

Umeå Universitet  
Institutionen för Datavetenskap  
Emergenta system C, ht -06  
Projektrapport

# Folksonomi

Power to the people

Fredrik Johansson ([dit02fjs@cs.umu.se](mailto:dit02fjs@cs.umu.se))  
David Markusson ([bio02dmn@cs.umu.se](mailto:bio02dmn@cs.umu.se))

Handledare:  
Jonny Pettersson

### **Abstract**

In recent years there has been a development of Web 2.0 services which supply a simpler way of communicating and sharing information over the Internet. Earlier, Web 2.0 has been a synonym for the semantic web which is a universal medium for giving web content a meaning for computer machines. Out of Web 2.0 services Folksonomies has been developed which is a way for users to tag their own Internet material for later and easier information retrieval.

Folksonomies are useful in that way it categorizes information instead of how the earlier semantic web classified web pages and its information. It uses a kind of bottom-up strategy created by users own intuitive way of organizing information. A strategy, where the human intelligence is embedded in the system.

## Innehållsförteckning

1. Introduktion .....	sida 4
2. Metodbeskrivning .....	sida 4
3. Semantiska webben.....	sida 5
4. Taxonomier.....	sida 5
5. Folksonomi .....	sida 6
5.1 Två typer av Folksonomier.....	sida 7
5.2 Komplexa system.....	sida 8
5.3 Emergens.....	sida 9
5.4 Fördelar med taggning.....	sida 10
5.5 Nackdelar med taggning .....	sida 10
6. Experiment .....	sida 10
6.1 Hypotes .....	sida 10
6.2 Resultat.....	sida 10
6.3 Experiment: Diskussion.....	sida 10
7. Diskussion .....	sida 11
8. Referenser .....	sida 12



Detta projekt kommer att behandla begreppet folksonomi och vad det innebär för såväl användare som Internet i helhet. Ett webbaserat folksonomiexperiment har vid sidan om projektet startas med förhoppningen att vid ett senare skede kunna verifiera huruvida dessa system kan anses ha ett visst emergent beteende [9]. Eftersom testperioden för experimentet sträcker sig längre än projektperioden måste detta tas i åtanke då det resultat som visas endast är snitt ur systemets livscykel.

Enligt Halpin 2006 har system som använder sig av folksonomi ett komplext beteende med ett stort antal enheter (användare), social feedback (från de taggar som används), möjligheten att interagera mellan användare varandra, ingen central ledare och uppvisandet av ett icke linjärt system. Dessa beskrivningar fyller kraven som ställs på ett emergent system och därför har experimentet som nämndes ovan startats för att verifiera eller förkasta detta påstående.

## **2. Semantiska webben**

Den semantiska webben som projekt startades i början av 2000-talet (leddes av Tim Berners-Lee) för att underlätta informationsutbyte mellan system. Detta görs genom att skapa semantiska kopplingar till data på webben som gör det lättare för dataprocesser att hitta och återfå information om dessa data. Semantiska webben använder sig av Resource Description Framework (RDF) och Web Ontology Language (OWL) vilka är två tekniker för att skapa just dessa semantiska kopplingar i form av s.k. metadata [1]. Metadata är data om det egentliga datat och representerar vad det egentliga datainnehållet handlar om och dess väsentliga betydelse (semantiken). De två ovan nämnda teknikerna använder sig av olika metoder för att skapa dessa metadata. RDF är en metadatamodell som bygger på att skapa ett påstående om datat i form av subject-predikat-objekt. OWL har designats just för att datats innehåll ska processas av olika applikationer [2]. Semantiska webben blir därför ett verktyg som förlänger webben genom användandet av dessa tekniker för att hitta information på webben, en så kallad taxonomi.

## **3. Taxonomier**

Taxonomi är användandet och läran om klassificering och i detta fall klassificering av data. Semantiska webben och bibliotek kan ses som system där man klassificerar data eller böcker efter vilken kategori dom tillhör. Systemet byggs upp av en administratör som bestämmer vilket data som skall tillhöra vilken kategori. Data och böcker blir därefter klassificerade att tillhöra en viss kategori och det är efter denna förutbestämda kategori man får söka efter för att hitta det man söker.

