71).

Huruvida en kontrollerad vokabulär endast utgör ett hjälpmedel som höjer indexeringskonsistensen går att diskutera. Lancaster (1998) tar upp en studie som visar på det motsatta. I studien indexerar försökspersonerna först genom att använda termer ur dokumenten när de senare ska översätta dessa termer till de termer som finns i en kontrollerad vokabulär minskar konsistensen avsevärt. Personerna var inte förtrodda med ämnet och inte heller dess terminologi. Följden blev att olika termer ur den kontrollerade vokabulär tillskrevs för samma fria term. (Lancaster 1998, s. 73)

### **3.2.2 Selektiv och uttömmande indexering**

En faktor som bör diskuteras är huruvida indexeraren har valt att begagna sig av selektiv eller uttömmande indexering. Är det en som har använt sig av selektiv indexering och en annan har valt uttömmande indexering får det genomslag i indexeringskonsistensen.

Selektiv indexering innebär att ett mindre antal termer, vanligtvis fem, tillskrivs objektet. De tillskrivna termerna är generella och följderna blir att representationen av objektet blir generell. Uttömmande indexering betyder att fler termer tillskrivs objektet men att dessa har en mer specifik betydelse. (Lancaster 1998, s. 8-9) Således innebär det att om uttömmande indexering används, tillskrivs mer termer vilket enligt Lancasters lista påverkar indexeringskonsistensen negativt då fler termer måste stämma överens för att få samma konsistens som vid selektiv indexering.

### **3.2.3 Bra indexering**

För att en hög konsistens ska kunna uppnås krävs det en bra indexering. Indexering bygger på att den som indexerar ska förstå textens budskap och målgrupp. Det fordras också att indexeraren ska kunna bedöma vad i dokumentet som är av intresse för återvinning och vilka ord som bäst stämmer överens med det. En bedömning rörande vilka ord de tänkta användarna skulle söka med behövs också, det är inte alltid som de bäst beskrivande orden stämmer överens med vad användarna skulle söka på.

Samtidigt är det ingen indexerare som indexerar exakt som en annan. Det beror på att de ser olika aspekter på dokumentet, möjligtvis indexerar de också för olika målgrupper. Vidare fordras det att indexeraren kan uttrycka i ord vad dokumentet handlar om. Oftast tas dessa ord från dokumentet, men de ord som används vid indexering är ofta deskriptiva, vilket innebär att de skrivs som substantiv och då prefereras pluralformen. Det betyder att många av de ord som tillskrivs dokumentet oftast måste skrivas om från formen de hade från början. (Booth 2001, s.15). Indexering är något som kräver en bedömning av den som utför den. Därför är det av största vikt att indexeraren har kunskap om ämnet och även om de användare som kan tänkas vilja återvinna dokumentet.

### **3.2.4 Taggning i en folksonomi**

Då det är många människor som begagnar sig av folksonomi är det troligtvis så att de innehar både olika ämneskunskaper och olika mycket insikt i hur taggar bör tillskrivas ur ett indexeringsperspektiv. Detta bör då leda till att systemet innehåller en del inkonsekvenser gällande konceptuell analys och översättning samt selektiv och uttömmande indexering. Om användaren som tillskriver taggar endast besitter ytlig eller ingen kunskap om det ämne som webbsidan handlar om kan han/hon inte bedöma vad som är värt att tagga. Det leder antagligen till att användaren tillskriver taggar som tas från webbsidan och att användaren inte har någon större förståelse för de ord som blivit tillskrivna. Om sedan en användare med stor förtrogenhet inom ämnet tillskriver taggar blir det antagligen helt andra ord som väljs. Detta leder till att de tillskrivna taggarna blir olika i djup och att kvalitén på hur de olika användarna tillskriver taggar blir väldigt skiftande.



## 4 Metod

Min undersökning genomfördes med material från Delicious, som i dagsläget är en av webbens ledande webbsidor, bland de som begagnar sig av folksonomi och som är inriktad på bokmärkeshantering.

Delicious, vilken är ett exempel på en sajt som använder en bred folksonomi, lämpar sig väl för min undersökning. Den möjlighet en bred folksonomi erbjuder i form av att olika användare kan tagga samma objekt gör det möjligt för mig att kunna undersöka indexeringskonsistensen mellan användare. En smal folksonomi ger inte den möjligheten. Alternativet hade då varit att titta på liknande objekt eftersom det endast går att tillskriva objektet en enskild tagg en gång. Användarna i en smal folksonomi kan inte tagga objektet oberoende av varandra. Att använda en smal folksonomi såg jag därför inte som något alternativ. Samtidigt erbjuder en bred folksonomi möjligheten att kunna kontrollera felkällor på ett bättre sätt, då objektet är exakt detsamma för alla och att alla användare som sparar ner den kan tillskriva egna taggar oberoende av vilka taggar som redan har blivit tillskrivna. (Vander Wal 2005)

I den här studien undersöks, som konstaterat, klassificeringen av fem objekt. Dessa objekt var de fem första i listan 'popular bookmarks' den nionde juni 2007. För att kunna undersöka hur väl de taggar som användarna konstruerar överensstämmer med varandra tvingades jag, som en konsekvens av utrymmet för den här studien, att begränsa antalet användare och deras tillskrivna taggar. Jag har då valt att undersöka de tjugo senaste användarnas taggar för objekten.

Material för att besvara samtliga frågor i min frågeställning har hämtats från de olika taggar som användare konstruerat för att beskriva objekten. Med olika taggar menar jag att jag endast ser till de olika ord som användarna har nyttjat för att beskriva objektet. När jag pratar om olika taggar tar jag således ingen hänsyn till hur många gånger respektive tagg har begagnats. För att undersöka hur användarna hanterar böjningsvarianter kommer ett kodningsschema konstrueras, där de olika taggarna placeras in beroende på vilken ordklass och böjningsvariant de står i. Vad gäller meronymer och hyponymer konstruerade jag även här ett kodningsschema, se 4.4.1. De egenkonstruerade ord som fanns placerades i en lista som jag undersökte närmare genom att titta på hur frekvent de används och hur de påverkar indexeringskonsistensen. För att studera hur olika användares taggar överstämmer med varandra använde jag mig utav Rollings mått, se 4.3.1.

Jag har också valt att utesluta taggen *system:unfiled* då den innebär att användaren inte har valt att själv tillskriva något. Därför anser jag inte heller att jag kan se den som en tagg och som en följd av det är den inte relevant för undersökningen. Jag har dock i början på varje objekt nämnt hur många användare som har valt denna lösning. Då taggarna, som jag har undersökt, är skrivna på engelska har jag även valt att utesluta singular bestämd form eftersom denna skapas med prefixet 'the'. Vidare har jag också upptäckt att ett fåtal taggar inte passar in i de kodningsscheman som jag använt mig av. Det är exempelvis pronomen och adverb samt ord som jag inte kunnat finna en förklaring på. Dessa har jag därför valt att utesluta ur undersökningen. Dock är de fortfarande med i beräkningen vad gäller indexeringskonsistens. Detta för att ge en bra bild av verkligheten.

## 4.1 Ett semiotisk perspektiv

Det perspektiv på betydelse som jag valt att begagna mig av är det semiotiska perspektivet vilket Goddard även benämner som ”translational”. Innebörden i detta perspektiv är att betydelse inte är en sak och att det inte är värt att diskutera vad betydelse innebär i sig själv. Det semiotiska perspektiven hävdar istället att betydelsen av ett ord beskrivs genom användningen av andra, lättare ord som förklarar ordet ifråga. (Goddard 1998, s 10-11)

Två användare på Delicious som tillskriver en källa taggen *ajax* kan ha skilda åsikter angående betydelsen av *ajax*. *Ajax* kan för en individ innebära fotbollslaget Ajax medan en annan avser webbtekniken. Enligt det semiotiska perspektivet förklarar användarna med andra ord vad de avser, exempelvis fotboll respektive webbteknik eller IT. I min studie blir följden att jag endast kan ta hänsyn till de taggar som användarna har valt att tillskriva och inte vad användarna lägger för betydelse i de enskilda orden. Samtidigt är detta en viktig del av problematiken inom folksonomi och därför kommer en diskussion kring problemet tas upp senare.

### 4.1.1 Språk, kultur och mening

Semantiken syftar till att studera språket och dess betydelse samt hur språk och kultur interagerar med varandra. Språk avser att förmedla information mellan människor. Med språket som verktyg uttrycker och utbyter människor tankar, idéer och åsikter med omgivningen. Ett antal definitioner av betydelse har framförts men jag nöjer mig här med att konstatera att språket bär en betydelse. Språket förmedlar inte enbart betydelse utan speglar och förmedlar även ett lands kultur och värderingar. Detta hävdar Cliff Goddard (1998) i sin bok *Semantic analysis* där han även konstaterar att barn tillägnar sig kulturella värderingar och tankemönster genom språket. (Goddard 1998, s. 1)

Efter att ha konstaterat kulturens inverkan på språket ter det sig begripligt att de skillnader som föreligger mellan språk är en konsekvens av olika kulturer. Goddard anser att skillnaderna mellan språk, med avseende på ords betydelse, är stora (Goddard 1998, s. 2). Ord kan ha skilda betydelser beroende på kultur och inte sällan existerar ord på ett språk utan att det finns någon ekvivalent på ett annat. Detta kan exemplifieras genom eskimåerna på vars språk Franz Boas under tidigt 1900-tal identifierade fyra ord för snö, däribland *aput* som betecknar snö liggande på marken och *gana* som betecknar snö vilken allttjämt faller (Pullum 1991, s. 163). På engelska liksom på svenska krävs mer än ett ord för att beskriva såväl *aput* som *gana*, då ord som är ekvivalenta med *aput* och *gana* inte existerar i någondera av språken. Svenskans *det snöar* är en sammansättning som motsvarar eskimåernas *gana*. Ett flertal sammansättningar kan bildas, med hjälp av roten snö, som motsvarar eskimåernas ord men ekvivalenta ord existerar alltså inte.

Elizabeth Johnston, verksam vid Institutionen för psykologi på Sarah Lawrence College i New York, anser att ett koncept erhåller ett speciellt ord när de figurerar ofta i språket. Johnston exemplifierar detta genom det relativt nya ordet *television* som härstammar från *vision*, dvs syn, och *tele* som avser något som verkar på avstånd (Johnston 1997). Johnstons uppfattning stödjer tanken att språk uppkommer i en kontext och att språk inte enbart förmedlar mening utan även kultur. Att eskimåerna, åtminstone i början av

1900-talet, har fyra ord för snö skulle således vara en produkt av ett frekvent behov av att uttrycka de koncept orden ger uttryck för.

Men går det verkligen att förutsätta att vi avser samma koncept genom användandet av ett specifikt ord? John Lyons, lingvistikern och författaren till *Language and linguistics* anser att ett ords innebörd delvis kan lösas genom att ett ord förknippas med mentala bilder. Ord som *allt* och *kan* är dock svåra att förknippa med mentala bilder och även egennamn, t ex. *Anna* och *Amnesty*, utgör ett problem då de mentala föreställningarna varierar från individ till individ. (Lyons 1981, s. 137) *Anna* och *Amnesty* bär olika mening för olika personer likt åtskilliga andra ord i språket. *Frihet* är ett annat exempel på detta. De mentala föreställningarna kring *frihet* varierar förmodligen från person till person och från en kultur till en annan. Att definiera betydelse som en mental bild är således ingen helt lyckad definition, främst pga de svårigheter det innebär att framkalla en bild av ett ord som *kan*. (Lyons 1981, s. 137) De slutsatser som dock kan dras av ovanstående resonemang är att människor har föreställningar kring ords betydelse och att dessa föreställningar inte alltid överensstämmer.

Jag har i detta kapitel strävat efter att presentera vissa elementära delar av semantiken som är av intresse för denna uppsats. Avsnittet syftar till att visa på den problematik som föreligger kring språkbruk och hur individuella och kulturella skillnader inverkar på förekomsten av ord, dess betydelse och hur ord kan innehålla ett visst koncept.

#### 4.1.2 Hyponymi och Meronymi

Hyponymi och meronymi är två närbesläktade begrepp inom lingvistik. Hyponymi uttrycker vad språkvetare skulle benämna som hierarkiska relationer. Hyponymi uttrycker relationen mellan t ex. *björn* och *isbjörn* där *björn* är den överordnade klassen till *isbjörn*. En generell regel som tillämpas för att utröna om en hierarkisk relation existerar eller inte är att testa om följande utsago gäller eller ej: *isbjörn är en björn* och *isbjörn är en typ av björn*. Undantag finns, t ex en hingst är en häst men knappast en typ av häst. (Cruse 2004, s. 149-150) Meronymi är en relation som uttrycker del och helhet, t ex att (a) *Kronobergs län är en del av* (b) *Sverige* och (c) *Växjö är en del av* såväl *Kronobergs län* som *Sverige*. Det är dock inte självklart att alla delar relaterar till varandra som i exemplet ovan där (c) utgör en del av såväl (b) som (a) (Cruse 2004, s. 150-51). Ett exempel är hus, dörr och dörrhandtag där en dörr utgör en meronym till både hus och dörrhandtag. Dörrhandtaget är däremot inte en direkt del av huset utan mer en del av dörren.

#### 4.1.3 Egenkonstruerade ord i Delicious

För en del objekt krävs det att användaren begagnar sig av flera ord för att uttrycka innehållet i objektet. Några exempel kan vara 'New York', 'information reatrival' eller 'London Bloomsbury'. Problemet är dock att många folksonomier inklusive den som Delicious använder sig av använder blanksteg som en särskiljare mellan de taggar som användarna tillskriver. Det medför att koncept som uttrycks med mer än ett ord i praktiken blir uppdelade i två taggar.

I Delicious förordas det att dessa uttryck och koncept skrivs ihop och att användarna endast begagnar sig av gemener. Exempelvis skrivs då 'New York' istället *newyork*. Det innebär att koncept som 'to read' och 'how to' ger taggarna *toread* och *howto*. Vill däremot användarna uttrycka stadsdelar i London eller två olika koncept som hänger ihop innebär det större problem. Här är användarna inte helt överens om hur dessa ska

skrivs. Det finns exempel på att både understreck, bindestreck och plustecken används. Övriga ord som användarna kan sägas konstruera på egen hand är taggar som endast består av tecken eller ord som har diverse förtecken eller tecken efter sig, exempelvis "ice och cream". Även om dessa ord redan har blivit tillskrivna utan tecken före och efter så gör tecknen att det blir ett annat ord och därför blir det också en helt ny tagg.

## 4.2 Urval

Då Delicious numera har över 5,3 miljoner användare innebär det att ett stort antal användare intresserar sig för och klassificerar samma webbsida (Hood 2008). Konsekvensen blir att en webbsida ofta har över två hundra användare.

Jag har valt att titta på fem bokmärken i Delicious databas för att se hur användarna taggat dessa. Dessa fem bokmärken är hämtade från Delicious lista 'popular bookmarks'. Det innebär att jag inte kommer att titta på alla användare som har taggat utan väljer de tjugo första användarna som tillskrivit taggar från varje bokmärke. Viktigt att veta är att dessa tjugo användare inte är samma för samtliga bokmärken.

En följd blir då att samma användare inte kommer vara med på alla sidor, dock är det ingenting jag har strävat efter då jag vill ha en bred undersökning med avseende på ämnen och användare. Vad gäller semantikdelen, 4.1, av undersökningen har alla taggar från de fem bokmärkena använts, dock använde jag de olika taggarna senare i kodningen som berör böjningsvarianter, hyponymer, meronymer och egenkonstruerade taggar

Med över fem miljoner användare världen över är det svårt för mig att kunna göra någon form av representativt urval eftersom det finns för många faktorer att ta hänsyn till vid ett sannolikhetsurval samt att det blir för tidskrävande för att kunna göras i en magisteruppsats. Det skulle också vara en ekonomisk omöjlighet då sannolikhetsurval även tar mer pengar i anspråk. (Halvorsen 1992, s.100) Mina olika urval har därför skett genom ett så kallat bekvämlighetsurval. Jag har inte haft de resurser som krävs för att göra någon annan typ av urval och jag är också medveten om att det inte kommer att är möjligt att dra några generella slutsatser. Utifrån undersökningen ska jag istället försöka urskilja mönster i de taggar som ligger till grund för undersökningen. Detta beror dels på studiens omfattning och dess urval men även på saker som vilken nationalitet användarna har, ålder och kön samt erfarenhet. Då möjligheten inte har funnits för mig att ta reda på detta gör det att jag inte vet hur representativt urvalet är i förhållande till alla användare. (Bryman 2001, s.114-115, 119-120)

Mitt urval är baserat på Delicious lista över de mest populära bokmärkena för tillfället, vilka hämtades den 9 juni 2007. Jag valde att använda mig av de fem högst rankade bokmärkena. När det gäller undersökningen av förekomsten av böjningsvarianter, hyponymer och meronymer samt egenkonstruerade taggar har jag använt mig av samtliga användare som har tillskrivit bokmärket taggar. Indexeringskonsistensen är beräknad på de 20 senaste användarna för varje sida som har tillskrivit bokmärket taggar. Delicious använder sig inte av något annat rangsystem än att den användare som senast taggade bokmärket hamnar högst upp på listan. Därför ansåg jag inte att jag behövde göra ett slumpmässigt urval från dessa listor.

## 4.3 Mätning av indexeringskonsistens

För att mäta indexeringskonsistensen har jag begagnat mig utav Rollings mått. Taggar som uttrycker samma ska anses dock inte vara ekvivalenta om de inte är identiska i stavning och liknande. Exempel från undersökningen är *ice\_cream* och *icecream* som jag då inte ser som ekvivalenta.

### 4.3.1 Rollings mått

Trots att definitionen av indexeringskonsistens är vedertagen så finns det ingen vedertagen metod att räkna ut den. Som en följd av detta finns det olika beräkningsmodeller. Den vanligaste är Hoopers modell, dock finns det invändningar mot denna (Leininger 2000). Rolling menar att det är viktigt att beräkningen visar om någon eller några av termerna har använts två gånger (Rolling 1981). Hoopers modell tar inte hänsyn till detta vilket är den största orsaken till att jag istället valde Rollings mått även om Hoopers är mer använt. Lancaster (1998) menar att måttet bör visa vilka termer som är mest väsentliga och ta med det i beräkningen för att ge en mer sanningsenlig bild av konsistensen (Lancaster 1998, s.63). Rollings mått tar dock inte med denna aspekt men jag ansåg ändå att Rollings mått fungerar tillfredsställande för min uppsats.

