

MAGISTERUPPSATS I BIBLIOTEKS- OCH INFORMATIONSVETENSKAP  
VID INSTITUTIONEN BIBLIOTEKS- OCH INFORMATIONSVETENSKAP/BIBLIOTEKSHÖGSKOLAN  
2007:66  
ISSN 1654-0247

## PennTags

En kvantitativ empirisk studie av en read/write OPAC

CAMILLA ERIKSSON-ÅHL



HÖGSKOLAN I BORÅS  
VETENSKAP FÖR PROFESSION

© **Camilla Eriksson-Åhl**  
Mångfaldigande och spridande av innehållet i denna uppsats  
– helt eller delvis – är förbjudet utan medgivande.

Svensk titel: PennTags: En kvantitativ empirisk studie av en read/write OPAC

Engelsk titel: PennTags: A quantitative empirical study of read/write OPAC

Författare: Camilla Eriksson-Åhl

Kollegium: 2

Färdigställt: 2007

Handledare: Ingrid Johansson

Abstract: The aim of this thesis is to study the phenomenon of *read/write OPAC* by looking at how students at Pennsylvania University use *PennTags* in order to influence document descriptions in the library catalogue. The main questions to be answered are to what extent *PennTags* is being used in different subject areas and what users contribute to document descriptions through tagging in different subject areas. Studying the posts created by a sample of *PennTags* users from a quantitative approach I try to find out if there is reason to believe that the activity of users and users' possibility to make meaningful contributions are dependent of the subject matter being dealt with in documents described, as anticipated by the theory of *pace layering*.

I find that user interest in influencing the library catalogue through *PennTags* appears to be low in general, with exception for in the subject areas of the main classes *Language and literature* and *Social sciences* in *Library of Congress Classification*. Users do however make meaningful contributions to the document descriptions through tagging in 59 % of the cases. The results support only vague evidence of the connection anticipated by the theory of *pace layering* and I suggest that other explanatory models must be sought in order to understand user behaviour in systems like *PennTags*.

Nyckelord: taggning, OPAC, universitetsbibliotek, folksonomi, kontrollerad vokabulär

# INNEHÅLLSFÖRTECKNING

1. INLEDNING.....	1
1.1. Syfte och forskningsfrågor.....	2
1.2. Disposition .....	2
1.3. Begrepp och definitioner.....	3
1.4. Avgränsningar.....	4
2. BAKGRUND.....	5
2.1. Webb 2.0.....	5
2.1.1. Taggning och folksonomi .....	5
2.2. Bibliotek 2.0.....	7
2.3. PennTags.....	7
2.4. Kontrollerad vokabulär .....	8
2.4.1. Library of Congress Subject Headings .....	8
3. RELATERAD FORSKNING.....	10
4. TEORETISKA UTGÅNGSPUNKTER .....	13
4.1. Vad skulle taggning kunna tillföra dokumentbeskrivningar?.....	13
4.2. Pace Layering.....	14
5. METOD .....	16
5.1. Datainsamling och urval .....	16
5.2. Kodning och analys av materialet.....	17
6. RESULTAT .....	20
6.1. Aktivitet och ämnesområde .....	20
6.1.1. Att lägga till en post från Franklin till PennTags.....	21
6.1.2. Att lägga till en post från Franklin i ett projekt på PennTags .....	24
6.1.3. Att annotera en post från Franklin .....	25
6.1.4. Att tagga en post från Franklin .....	26
6.2. Tillförandegrad och ämnesområde .....	27
7. SLUTSATSER.....	29
8. ANALYS & DISKUSSION .....	30
9. SAMMANFATTNING.....	36
KÄLL- OCH LITTERATURFÖRTECKNING .....	38
Bilaga 1 – Tillförandegraderna .....	42

## 1. INLEDNING

Automatisk indexering med tillhörande fulltextsökning har länge framstått som det enda möjliga alternativet på webben eftersom den på grund av sin omfattning betraktas som omöjlig att indexera manuellt av en enda institution eller utifrån en enda enhetlig standard. Under slutet av 1990-talet dök emellertid webbtjänster upp där människor uppmuntrades att beskriva dokument och skapa metadata på sätt som kändes relevanta för dem själva. Fenomenet växte och fick med tiden ett namn – taggning – och den vokabulär som dessa taggar skapade kom att kallas folksonomi. Genom taggningen skapas informationssystem där exempelvis länkar till intressanta webbsidor utbyts mellan användarna. Det kanske mest kända exemplet på detta är *del.icio.us* (hädanefter kallad Delicious) men även *Flickr* för organisering av bilder.

Dessa taggningssystem kan vid en första anblick besanna precis varje farhåga en anhängare av vokabulärkontroll skulle kunna komma på. Marieke Guy och Emma Tonkin (2005) betraktar taggningssystemet Delicious och finner att taggarna ofta är otydliga, felstavade, mångtydiga och dessutom ofta för personliga för att ha ett allmänt värde. Synonym- och homonymkontroll är inte att tala om och termer i plural och singular blandas hur som helst. Fenomenet blåser på så vis liv i en gammal diskussion där kontrollerade och fria vokabulärer ställs mot varandra och taggning framställs som raka motsatsen till den indexering som till exempel bibliotekarier ägnar sig åt. Clay Shirky är förmodligen en av de allra största förespråkarna av taggning och i sitt kända tal *Ontology is overrated* går han till hård attack mot kanske särskilt bibliotekarier, men även andra som ägnar sig åt att kategorisera världen med hjälp av taxonomier. För Shirky representerar taggning ett radikalt brott med alla tidigare former av kunskapsorganisation och han menar bibliotekarier med hierarkiska kategoriseringar påtvingar människor någonting som de egentligen inte vill ha – och som de dessutom inte längre behöver.

Under de senaste åren har taggning närmast sig biblioteksvärlden då man i diskussionen om *Bibliotek 2.0* har föreslagit att användarna borde få vara delaktiga i indexeringsprocessen genom taggning eller på andra vis direkt påverka innehållet i OPAC: en. Fenomenet kallas *read/write OPAC* och ett bibliotek som redan testat denna modell är biblioteket vid University of Pennsylvania som genom systemet *PennTags* har låtit användare påverka OPAC: en Franklin genom att tagga, skriva annotationer och skapa projekt sedan 2005. Ett annat exempel på detta är hur möjligheten att kommentera, diskutera och tagga poster är central i Stockholms stadsbiblioteks *Biblioteket.se* som släpptes i en testversion våren 2007.

Ämnesingångarna i katalogen verkar vara hetare än någonsin, åtminstone om man studerar debatten om taggning där en vanligt förekommande åsikt är att det skulle vara en djup orättvisa – eller rentav ett tecken på odemokrati – att bibliotekarier ensam haft makten över indexeringen. Ser man till ämnesingångarnas historia är detta ett ganska intressant faktum. Ämnesingångar var egentligen aldrig särskilt intressanta innan Cutter så sent som 1876 skrev de första formella riktlinjerna för hur ämnesord skulle sättas och i sina kända *Rules for a printed dictionary catalogue* menade att ett av katalogens syften var att möjliggöra för en användare att finna en bok om bokens ämne var känt. Panizzi hade under 1800-talet mött hårt motstånd när han hävdade detta och hans meningsmotståndares huvudargument var att om man var

intresserad av ett ämne så kände man till de relevanta författarna. (Taylor 2004, s. 34, 55-59) Redan under Antiken hade bibliotek dock börjat laborera med ämnesingångar men länge ansågs ett fåtal breda kategorier tillräckligt, som att det exempelvis under medeltiden notera om en bok var "biblisk" eller "humanistisk" (Strout 1956, s. 256-259).

Många pratar om read/write OPAC och det är oftast i mycket entusiastiska ordalag. Jack M. Maness (2006) går i artikeln *Library 2.0 theory* så långt som att kalla det för biblioteksvetenskap när den är som allra bäst. Några empiriska studier har dock inte gjorts på hur en read/write OPAC skulle fungera i praktiken. Vad händer egentligen när användare ges möjlighet att påverka OPAC: en? Jag ska i denna studie genom en empirisk undersökning undersöka hur PennTags används för dokument inom olika ämnesområden för att studera hur användarna väljer att tillföra till dokumentbeskrivningarna i OPAC:en och skapa nya sökingångar till dokument. En vanligt förekommande åsikt – vid sidan om att folksonomier och kontrollerade vokabulärer är varandras raka motsatser – är att dessa två skulle kunna komplettera varandra på ett bra sätt om de kombinerades i ett system som PennTags, exempelvis genom vad som man inom informationsarkitekturen kallar *pace layering*.

### 1.1. Syfte och forskningsfrågor

Syftet med denna studie är att undersöka hur en tillämpning av read/write OPAC kan fungera i praktiken genom att studera hur användarna vid universitetsbiblioteket vid Pennsylvania University använder PennTags. Detta för att ta reda på hur användarna utnyttjar möjligheten att påverka innehållet i OPAC: en Franklin och skapa sökingångar till dokument som inte täcks in av den professionella indexeringen.

För att uppnå detta syfte ska jag arbeta med forskningsfrågorna:

- I hur stor utsträckning utnyttjas PennTags tekniska möjligheter för poster från Franklin inom olika ämnesområden?
- Vad tillför användarna genom sin taggning till dokumentbeskrivningar i Franklin inom olika ämnesområden?