Som Internet till största del ser ut idag använder sig sökmotorer av klassificeringen av data. Detta sägs vara en sorts top-down strategi, men kan även kallas för Hierarchical-enumerative. Detta innebär en uppbyggnad av en trädstruktur som skapas av att dela in information i smalare och smalare områden som då utgör grenarna (Quintarelli, 2005). Klassificering är en metod som har ett konstant ramverk vilket leder till svårigheter för användaren att intuitivt veta var i trädet informationen som denne söker efter finns.

Hela tiden växer det dock fram nya metoder för att strategiskt kunna lösa problemet. En strategi är taggningssystem där användaren själv taggar sitt material för att sedan lättare hitta tillbaka till informationen.

Halpin 2006 tar i sin artikel upp att taggningssystem, så som folksonomi, inte använder sig av en ontologi (datarepresentation) som den ovan där det rör sig om ett system som klassificerar data. Halpin menar istället på att taggningssystem är ett system som kategoriserar data och att all kategorisering sker efter hand, s.k. "post-hoc kategorisering". Inom folksonomi är det ju användaren som kategoriserar allt data själv och det finns alltså inga förutbestämda kategorier.

Kategorisering på detta vis skapar en obestämbär ontologi som kan fortsätta utvecklas mellan användare och hitta gemensamma nämnare efter hur data är tolkat och då placerat i grupper/kategorier (Jacob 2004).

Taxonomier har begränsningar och passar inte i alla domäner. Där de passar bäst är då domänen har ett relativt begränsat innehåll med förbestämda kategorierna och stabilt och kontrollerat data (Quintarelli, 2005). Ett bra exempel är just ett bibliotek som är konstant till sin storlek och har väldefinierade kategorier varefter böcker sorteras och sedan placerar i uppmärksamma hyllor. Men som synes är platsen begränsad och det går inte att ta in hur mycket information som helst då det inte finns plats för alla böcker på hyllorna.

#### 4. Folksonomi

Ordet Folksonomi definierades 2004 av Thomas Vander Wal för att beskriva ett fenomen som uppkommit på webben. Vander Wal beskrev folksonomi som "tagging that works" (Koivunen, 2005). Som tidigare nämnts är folksonomi en informationsmetodik för att återfinna information på webben. Denna styrs av systemets användare då de "kategoriserar" informationen med hjälp av taggar för att sedan kunna hitta informationen lättare. Den mest nämnda webbsidan som påvisar folksonomi är del.icio.us.

*("A social bookmarks manager. It allows you to easily add sites you like to your personal collection of links, to categorize those sites with keywords, and to share your collection not only between your own browsers and machines, but also with others" (Schachter, 2004, om del.icio.us))*

På del.icio.us kan man som användare bygga upp sitt eget nätverk av länkar till webbplatser och annan information sparad på nätet genom att tagga länken med nyckelord. Taggarna går sedan att söka på likt en sökmotor för att sedan få tillbaka information kopplad till just denna tag. Det centrala i detta system är att det är användaren som bestämmer hur informationen skall taggas, vilket gör att ett kategoriseringssystem byggs upp utan någon högre påverkande enhet [3]. Data kan därför ges en fri tolkning av hur man upplever informationen och ger olika synvinklar som kan stämma överens med fler användare. Varje informationsobjekt visar alltid upp sina mest vanliga taggar vilket representerar användares olikheter i tolkningar av informationen. Detta gör det möjligt för likasinnade användare att

snabbt hitta den sökta informationen då olika tolkningar kan leda till samma informationskälla [5].

System som det som ovan beskrivits bygger upp en social struktur med sociala interaktioner där användaren kan dela med sig av sina taggar och den information som är relaterat till dess. Allting är uppbyggt av beteendet mellan användare och informationen som även bygger upp en konceptuell och språklig struktur i användarsamhället (Marlow, 2006).

Namnet folksonomi är sammansatt av orden folk och taxonomi vilket skall beskriva det som en taxonomi gjord av folket, dock skall det inte tolkas som en ren taxonomi då det inte finns någon ensam styrande enhet och det kan dessutom förekomma flera tolkningar (metadata) för samma sort data.