Anledningen till att båda måtten inte använts, vilket skulle göra resultatet mer jämförbart med andra studier, beror på att jag framför allt ville se hur konsistensen i en folksonomi ser ut och om det finns några mönster i taggningen. Vidare menar jag att mitt resultat bara ska ses i förhållande till min uppsats. Rolling (1981) påpekar att konsistensens främsta mål är att visa mönster. Rollings mått ser ut som följer (Rolling 1981, s 70):

$$\text{Genomsnittlig konsistens} = \frac{2C}{(A+B)}$$

C står för ”consistency”, konsistens, mellan två indexerare. Det totala antalet gemensamma termer för de båda indexerarna (C) multipliceras med två, då Rolling menar att det är av vikt att trycka på att termen/termerna förekommer två gånger. I nämnaren räknas det totala antalet termer ut, A och B står för antalet termer respektive indexerare begagnat sig av. Rollings mått beräknar konsistensen mellan två indexerare.

Trots att Rollings mått endast avser att beräkna konsistensen mellan två indexerare valde jag att begagna mig av det, då jag ansåg att det lämpade sig bäst för min undersökning. Följden blev att jag beräknade medelvärdet för alla tänkbara kombinationer av indexerare. För att kunna göra det utformade jag en ny beräkningsmodell vilket innebar att jag beräknade konsistensen mellan varje användare gentemot alla andra användare. Därefter räknade jag ut ett medelvärde på konsistensen med hjälp av de tidigare beräkningarna jag utfört. Till hjälp utformade jag en tabell som den nedan, Tab. 1. Denna beräkningsmodell, som jag konstruerat, innebar att det totalt genomfördes cirka 1 050 beräkningar.

2+1	1	10	0,2
3+1	1	11	0,18
4+1	1	6	0,33
5+1	0	4	0
6+1	1	4	0,5
7+1	1	3	0,67
8+1			
9+1	0	4	0
10+1	0	3	0
11+1	2	8	0,5
12+1	2	6	0,67
13+1	1	7	0,29
14+1	1	6	0,33
15+1	1	5	0,4
16+1	2	6	0,67
17+1	2	10	0,4
18+1	1	7	0,29
19+1	1	6	0,33
20+1	2	15	0,27
21+1	0	4	0
			0,317368

1+2	1	10	0,2
3+2	4	17	0,47
4+2	2	12	0,33
5+2	1	10	0,2
6+2	2	10	0,4
7+2	1	9	0,22
8+2			
9+2	1	10	0,2
10+2	0	9	0
11+2	5	14	0,71
12+2	3	12	0,5
13+2	3	13	0,46
14+2	3	12	0,5
15+2	3	11	0,55
16+2	1	12	0,17
17+2	5	16	0,63
18+2	3	13	0,46
19+2	4	12	0,67
20+2	5	21	0,48
21+2	1	10	0,2
			0,386842

Tab. 1. Här är uträkningen för användare 1s och 2s konsistens på objekt D

Jag har, vilket framgår av tabellen, beräknat indexeringskonsistensen mellan varje parkonstellation. Denna konstellation är synlig i de båda kolumnerna som inleder varje beräkning. Kolumnerna därefter visar hur många taggar som är gemensamma respektive hur många taggar som de båda användarna totalt har tillskrivit objektet. Den sista kolumnen visar vilken konsistens de båda uppnådde mellan sig och längst ner i den kolumnen visas de medelvärden som har uppnåtts för användare 1 respektive användare 2. Den tomma raden, vilken representerar användare 8, visar att användare 8 valt att inte tillskriva några taggar. Därför har jag också använt mig av användare 21 vid beräkningen av konsistensen för objekt D.

Den genomsnittliga konsistensen för varje objekt, som jag använder mig av i resultat och analysdelen, är beräknad på de tjugo användare som jag har beräknat konsistensen för. Genom att beräkna genomsnittet för de användare som jag beräknat konsistensen för har jag fått fram den genomsnittliga konsistensen för varje objekt. Då det i denna uppsats endast är det tjugo första användarna som har nyttjats för beräkandet av indexeringskonsistensen ansåg jag att även objektets genomsnittliga konsistens skulle beräknas med hjälp av dessa tjugo användare.

#### 4.4 Innehållsanalys

Vid bestämningen av olika semantiska relationer, ansåg jag att innehållsanalys var en bra metod. Den gav en möjlighet att undersöka grupper som annars hade varit svåra att få tag på. Det är också en metod som är lätt att göra antingen replikationer eller uppföljningar av, vilket ökar forskningsvärdet i studien. Vid en innehållsanalys är ett kodningsschema ett bra hjälpmedel, det ger en bra överblick av materialet som senare ska analyseras och kan presenteras på ett lättförståeligt sätt. I ett kodningsschema förs all relevant information för ett objekt in. Därefter kodas de olika aspekterna av objektet och de skrivs sedan in i kodningsschemat. (Bryman 2001, s.199, 203) Då en

innehållsanalys görs är den beroende av kvaliteten på den text som den görs på. Jag kan inte garantera att de taggar som blir föremål för undersökningen är av bra kvalitet. Fallet är istället att taggarna säkerligen kommer vara av skiftande kvalitet, men då undersökningen endast koncentreras på olika böjningsvarianter, hyponymer och meronymer kommer inte kvalitén på taggarna att spela en avgörande roll för mitt resultat (Bryman 2001, s 203). Det som kommer att ge avtryck i undersökningen är i stället hur många olika taggar som användarna nyttjar.

#### 4.4.1 Utformning av kodningsschema

Vid utformandet av ett kodningsschema är det viktigt att varje grupp utesluter varandra för att det inte ska uppstå någon osäkerhet om ordens placering vid kodningen (Bryman 2001, s.202). Jag upptäckte omgående att en del ord kan tillhöra olika ordgrupper. Lösningen blev att titta på de andra taggarna som användaren använt sig av för att se vilken ordgrupp som avsågs.

När jag utformade mina kodningsscheman avseende böjningsvarianter utgick jag från de tre ordklasserna: substantiv, adjektiv och verb. Jag undersökte vilka olika böjningsvarianter dessa ordgrupper innehåller och vilka böjningsvarianter som är vanligast i vardagligt tal i det engelska språket. Utifrån det formade jag mina kodningsscheman och fick nio olika kategorier att jobba med. Då majoriteten av taggarna är skrivna på engelska har jag tagit bort substantiv bestämd form. Anledningen är att då dessa kategorier skapas med hjälp av artikeln *the* blir följden att två stycken taggar kommer att skapas. Vid inplaceringen av taggarna i kategorierna som bygger på de ordklasser som finns har jag tillämpat vedertagna regler vad gäller ordklasser och böjningsvarianter. Som hjälpmedel har jag använt mig av Oxford advanced English Dictionary.

Kodningsschemat som visar på användningen av de hierarkiska relationerna mellan de använda taggarna gav mig kategorierna meronymi och hyponymi. Dessa placerades in i mitt kodningsschema i enlighet med vad jag angett i avsnitt 5.1.2. Vad gäller de egenkonstruerade taggarna har jag skapat en lista, vilken anger vilka de aktuella taggarna är och hur frekvent de används. Dessa listor har jag sedan studerat för att kunna se hur användarna nyttjar egenkonstruerade ord och om det finns något mönster i användandet.

En stor utmaning var att kategorisera hyponymer och meronymer. Det räckte inte med att jag hade en yttlig förståelse av orden utan jag tvingades läsa på kring deras innebörd och betydelse i ett brett perspektiv. Ett exempel är *network*. För att kunna placera in hyponymer och meronymer tvingades jag läsa om hela konceptet nätverk vilket är betydligt mer omfattande än hur begreppet används i vardagligt tal. Problem uppstod också vad gäller meronymer och dess relation 'del av'. För att placera ett ord under kategorin meronymer skulle inte förhållandet vara 'kan vara en del av' utan 'är en del av'. Samma problematik föreligger inte för hyponymer då dessa alltid är 'en typ av' någonting.

Under de egenkonstruerade taggarna återfinns taggar av väldigt skiftande karaktär. Det är allt från taggar som är hopskrivna i enlighet med vad Delicious förespråkar till taggar som endast består av enskilda bokstäver eller andra typer av tecken. Som jag tidigare sagt utgick jag från listor, som jag själv skapat, över de aktuella taggarna där även deras frekvens var angiven.

## 4.5 Felkällor

Under arbetets gång har det framkommit en del fakta, vilken tas upp här nedan, som kan komma att påverka mitt resultat. Jag återvänder till denna problematik i diskussionskapitlet.

Då materialet som ligger till grund för uppsatsen baserats på ett bekvämlighetsurval har detta fått genomslag i resultatet. Det innebär att jag kan få en vinklig på mitt resultat som jag inte är medveten om eftersom att jag inte vet vilka som har taggat (Bryman 2001, s. 114-115, 119-120). Vidare har jag inte heller använt mig av en beräkningsmodell som tar hänsyn till hur viktig taggen är. Det innebär att alla taggar är värda lika mycket i min beräkning. (Lancaster 1998, s 63) Ytterligare en felkälla som jag vill ta upp är att jag valt att hoppa över taggen *system:unfiled*. Följden är att användare har blivit uteslutna ur studien då jag har ansett att de egentligen inte har taggat objektet och därför menar jag att jag inte har en tagg att beräkna konsistensen från.

Vidare finns det också några taggar som jag inte har kunnat avgöra i vilken ordklass de hör hemma. Det gäller framför allt taggarna *network* och *networkings* vilka är de som används mest frekvent genom min undersökning. Då de är så pass frekvent använda har jag valt att placera in dem både som substantiv och verb. Detta för att jag anser att det är av vikt att inte utesluta någon möjlighet då användarna begagnar sig mycket av dem. Samma problematik föreligger vad gäller taggarna *blog*, *hack* och *cooking*. Dock har jag här valt att endast placera in dem som verb då jag anser att det är troligare att det är verben som avses.



## 5 Delicious

Delicious, delicious.com, påbörjades redan 2001 av analytikern Joshua Schachter. Han skrev ett program som innebar att han kunde organisera sin egen samling av webbsidor med hjälp av taggar. Detta visade sig så populärt att han 2003 lanserade programmet under namnet Delicious. 2005 såldes Delicious till Yahoo!. (Surowiecki 2006)

Delicious är en webbplats som begagnar sig av den indexeringsteknik som benämns folksonomi, där tyngdpunkten ligger på det sociala nätverk som den här typen av system bygger på (Delicious 2009). Webbplatsen är inriktad mot vad som brukar kallas bokmärkeshantering. Detta ger användarna en möjlighet att lättare kunna återvinna tidigare hittade objekt. För att kunna spara ner och tagga ett objekt krävs att det aktuella objektet har en URL. Det är denna väg till objektet som sparas och inte själva objektet. Det går att dra vissa paralleller till hur datafunktionen 'favoriter' fungerar. Fördelen här är dock att de bokmärken som sparas genom Delicious inte är knutna till en specifik dator vilket är fallet vad gäller favoriter (Delicious 2009). På Delicious har varje användare sin egen sida där alla objekt som användaren sparar är synliga. Alla de taggar som användaren har begagnat sig av ligger som en lista i ena kanten av dataskärmen. Det är även här som så kallade 'tag bundles', ett sätt för användaren att organisera upp sin taggar, syns om användaren har valt att begagna sig av den funktionen.

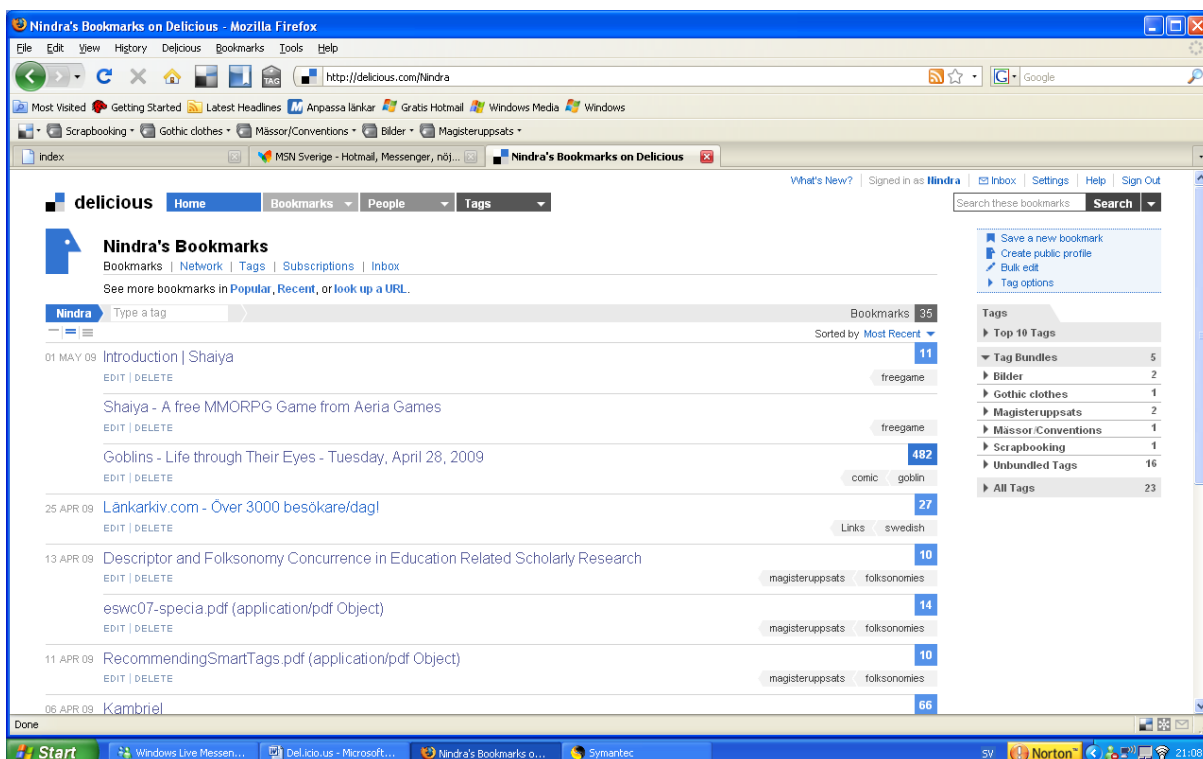


Fig. 1. En användares egen sida. De nedsparde bokmärkena är till vänster och 'tag bundles' är till höger i bilden.

Det finns även en del intressanta funktioner som är till för att underlätta för användarna. Det går till exempel att söka på en tagg och hitta de objekt som är sparade i Delicious och är taggade med den specifika taggen. Samtidigt går det att genom funktionen 'subscriptions' lägga upp taggar som är av intresse. Dessa genererar sedan de senaste objekten som blivit taggade med den/de aktuella taggarna. Dessutom kan användarna

ordna sina egna taggar i så kallade 'tag bundles' vilket gör att en form av ordning kan uppnås. Dessa fungerar genom att användaren kan skapa en form av mapp och bestämmer vilka taggar som ska ingå i den mappen exempelvis *shopping* och *books*. Följden är att alla objekt som användaren taggar med *shopping* eller *books* hamnar under den fliken i användarens egen tagglista.

Dessutom finns det även en funktion som gör det möjligt för användare att skapa egna nätverk, vilket gör det möjligt att se exempelvis vilka objekt som ens vänner senast har sparat ner. För att skapa ett nätverk måste användaren lägga till de användare som han/hon anser som intressanta, till exempel vänner eller en användare med samma intressen. Därefter går det att se vilka de senaste nedsparade objekten är för de användare som ligger inlagda i nätverket. (Delicious 2009)

Det finns också olika typer av listor vilka kan underlätta för användarna. Två exempel på listor är 'popular bookmarks' och 'popular tags'. 'Popular bookmarks' visar vilka objekt som har blivit taggade av flest användare medan 'popular tags' visar vilka taggar som användare begagnar sig av mest.

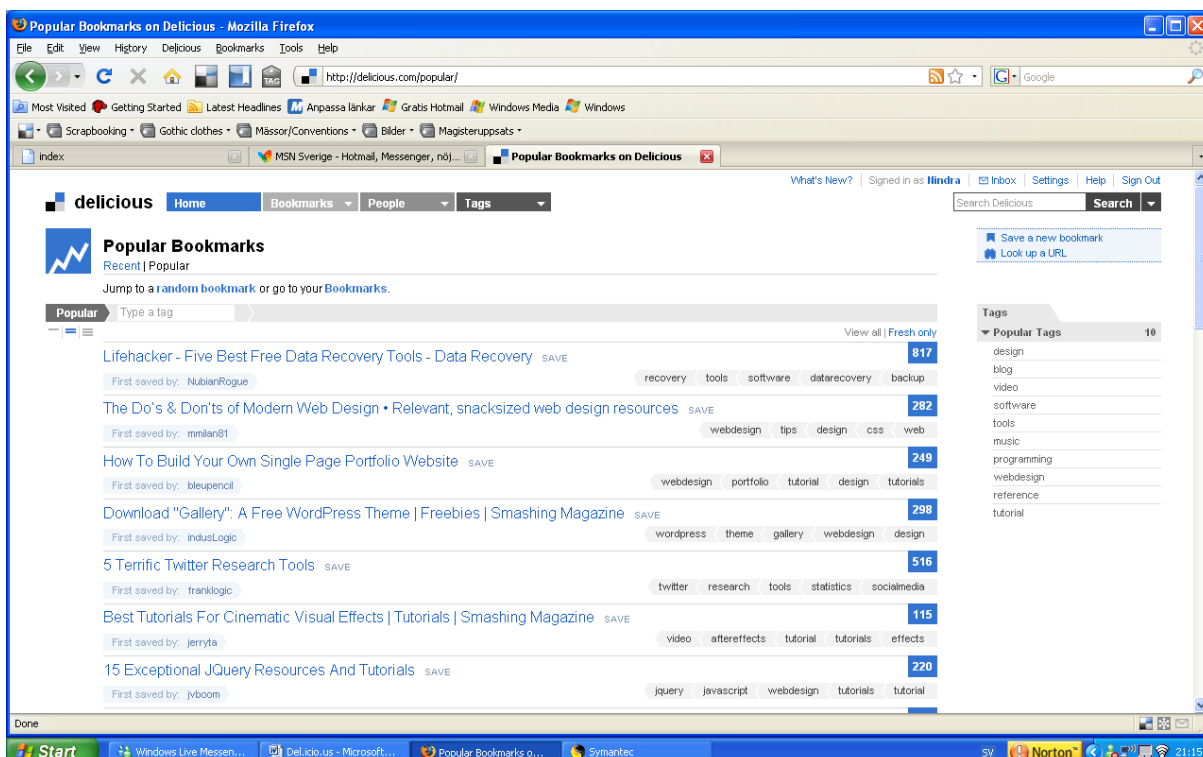


Fig. 2. Listan 'popular bookmarks' är till vänster i bild och listan över 'popular tags' är till höger. Här syns också hur bokmärkena är uppbyggda.

Vad gäller de bokmärken som genereras syns det vad objektet heter, hur många som har sparat ner det, samt vilka de fem vanligaste taggarna för objektet är. Genom att klicka på antalet personer som taggat objektet går det att se vilka taggar de enskilda användarna har begagnat sig av. Vidare går det också att se vilka objekt som blivit taggade med vilken tagg. Genom att klicka på någon av de fem vanligaste taggarna, för ett objekt, genereras en lista över de objekt som blivit taggade med den aktuella taggen.