### 1.2. Disposition

I kapitel 2 ges en beskrivning av den kontext som jag menar att read/write OPAC bör ses i. Webb 2.0, Bibliotek 2.0, PennTags och den kontrollerade vokabulären Library of Congress Subject Headings beskrivs.

I kapitel 3 tecknas en bild av forskningsläget omkring taggning och då särskilt forskning som jämför taggning med manuell indexering med kontrollerad vokabulär.

I kapitel 4 beskrivs de teoretiska utgångspunkter som använts för att konstruera instrument för att mäta vad taggningen tillför samt teorin om *pace layering* som ett sätt att förstå hur PennTags skulle kunna fungera i samspel med Franklin.

I kapitel 5 redogörs för metoden för datainsamling, kodning och analys.

I kapitel 6 redovisas resultatet av den empiriska undersökningen.

I kapitel 7 sammanfattas de svar som det empiriska materialet ger på studiens forskningsfrågor.

I kapitel 8 analyseras resultaten av den empiriska undersökningen utifrån teorin om *pace layering* samt diskuteras i relation till relaterad forskning.

### 1.3. Begrepp och definitioner

Universitetsbiblioteken vid Pennsylvania University kommer härnäst kallas för *Biblioteket*.

*Taggar* är keywords i naturligt språk avsedda för återvinning som används för att beskriva dokument i taggningssystem. *Folksonomi* är den uppsättning taggar som förekommer i ett taggningssystem. Samma relation finns mellan *ämnesord* och *kontrollerad vokabulär*.

Termen *taxonomi* brukar i diskussionen användas för att beteckna motsatsen till folksonomi. Närmast förknippas taxonomi med hur man inom biologin klassificerar och namnger organismgrupper men termen kan också användas som ett samlingsnamn för sätt att kategorisera företeelser som bygger på att varje bestämd företeelse har en bestämd plats i kategorisystemet. (Peterson 2006)  
Klassifikationssystemen SAB och DDC är i den bemärkelsen exempel på taxonomier och det är i den bemärkelsen termen kommer att förekomma i detta arbete.

*Library of Congress Subject Headings* är den kontrollerade vokabulär som används vid indexering på Biblioteket. Denna kommer härnäst förkortas *LCSH*. När förkortningen används är det systemet som åsyftas. Termen *subject heading* kommer användas för att beteckna ämnesord ur denna vokabulär.

*Library of Congress Classification* är det klassifikationssystem som Biblioteket i huvudsak använder. Detta kommer härnäst förkortas *LCC* och enskilda klassifikationskoder från systemet kommer kallas *LCC-koder*.

*Franklin* är Bibliotekets största OPAC. Franklin är samlingskatalog för alla filialer förutom filialen för juridik som har sin egen katalog (Penn Libraries).

Termen *användare* syftar i denna studie när PennTags diskuteras på en person som skapat en post eller ett projekt på PennTags, något som på PennTags benämns *owners*.

*Pace layering* är en modell som används inom bland annat informationsarkitekturen för att för att förstå hur komplexa system fungerar. Kortfattat menar denna modell att sådana system kan fungera genom att delar av konstruktionen har olika förändringshastighet vilket är bra för konstruktionen som helhet då de trögare delarna ger stabilitet medan de snabbare delarna driver innovation och nytänkande. (Morville 2005, s. 139-141). Detta kommer beskrivas närmare i kapitel 4.2.

#### 1.4. Avgränsningar

I denna studie relateras taggning och folksonomi till manuell indexering och kontrollerad vokabulär genom en empirisk undersökning av relationen mellan taggningssystemet PennTags och OPAC: en Franklin. Vanligt är att taggning och folksonomi relateras till ontologier eller automatisk indexering, men relationen till dessa fenomen kommer inte att undersökas närmare i denna studie.

Förutom möjligheten att utöka den beskrivning av ett dokument som finns i OPAC: en möjliggör PennTags för användarna att skapa poster för webbsidor, bibliotekets elektroniska resurser eller poster från Bibliotekets OPAC för videomaterial (VCat). Jag ska inrikta mig på hur användarna använder PennTags med avseende poster från Franklin eftersom jag är intresserad av att se hur användarna utnyttjar möjligheten att göra dokument sökbart på sätt som inte täcks in av indexeringen i LCSH. Att OPAC: en VCat inte inkluderas i studien beror på poster från denna förekommer i mycket liten utsträckning i mitt empiriska material samt att jag uppfattar problematiken med beskrivning av fiktivt material som delvis annorlunda.

## 2. BAKGRUND

Idén om att den vanliga biblioteksanvändaren skulle kunna ha något meningsfullt att tillföra en bibliotekskatalog är om man ser till bibliotekens historia på intet vis given. Eftersom jag menar att tankarna bakom fenomenet read/write OPAC måste ses i samband med utvecklingen mot Webb 2.0 och dess motsvarighet i biblioteksvärlden, Bibliotek 2.0, beskrivs dessa fenomen i kapitel 2.1. respektive 2.2. PennTags presenteras närmare i kapitel 2.3.

I kapitel 2.4. beskrivs den kontrollerade vokabulärens kännetecken och i kapitel 2.4.1. sägs något om den kontrollerade vokabulären LCSH.

### 2.1. Webb 2.0

Begreppet Webb 2.0 myntades av Tim O'Reilly 2005 i en artikel där han menade att de mest framgångsrika webbplatserna och företagen på webben hade vissa gemensamma drag. Dessa gemensamma drag definierades i artikeln som Webb 2.0. Peter Giger (2006) som forskar i vid Blekinge tekniska högskola identifierar i sin licentiatavhandling fyra centrala komponenter i Webb 2.0 –*webben som en plattform, ajax-gränssnitt, kollektiv intelligens* och *folksonomi*. De första två rör framförallt tekniska aspekter och kommer därför inte att beröras närmare här. De två senare är däremot av mer filosofisk natur och handlar om ett ökat fokus på delaktighet och vad som brukar diskuteras som en demokratisering av webben. (s. 24-29) Ett exempel som brukar lyftas fram på hur Webb 2.0 bygger på kollektiv intelligens och delaktighet är Wikipedia, där användarna tillsammans producerar och redigerar innehållet i en encyklopedi. Ett annat exempel som brukar framhållas är Googles rankningsalgoritm *PageRank* som frågar hela webben vilka sidor som är intressanta genom att beakta länkar till sidor i relevansbedömningen. Ytterligare ett exempel är nätbokhandeln Amazon som både explicit och implicit drar nytta av sina användare. Explicit sker det genom att man uppmuntrar användare att skriva recensioner och betygsätta böcker som sedan används i syfte att marknadsföra produkter, implicit genom så kallad *data mining* där användares beteende registreras för att sedan användas för att tipsa andra användare om produkter. (O'Reilly 2005) När webben gick från 1.0 till 2.0 innebar det bland annat att den, genom introduktion av applikationer som dessa, förändrades från ett medel för envägskommunikation via mer eller mindre statiska hemsidor till att handla om snabbare informationsflöden i flera riktningar – vad som brukar kallas *read/write web* (Macaskill & Owen, 2006).

#### 2.1.1. Taggning och folksonomi

Under slutet av 1990-talet och början av 2000-talet började webbtjänster där människor uppmuntrades att tagga och på andra vis beskriva dokument dyka upp på webben. Att taggningssystem är barn av 2.0 eran är uppenbart och Delicious lyfts ofta fram som en symbol för allt det som är Webb 2.0. (Giger 2006, s. 27) Ett av de första systemen för taggning var Bitzi som dock enligt Thomas Vander Wal (2007) drogs med en hel del problem, exempelvis att det var svårt att förstå innebörden av en tagg eftersom man inte fick reda på vem som satt taggen och därigenom inte kunde förstå vad taggen betydde för henne eller honom. När Joshua Schater skapade Delicious 2003 inkluderades användaridentitet och taggningen blev då mer meningsfull när man kunde ta del av vem som skapat taggen och vilka andra dokument hon eller han

använt taggen för. Professor Liz Lawley (2005) menar att det är här taggningens styrka ligger, i att man kan se vilka resurser personer man anser kunniga eller intressanta inom ett visst område har taggat.

Delicious är ett redskap för *social bookmarking* vilket innebär att användare lägger upp bokmärken, det vill säga länkar till intressant material på webben, och beskriver materialet med hjälp av taggar. Genom att länkarna postas på Delicious blir de tillgängliga från flera datorer och med hjälp av en sökfunktion kan man lokalisera bokmärken som beskrivits med en viss tagg eller en av en viss användare. Inte långt efter Delicious dök Flickr upp, en webbtjänst där användare kan ladda upp sina eller andras bilder och beskriva dem med hjälp av taggar. Vid sidan av dessa finns en rad andra taggningssystem. (Vander Wal 2007)

Folksonomi är den uppsättning termer som skapas genom fri taggning i en social miljö. Termen ”folksonomi” myntades av Vander Wal 2004 och är en kombination av termerna ”folk” och ”taxonomi”. Vander Wal (2007) inkluderar i sin egen definition av folksonomi att taggningen ska ske utifrån ett personligt kunskapsorganiserande syfte. Ulises Ali Meijas (2005) som är doktor i pedagogik vid Colombia University pekar dock på hur folksonomins syfte som kunskapsorganiserande system är tudelat. Visst handlar det om att människor organiserar dokument för egen del på ett sätt som känns rimligt för dem själva. Men det handlar minst lika mycket om att faktiskt dela med sig av sina taggar och på så viset skapa ett system som kan förmedla information mellan människor. Melanie Feinberg (2006) som är doktorand vid University of Washington menar däremot att någon kollektiv intelligens aldrig uppstår utan liknar istället att söka i taggningssystem vid att leta igenom andras bokhyllor och menar att folksonomi är resultatet av en mängd individuella beslut på personlig nivå (s. 5 f).