Shirky, 2005, lägger fram argument för en folksonomi som ska utveckla den semantiska webben såväl som underlätta sökandet av information på Internet som; en ontologi är överskattad då dagens taggningssystem är så mycket mer organiska än de tidigare använda organiseringsystemen.

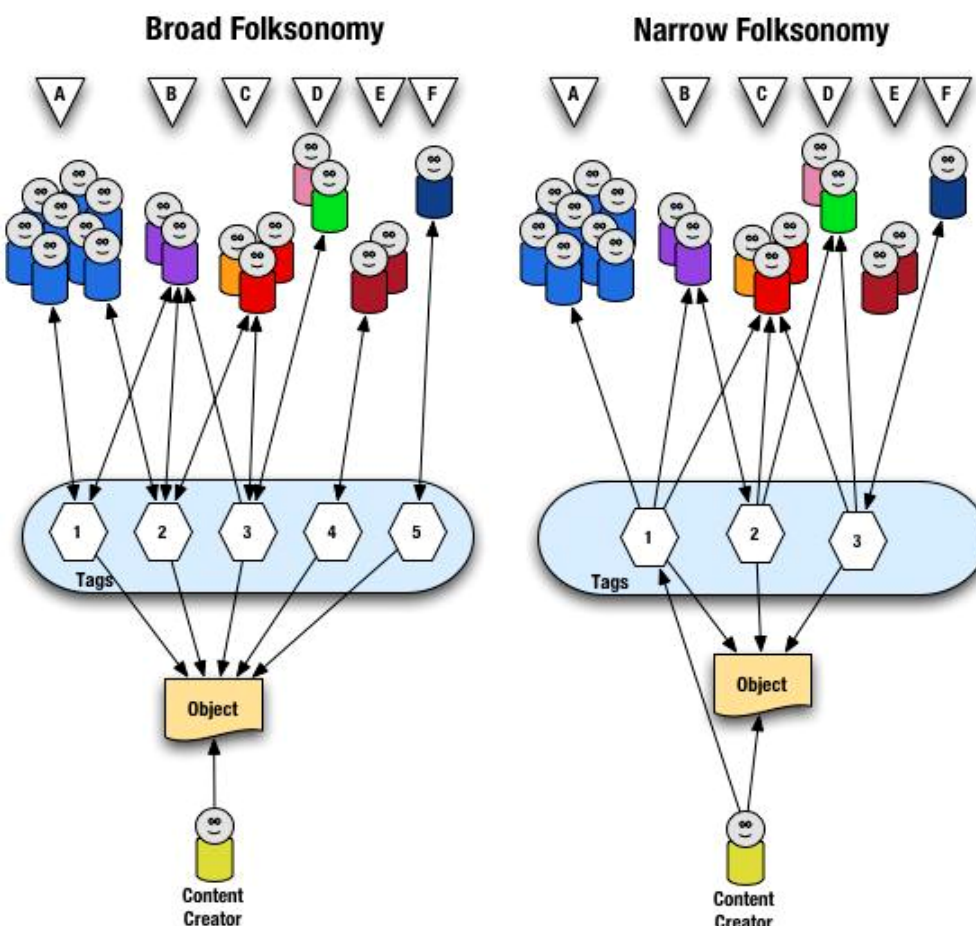
Tidigare sätt att klassificera data använder sig av en top-down metod medan taggning system har börjat använda sig av en bottom-up metod.

En down-up strategi innebär att subjekt eller data istället bryts ner till individuella koncept vilka bildar regler för att bygga upp och konstruera rubriker för samansatt komplexa subjekt (Quintarelli, 2005). Vi bryter alltså ner scheman till mindre delar och grupperar dessa för att få en högre grupp eller kategori. Nya delar i systemet kan läggas till och skapa nya koncept och rubriker utan att förstöra övrig struktur.

#### **4.1 Två typer av Folksonomier**

Thomas Vander Wal som myntade ordet Folksonomi har tagit fram två typer av Folksonomier, vardera med olika egenskaper och användande. Dessa benämns som bred respektive smal folksonomi.