Vad gäller hur taggar tillskrivs finns det även här olika funktioner som är till för att hjälpa användaren. Exempelvis visas det hur taggarna bör skrivas och det ges en möjlighet till användaren att se vilka taggar som denne redan har begagnat sig av. På det sättet går det att välja en tagg som redan har använts. Detta minskar risken för att objekt som handlar om samma sak får vitt skilda taggar. Samtidigt finns det även en funktion som kallas för 'popular tags'. Det innebär att användaren får upp förslag på tidigare använda taggar som passar till objektet. Dessa taggar i sin tur genereras från de mest frekventa taggarna för objektet.

## 6 Resultat och analys

Det resultat som jag redovisar i följande kapitel visar endast på en ögonblicksbild som gäller för de objekt som ligger till grund för undersökningen. Då analysen grundar sig på mitt resultat kommer också denna att endast kunna visa på en ögonblicksbild. De mönster som jag kan se behöver alltså inte vara gällande vid läsandets stund.

Resultatet presenteras bland annat att genom användandet av diagram för att ge läsaren en överblick av indexeringskonsistensen. Då jag exemplifierar med existerande taggar har jag skrivit dessa i kursivt för att tydliggöra att det rör sig om en tagg från objektet. Vad gäller böjningsvarianter, meronymer och hyponymer och sammansatta ord redovisar jag resultat efter att ha tittat på samtliga användare som taggat objektet, dock är indexeringskonsistensen beräknad på de tjugo första användarna.

### 6.1 Objekt A

Objekt A är en artikel från webbplatsen Atomic PC. Artikeln behandlar videokodeks och har den passande titeln *Everything you ever wanted to know about video codecs*.

Av Delicious användare har 190 personer valt att spara artikeln som ett bokmärke. 19 användare har valt att inte tillskriva artikeln några som helst taggar vilket genererat taggen *system:unfiled*. Övriga användare, 171 stycken, har genererat 639 taggar av vilka 64 taggar var olika. 64 olika taggar av 639 innebar att de olika taggarna enbart utgjorde 10 % av den totala mängden taggar. Detta medförde att variationen av taggar inte var speciellt omfattande.

#### 6.1.1. Böjningsvarianter

Efter att ha studerat de taggar som genererats för artikeln har jag konstaterat att substantiv är vanligt förekommande. Trots möjligheten att sätta substantiv i såväl singular som plural har användarna, vars taggar jag tagit del av, främst valt att placera sina taggar i singular. Fyra undantag står att finna och dessa är *articles*, *tips*, *codecs* och *codecses*. Den sistnämnda taggen, *codecses*, är dessutom en alternativ böjning i plural av *codec* och förekom enbart hos en användare. Nio användare har valt att begagna sig av taggen *articles* men ingen användare har valt singularformen *article*. Taggen *tips* har tillskrivits fem gånger för objekt A och likt *articles* förekommer *tips* enbart i plural och inte i singular.

Väljer jag däremot att titta på *codecs* ser jag att hela 102 användare begagnat sig av taggen medan singularformen endast begagnats av 43 användare. Jag har i mitt material kunnat konstatera att det förekommer fall där användare begagnar sig av såväl plural- som singularformen. Hela 9,9 % av användarna nyttjar såväl *codec* som *codecs* vid tillskrivandet av taggar. Värt att uppmärksamma är att ett flertal av de användare som tillskriver såväl singular som plural har tillskrivit ett större antal termer totalt. Bland andra har jag funnit att användare fyra och fem tillskrivit 14 termer för att beskriva objekt A.

Verb som taggar förekommer sparsamt för objekt A. Bland 64 taggar figurerade enbart åtta verb. Utmärkande för dessa var att de vanligtvis var placerade i infinitiv eller presens. Exempel på detta visas bland annat genom *encode* och *encoding*. *Encoding* förekommer som tagg nio gånger medan infinitivformen *encode* enbart figurerade tre gånger. Något tydligt mönster av vilken böjningsform användarna begagnade sig av

finner jag inte. Jag konstaterade dock att imperfekt enbart förekommer en gång, detta hos användare 19 i form av taggen *wanted*. Användare 19 skilde sig från övriga användare då personen ifråga valt att skriva titelns ord, *Everything you ever wanted to know about video codecs*, som enskilda taggar.

### 6.1.2. Hyponymer och Meronymer

Såväl hyponymer som meronymer var inte helt ovanliga. Särskilt hyponymer var vanligt förekommande och jag exemplifierar här genom *codec*. Bland användarnas taggar för objekt A har jag funnit tre hyponymer till *codec*. De tre hyponymerna till *codec*, *DIVX*, *XVID* och *XDIV* är alltså samtliga en typ av kodek. *DIVX* förekommer fem gånger medan *XVID* och *XDIV* förekommer två respektive en gång.

*Codec*, och dess böjningsvarianter, var en av de mest frekvent använda termerna bland taggarna. *Video* var dock som tidigare konstaterats den vanligast förekommande termen. Bland hyponymer till *video* fann jag de två videoformaten *AVI* och *MPEG*. Likt de olika typerna av kodek förekom de båda videoformaten sparsamt som taggar. *AVI* förekom enbart hos användare 167 medan *MPEG* återfanns hos användare 148 och 169.

Vid skapandet av en MPEG- eller AVI-fil tillämpas komprimering (eng. *compression*). Jag har valt att betrakta *compression* som en meronym till *MPEG* och *AVI*. *Compression* var en tagg som återfinns femtio gånger för objekt A och kan därmed betraktas som en frekvent tagg. En annan tagg som återfanns för objekt A är *encoding* vilken förekommer nio gånger. *Encoding* är en meronym till *compression* men inte till vare sig *MPEG* eller *AVI*. Här är ett exempel på en relation där samtliga delar inte relaterar till varandra likt exemplet kring meronymer i kapitel 5.1.2.

Det jag har kunnat konstatera genom att titta på meronymer och hyponymer är att båda ordklasserna förekommer. *DIVX* och *XDIV* är två termer vilka hierarkiskt är överordnade av *codec* medan *encoding* ingår som en del av *compression*. Hur nyttjandet av meronymer och hyponymer, samt övriga ordklasser, påverkar överensstämmelsen mellan taggar kommer att diskuteras och tydliggöras i min diskussion, se kapitel 8.

### 6.1.3 Egenkonstruerade taggar

För objekt A fanns det 13 egenkonstruerade taggar vilket utgjorde hela 20 % av de olika taggarna. Av dessa 13 taggar var det två som förekommer mer än en gång, *howto* och *toread*, vilka användarna begagnat sig av sju respektive åtta gånger. Dock fanns det ytterligare tre taggar, *readlater*, *to.read* och *to-read*, som med fördel skulle kunnat taggas med *toread* för att öka indexeringskonsistensen. Denna typ av tagg är framför allt till för användaren som nyttjat den. Ingen annan än användaren kan få någon behållning av att veta att objektet ska läsas. Vidare har användare 22 begagnat sig av taggen *videocodecs* medan användare 137 har valt att tillskriva taggen *video\_codecs*. Båda taggarna hade istället kunna delas upp i två taggar, *video* och *codecs*, då dessa förekom frekvent genom hela objektet. Det hade avsevärt ökat dessa användares indexeringskonsistens.

Sex stycken användare har valt att konstruera taggar i enlighet med vad Delicious rekommenderar, hopskrivna utan några andra tecken mellan orden. Om alla användare nyttjat detta sätt hade minst fyra egenkonstruerade taggar ändrats. Likväl är det inte många användare som har utnyttjat möjligheten att skapa taggar genom användandet av skiljetecken mellan orden, endast fem stycken har gjort detta. Däremot har användare

92 valt den något udda taggen ????. Här går det att konstatera att användaren enbart har taggat för sin egen skull. Taggen innehåller inget som kan ge en mening till någon annan användare.

Det går inte att skönja om användare prefererar en vis typ av egenkonstruerade taggar. Taggarna är jämt fördelade mellan hopskrivna ord och taggar med skiljetecken. Även förhållandet mellan taggar för eget bruk och taggar som också är till för andra är jämt fördelade. Dock tycks det som att användarna är relativt överens om hur tillskrivandet av taggarna *howto* och *toread* ska göras. Vissa avvikelser finns vad gäller *toread* men användarna prefererar ändå den hopskrivna taggen.

#### 6.1.4 Indexeringskonsistens

Användarna har huvudsakligen enats kring tre termer de anser uttrycker innehållet för objektet bäst. Dessa tre termer är *video* och *codec* samt pluralformen *codecs*. Den genomsnittliga konsistensen för objekt A var 25 %. Diagrammet nedan återger den genomsnittliga konsistensen för varje användare.

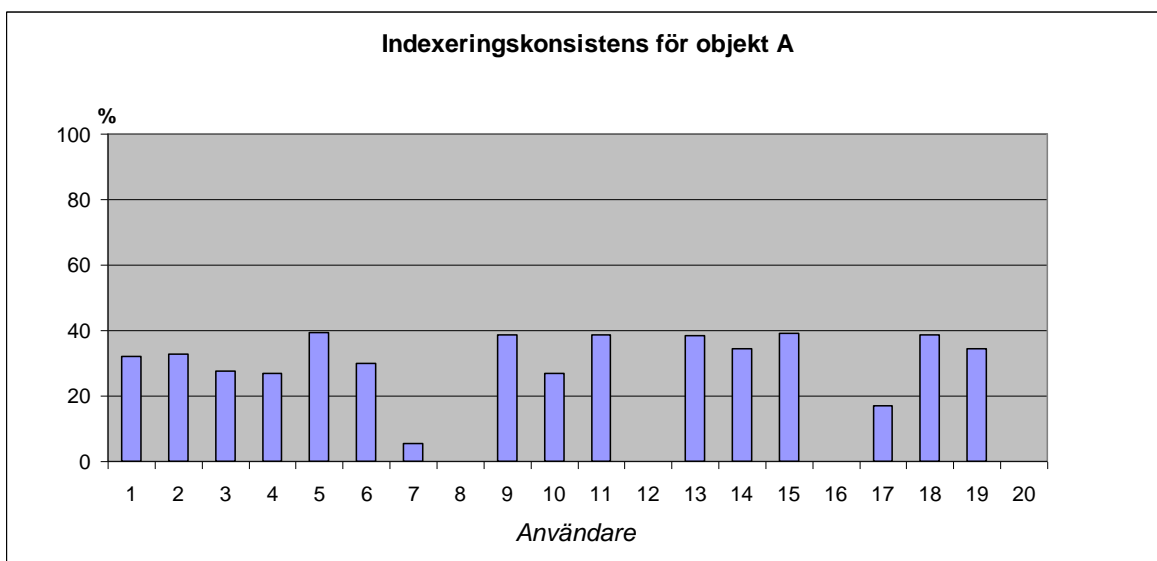


Diagram 1: Diagrammet återger den genomsnittliga konsistensen för användare 1-20<sup>1</sup>

Min insamlade data visar att användare i stor utsträckning valt att tillskriva tämligen olika taggar utöver de tre frekvent använda taggarna *codec*, *codecs* och *video*. Av de tjugo användare vars taggar indexeringskonsistensen har beräknats på, har sexton stycken valt att begagna sig av taggen *video*, vilket innebär att 80 % begagnat sig av den termen. Även när frekvensen av samma tagg beräknas på det totala antalet användare råder samstämmighet. Hela 81,3 % av användarna är överens om att termen *video* bör användas för att beskriva innehållet för objekt A.

En hög indexeringskonsistens innebär att användarna är överens kring vilka termer som ska användas som taggar. Ju vanligare en tagg är bland användare desto högre blir även den genomsnittliga indexeringskonsistensen för de användare vilka begagnar sig av den

<sup>1</sup> X-axeln visar användarnas nummer medan Y-axeln visar indexeringskonsistensen i procent. Av diagrammet går således att utläsa att användare 1 ligger på en konsistens runt 30 %. Resterande diagram avläses på samma sätt.

specifika termen. Användare 1 har uppnått en konsistens på över 30 %, se diagram 1, genom att enbart tillskriva termen *video*. Som tidigare konstaterats är även termerna *codec* och *codecs* vanligt förekommande och de används av sju respektive tio användare. I analysen kan jag konstatera att de tre främst använda termerna självfallet ger utslag på konsistensen. De flesta användare begagnar sig av en eller flera av termerna *codec*, *codecs* och *video*. Användarnas preferens av dessa tre termer genererar dels, som konstaterats, en högre konsistens för det indexerade objektet och dels en jämnare konsistens mellan användare. Av diagrammet framgår att 14 av tjugo användare har en konsistens mellan 25 och 40 %.

Diagrammet avslöjar att fyra användare uppnått ett nollvärde vad gäller konsistens. Gemensamt för dessa fyra användare är att de enbart valt att begagna sig av en term och denna term är för användaren unik bland de tjugo första användarna. Användare åtta har t ex valt att begagna sig av termen *read* vilken är unik och enbart förekommer hos denna användare. Användare 16 har valt taggen *programming* vilken visserligen är unik bland de tjugo första användarna men därutöver förekommer taggen ytterligare en gång.

Användare sju är en användare som genererat en låg konsistens vilket främst beror på att användaren ifråga valt att enbart tillskriva en term. Vad som dock skiljer användare sju från användare åtta och 16 är att användare sju har valt en term som faktiskt används av andra användare. Användare sju utgör därmed ett exempel på en användare som trots valet av en term fortfarande är relativt överens med övriga användare kring objekt As innehåll och vilka termer som bör representera detta innehåll.

Användare nio är en av fyra användare som utmärker sig genom att ha genererat en hög konsistens, se diagram 1. Vad användare nio gjort är att enbart välja två termer vilka dessutom tillhör de mest frekvent använda termerna. En användare som tillskriver dessa två termer kan, trots att övriga tillskrivna termer är unika för användaren ifråga, innebära att en relativt hög konsistens fortfarande kan uppnås. Användare 17 är ett exempel på en användare som tillskrivit nio termer varav samtliga är unika förutom *video* och *codecs*. Tillskrivandet av *video* och *codecs* innebär att en konsistens på 16,9 % trots allt har genereras.

## 6.2 Objekt B

Det taggade objektet är en artikel från tidskriften *Men's Health* vilken är publicerad av tidskriftens online på dess webbsida. I artikeln *Grill the best steak ever* ger den välkände amerikanske kocken Bobby Flay tips och recept för grillen. Flay delger hur en stek bäst tillagas och ger även i artikeln recept på Chimichurri. Chimichurri är en kombinerad dipsås och marinad från Argentina som vanligtvis kombineras med grillat kött.

Antalet användare som taggat objekt B är 86 stycken. Av dessa 86 är det fyra användare som har valt att inte tillskriva objektet taggar vilket innebär att taggen *system:unfiled* har genererats. De 82 användare som har valt att indexera objekt B har sammantaget genererat 288 taggar varav 29 stycken är olika vilket innebär att cirka 10,1 % av det totala antalet taggar för objektet utgjordes av de olika taggarna. Värt att notera är att tio av dessa taggar enbart figurerade hos en användare. Sex taggar förekommer enbart hos två eller tre användare. *Delicious*, *testing* och *advice* är exempel på taggar med låg kvantitet.

### 6.2.1 Böjningsvarianter

Böjningsvarianter förekommer sparsamt bland de användare som har valt att indexera objekt B. När jag studerade förekomsten av singular och plural såg jag att användarna var rörande överens om att singular är den form som bör användas. Bland tjugo substantiv var hela 18 stycken i singularform. De två substantiv som förekommer i plural är *tips* och *recipes*. *Tips* begagnades enbart av två användare medan *recipes* förekommer bland 34 användare.

Även om jag har konstaterat att singular är den form som dominerade bland användarna förekommer ett undantag. Bland de fyra taggar som användarna tillämpade mest frekvent fann jag *recipes*. Termen *recipes* utmärkte sig bland dessa fyra termer genom att vara den enda termen i plural. Vidare förekom såväl *recipe* som *recipes* som taggar för objekt B. Bland samtliga taggar som tillskrivits objekt B är *recipe* och *recipes* de enda taggar som förekommer både i singular- och pluralform. Fördelningen mellan användandet av *recipe* och *recipes* är ojämn. *Recipe* användes enbart sju gånger medan *recipes* förekom hela 32 gånger. Bland användarna för objekt B är det enbart användare 19 som valt att tillskriva både singular- och pluralformen, *recipe* och *recipes*. Användare 19 har begagnat sig av åtta termer för att beskriva objekt B. Åtta termer kan dock inte betraktas som ett stort antal termer då antalet är relativt vanligt förekommande.

Adjektiv är en ordklass som ytterst få användare väljer att begagna sig av. Bland 28 olika taggar är enbart tre adjektiv: *delicious*, *doneness* och *interesting*. De tre adjektiven förekom sju gånger bland sex användare. Användare 13 har valt att begagna sig av två adjektiv då han tillskrivit både *delicious* och *interesting* som taggar. Användare 13 är ett exempel på en användare, vilken valt att göra en subjektiv klassifikation där de egna uppfattningarna kring det indexerade objektets värde fått styra.

Såväl verb som böjningsformer av verb är sparsamt förekommande bland användarna. I de fall böjningsvarianter förekommer är infinitiv och presens vanligt förekommande. Användare 60 är unik i min undersökning då den valt att tillskriva *done* som en tagg.



Användare 60 är också den ende användare som valt att begagna sig av böjningsformen perfekt particip av ett ord.

### 6.2.2 Hyponymer och Meronymer

För objekt B förekommer enbart hyponymer och inga meronymer. De hyponymer som förekommer var få till antalet och flera var hyponymer till samma ord.

*Guide* och *recipe* är två taggar vilka är hyponymer till *advice*. *Guide* och *recipe* förekommer åtta respektive sju gånger medan taggen *advice* förekommer tre gånger. Den mest frekvent förekommande taggen för objekt B var *cooking* vilken förekommer 54 gånger och gav en konsistens på 19 %. De tre taggarna *bbq*, *barbecue* och *grill* utgör samtliga hyponymer till *cooking*. Om de användare, vilka valt att begagna sig av någon av taggarna *bbq*, *barbecue* eller *grill*, istället valt den överordnade termen *cooking* hade indexeringskonsistensen påverkats positivt.

Taggarna *marinade* och *chimichurri* utgör hyponymer till taggen *food*. *Chimichurri* är dock inte enbart en typ av mat utan även en typ av *marinade*. Detta innebär således att taggen *chimichurri* utgör en hyponym till *marinade*. *Marinade* är således hierarkiskt överordnad *chimichurri* men underordnad *food* vilket bäst visualiseras genom följande: *food* > *marinade* > *chimichurri*. De två användarna, vilka valt att begagna sig av taggarna *chimichurri* (användare 9) och *marinade* (användare 64) hade onekligen kunnat välja den överordnade termen *food* för att öka konsistensen. Ett annat alternativ är att användare 9 hade tillskrivit objektet med taggen *marinade* i stället för *chimichurri*.