I folksonomin tillämpas generellt ingen form av vokabulärkontroll (Macgregor & McCulloch 2006). Detta resulterar enligt bland annat Peter Merholz (2005) och Ellyssa Kroski (2005) i att användare missar relevant information vid sökning samtidigt som alltför många irrelevanta träffar returneras, det vill säga låg *recall* och låg *precision*. Clay Shirky (2005c) menar emellertid i likhet med många andra förespråkare av folksonomi att detta tvärtom är en fördel. Ett av Shirkys mest kända exempel på problemet med synonymkontroll handlar om hur människor som är intresserade av film helt enkelt inte vill träffa folk som är intresserade av bio då det inte är samma fenomen för dem och samma sak antas gälla när de söker information om dessa företeelser. David Weinberger (2005) kallar recall och precision för ”hjärnspöken” och menar att dessa mått saknar relevans i en tid av informationsöverflöd med ett exempel om att det inte gör något om du missar 50 000 bilder på London om du ändå hittat 250 000.

I den kontrollerade vokabulären får varje ämne en plats – och endast en plats – genom att relationerna mellan termerna definieras (Giger 2006, s. 35f) Folksonomin kännetecknas istället av att det inte finns några tydligt definierade relationer mellan termerna och termlistan är ”platt” (Spiteri 2006, s. 77). Detta innebär att det inte finns någon ”rätt” plats för ett visst ämne, utan ett ämne kan ingå i en uppsjö av sammanhang samtidigt, allt utifrån användarnas tycke och smak (Giger 2006, s. 35f)

Elaine Peterson (2006) som forskar vid Montana State University menar att det största problemet med folksonomi ligger i just detta faktum och bör ses som en konsekvens av vad hon menar är en relativistisk inställning till kategorisering. I en taxonomi bestämmer man att A är A och inte är B men i en folksonomi kan A vara både A och B samtidigt – och dessutom vara A samtidigt som det inte är A. Weinberger (2005) pekar på demokratiaspekten i detta och menar att det må vara rörigt och ineffektivt men att det är ”bra nog” framförallt för att de reflekterar användarnas behov. Shirky (2005a, 2005b) beskriver ibland taggning som en framtvingad rörelse i brist på bättre alternativ och säger att den stora fördelen med folksonomi inte är att de är bättre än kontrollerade vokabulärer utan att de är bättre än ingenting och att kontrollerade vokabulärer inte är ett alternativ på webben.

Folksonomin brukar visuellt visas i form av ett tagg moln (från engelskans *tag cloud*) där taggarnas relativa popularitet visas med hjälp av teckenstorlek (Giger 2006, s. 36)

## 2.2. Bibliotek 2.0

Termen Bibliotek 2.0 sägs ha myntats på Michael E. Caseys blogg *LibraryCrunch* men vad det innebär diskuteras fortfarande (Maness 2006). Michael E. Casey och Laura C. Savastinuk (2006) menar att själva hjärtat i Bibliotek 2.0 är användaranpassad förändring. Bibliotek 2.0 beskrivs som en process för att skapa ett fysiskt och virtuellt bibliotek på användarnas villkor genom att användarna kontinuerligt inbjuds till delaktighet i utformningen.

Michael C. Habib (2006) beskriver i sin magisteruppsats i biblioteks- och informationsvetenskap Bibliotek 2.0 som en uppsättning bibliotekstjänster som utvecklas för att möta de nya behov som användarna har som en direkt eller indirekt följd av Webb 2.0. För att möta dessa potentiellt radikalt annorlunda förväntningarna hos delar av användarskaran menar Habib att det finns olika vägar att gå. Biblioteken skulle till exempel kunna använda sig av existerande Webb 2.0-applikationer för att nå ut på ett nytt sätt. Ett exempel på detta skulle kunna vara att använda en kommersiell Webb 2.0-applikation som Flickr för att tillgängliggöra sin bildsamling eller att skaffa ett konto på en populär online community för att tipsa om olika evenemang som biblioteket ordnar. Biblioteken skulle också kunna hämta inspiration från kommersiella Webb 2.0-tjänster i utformningen av sina egna webbtjänster. Read/write OPAC är ett exempel på detta. Det har också föreslagits att man med hjälp av tekniken data mining ska kunna samla information om användares beteende i OPAC: en och med hjälp av den informationen kunna utforma rekommendationssystem. (s. 22 f)

Giger (2006) påpekar att Bibliotek 2.0, precis som Webb 2.0, har både filosofiska och tekniska aspekter. Det handlar om att anpassa bibliotekssystem – men även vårt sätt att se på bibliotekssystemens funktion – till den anda av delaktighet som Webb 2.0 sägs kännetecknas av. (s. 51)

## 2.3. PennTags

PennTags skapades 2005 och är avsedd för vad som kallas *Penn community*, det vill säga studenter, lärare och bibliotekarier vid Pennsylvania University i Philadelphia, USA. Penn Community består enligt Laurie Allen och Michael Winkler (2006) som

jobbar vid Biblioteket av drygt 48 000 personer, varav nästan 24 000 är studenter. För att försäkra sig om att ingen annan använder systemet krävs ett särskilt login, ett så kallat *Pennkeylogin* för att kunna skapa nya poster på PennTags. Vem som helst har dock tillgång till de poster som lagts upp på PennTags.

PennTags är ett redskap för social bookmarking. Användare kan skapa poster på PennTags för material från webben, för Bibliotekets elektroniska resurser och för poster från Franklin och VCat (Bibliotekets OPAC för videomaterial). Allen och Winkler (2006) uppger att 27 poster genomsnittligt skapas per dag och att varje post i genomsnitt ges 4 taggar. Den folksonomi som har skapats innehåller 7304 unika taggar som tillsammans använts vid 26714 tillfällen.

När en användare vill lägga till en post från Franklin till PennTags måste användaren först söka upp posten i Franklin och därifrån trycka på en länk. När användaren väljer att lägga till posten till PennTags får denne också möjligheten att ge dokumentet ett valfritt antal taggar, skriva en annotation och/eller placera posten i ett projekt. En post skapas nu på PennTags för Franklinposten. I posten i Franklin syns då de taggar och/eller annotationer dokumentet getts på PennTags samt om posten lagts i ett projekt på PennTags. Via ett klick på en länk i Franklinposten kan man sedan ta sig till posten på PennTags och på motsvarande sätt från posten på PennTags till posten i Franklin.

*Projekt* på PennTags är ett slags foldrar där användarna samlar länkar till resurser inom exempelvis ett ämne eller till ett visst skolarbete. En *annotation* är en kortfattad beskrivning av ett dokumentets innehåll i naturligt språk. Annotationen syftar också på att lägga till en kommentar om dokumentets innehåll. *Taggar* kan bestå av ett enda ord eller kombineras som fraser vilket användarna uppmanas att göra med hjälp av understreck (ex. *film\_noir*). PennTags hjälpsidor innehåller riktlinjer för hur taggar ska skapas både när det gäller syntax och språkformer och vad som kan vara vettiga aspekter att fokusera på.

## 2.4. Kontrollerad vokabulär

En kontrollerad vokabulär är i sin enklaste form en lista som fastställer vilka termer som kan användas vid indexering. Den kontrollerade vokabulärens syfte är tudelat. Det ena syftet är att öka *consistency*, det vill säga samstämmigheten i val av termer mellan olika indexerare och mellan sökare och indexerare. Det andra syftet är att länka ihop relaterade ämnen genom att definiera relationer mellan termer. Den kontrollerade vokabulären kännetecknas av att synonymkontroll, homonymkontroll och kontroll av språklig form används i syfte att öka återvinnings effektiviteten. Klassifikationssystem, ämnesordslister och tesaurer är de tre främst manifestationerna av den kontrollerade vokabulären (Macgregor & McCulloch 2006). De nackdelar som förknippas med kontrollerade vokabulärer är att indexeringen blir dyrare, att det kan ta tid för nya termer att komma in i systemet och att man därför ofta når en lägre grad av specificitet och uttömmandegrad (Benito 2001, s. 256; Rowley 2000, s. 128).

### 2.4.1. Library of Congress Subject Headings

LCSH var från början ett sätt att skapa ämnesingångar till samlingarna vid Library of Congress i USA men är idag den mest använda ämnesordslistan på amerikanska

bibliotek. LCSH används även på bibliotek i många andra länder och anses vara en av världens mest använda kontrollerade vokabulärer. (Chan & Hodges 2000, s. 226). LCSH är också en av världens största kontrollerade vokabulärer med ungefär 270 000 termer (Fischer 2005, s. 64).