Ett typiskt exempel för bred Folksonomi är inom webbsidan del.icio.us där objekt kan ha samma tag från flera olika användare. Detta innebär att ett objekt kan ha flera taggar som är precis samma och som vilket i slutändan gör att användare till slut kommer överens om ett visst antal taggar men där det även finns en grupp som håller sig till de mer ovanliga taggarna (Quintarelli, 2005).



En smal folksonomi kan istället hittas på *Flickr* där den som postar objektet även postar taggarna till objektet. Detta kan innebära en tag eller flera (Quintarelli, 2005) Fördelen med smal folksonomi trots det faktum att bred folksonomi är något rikare i sin funktion är att taggarna är direkt kopplade till objektet och sidor som flickr där objekten är foton skapar detta direkt något konkret att söka efter då det enda sättet att beskriva objektet är genom taggar [6].

#### 4.2 Komplexa system

Harry Halpin tar i sin artikel upp hur ett taggande system kan ta sig likheter med ett komplext system. Detta eftersom att varje användare kan ses som en agent som vill öka möjligheten att hitta data i framtiden (för din egen vinning) men även att dela sitt taggade material med andra för att minimera sin egen tid för taggande. Detta för att i gengäld få taggat material tillbaka.

Dessutom säger Udell 2004, hur tagging system ger en direkt feedback till användaren efter att användaren lagt till en tag. Genom tag clouds ser man hur andra har taggat då man lagt till sin egen tag och var sin egen tag placerades och vad man beskriver för andra användare. Om man ser att man beskrivit innehållet med en "felaktig" eller missledande tag så kan man genast ändra den till något som passar bättre.

Tag clouds är genom vilken feedbacken ges till användaren efter att användaren har lagt till en tag. Detta visas ofta i form av *tag clouds*, som är ett visuellt sätt att visa

vilka taggar som är populära eller har en betydande mening i sammanhanget. I ett moln som detta visas de viktigaste taggarna tydligare. Kan liknas med ett informationsträsk där den viktiga informationen har flytit upp till ytan.

Både del.icio.us och Flickr som är de mest erkända sajterna för folksonomi använder sig av dessa *tag clouds* vilket för användaren gör att hon kan snabbt se vilka taggar som är populära och även vilka taggar hon själv använder mest. Genom att använda sig av taggande inför man intelligens i systemet och denna intelligens är människans (Gruber, 2005).

.net adult advertising ajax apple architecture art article audio blog blogging blogs book books business car community computer cool css culture daily design development download downloads dvd education entertainment fanfic fashion fic finance firefox flash flickr food forum free freeware fun funny furniture gallery game games google graphics gtd hardware health history home howto html humor illustration imported inspiration insurance interesting internet japan java javascript jobs language library life lifehacks links linux mac maps marketing media microsoft mobile money movie movies mp3 music news online opensource osx personal photo photography photos photoshop php podcast politics porn portfolio productivity programming python radio rails recipes reference research resources ruby science search security seo sex sga shop shopping social software tech technology tips tool tools toread travel tutorial tutorials tv ubuntu video videos web web2.0 webdesign webdev wiki windows wordpress work writing youtube

*Exempel på ett "tag cloud" från del.icio.us*

### 4.3 Emergens inom Folksonomi

Adam Mathes 2004 tar upp ett antal exempel på hur en tag inom folksonomi kan starta en diskussion om begrepp och deras definition. Han vill också visa genom dessa exempel att det finns en sorts kommunikation mellan användarna och hur metadata kan förenkla en ad-hoc gruppering där information läggs till eftersom, och har en gemensam nämnare.

Ett exempel är taggen 'Irak' där foton om kriget i Irak, amerikanska trupper och demonstrationer om kriget har lagts upp. Genom detta har det startat en bildlig konversation och en definition om termen och dess innebörd.

Likaså från tagen 'sometaithurts' (står för 'so meta it hurts') under vilken bilder om användandet av och runt om webbsidan Flickr lagts upp och fortsätter läggas upp. Det kan t ex. vara screenshots av bilder i Flickr eller människor som använder Flickr. Detta visar på att taggar har blivit ett kommunikationsredskap då man på detta vis skapat en konversation och enda sättet att ta del av konversationen är att tagga sin egen bild med samma tag som inlägg.