### 6.2.3 Egenkonstruerade taggar

Endast fem olika egenkonstruerade taggar förekommer i objekt B. De utgjorde dock 18 % av de olika taggarna men endast 10% när jag tittade på det totala antalet taggar för objektet. Två av dessa fem taggar har blivit tillskrivna på det sätt Delicious rekommenderar, *howto* och *menshealth*. Taggen *menshealth* var troligtvis tillskriven som en följd av att objektet kommer från en tidskrift med namnet Men's health. Vidare var det två användare som har begagnat sig av taggar med skiljetecken samt två som kan sägas ha taggat för sin egen del.

Två av taggarna har använts mer än en gång, *howto* och *recepess--dinner*. Anledningen till att den sistnämnda taggen har nyttjats mer än en gång beror, vid en närmare granskning, på att det är samma användare som har taggat objektet två gånger. Vidare hade användaren ur ett konsistensperspektiv vunnit mer på att dela upp taggen i två, då de två enskilda taggarna förekommer i objektet varav *recepess* är en frekvent förekommande tagg. Vad gäller *howto* är det hela 23 användare som begagnat sig av den. Förklaringen kan vara att ett recept visar hur du går tillväga för att göra något. Användarna tycks också vara väldigt överens om hur taggen ska tillskrivas. Det fanns ingen tillskrivning som avviker från den som Delicious rekommenderar, alla har skrivit *howto*.

För objekt B var antalet egenkonstruerade taggar jämt fördelade vad gäller begagnandet av skiljetecken och hopskrivning. Detsamma gäller även vem taggarna var skrivna för. Intressant är dock att det är så många användare som har valt att begagna sig av *howto* och att de var överens om hur det ska tillskrivas.

## 6.2.4 Indexeringskonsistens

Bland användarna för objekt B förekom en förhållandevis hög indexeringskonsistens. Den genomsnittliga konsistensen för objekt B ligger på hela 41,3 %. Ett antal termer dominerade bland de 20 användare vilka indexeringskonsistensen beräknats på. De mest frekvent förekommande taggarna var *cooking*, *steak* och *recipes*, vilka 19 av 20 undersökta användare valt att begagna sig av. Undantaget utgörs av användare 6 som istället har valt att använda taggen *food* för att beskriva objekt B.

Bland de 20 användare, vilken indexeringskonsistensen beräknats på, förekommer taggen *food* nio gånger. När jag däremot tittade på samtliga användare och taggar bland mitt material var *food* den tagg vid sidan av *cooking* som användare prefererade vid tillskrivandet av taggar. Även om *food* inte är lika vanligt förekommande bland de 20 första användarna som bland samtliga användare lyckades användare 6 uppnå en konsistens på 12,3 % genom att tillskriva taggen *food* för objekt B. Konsistensen som användare 6 uppnår var avsevärt lägre än den konsistens vilken genererats för övriga användare. I diagrammet nedan framgår tydligt den skillnad som föreligger mellan användare 6 och övriga användare.

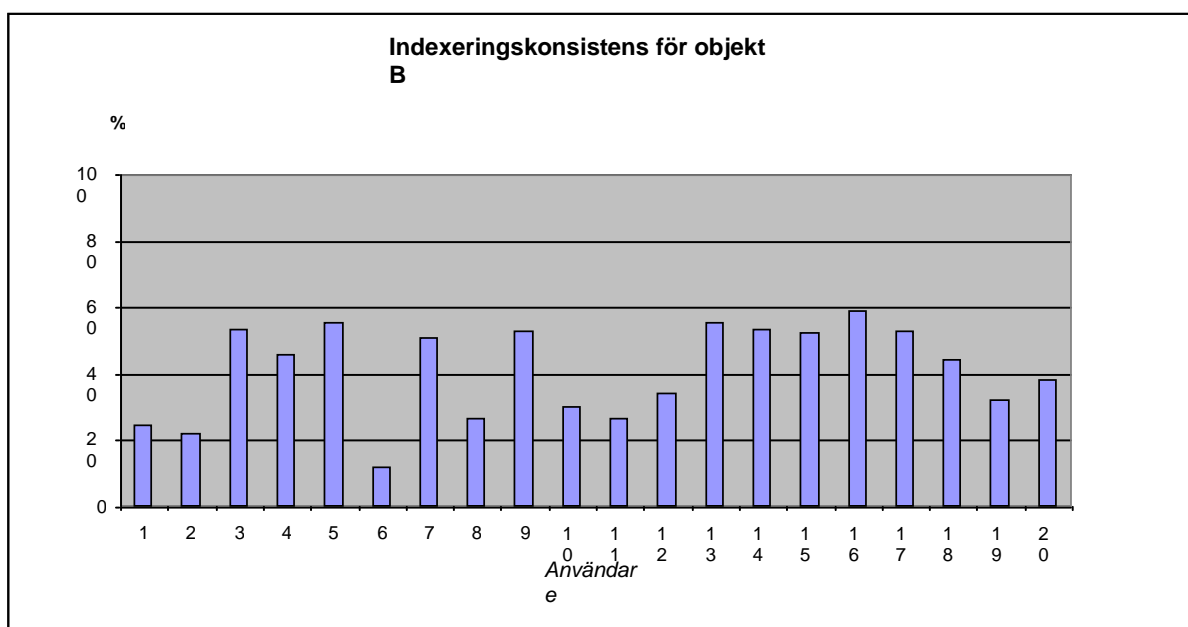


Diagram 2: Diagrammet återger den genomsnittliga konsistensen för användare 1-20.

Förutom användare 6 utmärker sig även användare 16 vilken genererat den högsta konsistensen, 59,1 %, bland användarna. Tittade jag på de taggar som användare 16 valt att tillskriva, fann jag att den höga konsistensen genererats med anledning av följande taggar: *cooking*, *steak*, *food*, *recipes* och *bbq*. Av dessa fem taggar utgör *cooking*, *steak*, *food* och *recipes* de taggar vilka användarna för objekt B prefererade vid taggning. *Bbq* var en förhållandevis populär tagg bland användarna. Taggen förekom tjugo gånger bland samtliga användare och sju gånger bland användarna för vilken indexeringskonsistensen beräknats. Användare 16 har helt enkelt genererat en hög

konsistens då han uppfattat och valt att uttrycka konceptet för objekt B som majoritet av användarna.

Som framgår av diagrammet ligger de flesta användare på en konsistens runt 40 %. Detta beror sannolikt på att fyra taggar, *cooking*, *food*, *steak* och *recipes*, är dominerande bland användarna. *Cooking* är, som tidigare nämnts, den mest frekvent förekommande taggen och utgör 18,8 % av de tillskrivna taggarna. *Food* utgör 16,4 %, *steak* 14,6 % och *recipes* står för 11,8 % av de genererade taggarna. Tillsammans utgör dessa fyra dominerande taggar hela 61,6 % av taggarna bland användarna, vilket kan förklara den förhållandevis höga konsistensen bland användarna.

## 6.3 Objekt C

Det tredje objektet utgör en del av webbplatsen Karboose. Webbplatsen riktar sig främst till mammor och behandlar bland annat föräldraskap, mat och hälsa. Objekt C har titeln *Make ice cream in a plastic bag* och är ett recept på glass. Det framgår av titeln att receptet riktar sig till föräldrar medan produkten främst är tänkt för barnen.

Av Delicious användare har 233 valt att spara artikeln. Av dessa användare har 207 stycken valt att tillskriva objektet taggar. Ett antal användare, 26 stycken eller 11,2 %, har dock valt att inte tillskriva några taggar för objekt C vilket genererat taggen *system:unfiled*. Dessutom har en av användarna tillskrivit taggar på tyska vilket gör att den användaren inte har räknats med. De användare, 206 stycken, vilka valt att tillskriva taggar på engelska för objekt C var dock i klar majoritet. Antalet taggar som genererades var 691 och av dessa var 90 stycken taggar olika. Det betyder att 13 % utgjordes av de olika taggarna.

### 6.3.1. Böjningsvarianter

Jag har för objekt C kunnat konstatera att substantiv var vanligt förekommande och att singular var den form vilken användare främst valde att begagna sig av.

Anmärkningsvärt för objekt C är dock att användarna i större utsträckning än tidigare har valt att begagna sig av substantiv i plural. Av 90 olika taggar och 42 substantiv var 12 stycken, eller 28, 6 %, skrivna i pluralform.

Av de tolv förekommande substantiven i plural förekommer bland andra *articles*, *children*, *kids* och *lifhacks*. Tre taggar var dominerande bland användarna och samtliga var substantiv, men enbart ett substantiv förekommer i pluralform. Nämda substantiv är *recipes* vilket utgjorde 13,7 % av det totala antalet taggar. *Food* var den mest frekvent förekommande taggen och utgjorde 15,2 % medan taggen *icecream* stod för 10,1 % av de tillskrivna taggarna. En annan populär tagg bland användarna för objekt C var *recipe* vilken utgjorde 8,9 % av det totala antalet taggar. Såväl *recipe* som *recipes* är alltså två frekvent förekommande taggar.

Utöver *recipe* och *recipes* återfanns enbart ytterligare en tagg i såväl singular som plural. Två användare har valt att begagna sig av taggen *tutorial* medan fyra användare valt pluralformen *tutorials*. De två böjningsformerna, *tutorial* och *tutorials*, förekommer hos sex användare. När jag däremot tittade på *recipe* och *recipes* såg det annorlunda ut. Av de 59 användare, vilka valt att tillskriva taggen *recipe*, har 21 stycken, eller 35,6 %, även valt att begagna sig av *recipes* vid beskrivandet av objekt C. Användare vilka valt att begagna sig av både *recipe* och *recipes* var bland andra användare 48, 85, 92, 115 och 179. Användare 85 utmärkte sig bland användarna då enbart de båda taggarna *recipe* och *recipes* begagnats för att beskriva objekt C.

Tre användare har valt att begagna sig av ett större antal taggar för att beskriva objekt C. Användare 115 har tillskrivit objekt C 11 taggar medan användare 103 och 150 begagnat sig av 14 respektive 13 taggar. Övriga användare för objekt C begagnar sig vanligtvis av mellan fem och sju taggar. Vidare förekom det sex unika adjektiv vid beskrivandet av objekt C. Intressant är att flera användare valt att begagna sig av de adjektiv som tillskrivits objekt C. Hela 43 användare har valt att tagga med termen *fun* medan 16 användare valt *cool* för att beskriva objekt C. Övriga taggar som förekom var *interesting*, *awesome*, *sweet* och *keytarded*.

Verb förekommer, som för tidigare objekt, sparsamt. Fyra verb återfanns bland användarna och dessa fyra var följande: *blog*, *cooking*, *hack* och *make*. Både *blog* och *hack* förekom tre gånger medan *cooking* var det dominerande verbet och förekom 43 gånger. Som jag tidigare angett ligger det utanför ramen för min uppsats att undersöka, när tveksamhet råder, vilken ordklass användare anser sina taggar tillhöra. Det är dock värt att notera att tveksamhet kan förekomma vad gäller såväl *blog* och *hack* som *cooking* men att jag har valt att placera in dessa taggar som enbart verb i min undersökning.

### 6.3.2. Hyponymer och Meronymer

För objekt C har en mindre mängd hyponymer och meronymer konstaterats. Fyra hyponymer och sex meronymer förekom. Bland hyponymerna förekom två som utmärkte sig genom att inneha en förhållandevis stor kvantitet.

En tagg som jag återfann för objekt C är *crafts* vilken förekommer sju gånger. Till *crafts* har jag identifierat en hyponym, *cooking*. *Cooking* var en av de två hyponymer vilka förekom med en förhållandevis stor kvantitet. *Cooking* har tillskrivits, som tagg, 43 gånger för objekt C. Den andra hyponymen är *icecream*. Då glass är en typ av dessert innebär detta att *icecream* utgör en hyponym till *dessert*. *Icecream* förekom som tagg hela 67 gånger för objekt C. Utöver *cooking* och *icecream* förekom ytterligare två hyponymer för objekt C. *Cream* och *dessert* utgör båda hyponymer till *food* och de förekom sju respektive fyra gånger.

Bland de meronymer som jag hittat förekom *cream* som en meronym till *icecream* och *math* som meronym till *science*. Utöver dessa två meronymer har jag identifierat ytterligare fyra stycken *children*, *kid*, *kids* och *nephews*. Dessa utgör samtliga fyra meronymer till *family*.

### 6.3.3 Egenkonstruerade taggar

Objekt C avvek från tidigare objekt vad gäller begagnandet av egenkonstruerade taggar, hela 36,4 % av de olika taggarna utgjordes av egenkonstruerade taggar, vilket var betydligt mer än vad jag har sett för tidigare objekt. 19 av dessa taggar innehöll någon form av skiljetecken. Av dessa skiljetecken var det några som användarna för de tidigare objekten inte har begagnat sig av, bland annat @ och kommatecken, vilka återfanns i 13 av de tillskrivna taggarna. Två användare, 30 och 154, har valt att begagna sig av taggar som endast innehöll kommatecken. Dessa taggar utgjorde 61,5 % av de 13 taggar som jag tidigare nämnt.

Fyra taggar förekom mer än en gång och den tagg som utmärkte sig är *icecream* vilken har blivit tillskriven 43 gånger. Förklaringen är att objektet rör sig om ett recept på glass. Däremot fanns det tre olika varianter på hur taggen har tillskrivits varav en har nyttjats två gånger. Om de användarna istället hade begagnat sig av den mer frekvent använda taggen *icecream* hade deras konsistens ökat avsevärt. Vidare fanns det 10 taggar som innehöll lite eller ingen innebörd för andra användare. Ett exempel var taggen *thorberg:inlog* vilken inte innehöll någon information om vad objektet handlar om.

Hos användare 180 återfanns taggen *foooooood* vilket var ett exempel på ett frekvent använt ord och som om det hade varit rättstavat gett användaren en relativt hög

konsistens. Ytterligare ett exempel på hur språk och stavning påverkar konsistensen är taggen *baggie* som är slang för *bag*. Taggen *bag* nyttjas här endast av en användare och i detta fall hade ett byte av taggen *baggie* inte haft någon större inverkan. Samma problem föreligger vad gäller taggarna *totry* och *todo*. Användarna hade med fördel endast behövt tillskriva en av dessa taggar då de egentligen uttrycker samma koncept.

Vad gäller detta objekt konstaterade jag att fördelningen mellan hopskrivna taggar och taggar med skiljetecken inte var jämt fördelade vilket framför allt beror på de användare som har använt sig av kommatecken. Vidare var användare även här överens om hur *howto* ska tillskrivas. Även *toread* återfanns endast i en variant.

### 6.3.4 Indexeringskonsistens

Den genomsnittliga indexeringskonsistensen för objekt C uppgick till 14,1 %, vilket jag anser vara en låg konsistens jämfört med tidigare objekt. Tre användare med låg indexeringskonsistens figurerade, vilka sänkte den genomsnittliga konsistensen rejält. Genom att titta på diagrammet nedan såg jag att användare 10, 12 och 20 samtliga har genererat en indexeringskonsistens på 0 %.

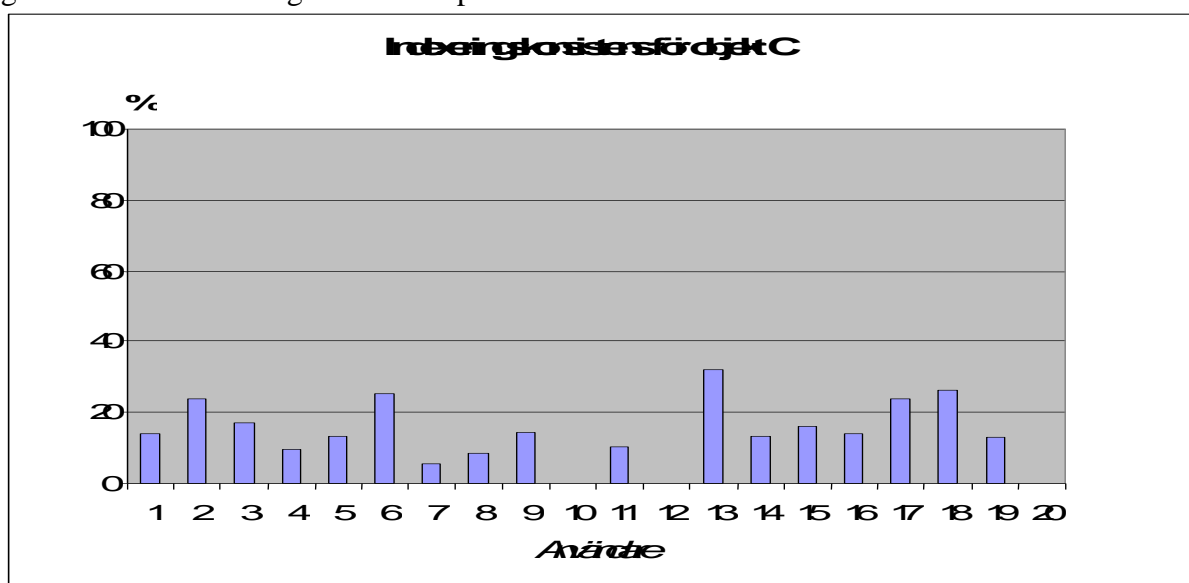


Diagram 3: Diagrammet återger den genomsnittliga konsistensen för användare 1-20.

En indexeringskonsistens på 0 % innebär att ingen av de termer användarna valt överensstämmer med övriga användares. Användare 10 har valt att enbart begagna sig av taggen *schon* vilken alltså inte förekommer hos övriga användare. Detsamma gäller för användare 12 och 20 vilka begagnat sig av *learningbank* respektive *lifhack*. När jag undersökte samtliga 206 användare vilka taggat objekt C kunde jag konstatera att *schon*, *learningbank* och *lifhacks* bara förekom hos användare 10, 12 och 20. Om nämnda användare hade valt att tagga objekt C med mer än en tagg är det troligt att indexeringskonsistensen skulle ha överstigit 0 %. Då fler ord tillskrivs som taggar ökar också sannolikheten att de förekommer hos andra användare. Dock ökar den inte i takt med att fler taggar tillskrivs. Det finns en gräns då indexeringskonsistensen istället börjar minska.

Den användare vilken genererat den högsta konsistensen för objekt C är användare 13. Användare 13 har valt att tillskriva följande taggar: *food*, *icecream*, *recipes* och *cooking*. Dessa fyra termer var även de mest frekvent använda taggarna bland de 20

första användarna. *Recipes* förekom fem gånger, *cooking* och *food* sex gånger var och *icecream* förekom sju gånger. Genom att välja just dessa termer har användare 13 uppnått en genomsnittlig konsistens på 32,4 % jämfört med övriga användare.