Historien bakom LCSH kan sägas starta 1898 då Library of Congress lade om sitt katalogsystem och skapade en ämneskatalog. 1902 startade *Library of Congress* tjänsten försäljning av katalogkort till andra större amerikanska bibliotek vilket skapade en efterfrågan av att få ta del av ämnesordslistan i sin helhet. Arbetet med en första tryckt version av LCSH påbörjades 1909 och gavs ut första gången 1914. (Stove 2000, s. 2) LCSH har under åren getts ut i ett antal olika upplagor och uppdateras numera dagligen i online-versionen (Taylor 2004, s. 59).

I LCSH skapas ämnesord i form av prekoordinerade strängar. Detta har varit föremål för diskussion under åren men man står fast vid en övertygelse om att detta är effektivare utifrån återvinningssynpunkt än att använda sig av ämnesord som postkoordineras med hjälp av exempelvis booleska operatörer. (Stove 2000, s. 5)

LCSH har fått utstå en hel del kritik genom åren. Karen S. Fischer (2005) sammanfattar denna kritik i *Critical views of LCSH* och menar att den kritik som hängt kvar under hela andra halvan av 1900-talet handlar om syntaktiska svårigheter, ideologiska undertoner i vilka ämnesord som finns tillgängliga och vilka namn ämnen ges samt problem med specificiteten i vokabulären. Till de problem som Fischer menar att fokus har flyttats till under de senare åren hör problem med underindelningar, consistency och vid internationell implementering av systemet. (s. 103-104)

### 3. RELATERAD FORSKNING

Eftersom taggning och folksonomi är nya fenomen består den största delen av det som finns skrivet om dem fortfarande av mer eller mindre välgrundade uppfattningar. En del forskning om taggning och folksonomi har dock producerats under de senaste åren inom biblioteks- och informationsvetenskap och datavetenskap. Majoriteten av forskningen som finns om taggning består av olika former av statistiska analyser av exempelvis spridning av taggar i ett visst system (från engelskans *tag distribution*) eller hur taggningen utvecklats över tid. De taggningssystem som främst förekommer i studierna är Delicious och Flickr. Jag ska för denna översikt inrikta mig mot forskare som relaterar taggning och folksonomi till manuell indexering med kontrollerad vokabulär då detta ligger närmast studiens syfte.

Flera forskare har undersökt huruvida taggning representerar ett radikalt brott med tidigare praktiker för ämnesrepresentation och delvis kommit till motstridiga slutsatser. Joseph T. Tennis (2006) som är doktor i biblioteks- och informationsvetenskap vid University of Washington jämför indexering med taggning med hjälp av en ramverksanalys. Tennis finner att det rör sig om två tämligen olika praktiker som bland annat har olika syften och fyller olika funktioner. Indexering har till syfte att möjliggöra för användare att finna dokument inom ett visst ämne och fyller en funktion i att hålla samman dokument inom ett ämnesområde. Taggning handlar istället om att dela resurser och att hitta nya sätt att organisera dokument. Vid taggning har den som indexerar en personlig relation till materialet och beskriver det utifrån egna intressen medan indexeraren har ett professionellt förhållningssätt till materialet och strävar efter att finna termer som en så stor del av de potentiella användarna skulle finna användbara.

Marieke Guy och Emma Tonkin vid University of Bath utförde en mindre studie 2006 för att undersöka taggningskompetensen (från engelskans *tag literacy*) hos användarna på Delicious och Flickr. De taggar användarna valde ansågs generellt vara undermåliga ur flera olika synvinklar, bland annat var hela 40 % av taggarna på Flickr felstavade eller på annat vis konstruerade på ett sätt som inte kunde tydas av den mjukvara som materialet analyserades med hjälp av.

När K. F. Lawrence och M. C. Schraefel (2006) som är forskare vid University of Southampton däremot undersökte en grupp av läsare/författare av *fan fiction* som under en avsevärt längre tid, sedan 1960-talet, själva satt ämnesord fann man att användargruppen med tiden utvecklat en stabil vokabulär med endast små variationer i hur ämnesord sattes mellan olika delar i gruppen.

Scott A. Golder och Bernardo A. Huberman (2006), forskare vid det privata HP Labs, undersökte data från Delicious och fann vissa regelbundenheter i hur bokmärken associerades med taggar. Ju populärare ett bokmärke var, desto sannolikare var det att det utvecklades stabila mönster kring hur bokmärket beskrevs. Detta förklarar författarna med att användarna tar efter varandra samt att en gemensam referensram byggs upp med tiden. Golder och Huberman identifierar i sin studie sju olika funktioner som en tagg kan ha för ett dokument. Dessa funktioner är:

1. Identifiera vad (eller vem) bokmärket handlar om (ex. *folksonomy*)

2. Identifiera vad det bokmärkta dokumentet är (ex. *film*)
3. Identifiera upphov för bokmärket (ex. *Liz\_Lawley*)
4. Förfina eller vidareindela existerande kategorier (ex. *25, 100*)
5. Identifiera kvaliteter eller karaktäristika för bokmärket (ex. *funny*)
6. Visa en relation till personen som taggar genom självreferens (ex. *mystuff*)
7. Organisera en uppgift, samla bokmärken som är användbara för en viss uppgift (ex. *for\_work*)

Forskarna Margaret E. I. Kipp och Grant D. Campbell (2007) vid avdelningen för biblioteks- och informationsvetenskap vid University of Western Ontario genomförde en analys av samförekomster mellan termer (co-word analysis) på data i form av taggar från Delicious i syfte att undersöka om några mönster som liknade traditionella sätt att beskriva dokument framträdde. Kipp och Campbell fann att taggning till viss del följde liknande mönster som traditionell indexering, exempelvis att konsensus förefaller råda kring hur vissa termer ska användas. Kipp och Campbell fann även, liksom Golder och Huberman, att en del taggar skapar en dimension som inte är närvarande i traditionell ämnesordssättning. Denna dimension kallas för en tidsdimension och genom att sätta taggar som "To Read" eller "Cool" bryter man mot fler viktiga principer som indexering bygger på.

*Exploring the characteristics of social classification* (2006) består av tre empiriska undersökningar där forskare inom informationsvetenskap vid Drexel University (Lin et al) försöker karaktärisera fenomenet folksonomi. I en av dessa jämförs taggning med indexering med kontrollerad vokabulär och med automatisk indexering. 45 dokument från Connotea (ett taggningssystem för vetenskapligt material) som tidigare indexerats med tesauren MESH valdes ut. Från dessa dokument extraherades sedan termer genom automatisk indexering från titelfältet. Jämförelsen mellan taggarna och ämnesorden skedde på termnivå. Ingen form av normalisering av termerna gjordes innan analysen men stoppord, symboler, prepositioner och konjunktioner uteslöts. Lin et al fann att överlappningen mellan taggarna och ämnesorden från MESH endast var 11 %. Detta försöker Lin et al bland annat förklara genom att notera att taggar inte alltid används för att beskriva dokumentets ämne och att användarna inte förefaller intresserade av att beskriva dokumentet till sin helhet. Lin et al tror att användarna istället taggar de aspekterna av ett dokument som är intressanta för dem själva och dokument beskrivs därför med väldigt specifika termer. Överlappningen mellan taggarna och de automatiskt generade ämnesorden var något högre (19 %) och Lin et al menar att detta tyder på att användarna ofta väljer termer ur titelfältet vid taggning. Det förklaras också med att taggarna är mer lika naturligt språk än kontrollerad vokabulär.

Frågan om hur tydliga riktlinjer eller annan styrning påverkar taggning diskuteras av ett antal forskare (Bar-Illan et al) inom informationsvetenskap vid Bar-Illan University i artikeln *Structured vs. unstructured tagging* (2006). Resultatet av ett experiment där en grupp studenter fick i uppgift att sätta ämnesord till ett antal bilder redovisas i artikeln. Den ena halvan av gruppen fick fria händer medan den andra halvan fick en tydlig mall för vilka aspekter som var viktiga att ta upp i beskrivningen. Studien visar att strukturerad taggning generellt resulterar i mer detaljerade beskrivningar men att en för tydlig struktur kan verka avskräckande. Guy och Tonkin (2006) ger i sin studie några tips på hur taggningskompetensen kan förbättras genom att riktlinjer sätts upp i

ett läge när användargruppen är redo för det. Samtidigt ifrågasätter Guy & Tonkin nyttan av att sätta upp allt för strikta regler eftersom systemet då kanske förlorar sina styrkor.

Louise F. Spiteri (2006) som är doktor i biblioteks- och informationsvetenskap vid Dalhousie University utreder hur taggningssystem skulle kunna fungera i en folkbiblioteksmiljö och menar att kombinationen av folksonomi och kontrollerad vokabulär skulle kunna vara ett värdefullt redskap för att användaranpassa biblioteket. Spiteri ser tre olika funktioner som taggning skulle kunna fylla:

1. att ge användarna möjlighet att organisera resurser på en egen användare
2. att skapa möjligheter för grupper med gemensamma intressen att forma online communities
3. att fungera som ett komplement till indexering med kontrollerad vokabulär eftersom dessa ämnesord inte alltid känns naturliga för användarna

Habib (2006) skapar i sin magisteruppsats i biblioteks- och informationsvetenskap en metodologi där fyra set av frågor används för att undersöka Webb 2.0: s effekter på bibliotek. Metodologin används sedan för att skapa en vision om hur ett Bibliotek 2.0 i den akademiska världen skulle kunna se ut. I denna spelar read/write OPAC en viktig roll. De användningsområden Habib ser för en read/write OPAC är att studenter ges möjlighet att diskutera, kommentera och tagga poster. Taggningen har enligt Habib sin största potential i att vara ett sätt för studenter att tipsa varandra om relevant material till en viss kurs eller till ett visst projektarbete.