Peter Merholz 2004 ser Taggning som ett sätt att få ordning på kaoset, där han menar kaoset av information på Internet. Han säger att potentialen av att fritt kunna tagga det material man kommer över skapar en stig över hur användaren rör sig genom landskapet (i detta fall webbsidor på Internet) vilket vanligen inte är efter de vanliga stigarna. Vad Merholz också vill säga är att en bra landskapsdesigner låter användare skapa deras egna vägar genom rymden för att använda, och sen efter dessa göra vägarna bestående likt att lägga plattor där någon har gått. Detta just för att försäkra optimal användbarhet då man navigerar genom information på nätet

#### 4.4 Fördelar med taggning

En stor anledning till att taggning blivit så populärt och genomslående är att användaren slipper alla överenskommelser och regler för att kunna "klassificera" det data som intresserar denne (Halpin, 2006). Den semantiska webben har taxonomier och regler som måste följas för att metadatat eller taggen som blir knytet till datat blir korrekt. Detta är begränsningar som inte behövs inom folksonomin då den egna tolkningen räcker för att placera ut en tag och bidra till ett delat värde, (Mahes 2006). Med Folksonomi behöver en grupp av användare inte komma överens om en hierarki för taggar eller en detaljerad taxonomi.

Folksonomier är också "*Inklusiva*". Med detta menas att det inkluderar allas vokabulär och reflekterar alls behov utan social, kulturell eller politisk påverkan. Eftersom Folksonomier inkluderar alternativa åsikter tillsammans med populära så skapar de möjligheten att utforska en s.k. "Long tail". Med detta menas att det finns ett en statistisk normalfördelning, och att det från detta växer det ut en svans som representerar minoriteten. Alltså de bägge sidorna på normalfördelningen [8].

Ännu en fördel med användandet av taggningssystem är just distribueringen av arbetet för att bygga upp systemet. Varje enstaka användare bidrar hela tiden med sina taggar för att bygga ut systemet ytterligare, (Marlow, 2006).

#### 4.5 Nackdelar med taggning

Som ovan beskrivits finns det många fördelar med taggning, dock måste det även nämnas att det även finns nackdelar. Taggar är inte så avancerade att de känner igen synonymer och ord som betyder precis samma sak. T ex. att söka efter New York så kan det behövas sökningar efter NY, nyc och newyork även då alla dessa betyder samma sak.

En annan svårighet med metadata eller taggar är att enstaka ord sällan uttrycker hela det semantiska innehållet. Och läggs fler ord till i beskrivningar blir det svårare att systematisera. Svårigheten med metadata är alltså att göra ett system som är lagom komplext, men ändå flexibelt och användbart för exempelvis sökning.

Ett sätt att brygga detta problem är att skapa ordlistor som begränsar vilka ord man kan använda till taggningen. Exempelvis kan man för en bild på en liggande, brun hund använda taggen "brun" men inte "ljusbrun" och "ligga" men inte "liggande". En sådan ordlista där endast vissa ord är tillåtna brukar kallas för ett "collabulary" (collaborative vocabulary). På svenska kan det sägas betyda ungefär "samarbetsordlista". Det är en ordlista där användarna lägger till ord, som sedan kan användas för taggning. Här kan man också bygga in kontrollsystem som gör att orden måste godkännas av en administratör eller genomgå en automatiserad verifieringsprocess. Exempel på "collabularies" är wikisidor, som Wikipedia.

Ytterligare ett problem är då samma tag kan beskriva olika objekt (Merholtz 2004), taggar blir därför överflödiga då t ex. en webbsida om trummor måste taggas som både 'trummor' och 'musik' även fast trummor handlar om musik (Halpin, 2006).

Taggar kan dessutom inte ge någon större precision om vad webbsidan eller datat specifikt handlar om. Taggarna till datat kan ge en grov översikt om vad det handlar om men några detaljer kan vara svårare att se. Låt om säga att det finns en webbsida om ett regeringsval. Taggarna för denna sida kanske säger just 'val' men inte något om hur valet slutade eller hur många mandat respektive parti fick (Halpin, 2006).