När de taggar som samtliga 206 användare genererat för objekt C undersöktes såg jag att de populäraste taggarna överensstämde bra med de populäraste taggarna för de 20 första användarna. *Food*, *recipes*, *icecream* och *DIY* var de fyra mest frekvent förekommande taggarna för samtliga användare. Det som skiljde de 20 första användarna var termen *DIY* vilken intressant nog enbart existerade två gånger bland de 20 första användarna.

En majoritet av de 20 första användarna har genererat en konsistens på mellan 12 och 17 %. Det förekommer dock fem användare vilka har genererat en något högre konsistens. Användare 17 har t ex genererat en konsistens på 23,9 % och de fyra övriga användarna har genererat liknande siffror för indexeringskonsistensen. Vid genomgång av de taggar dessa fem användare tillskrivit objekt C såg jag att användarna har begagnat sig av de mest populära taggarna. Användare 6 har t ex valt att begagna sig av *food*, *recipe* och *icecream* och därmed uppnått en konsistens på 25,5 %. Användare 13 är den som har uppnått den högsta genomsnittliga konsistensen, 32,4 %, vilket jag konstaterat i tidigare stycke.

## 6.4 Objekt D

Det taggade objektet är producerat av bloggen Lifehacker och bär titeln *Lifehacker top 10: Top 10 network utilities – Lifehacker*. Objekt kommer från webbplatsen Lifehacker och tipsar om tio stycken gratis webb- och mjukvaruapplikationer för nätverk.

För att få en förståelse för de termer många användare valt att tillskriva objektet anser jag att en förklaring av termen lifehack kan vara på sin plats. Lifehack är en term myntad av den engelske teknikjournalisten Danny O'Brien för att beskriva de skript många programmerare skriver och nyttjar för att underlätta sin vardag och sitt arbete. Ett lifehack kan t ex innebära en applikation, vilken sorterar bland inkomna meddelande och sorterar bort spam. (Caplan 2007) I fallet med mitt taggade objekt figurerar lifehack i betydelsen av en applikation tänkt att underlätta och optimera användandet av nätverk.

Bland Delicious användare har 197 stycken valt att spara objekt D. 12 av användarna för objekt D har valt att inte tillskriva några taggar vilket genererat taggen *system:unfiled*. Användarna har genererat 776 taggar. Av taggarna är 55 stycken olika och en del av dessa förekommer enbart hos en eller ett fåtal användare. Det innebar att variationen av taggar inte var omfattande, endast 7,1% av taggarna utgjordes av de olika taggarna. Av de 55 olika taggarna förekom 20 stycken enbart hos en användare. Exempel på taggar vilka enbart förekom hos en användare var *projects*, *temp*, *net* och *systemadministration*. Andra taggar med låg kvantitet var *link* och *webapps* vilka förekom hos två respektive tre användare.

### 6.4.1. Böjningsvarianter

Substantiv är den typ av ord som var mest frekvent nyttjade. Av de 36 substantiv vilka studerats för objekt D var 26 stycken skrivna i singular. De tio övriga substantiven var följaktligen skrivna i plural. För tidigare objekt har användarna aldrig genererat så många substantiv i pluralform. Fyra av de tio substantiv som förekom i plural förekom dessutom i singular.

De fyra substantiven var följande: *computer/computers*, *list/lists*, *network/networks* och *tool/tools*. Taggen *network* förekom hela 126 gånger medan pluralformen *networks* enbart förekom fyra gånger. En stor skillnad förelåg även mellan valet av *tool* och *tools*. Taggen *tool* har tillskrivits av sju användare medan 98 stycken har valt att begagna sig av pluralformen *tools*. Taggen *tools* utgjorde 12,8 % av samtliga tillskrivna taggarna för objekt D, vilket innebar att *tools* var en av de vanligaste förekommande taggarna. Den mest frekvent förekommande taggen var *network* vilken utgjorde 16,5 % av den totala mängden taggar.

Likt för tidigare objekt förekom det användare vilka valt att tagga med såväl singular- som pluralform av ett ord. Ett exempel för objekt D var användare 20 vilken valt att tillskriva både *list* och *lists*. Även användare 35 har valt att tillskriva både *list* och *lists* samt även både *tool* och *tools*. Exempel på andra användare vilka valt att begagna sig av såväl singular- som pluralform av en tagg var användare 29, 51 och 91.

För objekt C konstaterade jag att de användare vilka valt att tillskriva såväl singular- som pluralform av ett ord generellt sett begagnade sig av ett större antal taggar. För objekt D gick det att se ett samband mellan ett större antal taggar och begagnandet av



singular- och pluralform av en tagg. Ett exempel på en användare vilken valt att begagna sig av ett större antal taggar är användare 35 vilken valt att tillskriva 14 taggar.

Vid en genomgång av förekomsten av adjektiv återfanns enbart två stycken. Tre användare har valt att använda sig av *free* medan en användare, användare 76, valt att tagga objekt D med adjektivet *temp*. De tre användare vilka använt *free* vid taggningen är användare 79, 81 och 84. Det är intressant att notera att begagnandet av *free* var koncentrerat till tre användare, vilka ligger nära varandra i den lista som har genererats för objekt D. Med tanke på att *free* enbart förekom tre gånger och begagnas av tre användare är det möjligt att användare 81 och 84 påverkats av tidigare taggar och därav valt att begagna sig av *free*.

Bland taggarna för objekt D förekommer enbart ett verb. Verbet *network*, att nätverka, samt dess böjningsvariant *networking*. Bland de 126 användare vilka valt att begagna sig av *network* kunde jag dock inte avgöra huruvida det var verbet eller substantivet *network* som avsågs. Samma problematik förelåg kring användandet av *networking*. Används *networking* i bemärkelsen ”att nätverka” eller avser *networking* datorrelaterade frågor? Detta är en fråga som låg utanför ramen för min uppsats och som jag därmed inte kan besvara. Dock har jag här valt att placera in *network* både som ett substantiv och ett verb. Detta för att taggen är en av de mest frekvent använda för objektet.

#### 6.4.2 Hyponymer och Meronymer

Det förekommer ett relativt stort antal meronymer och hyponymer för objekt D. Främst var hyponymer vanligt förekommande. Meronymerna däremot var enbart tre till antalet och dessa redogörs för nedan.. Den första meronymen var *Mac* vilken är en meronym till *Apple*. *Mac* är en förkortning för *Macintosh* och företaget *Macintosh* ägs av *Apple* vilket innebär att *Mac/Macintosh* är en del av *Apple*.

Då datorer som bekant ingår i ett nätverk utgör termen *computer* en meronym till såväl *wifi* som *network*. *WI-FI* står för ”wireless fidelity” och är en term som används för att benämna så kallade 802.11-nätverk. Tekniken i denna typ av nätverk grundar sig på en standard vid namn 802.11 som är utvecklad för trådlösa nätverk av IEEE, Institute of Electrical and Electronic Engineers (*WI-FI Alliance* 2007). Då *WI-FI* används som benämning för 802.11 nätverk innebär det att *WI-FI* kan sägas vara en typ av nätverk och därmed utgjorde taggen *wifi* en hyponym till *network*.

Två andra hyponymer som jag hittat var *link*, vilken utgör en hyponym till *reference* och *cybersecurity* som i sin tur utgör en hyponym till *security*. Jag skrev tidigare, se föregående stycke, att *Macintosh* är ett företag ägt av *Apple*. *Macintosh*, eller *Mac*, är dock även en dator, vilket innebär att taggen *Mac* även kan betraktas som en hyponym till *computer*. Jag har även valt att betrakta datorn som en typ av verktyg med följden att *computer* är en hyponym till *tool*.

En annan hyponym till *tool* är *software*. Till *software* i sin tur förekommer den största mängd av hyponymer jag sett för en tagg. Jag har identifierat fem stycken hyponymer till *software*. *Freeware*, *webbapplications* och *lifehacks* är alla en typ av program likt de två operativsystemen *OSX* och *Ubuntu*. Dessa fem listade taggar är dessutom samtliga en typ av verktyg och utgör därmed även hyponymer till *tool*.

### 6.4.3 Egenkonstruerade taggar

Även för objekt D var begagnandet av egenkonstruerade taggar högt. 35,8 % av de olika taggarna var egenkonstruerade, det vill säga 19 av 53 taggar. Av dessa 19 nyttjades fyra stycken mer än en gång men ingen av dessa taggar användes mer än fem gånger vardera. Även här prefererade användarna taggarna *toread* och *todo*. Dock var användarna inte lika om hur taggarna skulle tillskrivas. Både användare 126 och användare 134 har valt att begagna sig av en annan variant av taggarna nämligen *2read* respektive *2do*. Vidare fanns det en användare som valt att begagna sig av taggen *to\_look\_at*. Denna tagg hade med fördel kunnat skrivas som *toread* för att öka konsistensen för användaren.

Totalt fanns det nio stycken taggar som innehöll andra tecken än bara bokstäver. Av dessa var det tre taggar som innehöll understreck och tre innehöll en tvåa. För detta objekt har jag också hittat nio stycken taggar som har tillskrivits på det sätt som Delicious rekommenderar, hopskrivna. De tre sista taggarna var dels en felstavning av *utilities*, vilken räknas som en egen tagg, dels endast en bokstav, *n* och sist taggen *june2007*. De två sistnämnda taggarna anser jag vara skrivna för användaren som taggat objektet då ingen av de två taggarna gav någon information om objektet. Ytterligare en tagg fanns vilken inte uppbar någon information angående objektet och kan sägas blivit tillskriven för endast en användare.

Utmärkande för det aktuella objektet var att det fanns få taggar som kan sägas vara tillskrivna för endast en användare, jag har endast hittat, tre stycken. Dock var fördelningen mellan hopskrivna taggar och taggar med skiljetecken jämn. En sak av intresse är dock att objekt D var det första objektet där jag sett att användare begagnat sig av taggar som innehåller siffror, då framför allt en tvåa istället för att skriva *to*.

### 6.4.4 Indexeringskonsistens

Den genomsnittliga indexeringskonsistensen för objekt D var 33,9 %. Vid granskning av diagrammet nedan, såg jag att ett flertal användare uppnår en konsistens mellan fyrtio och femtio procent. Enbart fyra användare, användare 5, 8, 9 och 20 har genererat en konsistens lägre än 20 %, vilket kan sägas vara anmärkningsvärt. Detta särskilt med tanke på det fria vokabulär som tillämpas och den specifika terminologi som ämnet innehar.

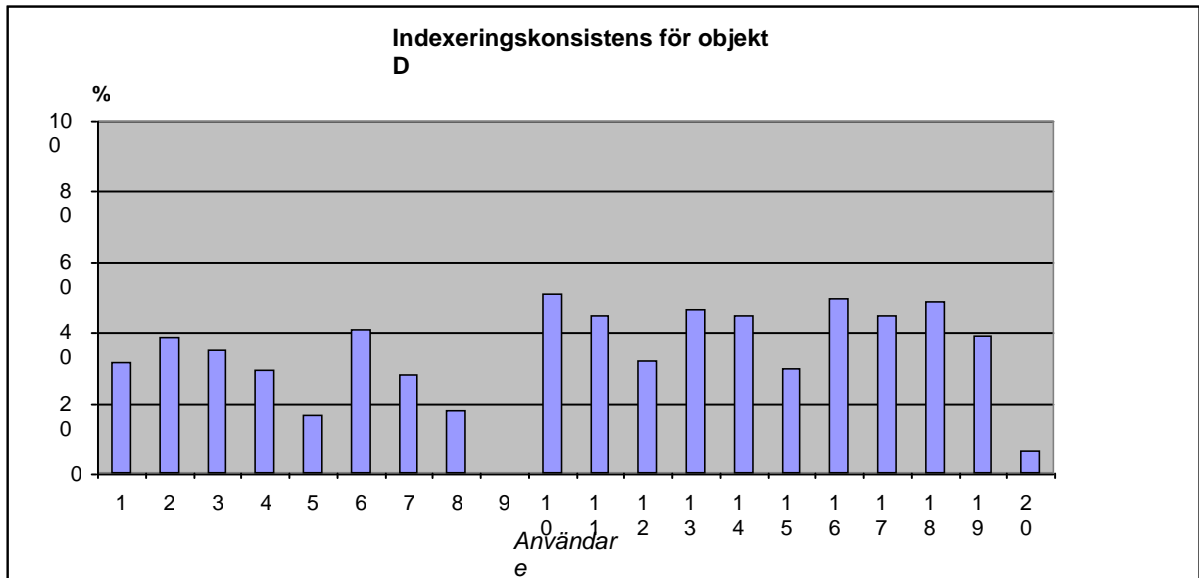


Diagram 4: Diagrammet återger den genomsnittliga konsistensen för användare 1-20.

Jag började med att titta närmare på de nio användare vilka uppnått en konsistens på över 40 %. Användare 10 var den som uppnått störst samstämmighet med övriga användare och därmed genererat en konsistens på 50,9 %. I tät följd kom användare 16 och 18, vilka genererat en konsistens på 49,6 % respektive 49,1 %. Dessa tre användare har uppnått en hög konsistens genom att, bland annat, välja taggarna *network* och *tools*.

*Network* och *tools* var de taggar de 20 första användarna främst prefererade vid tillskrivandet av taggar. Så många som 15 användare har valt att begagna sig av *network* och 13 stycken har valt att tillskriva *tools* som en tagg för objekt D. En användare, användare 6, har valt att enbart tillskriva de båda taggarna *network* och *tools* och uppnått en konsistens på 40,8 %. Användare 7 har valt att tillskriva *network* som enda tagg vilket medfört en konsistens på 28,3 %.

Användare 9 är den enda användare vilken genererat en konsistens på 0 %. Anledningen därtill är att användare 9 enbart valt att tillskriva en tagg, *todo*, vilken ingen av de övriga 19 första användarna valt att använda sig av. Användare 20 är en som genererat en låg konsistens. Med taggarna *software* och *redes\_seguridad* har användare 20 uppnått en konsistens på 6,8 %. *Redes\_seguridad* förekommer enbart som tagg hos användare 20. Däremot begagnas *software* som tagg av fem användare, utöver användare 20, vilket medför att användare 20 trots allt genererar en konsistens på 6,8 %.

## 6.5 Objekt E

Det taggade objektet var detsamma som objekt D. Objektet är således producerat av webbplatsen Lifehacker och bär titeln *Lifehacker top 10: Top 10 network utilities – Lifehacker*. Objektet rekommenderar gratis webb- och mjukvaruapplikationer för nätverk. Skillnaden mellan objekt D och E utgörs enbart av den URL som hänvisar till objektet. Det innebär att objekt D och E är samma objekt men att URLen är olika, de har alltså två skilda adresser, vilket har föranlett Delicious att behandla objekten som två olika objekt.

Av Delicious användare har 219 valt att spara objektet. Av dessa användare har 187 stycken valt att tillskriva objektet taggar. Detta innebär att 32 användare, eller 14,6 %, valt att inte tillskriva några taggar vilket genererat taggen *system:unfiled*. De användare som valt att generera taggar för objekt E har tillsammans konstruerat 794 taggar och av dessa är 73, eller 9,2 %, olika. Hela 37 stycken eller 51,4 % av de olika taggarna, det vill säga strax över hälften, förekom enbart hos en användare. Exempel på taggar vilka enbart förekom hos en användare för objekt E är *sort*, *temp*, *top10*, *vital\_apps* och *win*. Det förekom även ett antal taggar med låg kvantitet. Exempel på taggar med låg kvantitet är *applications* och *wifi* vilka förekom fyra respektive tre gånger.

### 6.5.1 Böjningsvarianter

För objekt E konstaterade jag att främst substantiv begagnas av användarna och att substantiv i singularform prefereras. Nio av de 43 existerande substantiven för objekt E förekom i pluralform. Substantiven i pluralform var följande: *applications*, *apps*, *computers*, *downloads*, *resources*, *tools*, *utilities*, *networkings* och *lifehacks*. Fem av de nio substantiven förekom dessutom i singularform. Termen *computer* har använts av fyra användare medan dess pluralform *computers* begagnats av sex användare. Ingen större skillnad förekom heller mellan användandet av *download* och *downloads* som tillskrivits som tagg av sex respektive fem användare.

Skillnaden mellan begagnandet av *networking* och *networkings* är intressant. Enbart en användare har valt att tillskriva *networkings* som tagg medan 76 stycken istället valt att begagna sig av singularformen *networking*. Vad gäller *utility* och *utilities* förekom *utility* enbart sex gånger som tagg medan *utilities* förekom hela 96 gånger.

För tidigare objekt har jag konstaterat att ett antal användare valt att begagna sig av såväl singular- som pluralform av ett ord men objekt E skiljer från tidigare objekt. Enbart två användare har valt att använda singular- och pluralform av ett ord. Användare 56 har valt att tillskriva tio termer för att beskriva objekt E och två av dessa termer är *utility* och *utilities*. Användare 126 har valt att tillskriva hela 18 termer och har likt användare 56 valt att använda både *utility* och *utilities*.

Två adjektiv förekommer för objekt E och dessa är *temp* och *free*. Användare 68 har valt att enbart tillskriva en tagg till objekt E och denna tagg var *temp*. Adjektivet *free* däremot förekom nio gånger bland användarna för objekt E. Fem av dessa nio användare, däribland användare 31, 126 och 140, har dessutom valt att använda taggarna *freeware* och *software*. Freeware innebär gratis mjukvara vilket betyder att taggarna *free* och *software* i praktiken är överflödiga då de speglar konceptet för *freeware*. Vidare valde sex av de nio användarna att kombinera taggarna *free* och *freeware* vilket även det är intressant då taggarna speglar så likartade koncept.

För objekt E förekom två verb: *network* och *utilise*. *Network* förekommer 109 gånger men huruvida användarna avsåg infinitivformen av verbet *network* eller substantivet *network* var oklart. Samma problem förelåg även för objekt D. Huruvida taggen *networking* avsåg substantivet eller presensformen av verbet var oklart men då dessa taggar är centrala för objekt D har jag valt att begagna mig av båda möjligheterna. Taggarna har jag därför placerat in både som substantiv och verb.

### 7.5.2 Hyponymer och meronymer

För objekt E förekom ett relativt stort antal meronymer och hyponymer och hyponymer var den ordklass vilken dominerade. *Computer* förekom tre gånger som meronym till olika termer. *Computer* förekom också som meronym till både *WI-FI* och *network* men för objekt E återfanns *computer* även som meronym till *ADSL*, som är en typ av nätverk. Två andra meronymer var *network* och *web* vilka förekommer tillsammans som meronymer till *internet*.

En annan meronym var *mac* som är en meronym till *apple* enligt förklaringen i kapitel 7.4.2. Utöver redan uppräknade meronymer förekom *windows*, *OS* (syftar här på operativsystemet Mac OS) och *OSX* som hyponymer till *OS* (här åsyftas förkortningen för operativsystem).