Ingen har mig veterligen gjort en empirisk undersökning där taggningens potential som komplement till en kontrollerad vokabulär i en bibliotekskontext utreds. Jennifer Trant (2006) som doktorerar vid University of Toronto undersökte däremot ett närliggande område då hon försökte finna svar på frågan om taggning skulle kunna skapa nya användbara ingångar till samlingarna på konstmuseet The Metropolitan Museum of Art. En serie experiment genomfördes där tränade katalogisatörers beskrivningar av ett antal bilder i museets samlingar jämfördes med beskrivningar som otränade katalogisatörer åstadkommit. I det sista och mest omfattande experimentet ombads 39 deltagare (bestående av frivilliga museibesökare och personal på museet som inte normalt ägnade sig åt katalogisering) beskriva 30 konstbilder. Deltagarna producerade tillsammans 3780 unika termer och 88 % av dessa fanns inte i de befintliga beskrivningarna av konstbilderna i museets databas. En grupp från av katalogiseringsexperter delade därefter upp de nya termerna i två kategorier beroende på om de ansågs vara passande eller opassande för bilden i fråga eftersom syftet med en eventuell introducering av taggning inte endast skulle vara att skapa nya sökingångar – utan användbara sökingångar. 77 % av termerna ansågs efter denna granskning vara användbara sökingångar.

## 4. TEORETISKA UTGÅNGSPUNKTER

I kapitel 4.1. beskrivs vad taggning kan förväntas tillföra ämnesrepresentationer skapade i kontrollerad vokabulär för att utifrån detta kunna konstruera ett redskap för att mäta vad det är PennTags användare tillför Franklin. Därefter presenteras teorin om *pace layering* i kapitel 4.2. som ett sätt att förstå hur en read/write OPAC skulle kunna kombinera fördelarna hos kontrollerade och fria vokabulärer.

### 4.1. Vad skulle taggning kunna tillföra dokumentbeskrivningar?

Förespråkare av taggning hävdar ofta att folksonomins natur är så skild från taxonomins att det inte går att jämföra de två eller på något vis tillämpa samma kriterier när man utvärderar dem. Samtidigt är det inte omöjligt att tänka sig att taggning skulle kunna uppvisa några av de fördelar som är förknippade med indexering i fri vokabulär. Det är heller inte omöjligt att tänka sig att taggning skulle kunna medföra åtminstone en del av de nackdelar som är förknippade med indexering i fri vokabulär, en tes som stöds av bland annat Spiteri (2006) som menar att folksonomier delar *alla* de problem som finns inbyggda i fria vokabulärer (s. 79).

De fördelar som är förknippade med indexering i fri vokabulär är att man når en högre grad av specificitet i indexeringen och därigenom bättre precision vid sökning. Ämnesorden kan vara mer aktuella om den som indexerar är bekant med utvecklingen inom ämnet i fråga. Kostnaden för indexeringen blir generellt lägre eftersom ingen kontrollerad vokabulär måste skapas och upprätthållas. Till nackdelarna med indexering i fri vokabulär hör att det kan bli svårare att söka i systemet eftersom användaren inte vet vilken av flera potentiella termer som använts vid indexering, både vad gäller synonymer men även språkvarianter (Benito 2001, 256 f; Rowley 2000, s. 128).

När fler ämnesord läggs till i en dokumentrepresentation kan detta enligt Frederick Wilfrid Lancaster (1998) antingen resultera i ökad uttömmandegrad eller ökad specificitet (s. 24). *Uttömmandegrad* anger i hur stor utsträckning dokumentets olika ämnesområden eller aspekter är närvarande i dokumentbeskrivningen. Alla dokumentbeskrivningar är förenklingar av dokumentets innehåll och uttömmandegrad handlar om hur grov denna förenkling är. (Rowley 2000, s. 99) Ökad uttömmandegrad innebär att en klarare bild ges av vad dokumentet handlar om vilket kan göra det lättare att avgöra om ett dokument är relevant eller inte utifrån ett givet informationsbehov. Ökad uttömmandegrad innebär också att fler sökingångar till dokumentet skapas. Vilken uttömmandegrad som används vid indexeringen är generellt en fråga om kostnad för den indexerande institutionen och därför brukar en viss policy finnas kring detta som anger ungefär hur många ämnesord som bör tilldelas varje dokument. (Lancaster 1998, s. 22 ff) *Specificitet* handlar om i vilken utsträckning en kontrollerad vokabulär kan specificera ett ämne. Högre specificitet innebär bättre precision vid sökning (Rowley 2000, s. 129-132) Specificitetsprincipen brukar kallas den enskilt viktigaste principen inom indexering. Den kan härledas tillbaka till Cutter som 1876 deklarerade att ett dokument ska indexeras med det mest specifika ämnesordet som är möjligt (Lancaster 1998, s. 28). Samtidigt måste specificitetsnivån i beskrivningen enligt Arlene G. Taylor (2004) alltid i slutändan bestämmas av behoven hos användarna och vad samlingen innehåller i övrigt (s. 262).

PennTags är något användarna ägnar sig åt på sin fritid av eget intresse och det är därför inte omöjligt att de genom taggning kan öka så väl uttömmandegraden som specificiteten i dokumentbeskrivningarna i Franklin. Samtidigt pekar flera på att de taggar som förekommer i folksonomier inte sällan skiljer sig radikalt från ämnesord, oavsett om dessa är skapade med en kontrollerad eller fri vokabulär. Golder och Huberman (2006) identifierade vid en undersökning av Delicious 7 olika funktioner som en tagg kunde fylla för ett givet dokument (se kapitel 3). Bland dessa funktioner ingår endast den första och den andra, det vill säga att beskriva dokumentets innehåll och form, i indexering. De övriga funktionerna bryter enligt Kipp och Campbell (2007) mot de principer som ligger bakom indexering på flera vis. För det första anger de inte vad dokumentet handlar om, utan snarare vad den som taggar tycker om dokumentet eller vad den som taggar ska göra med dokumentet. Den del av dessa funktioner som utgörs av värdeladdade ord är naturligtvis av subjektiv natur och en tagg som ”To Read” är ytterst känslig för tidens gång, för när dokumentet väl är läst är det tveksamt vilken funktion som taggen fyller. Taggar som speglar dessa funktioner kommer därför i denna studie antas spegla ett begränsat allmänt värde även om de naturligtvis kan fylla en funktion för den användare som använder dem.

## 4.2. Pace Layering

Konceptet *pace layering* är hämtat från arkitekturen och beskrivs av Stewart Brand (1997) som principen om att byggnader bör konstrueras i flera lager eftersom olika delar av byggnaden och olika material har olika lång livstid och därför måste byta ut olika ofta. En husgrund kan stå kvar hundratals år senare medan fasaden måste bytas redan efter något decennium. De långsamma lagren ger stabilitet åt konstruktionen medan de snabba lagren är föränderliga och därigenom utsatta för snabbare trender.

Inom informationsarkitekturen har man snappat upp tankarna om *pace layering* som ett sätt att förstå hur komplexa informationssystem kan fungera. Peter Morville (2005) som av vissa betraktas som informationsarkitektens fader menar att ett system där kontrollerad vokabulär fungerar i symbios med folksonomi kan fungera genom just *pace layering*. Den kontrollerade vokabulären skapar ett stabilt grundfundament som sen en mer trendkänslig och innovativ folksonomi kan byggas upp på. Så småningom sipprar nyheterna ner och tar sig in i de trögare lagren där användarnas idéer på så vis enligt Morville kan komma till användning. (s. 139-141) Grundtanken bakom *pace layering* är just att olika förändringshastighet faktiskt är bra för helheten (Brand 1999, s. 35-39).

Brand (1999) tar i *The clock of the long now* tankarna steget längre och beskriver hela samhället som en konstruktion av olika lager med olika lång livscykel eller förändringsbenägenhet. Brand identifierar sex sådana lager. Lagren, här presenterade från det snabbaste till det långsammaste, är:

*Mode och konst* (Fashion & art)

*Handel* (Commerce)

*Infrastruktur* (Infrastructure)

*Statsskick* (Governance)

*Kultur* (Culture)

## *Natur* (Nature)

Jag ska i denna studie undersöka Morvilles antagande om hur kontrollerade vokabulärer och folksonomi kan fungera i symbios och använda Brands modell (1999) över hur samhället är uppdelat i lager för att undersöka om användarnas aktivitet och förmåga att tillföra något nytt genom PennTags kan förklaras mot bakgrund av *pace layering*. Min bearbetning av Morvilles och Brands teorier säger att om folksonomi är ett snabbt lager borde det lämpa sig särskilt bra för att fånga upp snabba ämnena och trender i dem. Användarna borde således ha större incitament att lägga till poster till PennTags inom snabbare ämnesområdena än inom trögare och de borde dessutom kunna göra fler meningsfulla tillägg inom dessa områden genom taggning. När det gäller trögare ämnen borde taggningens lockelse och användbarhet vara mer begränsad. Den kontrollerade vokabulärens tröghet till förändring borde utgöra ett mindre problem och därför borde användarnas incitament själva beskriva dokument inom sådana ämnen vara mindre. Brand ger inte några omfattande beskrivningar om vad som innefattas i de olika lagren, förutom några kortfattade exempel, men jag tror ändå att en sådan analys låter sig göras.