## 5. Experiment

Beskrivning, metod, hypotes mm

## 6. Diskussion

Internet börjar gå mot en utveckling där användarna får mer och mer inflytande. Man har hittills försökt strukturera upp Internet för att hålla det ordnat och lättnavigerat men detta kanske är lättare sagt än gjort i och med att Internet fortsätter att växa och växa. Kanske är det så att system som använder sig av folksonomi kan lösa problem som strukturering av information på nätet. Hittills ha man ju sett att taggningssystem faktiskt ger nytta för användare och skapar en strukturering av information som man sedan själv använder och som även andra kan ha vinning av. Självklart finns det nackdelar som för tillfället kan missleda användare till fel information. Även svårigheter som att hitta mer specifik information då taggar ofta är ett övergripande ord om innehållet, ett sorts kategoriseringsnamn. Men detta betyder inte att det skulle överväga fördelarna.

Det ser ut som att dessa system inte bara skulle kunna göra nytta på nätet utan även inom företag som vill strukturera deras information eller för att se vilka trender som gäller för tillfället ute på marknaden. Ett exempel är IBM som har tänkt införa folksonomi som en metod för att strukturera information inom sitt intranät. Ett argument till att IBM vill införa detta är just att information inom deras intranät förändras så snabbt att det för en central enhet skulle ta alltför lång tid att omorganisera all information. Detta skulle likaväl kunna appliceras på t ex. marknadsundersökningar i samhället för att ta reda på trender som svänger snabbt. För vad skulle vara bättre än om användarna själva fick styra hur den nuvarande trenden skulle se ut (genom att applicera folksonomi).

System som detta är fascinerande då de helt byggs upp utav användare byggs som egentligen inte vet om det. Och den största fördelen är ju att de organiserar något precis så som användarna vill ha det organiserat och inte som tidigare hierarkiska kategoriseringscheman som man trodde var optimala. Folksonomi är måhända inte det ultimata inom klassificering eller kategorisering men visst verkar det vara en bit på vägen till struktur på något som börjar likna kaos.

## 7. Referenser

Tim Berners-Lee, Jim Hendler, and Ora Lassila *The Semantic Web Scientific American*, 2001

Harry Halpin and Hana Shepard, *Evolving ontologies from Folksonomies: Tagging as a complex system*, 2006

Elin K. Jacob *Classification and categorization : a difference that makes a difference*, 2004

M. Koivunen, *Annotea and Semantic Web Supported Annotation*, 2005

Mahes *Folksonomies – Cooperative classification and communication through shared metadata*, 2004

Cameron Marlow<sup>1</sup>, Mor Naaman<sup>1</sup>, danah boyd<sup>1,2</sup>, Marc Davis<sup>1,2</sup>, *Position Paper, Tagging, Taxonomy, Flickr, Article, ToRead*, 2006

Adam Mathes, *Folksonomies - Cooperative Classification and Communication Through Shared Metadata*, 2004

Peter Meroltz 2004 *Metadata for the masses*, 2004

Jon Udell 2004, *Collaborative knowledge gardening*, 2004

Joshua , Schachter. *Del.icio.us About Page*. <http://del.icio.us/doc/about>. 2004.

Clark Shirky, *Ontology is overated*, 2005

Daniel H. Pink, *Folksonomies*, New york times December 11, 2005

Emanuele Quintarelli, *Power to the people*, 2005

[1] [http://en.wikipedia.org/wiki/Semantic\\_web](http://en.wikipedia.org/wiki/Semantic_web)

[2] [http://en.wikipedia.org/wiki/Resource\\_Description\\_Framework](http://en.wikipedia.org/wiki/Resource_Description_Framework)

[3] <http://en.wikipedia.org/wiki/Del.icio.us>

[4] <http://www.adaptivepath.com/publications/essays/archives/000361.php>

[5] <http://en.wikipedia.org/wiki/Folksonomi>

[6] [http://www.personalinfocloud.com/2005/02/explaining\\_and\\_.html](http://www.personalinfocloud.com/2005/02/explaining_and_.html)

[7] [http://en.wikipedia.org/wiki/Web\\_2.0](http://en.wikipedia.org/wiki/Web_2.0)

[8] Ellyssa Kroski <http://infotangle.blogspot.com/2005/12/07/the-hive-mind-folksonomies-and-user-based-tagging/>

[9] Experiment: The Swedish Wishing Well, <http://wish.signar.se>