Då operativsystemet Mac OS är en typ av operativsystem innebär det att *OS* (Macintoshs produkt) är en hyponym till *OS*. Då även *Windows* och *OSX* är exempel på operativsystem innebär det att *windows* och *OSX* är hyponymer till *OS*. Ett operativsystem i sin tur är ett exempel på ett program med konsekvensen att *OS* är en hyponym till *software*. Även de tre operativsystemen (Mac OS, Mac OSX och Windows) är exempel på program och är således även de hyponymer till *software*. Även *freeware*, *application* och *utility* är hyponymer till *software*.

En annan meronym vilken även förekom som hyponym är *mac*. *Mac* förekom som hyponym till *PC* som står för 'personal computer', eller persondator på svenska. IBM PC var dock även den första persondatorn som producerades av IBM och idag används *PC* allmängiltigt om en generell persondator med Windows installerat. Med denna definition av *PC* är alltså *PC*, likt *mac*, en hyponym till persondator, *PC*. En persondator är en typ av dator vilket innebär att *PC* är en hyponym till *computer*. *Web* är en meronym till *internet*.

Ytterligare funna hyponymer är *sysadmin*, en förkortning för systemadministratör, vilken utgör en hyponym till *admin*. Utöver dessa förekom enbart tre hyponymer som jag ännu inte tagit upp. *Internet*, *WI-FI* och *ADSL* är samtliga en typ av nätverk och därmed hyponymer till *network*.

### 6.5.3 Egenkonstruerade taggar

Av de 72 olika taggarna var 20 stycken egenkonstruerade taggar. Det innebär att de utgjorde 27,8 % av de olika taggarna för objekt E. Vidare var det två av dessa 20 taggar som begagnas mer än två gånger, *howto* och *toread*, vilka förekom fyra respektive tre gånger. Dock existerade det två taggar med skiljetecken som med fördel hade kunnat skrivas som *toread* för att öka konsistensen för både användarna och objektet.

Av de 20 egenkonstruerade taggarna var det fem som har blivit tillskrivna i enlighet med vad Delicious rekommenderar. 11 stycken taggar har blivit tillskrivna med någon form av skiljetecken och av dessa har sex stycken valt att använda sig av understreck. Vidare kunde jag uppfatta nio stycken taggar som inte var till för andra användare än de som har taggat med de aktuella taggarna. Ett skiljetecken jag inte observerat innan begagnades här, nämligen forward slash. Denna tagg, vilken var *tips/info/diy*, hade också med fördel kunnat delas upp eller gjorts en omskrivning på, då det hade inneburit att konsistensen för användaren hade ökat. Ytterligare taggar som användarna skulle kunnat dela på är bland annat *network\_apps* och *network\_utilities*. De hade då i stället bildat fyra taggar som andra användare redan hade nyttjat för att beskriva objekt E.

För objekt E kan jag konstatera att understreck var det vanligaste skiljetecknet samt att användarna här prefererade begagnandet av skiljetecken vid tillskrivandet i stället för det sätt som Delicious rekommenderar. Vidare var det substantiv som är det som användarna använder sig mest av. Jag har också kunnat konstatera att om användarna delat ett par av taggarna hade konsistensen ökat något.

#### **6.5.4 Indexeringskonsistens**

Den genomsnittliga indexeringskonsistensen för objekt E var 27,4 %, vilket betydde att konsistensen var relativt hög, med tanke på att antalet olika taggar uppgick till hela 72 stycken.

För objekt E förekommer ett antal taggar vilka var vanligt förekommande bland de tjugo första användarna. Hälften av användarna har valt att begagna sig av *utilities* och *tools*. Taggen *software* förekom hos 11 användare. Andra termer i majoritet var *freeware* och *network* som förekom hos åtta respektive nio användare. De termer som var populära bland de tjugo första användarna var samma termer som var populära bland samtliga användare för objekt E.

Genom att titta på diagrammet nedan kan vi konstatera att enbart en användare, användare 17, genererat en konsistens som är radikalt annorlunda det övriga användare genererat. Användare 17 har genererat en konsistens på noll procent som en konsekvens av tillskrivandet av *downloads* som tagg. Ingen annan av de 20 användarna har valt att begagna sig av *downloads* och bland det totala antalet användare förekom *downloads* enbart fyra gånger.

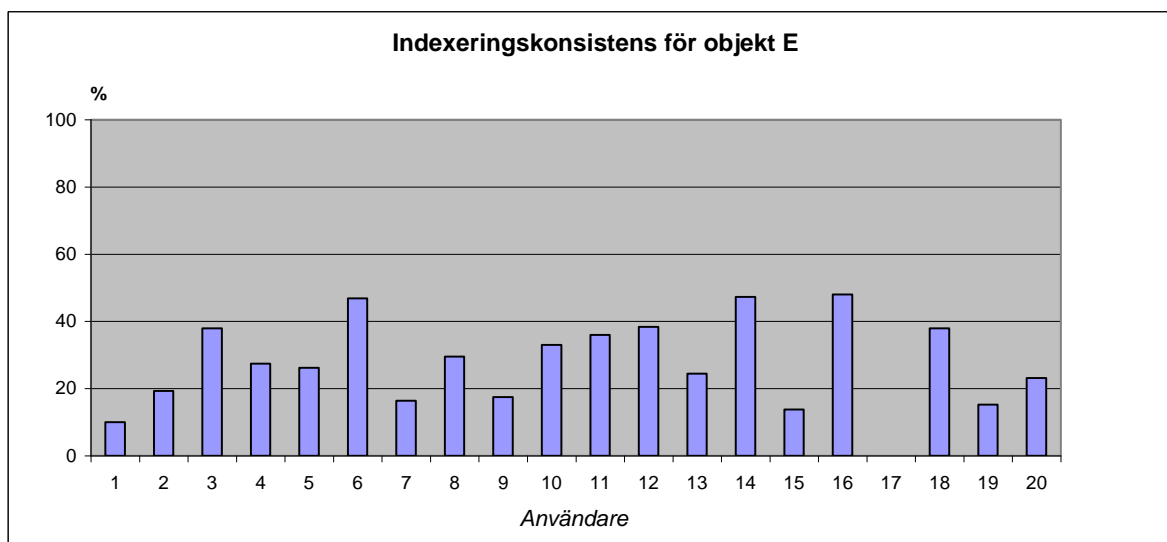


Diagram 5: Diagrammet återger den genomsnittliga konsistensen för användare 1-20.

För objekt E var det enbart tre användare som generat en konsistens på över fyrtio procent. Användare 16 har uppnått en konsistens på 48 % genom att tillskriva *network*, *utilities*, *networking*, *software*, *tools*, *freeware* och *lifehacker*. Användare sex har tillskrivit exakt samma termer som användare 16 och därmed råder en konsistens på 100 % mellan användare 6 och 16. Den genomsnittliga konsistensen för användare 6 är även den 48 %.

Den tredje och sista användaren, vilken uppnått en konsistens på över 40 % är användare 14 vilken generat en konsistens på 47,3 %. Användare 14 har tillskrivit samma termer som användare 6 och 16 men även använt sig av *security*. Gentemot användare 6 och 16 har användare 14 en konsistens på 93 %.

Som kan utläsas av diagrammet var användarna förhållandevis ense kring valet av termer för objekt E. Som noterat tidigare är det användare 17 som utmärker sig genom att ha genererat en konsistens på noll procent. Förutom användare 17 var användare 1 den användare vilken uppnått lägst konsistens. Användare ett har generat en konsistens på 10 %. Den i sammanhanget låga siffran är en konsekvens av att användaren enbart valt att tillskriva två termer varav den ena, *dev*, enbart förekommer hos användaren och den andra *networkings* enbart förekommer hos ytterligare en användare bland de 20 första användarna.

## 7 Övergripande analys

Totalt har objekten A till E genererat 3 188 taggar vilket gav mig 311 olika taggar att jobba med. Jag kan alltså konstatera att spridningen av ord som tillskrivits inte var speciellt stor, enbart 9,7 % över det totala antalet taggar vilket också speglar de resultat jag fått på de enskilda objekten, förutom objekt C vilken också sticker ut vad gäller indexeringskonsistensen, se 8.4. Vidare var det hela 142 av de olika taggarna som endast har blivit nyttjade en gång vilket innebar att över hälften av de olika taggarna bara har begagnats en gång totalt. Av de återstående taggarna har 41 stycken tillskrivits på mer än ett objekt och som en följd har de använts mer än en gång. Dock var det 42 taggar som har blivit begagnade tio gånger eller mer för samtliga objekt varav 17 stycken har tillskrivits mer än 45 gånger.

### 7.1 Böjningsvarianter

Av de olika böjningsvarianter som fanns var det substantiv singularform som var det vanligaste att användare begagnade sig av. Totalt var det 123 taggar som tillhörde kategorin substantiv, singular. Dock var det endast 28 stycken av dessa taggar som har blivit begagnade över tio gånger varav sex stycken har nyttjats mer än 100 gånger. Här fanns det en tagg som utmärkte sig, *network* förekom inte mindre än 234 gånger. När jag tittade på användningen av pluralformen figurerade 26 stycken taggar. I motsats till singularformen nyttjades 10 olika taggar mer än tio gånger. Användarna var relativt överens, vad gäller tillskrivningen av pluralformen, om vilka taggar som bör användas. Vidare fanns det 12 taggar som figurerade både i singular och plural. Intressant är att när båda formerna förekommer var det pluralformen som användarna prefererade. Det förekom dock tre undantag från detta i undersökningen, *networking*, *network* och *tutorial* har användarna i stället valt att preferera singularformen.

Vad gäller adjektiv nyttjades endast 11 olika taggar. Dessa har alla blivit tillskrivna i grundformen, neutrum, och endast fyra av taggarna nyttjas tio gånger eller mer. Den tagg som var mest frekvent är *fun* vilken 43 användare har valt att begagna sig av. Anledningen till att *fun* var den mest förekommande bland adjektiven anser jag bero på att de flesta människor har lätt att relatera till vad som är roligt.

Totalt förekom 19 verb bland taggarna i mitt material. Vanligast var att de är tillskrivna i grundformen infinitiv även om övriga former också används om än sparsamt, nio taggar har begagnats i en annan form än grundformen. Dock är det ingen tagg som har blivit tillskriven i formen imperfekt. Utöver grundformen var futurum den vanligaste, hela fem av de nio taggarna har tillskrivits denna form, vilken användarna har skapat genom begagnandet av 'to' i olika varianter. Vidare nyttjades denna typ av tagg endast av en användare men åtta stycken användes mer än en gång. Här utmärker sig taggen *network*, anledningen är att denna tagg både kan vara ett substantiv, nätverk, och ett verb, att nätverka. Då jag inte kunde veta vilket som användaren avsåg valde jag att ta upp denna tagg som både substantiv och verb.

Anledningen, som jag ser det, till att substantiv var det mest frekvent använda i min undersökning är att substantiv är den typ av ord som bäst kan beskriva objektet. Substantiv innehåller inga värderingar vilket gör de lätta att använda. Exempelvis beskriver taggen *computer* en dator, den innehåller ingen värdering kring begreppet dator vilket till exempel taggen *awesome* gör. Vidare syns det i en del fall att



användarna har utgått från objektet när de taggar. Jag exemplifierar med objekt B vilken handlar om hur du grillar en stek. Objekt B innehöll nämligen inte bara ett recept på grillad stek utan även recept på marinad vilket betyder att objekt B behandlade mer än ett recept. En av de vanligaste taggarna för objekt B var *recipes* men även *recipe* förekom. Jag vill med detta exempel påstå att materialet jag har för objekten visar att de användare som har taggat inte endast har gjort det snabbt och slarvigt, utan att de har tänkt efter hur de tillskriver taggarna. Samtidigt var det så att en hel del användare av Delicious har många objekt nedsparade vilket medför att de måste tagga objekten med taggar som beskriver objektet för att kunna lokalisera dem i sin samling.

Som jag har sagt innan talar ett adjektiv om hur saker och ting är vilket medför att adjektiv oftast har en värdering i orden. Vad bestämmer till exempel vad som är snyggt? Det är olika från person till person vilket innebär att en personlig värdering följer med taggen. En tagg som användes var exempelvis *interesting*. Taggen betyder 'här är det något intressant' men det kan omöjligt vara intressant för alla användare av Delicious. Därför är adjektiv, i den utsträckning de förekommer, taggar som är till för den som har taggat, vilket också förklarar varför de endast står för 4,4 % av det totala antalet olika taggar för objekt A till E.

Vad gäller förekomsten av verb låg den på 7,5 % av det totala antalet olika taggar. Verb är inte heller de lätta att tillskriva ett objekt. Ett verb talar om saker som går att göra, det går till exempel inte att "göra" objekt C. Däremot går det att göra det som objekt C beskriver, nämligen glass. Att verb ändå är något vanligare än adjektiv kan förklaras av att det finns verb som, visserligen inte beskriver objektet, men som går att genomföra. Som exempel begagnar jag mig av *read*. *Read* är ett verb och beskriver något som går att göra och alla objekt som har legat till grund för min uppsats går ju att läsa. Det är av största vikt att de är läsbara för annars hade de förmodligen inte blivit föremål för användarnas uppmärksamhet. Övriga exempel från mitt material är *tolearn*, *review*, *learn* och *toevaluate*.

Det fanns också taggar som inte har kunnat placeras in i någon av dessa kategorier. Det var bland annat olika adverb, pronomen och prepositioner. Även en del av de egenkonstruerade taggarna har inte placerats in. Det är taggar som oftast bestod av flera ord, vilka var hopskrivna på olika sätt. Ytterligare taggar som inte var med är sådana som jag inte har kunnat hitta någon betydelse på eller något sammanhang som klagör var de skulle kunna ha placerats in.

## 7.2 Hyponymi och meronymi

Av dessa båda är det mest hyponymer som användarna har begagnat sig av. Hyponymer visar på ett 'typ av' förhållande mellan taggarna, vilket betyder att hierarkiska relationer visar sig genom nyttjandet av hyponymer. Att användarna har använt denna typ av taggar visar att de oftast är överens om vad objekten handlar om i stort, däremot var de inte överens om hur specifika taggar som behövs tillskrivas. Jag anser att det kan beror på tre saker, användarna har olika förkunskaper, vilket innebär att de identifierar olika saker i objektet, deras egen samling med nedsparade objekt kräver specifitet i taggarna och vilken kunskap användarna har inom området som objektet behandlar. I materialet fanns dock exempel på hur användare valde att begagna sig av hierarkiska relationer, användare 50 och 128 för objekt D har båda valt att nyttja *network* och dess hyponym *wifi*.

Vad gäller meronymer visar de på förhållandet 'del av'. Det visade sig dock att det förhållandet som förelåg i mitt material oftast var 'kan vara en del av' vilket fick till följd att det var svårare att identifiera meronymerna och att dessa var relativt få till antalet, endast 11 stycken. Jag exemplifierar med *salt* och *cooking*. *Salt* ingår ofta i *cooking* däremot behöver det inte alltid ingå vilket medföljer att förhållandet mellan *salt* och *cooking* blir ett 'kan vara en del av' eller 'är oftast en del av'. Alltså är inte *salt* en meronym till taggen *cooking*. Ett tydligt meronymförhållande föreligger dock med taggen *family*. *Children*, *kid*, *kids* och *nephews* är alltid en del av taggen *family*. Då en familj består av barn och om begreppet utvidgas går det att benämna släkten som familj vilket medför att även brorson/systerson blir en del av begreppet familj. Värt att notera är också att jag endast för objekt C har kunna konstatera fyra taggar, vilka utgör meronymer till en och samma tagg, *family*, vars meronymförhållanden jag visat på ovan.

### 7.3 Egenkonstruerade taggar

En relativt stor andel av det totala antalet olika taggar utgjordes av egenkonstruerade taggar, hela 35,3 % eller 89 stycken. Dock var det bara sju av dessa som nyttjats mer än en gång vilket betyder att flertalet av dessa taggar endast användes en gång. Fyra av de sju taggarna utmärkte sig. *Icecream* var en av de taggar som blivit begagnad flest antal gånger. Anledningen är att objekt C handlar om att göra glass, vilket medför att *icecream* var en av de mest frekventa taggarna för det objektet. De tre kvarvarande taggarna vilka utmärkte sig var *todo*, *toread* och *howto*. Dessa tre taggar återfanns hos de flesta objekt som jag har undersökt, ibland endast ett fåtal gånger men de fanns. Vidare är det intressant att se att användarna för det mesta var överens om hur dessa taggar skulle tillskrivas, endast fyra avvikelser fanns, vilka har blivit begagnade en gång vardera. Dessa avvikelser drabbade oftast *toread*.

Nästan hälften av de egenkonstruerade taggarna innehöll någon form av skiljetecken, vilket innebar att det sätt att skriva ihop taggar som Delicious rekommenderar var mindre frekvent använt. Användarna prefererade vissa typer av skiljetecken vid tillskrivandet av taggar, framför allt var det understreck och kolon som begagnas. 11 stycken taggar innehöll understreck medan åtta använder sig av kolon som skiljetecken.

Intressant är att taggarna *todo*, *toread* och *howto* förekom i flertalet av mina objekt och att taggarna har blivit tillskrivna på samma sätt på de flesta objekten, med vissa tidigare nämnda variationer. Klart är att sättet de tillskrivits på är vedertagen, i alla fall i det material som ligger till grund för min uppsats. Vidare går det att fundera över varför egenkonstruerade taggar inte har förts vidare från användare till användare vilket jag sett tidigare, se 7.4.1. Troligtvis är det så att flertalet användare ansåg att de egenkonstruerade taggarna inte var det bästa sättet att uttrycka vad det aktuella objektet handlade om. Ytterligare en anledning kan vara att många användare tänker på att underlätta för de människor som söker i Delicious. Om objekt som handlar om 'video codecs' är av intresse är det troligare att sökningen görs genom taggarna *video* och *codecs* än *videocodecs*.

Vad gäller objekt D och E har det gått att göra en del intressanta iakttagelser då dessa båda är från samma webbsida, men har blivit nedsparad med olika URLer. Exempelvis förekom endast siffran tio i taggarna för objekt E. Detta i motsats till objekt D som vars taggar även innehöll siffran två i stället för 'to'. Vidare går det också att se att *howto* som begagnats fyra gånger för objekt E, inte finns som tagg för objekt D. Det fanns inte heller någon egenkonstruerad tagg som innehöll *network* för objekt D vilket det finns för objekt E. Dock har båda objekten blivit tillskrivna taggarna *toread*, *todo* samt *top10*.

## 7.4 Indexeringskonsistens

Jag anser att den genomsnittliga konsistensen var hög med tanke på att det troligtvis är flest människor utan utbildning i indexering som tillskrivit objekten taggar. Görs en beräkning på den genomsnittliga konsistensen för samtliga objekt uppgår den till 28,34 %. Viktigt att komma ihåg är dock att de beräkningar som jag har gjort för att fastställa indexeringskonsistensen endast är beräknad på de 20 första användarna, vilket innebär att resultatet hade kunnat bli ett annat om beräkningarna hade gjorts på samtliga användare för objekten.