## 5. METOD

För att besvara mina forskningsfrågor har jag utifrån en kvantitativ ansats studerat poster som skapats på PennTags för poster från Franklin. Kvantitativa forskningsansatser kännetecknas av att forskaren försöker mäta eller beräkna förekomsten av olika företeelser. Inom samhällsvetenskaplig forskning är mätningen ofta av mer abstrakt natur och man tar därför hjälp av olika indikatorer på ett begrepp och utvecklar mätinstrument som kan fånga upp dessa. (Bryman 2002, s. 81 f) Det är också vanligt att kvantitativa forskare undersöker samband mellan olika variabler. De samband jag ska studera är de eventuella sambanden mellan ämnesområde och aktivitet på PennTags och mellan ämnesområde och vad jag har valt att kalla tillförandegrad i taggningen. De variabler jag har valt att undersöka och de mätinstrument jag har utvecklat för att detta finns närmare beskrivna i kapitel 5.2.

Kvantitativa undersökningar försöker ofta skapa *generaliserbar kunskap* där man utifrån ett mindre empiriskt material kan säga någonting om den större helheten som detta ingår i. För att kunna generalisera sina slutsatser använder kvantitativa forskare därför olika former av *sannolikhetsurval*, det vill säga slumpmässiga urval, när de väljer ut en grupp att studera. (Ibid., s. 93) I min datainsamling har jag använt mig av sannolikhetsurval när jag skapat en sampel att studera för att i slutändan kunna säga någonting om hela populationen, det vill säga samtliga PennTags användare. Datainsamlingen beskrivs närmare i kapitel 5.1.

När man talar om kvaliteten på en kvantitativ studie brukar man tala om *reliabilitet* och *validitet*. Reliabilitet anger pålitligheten hos ett mått och kan exempelvis röra sig om ifall flera forskare kommer fram till samma resultat eller om man får samma resultat vid olika mättillfällen. Validitet handlar om huruvida måttet verkligen mäter det som det är avsett för att mäta, det vill säga om de indikatorer som har valts ut verkligen kan sägas spegla det begrepp som man vill komma åt genom mätningen. (Ibid., s. 86 f) Det är min förhoppning att genom att mäta både *hur mycket* aktivitet och *vilken sorts* aktivitet kunna komma till vettigare slutsatser om hur PennTags används. Jag har också genomfört en pilotundersökning när det gäller vad användarna tillför genom sin taggning och därefter förbättrat detta mätinstrument.

### 5.1. Datainsamling och urval

För att samla in data om hur PennTags används hämtades en lista över alla PennTags användare den 22 mars 2007 från PennTags. Populationen består av 834 användare och denna lista fick utgöra urvalsram för att skapa en sampel med hjälp av ett *systematiskt urval*. Ett systematiskt urval innebär att man slumpmässigt fastställer ett startnummer och därefter väljer ut enheter att studera med ett visst intervall i urvalsramen. Systematiskt urval kan endast användas om det inte finns någon särskild ordning i urvalsramen. (Bryman 2002, s. 105 f) Min urvalsram är alfabetiskt ordnad men jag menar inte att detta inte kan antas ge upphov till några skevheter i urvalet. Den sampel som åstadkoms består av ¼ av det totala antalet användare, alltså 208 användare.

De poster och projekt sampeln skapat på PennTags sparades ner den 22-23 mars 2007. Det finns en minimal risk att någon användare kan ha raderat eller ändrat något

medan datainsamlingen pågick. Insamlingen var dock inte genomförbar på något annat vis och jag ser heller ingen särskilt poäng i att ha samtidiga data eftersom frågeställningen inte rör någon tidsaspekt. Jag har dock inte tagit med några poster från PennTags som lagts till efter 22 mars för att ge användarna samma chans att bidra.

För att denna lista skulle bli hanterbar för vidare analys sattes en övre gräns på högst 100 poster per användare. För de användare som hade gjort fler bidrag valdes 100 av dessa ut genom ett *obundet slumpmässigt urval*. Ett obundet slumpmässigt urval innebär att man i en numrerad lista fastställer vilka enheter som ska studeras genom att hämta slumpmässigt bestämda tal, exempelvis från en slumpstalstabell eller en slumpstalsgenerator (Bryman 2002, s. 104 f). Jag har använt en slumpstalsgenerator och detta skedde för 5 användares PennTagsposter.

De poster och projekt sampelns skapat undersöktes därefter för att plocka fram de poster som var hämtade från Franklin. Att en PennTagspost kommer från Franklin avslöjas av att den URL som länken i posten pekar på innehåller /voyager/. Länkarna från PennTags till Franklin fungerade inte alltid och i dessa fall har jag själv sökt i Franklin för att finna posterna. Detta har dock inte inneburit några svårigheter eftersom de aktuella PennTagsposterna innehållit nog mycket bibliografisk information för att kunna avgöra vilken Franklinpost det är som åsyftas.

För att besvara frågeställningarna om vad det är användarna genom taggning väljer att tillföra dokumentbeskrivningar inom olika ämnesområdena används de *call number* som finns i Franklinposterna för att identifiera ämnesområden. Call number anger generellt ett dokumentets placering men på Biblioteket används vanliga klassifikationskoder från LCC som call number. För denna frågeställning undersöktes de 3 huvudavdelningar i LCC där taggning var mest frekvent. Från varje av dessa huvudavdelningar valdes poster att studera ut genom ett *stratifierat slumpmässigt urval* från listan med samtliga poster sampelns skapat på PennTags för poster från Franklin. Vid ett stratifierat slumpmässigt urval går man till väga som vid ett obundet slumpmässigt urval men har något kriterium för bland vilka enheter som urvalet ska ske (Ibid., s. 106 f). I mitt fall utgjordes dessa kriterier av 30 poster skulle komma från vardera av de tre undersökta huvudavdelningar och att de skulle vara taggade.

## 5.2. Kodning och analys av materialet

För att besvara frågeställningen om i hur stor utsträckning användarna utnyttjar PennTags inom olika ämnesområde står sambandet mellan ämnesområde och aktivitet på PennTags i centrum. Aktiviteten antas vara beroende av ämnesområdet utifrån de teoretiska utgångspunkter som finns beskrivna i kapitel 4.3. Listan med samtliga PennTagsposter sampelns skapat för poster från Franklin undersöktes.

Ämnesområden definieras med hjälp av call number som dokumenten tilldelats i Franklin. För 28 poster uppstod problem med att avgöra call number av olika anledningar. Eftersom detta bortfall var så pass stort valde jag att undersöka Marc-posterna för dessa dokument i syfte att finna en klassning i LCC. I de fall där jag hittade flera klassningar i LCC valde jag den avdelning som dokumentet

huvudklassats på. För 18 poster kunde ingen klassning i LCC identifieras och dessa sorterades som bortfall.

Att använda call number för att definiera ämnesområden innebär naturligtvis vissa begränsningar. Den kanske främsta invändningen är att en något statisk bild av materialet skapas eftersom ett dokument genom detta förfaringssätt bara kan ha *ett* ämne vilket sällan är fallet i verkligheten. Jag litar emellertid på att den huvudklassning som gjorts i LCC på ett korrekt sätt speglar åtminstone huvuddragen i dokumenten och jag menar att detta är bra nog när det gäller att visa ämnestillhörighet i ett så pass stort material. Ett annat problem är att huvudavdelningar i LCC inte på ett direkt sätt kan härledas till ett visst lager i den modell jag avser undersöka, men jag tror ändå att intressanta slutsatser kan dras utifrån denna uppdelning.

Användarnas aktivitet mäts med hjälp av fyra olika variabler som representerar de tekniska möjligheter som en användare kan utnyttja.

- 1) att lägga till en post från Franklin till PennTags
- 2) att lägga en post från Franklin i projekt på PennTags
- 3) att annotera en post från Franklin
- 4) att tagga en post från Franklin

Variablerna antar ett numeriskt värde som representerar det totala antalet användare som utfört aktiviteten i fråga.

För att det ska synas i Franklin att en post lagts till i PennTags måste användaren ge dokumentet åtminstone en tagg. Det förekommer att användare placerar posten i ett projekt och/eller lägger till en annotation men väljer att inte lägga till någon tagg och då syns varken annotationen eller att posten tillhör ett projekt i Franklin. I strikt mening kan alltså ett sådant tillägg inte sägas påverka Franklin eftersom det inte syns i Franklin. I denna studie jag har frånsatt detta eftersom systemet lika gärna skulle kunna fungera på ett sådant sätt att alla aktivitet på PennTags behandlades likvärdigt. Det är heller inte säkert att alla användare är bekanta med denna begränsning i PennTags och det vore olyckligt om en sådan faktor tilläts påverka studiens resultat.