Vad gäller indexeringskonsistensen var det endast objekt C som hamnade under en genomsnittlig konsistens på 25 %. Objekt C uppnådde endast en konsistens på 14,1 % vilket kan jämföras med objekt B som har uppnått en konsistens på hela 41,3 %. Tydligt är att de tjugo första användarna för objekt C har haft svårare att enas om vad objektet handlar om och hur det ska uttryckas än vad användarna för objekt B har. Jag har, vid min genomgång av de enskilda objekten, sett exempel på hur konsistensen minskats genom användandet av taggar som inte innehar någon information om vad det aktuella objektet handlar om. Vidare finns det också exempel på hur skiljetecken nyttjats och som gjort att konsistensen minskar. *Video\_codecs* och *videocodecs* är två taggar som med fördel hade kunnat delas upp i taggarna *video* och *codecs*, vilka var frekvent förekommande för objekt A. Om detta skett skulle följderna varit att indexeringskonsistensen ökat, särskilt med tanke på att originaltaggarna endast blivit begagnade en gång var.

### 7.4.1 Lancasters faktorer

Då jag ansåg att några av Lancasters sju faktorer inte var väsentliga att ta upp igen kommer jag bara att redogöra för de som jag genom mitt material menar har en större inverkan på konsistensen.

Antalet tillskrivna taggar menar jag är en viktig faktor. Det som spelar in är selektiv och uttömmande indexering men även och översättning och ämneskunskapen som användaren besitter. Beroende på hur mycket användaren kan om det ämne som objektet behandlar väljs taggar. Det innebär att olika böjningsvarianter samt synonymer kan bli tillskrivna. Om användaren besitter liten eller ingen kunskap om ämnet kan det få till följd att väldigt generella taggar tillskrivs eller att helt fel taggar tillskrivs. Framför allt är det olika böjningsvarianter som begagnas. Vidare så inverkar också ämnets terminologi, som jag redan berört lite, framför allt märkte jag av det på de tre tekniska objekten jag undersök. Det vanligaste var att användarna både begagnade sig av en övergripande tagg så som *videocodecs* och *software* och en mer specifik tagg så som *XVID* och *freeware*. Men det finns undantag, exempelvis på objekt B där två användare valt att begagna sig av de underliggande taggarna *chimichurri* och *marinade*.

Vem det är som tillskriver taggarna har också betydelse och även detta har jag berört innan. Detta gäller framför allt vilka ämnen som användaren är förtrogen med. Kan man mycket om ett ämne är det större chans att rätt taggar blir tillskrivna. Erfarenheten att tagga hos en användare kan visa sig genom att denna användare verkar mer medveten om sina egna taggar. Jag stärks i mitt antagande när jag i mitt material sett att några användare nyttjar både singular- och pluralformen av ett ord samt synonymmer för detta. Vidare begagnar de ibland även olika specifitet på taggarna.

För Delicious finns det också ett redskap som är till för att hjälpa användarna att tillskriva taggar på rätt sätt och även att välja rätt taggar. Detta rekommenderingsverktyg tar hjälp av de taggar som redan finns inom systemet. Det gör att verktyget förändras över tid men i samma takt som språket förändras vilket gör att även rekommenderingsverktyget hela tiden är aktuellt. Vad gäller tillskrivandet är det samma hela tiden. Rekommenderingsverktyget förklarar jag utförligt i kapitel 5.

## 7.5 Jämförelse mellan objekt D och E

Då objekt D och E är samma webbplats men har blivit nedsparade med två skilda URLer har det gett en möjlighet att jämföra dessa två objekt mer i detalj. Det har också gett mig en möjlighet att se om det förelåg stora skillnader i hur användarna valt att tillskriva taggar och vad de skillnaderna kan beror på. Genom att se likheter och olikheter kan jag se möjliga sätt att påverka användarnas tillskrivande av taggar och därigenom även påverka indexeringskonsistensen. Det kan då bidra till att jag kan göra mer kvalificerade reflektioner kring mitt resultat och även möjligtvis ge uppslag till vidare forskning inom detta område.

Användarna för objekt D har tillsammans tillskrivit objektet 55 olika taggar. Det var nästan 20 olika taggar mindre än vad användarna för objekt E har valt att begagna sig av. Av dessa totalt 128 olika taggar var det 28 stycken taggar som användarna för både objekt D och E har valt att begagna sig av. Vidare förekom det några stora skillnader vad gäller nyttjandet av taggar, till exempel taggen *internet* som 13 användare valt att tillskriva, men bara för objekt E. *Internet* fanns inte som en tagg för objekt D även om taggen *net* förekom en gång. Intressant är också förhållandet mellan *lifehacker* och *lifehacks*, För objekt E var det taggen *lifehacker* som dominerade medan det är *lifehacks* som var den mest använda taggen för objekt D. Anledningen till att *lifehacker* var dominerande för objekt E beror sannolikt på objektets titel vilken innehåller ordet *lifehacker*.

Vad gäller andra taggar var det svårare att förstå varför fördelningen såg ut som den gjorde. *Tool*, *tools* och *tutorial* var några taggar som oftare har blivit tillskrivna för objekt D än för objekt E. *Tools* var den tagg som nyttjats mest antal gånger av användarna för båda objekten, medan *tool* och *tutorial* endast använts av ett fåtal. Möjligen kan detta bero på att de flesta användarna anser att objektet tar upp flera olika hjälpmedel och följaktligen har de valt att begagna sig av pluralformen.

För båda objekten gällde samma fördelning rörande böjningsvarianter som jag tidigare har konstaterat, substantiv singularform är det som användarna begagnat sig av mest. Fördelningen mellan användningen av de olika böjningsvarianterna var snarlik när de båda objekten jämfördes. Vidare nyttjats fler hyponymer i objekt E än i D, vilket berodde på att användarna har tillskrivit fler hyponymer till samma tagg än vad som är

fallet för objekt D. Vad gäller meronymer så förekommer nästan exakt samma meronymer för båda objekten.

De egenkonstruerade taggarna var vanligt förekommande hos båda objekten, 19 stycken för objekt D respektive 20 för objekt E. Trots detta var det endast tre stycken som återfanns hos båda objekten, *todo*, *top10* och *toread*. Två av dessa var också taggar som jag har kunnat konstatera hos flertalet av de objekt som ligger till grund för undersökningen. Jag har också kunnat konstatera vissa avvikelser i stavning även om flertalet användare var överens om hur dessa taggar skulle tillskrivas.

Vad gäller indexeringskonsistensen har användarna för objekt D uppnått en högre konsistens än för objekt E. Orsakerna kan vara många, exempelvis har objekt D ett färre antal olika taggar än objekt E som har nästan 20 stycken fler olika taggar. Vidare var det också något fler användare som valt att spara ner objekt E, även om jag anser att det inte är tillräckligt för att ge något större genomslag på konsistensen. Som jag ser det var den största anledningen till skillnaden i konsistens de olika taggarna. Fler olika taggar ger en lägre indexeringskonsistens.

## 8 Diskussion

Syftet med denna uppsats har varit att undersöka indexeringskonsistensen i en folksonomi. Jag kommer på de närmaste sidorna att diskutera och reflektera kring resultaten av min undersökning. Inledningsvis kommer jag att ge en tillbakablick på vad som framkommit i kapitlen resultat och analys och övergripande analys. Jag kommer också förtydliga svaret på min frågeställning. Därefter kommer jag att reflektera kring användarnas taggar och konsistens i en folksonomi för att slutligen titta närmare på vilka förbättringar som skulle kunna ske för att uppnå en högre konsistens.

### 8.1 Sammanfattning av mitt resultat

*Hur konstrueras taggar med avseende på böjningsvarianter och hur vanligt förekommande är böjningsvarianter?*

Framför allt har det varit substantiv som tillskrivits, både i singular och plural även om singular varit det vanligaste. Vidare förekom både adjektiv och verb men de nyttjades sparsamt. Alla adjektiv jag har identifierat har varit skrivna i grundform. Vad gäller verben har de flesta varit skrivna i grundform, utöver det har futurum varit den mest använda. Genom att begagna sig av taggar vilka inkluderat 'to' har användarna kunnat skapa futurumtaggar.

*Vilken utbredning har närliggande ord i form av hyponymer och meronymer bland användarna?*

Användarna begagnade sig av både hyponymer och meronymer, men i relation till nyttjandet av böjningsvarianter förekommer de sparsamt. Vanligast var användandet av hyponymer vilka uttrycker en hierarkisk relation och är relativt lätt för användarna att begagna sig av. Meronymer är svårare att nyttja, för att en tagg ska vara en meronym krävs det att förhållandet alltid är 'del av'. Många taggar har inte det förhållandet utan i stället ett 'kan vara en del av'.

*I vilken utsträckning begagnar användarna sig av egenkonstruerade taggar och vilken betydelse har de för indexeringskonsistensen?*

Andelen egenkonstruerade taggar var relativt hög, 35,3 %. Vidare gick det att konstatera att de flesta av dessa taggar endast begagnas en gång, endast sju förekommer mer än en gång för samtliga objekt. Hälften av taggarna innehöll också någon form av skiljetecken vilket Delicious inte rekommenderar vid tillskrivandet av taggar. Tre taggar, *todo*, *toread* och *howto* utmärkte sig genom att, i de flesta fall, bli tillskrivna på exakt samma sätt för alla objekt där de förekom.

Indexeringskonsistensen var högre än jag vad jag först väntade mig, objekten har en genomsnittlig konsistens på 28,34 %. Vidare har jag sett några problem som bidrar till en minskad konsistens, bland annat är det möjligheten till att använda sig av skiljetecken och att en del taggar med fördel hade kunnat delas upp i två taggar i stället. Resultatet är dock endast baserat på beräkningar som utförts för de 20 första användarna för varje objekt.

Det som påverkar indexeringskonsistensen mest är de egenkonstruerade taggarna. Som jag tidigare konstaterat så begagnas denna typen av ord oftast bara av en eller ett par användare. Det betyder att fler taggar per objekt används vilket leder till att

konsistensen minskar. Jag har dock hittat vissa undantag, exempelvis *icecream* och *howto*.

Även användandet av meronymer och hyponymer spelar en vis roll, ett exempel som jag tagit upp innan är begagnadet av *marinade*, se 7.2.2. Detta är ett tydligt exempel på taggar som bidrar till en minskad konsistens. Meronymer och hyponymer visar också på hur förtroliga användarna är med ämnet. Att dessa typer av ord inte har använts mer kan ha två orsaker, antingen har användarna endast ytliga kunskaper inom ämnet eller också visar bristen på dessa ord att användarna tänker på möjligheten att återvinna materialet och nyttjar de taggar som objektet redan har blivit tillskrivet med. Vilket av dessa två det är kan jag ur mitt material dock inte avgöra.

De böjningsvarianter som jag har hittat är det om minst påverkar konsistensen. Visserligen bidrar begagnandet av adjektiv och verb till en minskad konsistens men flertalet användare prefererar substantiv i singularform och används pluralformen så är användarna oftast överens om att det är taggen som borde användas, exempel på det är *recepies*.

## 8.2 Indexeringskonsistens ur ett folksonomiperspektiv

Jag har i mitt resultat kunnat konstatera att indexeringskonsistensen var förhållandevis hög med undantag av objekt C vilken endast uppnått en konsistens på 14,1 %. En del författare, så som Golder och Huberman (2005) och Noruzis (2007), anser i grunden att den fria vokabulär som Delicious använder sig av innebär att konsistensen är låg, vilket också Lancaster ger uttryck för i sina sju faktorer som påverkar konsistensen. Jag menar dock att min undersökning kan visa på något annat. Indexeringskonsistensen som har uppnåtts på de objekt jag undersökt ligger i genomsnitt på 28.34 % vilket jag menar är ett förhållandevis högt resultat. Samtidigt bör förbättringar göras som kan leda till att konsistensen blir högre. Frågan är om det är rätt att försöka styra upp och implementera en kontrollerad vokabulär som innebär att användarna inte kan tillskriva de taggar de vill? Den typ av kontrollerade vokabulär som indexerare använder medför att användarna i en folksonomi exempelvis inte skulle kunna begagna sig av taggen *receptes*, en tagg som flertalet av användarna nyttjar. En professionell indexerare skulle inte nyttja denna typ av ord då det visar på en typ av dokument, inte dokumentens innehåll. Om användarna av en folksonomi tycker det är en viktig aspekt, menar jag att det måste gå att tillgripa denna typ av tagg.

Vad gäller frågan om begagnandet av singular eller plural prefererade användarna i min studie singularformen, dock fick jag uppfattningen om att pluralformen användes när användarna ansåg att det behövdes för att beskriva innehållet i objekten. Baserat på den konceptuella analys som användarna gör får översättningen taggar av både singular och plural, återigen är *recipes* ett bra exempel. Objektet innehåller mer än ett recept och därför anser jag att användarna väljer pluralformen av ordet.

Vidare är frågan om fri vokabulär inte helt entydig. Min undersökning omfattade totalt 3188 taggar, av dessa var det endast 311 olika taggar, vilket betyder att variationer av taggar inte var speciellt stor i mina objekt. Detta är också ett mönster som Golder och Huberman (2005) sett i sin studie. Det verkar alltså som användarna gör ett bra jobb både vad gäller konceptuell analys och översättning. Hade så inte varit fallet är jag övertygad om att jag hade sett ett större antal olika taggar och som en följd av det även en lägre genomsnittlig indexeringskonsistens.

Jag tror dock att det finns lösningar som kan öka konsistensen ytterligare, detta utan att förminska en av folksonomins största styrkor. Specia och Motta (2007) och Wu (2006) har i sina studier använt sig av kluster för att, dels underlätta för sökningar, och dels underlätta för navigering. Wu har skapat en taxonomi utifrån taggar som blivit genererade från folksonomin. Specia och Motta skapade kluster där taggarna relaterar till varandra, i förlängningen innebär det att användarna kan se både bredare eller smalare ord som de kan tillskriva som taggar. Samtidigt menar jag att det är av största vikt att bibehålla den fria vokabulär som folksonomi bygger på. En av folksonomins styrkor är att dess taggar förändras samtidigt som språket, det betyder att en folksonomi är föränderlig. Den aspekten är viktig att ha kvar och att användarna fortfarande kan tagga i enlighet med sin egen referensram. Implementeras någon form av taggenererad taxonomi är det viktigt att den tillåter att folksonomin är föränderlig, vilket i förlängningen betyder att den måste kunna uppdatera och förändra sig. Samtidigt måste det ske utan att det medför stora kostnader för dem som tillhandahåller folksonomin, då även den låga kostnaden oftast ses som en fördel.

Ytterligare en sak som en implementerad taxonomi måste ta hänsyn till är möjligheten att skriva ihop ord till en tagg. Då Delicious nyttjar blanksteg som en särskiljare mellan taggarna som tillskrivs är det viktigt att kunna uttrycka koncept som består av fler än ett ord. Vidare betyder det att om den möjligheten försvinner är det flera ämnen vars koncept inte längre går att uttrycka då deras terminologi kräver flera ord i samma tagg, exempelvis 'information retrieval' vilken utan möjlighet till hopskrivning skulle bli två taggar. Detta är en av Lancasters sju faktorer och även den som var tydligast i mitt material. Tydligt är att ett konkret ämne oftast får en högre konsistens än ämnen som kräver mer förkunskaper.

En viktig aspekt inom indexering är indexeraren vilken ofta har utbildning inom indexering. Utbildningen lär ut hur indexeringen utförs, vilka typer av ord som bör användas samt vad som bör vägas in, exempelvis målgrupp. De som utför indexeringen i en folksonomi har inte denna typ av utbildning. Det märks på att alla typer av ord används som taggar även om substantiv är det som nyttjas mest. Vidare visar Bruce (2008) i sin studie att detta inte behöver vara ett problem. Utbildade indexerare och användarna av en folksonomi tenderar att indexera annorlunda. Folksonomi i detta fall minskar klyftan mellan indexerarens taggar och användarens av systemet. Tydligt är dock att det råder en större enighet kring vilka taggar som bör användas om objektet är kortare och dessutom behandlar ett ämne som många kan relatera till utan att någon förkunskap krävs vilket även Lancaster har uttryckt, se 4.1.

Under arbetet med att redovisa och analysera mina resultat har jag gjort ett par egna reflektioner om hur konsistensen skulle kunna ökas, utan några större förändringar i systemet. Delicious ger rekommendation om hur ord bör skrivas ihop. Detta följs inte alltid utan användarna nyttjar skiljetecken av olika slag. Ett vanligt skiljetecken är understreck. Dessa taggar används sedan oftast inte mer än en gång och jag har sett exempel på likvärda taggar som utan skiljetecken har använts mer än en gång. Jag anser därför att några relativt enkla saker kan göras för att motverka användandet av skiljetecken. Om restriktionen att skiljetecken inte går att använda sig av genomförs, kommer inga taggar liknande *to.read* och *to-read* att kunna tillskrivas. Taggarna kommer alltså automatiskt att tillskrivas på det sätt som Delicious rekommenderar.



Vidare har jag, i min undersökning, kunnat konstatera att användarna i genomsnitt använde sig av 3 till 7 taggar var. Om det gick att begränsa antalet taggar som kan tillskrivas bör indexeringskonsistensen öka, vilket Lancasters första faktor visar, se 4.1. Att kunna tillskriva upp till sex taggar menar jag är rimligt. Eftersom indexeringen då blir vad som kallas selektiv indexering, innebär det att specifika taggar som *chimichurri* och *ADSL* inte längre tillskrivs i samma utsträckning, utan i stället används de bredare taggarna *marinade* och *network*. Det skulle alltså kunna innebära att begagnandet av hyponymer och meronymer minskar, men också att egenkonstruerade ord i förlängningen minskar då användarna har ett begränsat utrymme för sina taggar.

### 8.3 Metoden ur ett indexeringsperspektiv

Då uppsatsens syfte är att undersöka hur indexeringskonsistensen ser ut i en folksonomi är en central del att använda en beräkningsformel som möter uppsatsens behov. Jag valde att använda mig av Rollings mått. Anledningen är att måttet inte bara ser till hur många taggar som tillskrivits. Den tar också hänsyn till om taggarna har begagnats av båda användarna. Lancaster (1998) menar att formeln även ska visa och se om taggarna är av större eller mindre betydelse, till exempel som *codecs* och *MPEG*. Även om jag inte tyckte det var nödvändigt för min uppsats att begagna mig av en formel som visar vilka taggar som har en större betydelse, kan jag nu se att det finns ett värde i det. Jag har även under min examination blivit rekommenderad att använda mig av storlek istället för mängd när jag använde Rollings mått. Att jag inte har gjort det beror på att i den källan jag har för Rollings mått har inte måttet blivit begagnat på det sättet. Taggar med en större betydelse tenderar att vara taggar som är av en bredare karaktär, vilket även är fallet i mitt material. Många användare begagnar sig av selektiv indexering, men kan använda en eller ett par mer specificerade taggar.