För frågeställningen om vad användarnas taggning tillför den befintliga dokumentbeskrivningen undersöks sambandet mellan ämnesområde och vad jag har valt att kalla för tillförandegrad. Tillförandegrad antas liksom aktivitet vara beroende av ämnesområdet hos dokumentet som beskrivs. Listan med 90 poster fördelade över de tre mest frekvent taggade ämnesområdena undersöktes. De taggar en användare gett en post jämfördes mot de subject headings posten i fråga tilldelats i Franklin.

Tillförandegraden mäts med hjälp av tre variabler som antar ett numeriskt värde som speglar det totala antalet taggar som passar in på de kriterier som satts för varje tillförandegrad. Tillförandegraderna är konstruerade utifrån de för- och nackdelar som taggning förväntas kunna tillföra utifrån de teoretiska utgångspunkter som finns beskrivna i kapitel 4.1. Kriterierna för tillförandegraderna modifierades efter en pilotgenomgång av materialet och de tillförandegradernas slutliga utformning är alltså ett resultat av både tidigare forskning och teori och mitt eget empiriska material. En modifikation som görs efter pilotgenomgången är att definiera att taggar som är identiska med, eller språkvarianter av, textsträngar som redan är sökbara i Franklin skapar ett mycket begränsat värde (tillförandegrad 1). I detta fall går alltså

jämförelsen utanför relationen mellan taggarna och subject headings. Kriterierna för att en tagg ska passa in på en viss tillförandegrad finns i bilaga 1.

Kortfattat innebär tillförandegrad 1 att en tagg medför någon av de nackdelar som taggning förväntas kunna tillföra och tillförandegrad 3 att en tagg medför någon av de fördelar som taggningen förväntas kunna tillföra. Tillförandegrad 2 innebär att en tagg tillför synonymer till befintliga subject headings och därigenom skapa nya sökingångar, men på ett sätt som bryter mot principer som vokabulärkontroll vilar på. Tillförandegrad 2 och 3 kommer att betraktas som ett tillförande av meningsfulla nya sökingångar.

För att vara säker på att en tagg tillför de fördelar som taggning förväntas kunna tillföra, alltså ökar specificiteten eller uttömmandegraden i beskrivningen måste man naturligtvis egentligen undersöka dokumentet. Detta eftersom en term som förefaller vara mer specifik eller röra en ny aspekt av dokumentet skulle kunna vara felaktig i något avseende. Jag har emellertid inte gjort detta utifrån ett antagande om att användarna sätter termer som på ett korrekt sätt återger dokumentet som helhet eller en del av dokumentet.

Taggar, annotationer och projekttilhörighet är inte sökbara i Franklin. Detta innebär att ingen tagg skapar någon ny sökingång. Jag ska ändå analysera taggarna för att ta reda på vilken slags sökingångar dessa skulle skapa eftersom det är möjligt att skapa ett system där taggar är sökbara och i ett sådant fall kan det vara önskvärt att utreda vilka konsekvenser det skulle få.

## 6. RESULTAT

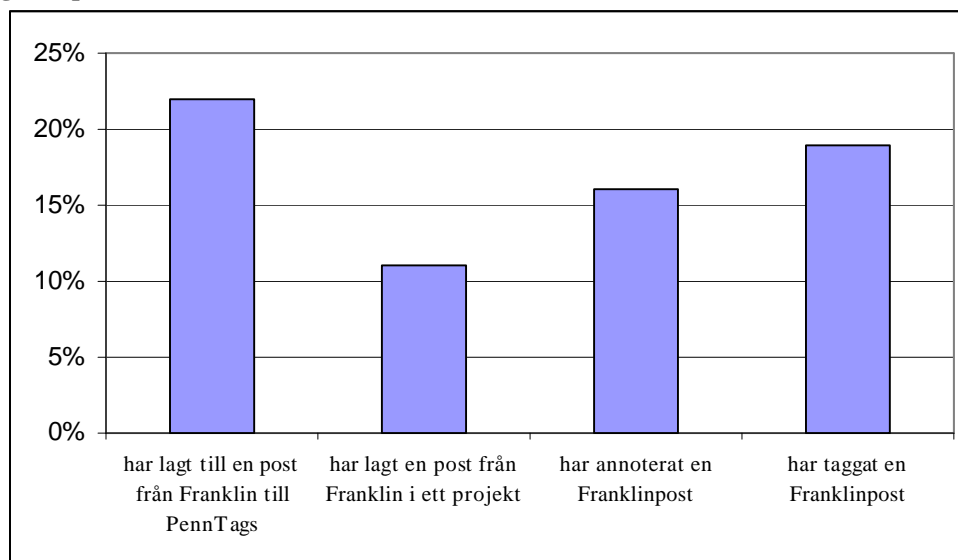
I resultatdelens första avsnitt presenteras hur användarna utnyttjat PennTags tekniska möjligheter och hur denna aktivitet är fördelad över huvudavdelningarna i LCC. Eftersom en av de aktiviteter som jag mäter förekomsten av – att lägga till en post från Franklin till PennTags – är en förutsättning för övrig aktivitet är detta naturligtvis den vanligast förekommande aktiviteten. För denna aktivitet kommer jag även att redovisa de populäraste underavdelningarna på en andra hierarkisk nivå i LCC för att på det viset kunna ge en klarare bild av vilka ämnesområden det är som är intressanta för användarna att lägga till poster ifrån. I resultatdelens andra avsnitt presenteras vad användarna tillför genom sin taggning inom de tre ämnesområden där taggningen är mest frekvent.

De översättningar av namn på avdelningar i LCC som förekommer är resultatet av eget kreativt arbete utifrån den översikt av systemet som återfinns på Library of Congress egen hemsida. Beskrivningar av avdelningars innehåll och omfång är på alla ställen i uppsatsen där de förekommer hämtade från *Guide to the Library of Congress classification* (Chan 1999).

### 6.1. Aktivitet och ämnesområde

Diagram 1 visar hur användarna valt att utnyttja PennTags tekniska möjligheter när det gäller poster från Franklin.

**Diagram 1. Andel av användarna som har utnyttjat PennTags tekniska möjligheter när det gäller poster från Franklin (i %)**



I diagram 1 kan man se att drygt en femtedel av användarna (22 %) har lagt till en post från Franklin till PennTags. Nästan alla användare som lagt till en post har också någon gång taggat en post (19 %). Nästan lika vanligt som att ha taggat är att ha lagt till annotationer (15 %). Att lägga en post från Franklin till ett projekt är en möjlighet som endast utnyttjats av drygt 10 % av användarna.

### 6.1.1. Att lägga till en post från Franklin till PennTags

45 av 208 användare har utnyttjat möjligheten att lägga till poster från Franklin till PennTags. Dessa användare har lagt till en post från Franklin vid 433 tillfällen. Hur aktiviteten är fördelad över huvudavdelningarna i LCC kan ses i tabell 1.

**Tabell 1. Antal tillfällen som en användare har lagt till en post från Franklin till PennTags från respektive huvudavdelning i LCC**

<u>LCC-kod</u>		<u>Antal</u>
P	Språk och litteratur	122
H	Samhällsvetenskap	63
B	Filosofi. Psykologi. Religion	43
R	Medicin	43
Q	Naturvetenskap	27
S	Lantbruk	26
D	Världshistoria	20
T	Teknik	12
E-F	De amerikanska kontinenternas historia	10
K	Juridik	10
M	Musik och böcker om musik	7
G	Geografi. Antropologi. Fritid	6
J	Statsvetenskap	6
Z	Biografi. B & I. Allmänna Informationsresurser	6
C	De historiska hjälpvetenskaperna	4
A	Allmänna arbeten	3
N	Konst	3
L	Utbildning/pedagogik	2
U	Militärväsen	2
V	Sjökrigsväsen	0
	Okänd kod	18
<u>Totalt</u>		433

Tabell 1 visar att användarna har lagt till poster från den populäraste klassen, *Språk och litteratur*, nästan dubbelt så ofta som från den näst populäraste klassen *Samhällsvetenskap*. Från 7 av 20 huvudavdelningar har användare lagt till poster till PennTags fler än 20 gånger. Från övriga 13 avdelningar har poster lagts till PennTags mycket sällan.

Dominansen hos framförallt huvudavdelningen *Språk och litteratur* men även i någon mån *Samhällsvetenskap* förstärks om man tar i beaktande hur många unika användare som lagt till minst en post från respektive huvudavdelning. Detta visas i tabell 2.

**Tabell 2. Antal användare som har lagt till poster till PennTags från varje huvudavdelning i LCC**

<u>LCC-kod</u>		<u>Antal</u>
P	Språk och litteratur	27
H	Samhällsvetenskap	18
B	Filosofi. Psykologi. Religion	7
E-F	De amerikanska kontinenternas historia	7
T	Teknik	7
D	Världshistoria	5
K	Juridik	5
R	Medicin	5
J	Statsvetenskap	4
M	Musik och böcker om musik	4
Q	Naturvetenskap	4
Z	Biografi. B & I. Allmänna informationsresurser	4
A	Allmänna arbeten	2
C	De historiska hjälpvetenskaperna	2
G	Geografi. Antropologi. Fritid	2
L	Utbildning/pedagogik	2
N	Konst	1
S	Lantbruk	1
U	Militärväsen	1
V	Sjökrigsväsen	0

I tabell 2 framstår *Språk och litteratur* och *Samhällsvetenskap* vara de enda huvudavdelning där något allmännare intresse för att använda PennTags finns. Endast från dessa avdelningar har fler än 10 användare lagt till poster. Från huvudavdelningen *Språk och litteratur* har 27 av totalt 45 användare valt att lägga till poster. Huvudavdelningen *Samhällsvetenskap* är inte fullt så populär men från denna avdelning till den tredje populäraste huvudavdelningen *Filosofi. Psykologi. Religion* sker ändå en markant förändring. Notera även hur de goda placeringarna i tabell 2 för huvudavdelningarna *Medicin* och *Naturvetenskap* men kanske ännu tydligare *Lantbruk* förefaller vara ett resultat av ett mindre antal användares stora intresse.