Vidare kan tyckas att det är väl djupgående att också titta på förekomsten av olika ordklasser. Det har ingenting med den beräknade konsistensen att göra. Dock anser jag det vara av intresse att kunna göra en jämförelse mellan hur indexerare tillskriver termer och hur Delicious användare tillskriver taggar. Booth (2001) visar på att indexering i professionell bemärkelse, framför allt görs med substantiv plural. Användarna i Delicious prefererade, i motsats till professionella indexerare, substantiv singular. Jag kan också konstatera att indexering är en fråga om bedömning och att olika indexerare inte bedömer samma dokument lika. Samma sak förelåg när det gäller användarna i min undersökning. Oftast kommer de överens om det övergripande ämnet i objekten och många användare nöjde sig med den typen av taggning, dock fanns det de som taggar mer specifikt. Dessa taggar begagnas inte alls i samma utsträckning som de bredare taggarna.

Vad gäller konsistensen så är det självklart att mitt urval och det faktum att jag endast har beräknat konsistensen på de 20 första användarna inte kan ge en sann bild av konsistensen. Hade konsistensen beräknats på alla användarna tror jag att resultatet sett lite annorlunda ut. Jag vågar dock påstå, med mitt material som grund, att några större skillnader inte skulle märkas. Genomgående tenderar användarna att tillskriva samma eller liknande taggar, vilket tydligt visar att de har tillgång till samma taggningsverktyg som användarna före dem, vilket i min undersökning är Delicious rekommenderingsverktyg.

## 8.4 Vidare studier

Jag har tidigare konstaterat att den genomsnittliga konsistensen över det totala antalet objekt uppgick till 28,34 %. Med stöd i mitt kapitel angående problematiken i folksonomier, se 2.2, anser jag att detta resultat är viktigt att lyfta fram. Då mitt utrymme är begränsat kan min undersökning endast visa en ögonblicksbild av förhållandet i Delicious. Därför menar jag att med tanke på den konsistens jag kunnat fastställa och med min diskussion som utgångspunkt vore det intressant med en liknande studie men som då är mer omfattande. Fler objekt analyseras, synonymer inkluderas och att indexeringskonsistensen beräknas på samtliga användare samt att en beräkningsmodell som tar hänsyn till de mer övergripande taggarna används. Det skulle kunna ge en tydligare och mer rättvisande bild av de förhållanden som faktiskt föreligger på Delicious.

Då den som taggar är av största vikt för taggningen och vilken kvalitet taggningen håller kan det också vara av intresse att studera om användare med olika bakgrunder taggar olika, och hur olika de i sådana fall taggar. Samtidigt anser jag att det skulle vara intressant att studera hur användarna tänker när de tillskriver taggar. Om de endast taggar för sin egen skull eller tänker de även på att andra med samma intresse kan vilja hitta objektet, och taggar de då utifrån det och om de tänker på att uppehålla indexeringskonsistensen när de tillskriver taggar.

Syftet med den här uppsatsen var att undersöka indexeringskonsistensen i en folksonomi och jag kan konstatera att resultatet blev att konsistensen, trots det fria vokabulären, kan ses som förhållandevis högt. Då utrymmet i en magisteruppsats är begränsat kan jag inte dra några generella slutsatser, däremot ser jag fram mot mer uttömmande studier där möjliga mönster kan skönjas, samt att dessa mönster kan appliceras på alla folksonomier. De mönster som ändock uppträtt är framför allt att användarna prefererar substantiv i singularform men också att de framför allt nyttjar övergripande taggar, samt att de i viss utsträckning väljer att begagna sig av möjligheten att skapa egenkonstruerade taggar.

# Sammanfattning av uppsatsen

Uppsatsen handlar om systemet folksonomi vilket är ett system som ger människor möjligheten att tillskriva objekt taggar för att underlätta återvinningen av objektet. En nackdel är dock tillämpandet av fri vokabulär. Användarna kan tillskriva vilket ord de vill, i vilken form de vill. Det gör det svårt med återvinningen då det inte finns något i systemet som talar om vilket ord som bör användas. Detta gör det intressant att undersöka indexeringskonsistensen mellan användare.

Syftet med uppsatsen är att undersöka indexeringskonsistensen i en folksonomi vilket görs genom frågeställningarna

1. Hur konstrueras taggar med avseende på böjningsvarianter och hur vanligt förekommande är böjningsvarianter?
2. Vilken utbredning har närliggande ord i form av hyponymer och meronymer bland användarna?
3. I vilken utsträckning begagnar användarna sig av egenkonstruerade taggar och vilken inverkan har de på indexeringskonsistensen?

En folksonomi är ett socialt nätverk som erbjuder användarna att spara ner objekt, enda kravet är att objektet har en URL. Då en fri vokabulär tillämpas vid taggingen innebär det att det inte finns några fasta relationer mellan taggarna och att exempelvis olika böjningsvarianter kan tillskrivas. Detta är också en av de stora nackdelar som har förts fram, den fria vokabulären försvårar återvinningen. Samtidigt är indexeringen i sig beroende av två steg, konceptuell analys och översättning. Indexeraren måste först bedöma vad objektet handlar om och sedan hitta ord som uttrycker detta vilket inte är helt lätt. Här uppstår den problematik som alltid föreligger vid bedömningsprocesser.

Lancaster (1998, s. 71) har identifierat sju faktorer som påverkar indexeringskonsistensen:

- |    |  |
|----|--|
| 1. | Antalet tillskrivna termer                   |
| 2. | Kontrollerad alternativt fri vokabulär       |
| 3. | Vokabulärens omfång och specifitet           |
| 4. | Ämnets utmärkande egenskaper och terminologi |
| 5. | Indexeraren                                  |
| 6. | Tillgängliga redskap vid indexeringen        |
| 7. | Det indexerade objektets omfång              |

Vidare påverkar också typen av indexering. Om användaren har valt att begagna sig av en selektiv indexering eller en uttömmande indexering.

De metoder som har använts var, innehållsanalys för att fastställa utformningen av de kodningsscheman som begagnades under undersökningen samt Rollings mått. Rollings

mått är ett beräkningsmått på indexeringskonsistensen mellan två användare. Måttet har dock använts på en grupp användare vilket innebär att indexeringskonsistensen beräknades i en parkonstellation där alla användarna medverkade. Sedan beräknades ett snitt av de resultat som framkommit.

Resultatet som framkommit kan sammanfattas enligt följande:

Framför allt är det substantiv som tillskrivs och då främst i singularform. De adjektiv som identifierats har nyttjats sparsamt och varit skrivna i grundform. De verb som förekommit har främst varit skriva i grundform eller futurum. Både hyponymer och meronymer begagnades även om de inte alls nyttjats i samma utsträckning som substantiv. Hyponymer var det som används mest då meronymer är svårare att begagna sig av eftersom förhållandet alltid ska vara 'del av'. Förekomsten av egenkonstruerade taggar låg på hela 35,3 % vilket bör anses som högt. Merparten av dessa taggar begagnades endast en gång, sju stycken taggar användes mer än en gång på samtliga objekt. Det fanns också tre taggar som utmärker sig, *todo*, *toread* och *howto* vilka var taggar som oftast blev tillskrivna på samma sätt.

Den genomsnittliga indexeringskonsistensen för samtliga objekt ligger på 28,34 %. Några stora problem är användandet av skiljetecken och att en del av de egenkonstruerade taggarna skulle kunnat delas upp i två taggar. Viktigt är dock att dessa beräkningar endast gjorts på de tjugo första användarna för varje objekt.

Några lösningar på hur indexeringskonsistensen skulle kunna ökas utan att större förändringar krävs är, exempelvis att göra det omöjligt att begagna skiljetecken vid tillskrivandet av taggar och att antalet taggar som gick att tillskriva begränsades.

## Källförteckning

Booth, Pat F (2001). *Indexing: the manual of good practice*. München: K G Saur.

Bruce, Robert (2008). Descriptor and folksonomy concurrence in education related scholarly research. *Webology*, vol. 5, no. 3. Tillgänglig som:  
<http://www.webology.ir/2008/v5n3/a59.html> [2009-04-13]

Bryman, Alan (2001). *Samhällsvetenskapliga metoder*. Malmö: Liber.

Caplan, Jeremy (2007). *hacking toward happiness*. Tillgänglig som:  
<http://www.time.com/time/magazine/article/0,9171,1635844,00.html> [2009-06-28]

Cruse, Alan D (2004). *Meaning in language: an introduction to semantics and pragmatics*. New York: Oxford university press.

Delicious (2009). *About Delicious*. Tillgänglig som: <http://delicious.com/about> [2009-04-20]

Folk (2001). *Online etymology dictionary*. Tillgänglig som:  
<http://www.etymonline.com/index.php?term=folk> [2009-06-13]

Goddard, Cliff (1998). *Semantic analysis: A practical introduction*. New York: Oxford university press.

Golder, Scott A & Huberman, Bernardo A (2005). *The structure of Collaborative tagging systems*. Tillgänglig som:  
<http://www.hpl.hp.com/research/idl/papers/tags/tags.pdf> [2007-02-01]

Google (2009). *Varför använda google?*. Tillgänglig som:  
[http://www.google.se/intl/sv/why\\_use.html](http://www.google.se/intl/sv/why_use.html) [2009-06-21]

Halvorsen, Knut (1992). *Samhällsvetenskaplig metod*. Lund: Studentlitteratur.

Hood, Stephen (2008) *Delicious is 5!*. Tillgänglig som:  
<http://blog.delicious.com/blog/2008/11/delicious-is-5.html> [2009-07-17]

Johnston, Elizabeth (1997). *Lecture 14: Thought & language*. Tillgänglig som:  
[http://pages.slc.edu/~ebj/IM\\_97/Lecture14/L14.html](http://pages.slc.edu/~ebj/IM_97/Lecture14/L14.html) [2007-04-27]

Lancaster, F. W. (1998). *Indexing and abstracting in theory and practice*. Third edition 2003 London: Library association publinsing.

Leininger, Kurt (2000). Interindexer consistency in PsycINFO. *Journal of Librarianship and Information Science*, vol. 32, no. 4, s. 4-8 Även tillgänglig som:  
<http://lis.sagepub.com/cgi/content/abstract/32/1/4> [2007-05-10]

Lyons, John (1981). *Language and linguistics: An introduction*. Cambridge: University press.

- Morville, Peter (2005). *Ambient Findability*. Sebastopol: O'Reilly Media.
- Morville, Peter & Rosenfeld, Louis (2006). *Information architecture for the world wide web*. Sebastopol: O'Reilly Media.
- Nauman, Mohammad & Khan, Shahbaz (2007). *Using personalized web search for enhancing common sense and folksonomy based intelligent search systems*. Tillgänglig som:  
<http://delivery.acm.org.lib.costello.pub.hb.se/10.1145/1340000/1331883/30260423.pdf?key1=1331883&key2=7762518321&coll=Portal&dl=ACM&CFID=27809589&CFTOKEN=34535284> [2009-03-27]
- Noruzi, Alireza (2007). *Folksonomies: Why do we need controlled vocabulary?* Webology, Volume 4, Number 2. Tillgänglig som:  
<http://www.webology.ir/2007/v4n2/editorial12.html> [2009-03-25]
- Pullum, Geoffrey K. (1991). *The Great eskimo vocabulary hoax and other irreverent essays on the study of language*. Chicago: University Of Chicago Press. S. 166-69. Även tillgänglig som: <http://web.mit.edu/jfitzpat/www/FoL/Pullum-Eskimo-VocabHoax.pdf> [2007-04-20]
- Recommending smart tags in a social bookmarking system* (2007). Tillgänglig som:  
<http://www.kde.ce.uni-kassel.de/ws/eswc2007/proc/RecommendingSmartTags.pdf> [2009-04-11]
- Rolling, L. (1981). Indexing consistency, quality and efficiency. *Information processing and management*, vol 17, no. 2, s. 69-76. Även tillgänglig som: doi: 10.1016/0306-4573(81)90028-5 [2007-03-01]
- Specia, Lucia & Motta, Enrico (2007). *Integrating folksonomies with the semantic web*. Lecture notes in computer science, Vol. 4519, p 624-639. Även tillgänglig som:  
<http://www.eswc2007.org/pdf/eswc07-specia.pdf> [2009-04-13]
- Stock, Wolfgang G (2007). Folksonomies and science communication. *Information services and use*, 27. s. 97-103. Även tillgänglig som: [http://wwwalt.phil-fak.uni-duesseldorf.de/infowiss/admin/public\\_dateien/files/1/1194272247inf\\_servic.pdf](http://wwwalt.phil-fak.uni-duesseldorf.de/infowiss/admin/public_dateien/files/1/1194272247inf_servic.pdf) [09-06-13]
- Surowiecki, James (2006). Joshua Schachter, 32: How tags exploit the self-interest of individuals to organize the web for everyone. *Technology review*. Tillgänglig som:  
<http://www.technologyreview.com/tr35/Profile.aspx?TRID=432&Cand=T&pg=1> [2009-07-17]
- Taxonomy (2001). *Online etymology dictionary*. Tillgänglig som:  
<http://www.etymonline.com/index.php?search=taxonomy&searchmode=none> [2009-06-13]
- URL (2009). *Wikipedia*. Tillgänglig som: <http://sv.wikipedia.org/wiki/URL> [2009-10-26]

Vander Wal, Thomas (2005). *Explaining and showing broad and narrow folksonomies*.  
<http://www.vanderwal.net/random/entrysel.php?blog=1635> [2007-02-01]

Vander Wal, Thomas (2007). *Folksonomy: Folksonomy soilage and definition*.  
Tillgänglig som: <http://www.vanderwal.net/folksonomy.html> [2007-05-24]

Wi-Fi Alliance (2007). *Knowledge center – what is 802.11b?*  
[http://www.wi-fi.org/knowledge\\_center\\_overview.php?docid=3266](http://www.wi-fi.org/knowledge_center_overview.php?docid=3266) [2007-12-03]

Wu, Harris (2006). *Harvesting social knowledge from folksonomies*. Conference on  
hypertext and hypermedia. Tillgänglig som:  
<http://delivery.acm.org.lib.costello.pub.hb.se/10.1145/1150000/1149962/p111-wu.pdf?key1=1149962&key2=7625209321&coll=Portal&dl=ACM&CFID=29119437&CFTOKEN=70037404> [2009-04-06]

# Bilaga

## Objekt A

- system:unfiled
- video
- codec compression video
- codec comparison encode compression encoding faq film howto internet computing technology reference codecs video
- av code faq howto internet film computing documentation video codecs compression reference codec technology
- video codecs reference
- video codecs comparison reference review tutorial technology
- technology
- read
- video codecs
- video encoding
- video codecs compression reference codec technology computing
- videoedicion
- video codecs reference
- video codec
- video codecs compression reference codec technology computing
- programming
- system:unfiled
- everything you ever wanted to know about video codecs
- video codecs
- video codec
- videocodecs
- software
- codecs
- video codecs reference
- video
- tech video reference
- howto interesting technology tips tutorial video code
- video
- system:unfiled
- codec video computing comparison
- system:unfiled
- codecs encoding reference toread video xvid compression
- howto video codec reference tips tutorial
- system:unfiled
- video codecs
- system:unfiled
- codecs compression reference
- system:unfiled
- codecs video
- video codecs compression reference
- school video
- system:unfiled
- video codec resource
- video codec
- system:unfiled
- articles codecs compression iptv video
- learn video codeces
- video compression codec code
- video codecs
- howto tutorial video codec
- codecs tips howto comparison
- compression video
- unread
- video codec tutorial



- video codecs reference technology
- video codecs compression reference codec computing technology internet howto
- documentation
- video codecs reference
- codecs
- codecs tutorial articles multimedia video readlater
- video video:reference
- video codecs reference computing
- video codecs compression reference codec technology computing via:atomicmpc
- video compression codec technology computing
- codecs
- tutorial video
- video
- system:unfiled
- code codecs video codec
- codecs internet toread tutorial video
- toread codec video tutorial
- video codec reference
- video codec
- video codecs compression reference codec technology computing
- review
- system:unfiled
- video codec tips reference
- video codecs interesting compression reference codec
- video codecs compression codec technology computing
- articles code film compression video codecs reference
- video codecs xdiv divx review
- codec tech video
- video codecs compression
- video codec
- video codecs
- codec computing technology reference compression codecs video
- video codecs reference computing codec technology
- video compression codec toread technology comparison encoding film multimedia review
- system:unfiled
- computing encode reference tech technology video codecs
- tutorial video ???+ software tips
- video multimedia codecs
- video codecs compression reference computing codec technology internet divx code film iptv
- video
- video compression codecs reference computing codec technology internet review interesting faq tech
- video codecs compression reference computing technology tutorial internet encoding comparison code
- articles codec to.read
- video codecs compression reference
- articles code codec codecs comparison encoding faq divx compression reference computing technology internet video
- video codecs codec
- system:unfiled
- codecs video compression reference
- articles codec comparison reference toread codecs
- video codecs compression reference technology
- system:unfiled
- video codecs
- divx toread technology
- reading
- codec bluray faq
- video codecs
- codecs reference reading
- video codecs reference
- video media
- compression encoding video
- video codecs compression computing mmw
- video codecs

- video tech media code motion
- video codecs
- system:unfiled
- av
- to-read
- codec comparison video technology reference
- articles codecs compression reference video
- video codecs reference
- internet codecs
- video codec compression confronto
- codecs compression video
- video codecs reference
- interesting technology video
- video
- codecs compression computing articles video
- video codecs compression reference internet
- video codecs
- video codecs tutorial
- codecs video
- video\_codecs
- compression video
- video codecs compression
- encoding
- codec compression film internet reference technology video codecs encode encoding
- video codecs
- codecs video technology comparison compression
- video reference
- video codecs
- reference articles
- compression reference video
- video codecs mpeg
- codec video reference
- video codec review
- video codecs compression computing film technology reference
- video codecs compression computing internet film
- technology
- system:unfiled
- programming
- video codecs
- codecs video
- video
- video codecs
- system:unfiled
- compression video computing comparison technology knowledge toread codecs
- video codecs film reference
- codecs video computing
- codecs compre compression computing video
- video compression film internet computing codecs toread
- codecs
- video codec avi xvid divx
- compression computing film internet video codecs iptv
- video codecs compression mpeg
- video codecs
- video codecs review
- video codecs compression
- compression video codecs
- video computing
- video internet compression codecs
- video
- codecs video
- video
- video codecs
- video codecs

- codecs video
- codecs video
- codecs
- codecs video interesting
- video codecs
- system:unfiled
- video codecs
- video codecs
- codecs film
- system:unfiled