Tittar man på spridningen av hur användarna lägger till poster på en andra hierarkisk nivå i LCC ser man att från de flesta underavdelningar har poster lagts till mycket sällan – eller inte alls. Från 11 underavdelningar på den andra hierarkiska nivån har poster lagts till vid fler än 10 tillfällen. Dessa visas i tabell 3.

**Tabell 3. Antal tillfällen som en användare har lagt till en post från Franklin till PennTags från de 11 underavdelningar i LCC där det är vanligast att lägga till poster till PennTags**

<u>LCC-kod</u>		<u>Antal</u>
PN	Litteratur (allmänna verk)	77
BV	Praktisk teologi	24
SF	Djuruppfödning	24
RC	Invärtes medicin	13

HD	Industrier. Markanvändning. Arbetskraft	12
HM	Sociologi (allmänna verk)	12
HQ	Familjen. Äktenskapet. Kvinnor	12
RA	Samhällsmedicin	12
DS	Asien (historia)	11
PS	Amerikansk litteratur	11
E151-889	USA (historia)	10

I tabell 3 placeras sig inte oväntat en underavdelning från *Språk och litteratur* i topp – Litteratur (*allmänna verk*). När användare har lagt till poster från *Språk och litteratur* har detta i nästan hälften av fallen skett från denna underavdelning. Går man dessutom ett steg längre ner i hierarkin ser man att i 72 av dessa 77 tillfällen så har posterna hämtats från underavdelningen *Film*.

Underavdelningen *Litteratur (allmänna verk)* är också mer än 3 gånger så populär som den näst största underavdelningen *Praktisk teologi*. Tillägg inom denna underavdelning står också för mer än hälften av det totala antalet tillägg från den aktuella huvudavdelningen som i detta fall är *Filosofi. Psykologi. Religion*.

Underavdelningar från den populära huvudavdelningen *Samhällsvetenskap* placeras sig ganska långt ner i tabellen men tre av de populäraste underavdelningarna, *Industrier. Markanvändning. Arbetskraft*, *Sociologi (allmänna verk)* och *Familjen. Äktenskapet. Kvinnor* kommer från denna huvudavdelning.

I tabell 3 utmärker sig även underavdelningen *Djuruppfödning* varifrån användare har lagt till poster nästan lika ofta som från underavdelningen *Praktisk teologi* som kommer från den något populärare huvudavdelningen *Filosofi. Psykologi. Religion*. Den huvudavdelning som denna underavdelning tillhör – *Lantbruk* – utgör den sjätte populäraste och vid 24 av totalt 26 fall när användare har lagt till poster från *Lantbruk* har posterna hämtats från denna underavdelning.

Bilden av dessa underavdelningars popularitet vacklar dock något när man undersöker antalet unika användare som lagt till poster från underavdelningarna i tabell 3. Detta visas i tabell 4.

**Tabell 4. Antal användare som lagt till poster från Franklin till PennTags från de 11 underavdelningarna där det är vanligast att lägga till poster till PennTags**

<u>LCC-kod</u>		<u>Antal</u>
PN	Litteratur (allmänna verk)	20
HD	Industrier. Markanvändning. Arbetskraft	9
E151-889	USA (historia)	7
HM	Sociologi (allmänna verk)	7
HQ	Familjen. Äktenskapet. Kvinnor	5
RA	Samhällsmedicin	5
PS	Amerikansk litteratur	5
DS	Asien (historia)	2
RC	Invärtes medicin	2

BV	Praktisk teologi	1
SF	Djuruppfödning	1

I tabell 4 blir det tydligt att populariteten hos underavdelningarna *Praktisk teologi* och *Djuruppfödning* som kunde ses i tabell 3 beror på en enda användares intresse för ämnena i fråga. Populariteten hos underavdelningen *Litteratur (allmänna verk)* består dock och nästan hälften av användarna som lagt till poster till PennTags har lagt till poster från denna underavdelning. Mer än en tredjedel (16 av 45) av dessa har lagt till poster från den tidigare nämnda underavdelningen *Film* på en tredje hierarkisk nivå. Av dessa underavdelningen är det endast från underavdelningen *Litteratur (allmänna verk)* som fler än 10 användare lagt till poster.

#### 6.1.2. Att lägga till en post från Franklin i ett projekt på PennTags

23 av 208 användare har utnyttjat möjligheten att lägga till poster från Franklin till projekt på PennTags. Dessa användare har tillsammans lagt en post från Franklin till ett projekt 173 gånger. Fördelningen över huvudavdelningarna för denna aktivitet visas i tabell 5.

**Tabell 5. Antal tillfällen som en användare har lagt en post från Franklin i ett projekt från respektive huvudavdelning i LCC samt antal användare som utfört denna aktivitet för poster från respektive huvudavdelning**

<u>LCC-kod</u>		<u>Antal tillfällen (antal användare)</u>
P	Språk och litteratur	65 (16)
H	Samhällsvetenskap	19 (7)
Q	Naturvetenskap	18 (2)
R	Medicin	18 (2)
S	Lantbruk	15 (1)
E-F	De amerikanska kontinenternas historia	7 (5)
K	Juridik	6 (3)
G	Geografi. Antropologi. Fritid	5 (1)
M	Musik och böcker om musik	5 (3)
T	Teknik	4 (3)
A	Allmänna arbeten	3 (2)
B	Filosofi. Psykologi. Religion	2 (2)
D	Världshistoria	2 (1)
J	Statsvetenskap	1 (1)
L	Utbildning/pedagogik	1 (1)
Z	Biografi. B & I. Allmänna informationsresurser	1 (1)
C	De historiska hjälpvetenskaperna	0 (0)
N	Konst	0 (0)
U	Militärväsen	0 (0)
V	Sjökrigsväsen	0 (0)
	Okänd kod	1
<u>Totalt</u>		173

I tabell 5 kan vi se att *Språk och litteratur* och *Samhällsvetenskap* även för denna aktivitet är de två huvudavdelningarna där aktiviteten sker mest frekvent. För denna aktivitet tappar dock *Filosofi*, *Psykologi*, *Religion* avsevärt i popularitet medan *Naturvetenskap*, *Medicin* och *Lantbruk* förefaller något populärare än i tidigare tabeller. Från övriga huvudavdelningar har aktiviteten skett färre än 10 gånger.

Tabell 5 visar dock även att ett mycket litet antal användare har lagt Franklinposter till projekt inom samtliga huvudavdelningar. Endast från avdelningen *Språk och litteratur* har fler än 10 användare lagt Franklinposter till projekt.

### 6.1.3. Att annotera en post från Franklin

33 av 208 användare har utnyttjat möjligheten att annotera poster i Franklin. Dessa användare har tillsammans lagt till 189 annotationer. Hur annotationerna fördelar sig över huvudavdelningarna i LCC kan ses i tabell 6.

**Tabell 6. Antal tillfällen som en användare har annoterat en post från Franklin från respektive huvudavdelning i LCC samt antal användare som annoterat poster från respektive huvudavdelning**

<u>LCC-kod</u>		<u>Antal tillfällen (antal användare)</u>
P	Språk och litteratur	110 (22)
H	Samhällsvetenskap	25 (13)
E-F	De amerikanska kontinenternas historia	9 (7)
K	Juridik	8 (2)
M	Musik och böcker om musik	5 (3)
R	Medicin	5 (2)
D	Världshistoria	4 (2)
A	Allmänna arbeten	3 (2)
J	Statsvetenskap	3 (2)
N	Konst	3 (1)
T	Teknik	3 (3)
B	Filosofi. Psykologi. Religion	2 (2)
L	Utbildning/pedagogik	2 (2)
Z	Biografi. B & I. Allmänna informationsresurser	2 (2)
C	De historiska hjälpvetenskaperna	1 (1)
G	Geografi. Antropologi. Fritid	1 (1)
Q	Naturvetenskap	1 (1)
S	Lantbruk	0 (0)
U	Militärväsen	0 (0)
V	Sjökrigsväsen	0 (0)
	Okänd kod	2
<u>Totalt</u>		189

I tabell 6 kan vi se samma mönster som tidigare och även när det gäller att annotera poster dominerar *Språk och litteratur*. Endast från en huvudavdelning utöver detta, *Samhällsvetenskap* har annotationer skrivits vid fler än 10 tillfällen. Närmast

avdelningen *Samhällsvetenskap* i popularitet ligger *De amerikanska kontinenternas historia* där annotationer skrivits vid 9 tillfällen.

Undersöker man antalet unika användare som annoterat poster från huvudavdelningarna och här förstärks denna trend. Endast två huvudavdelningar *Språk och litteratur* och *Samhällsvetenskap* har annoterats av fler än 10 användare. Fördelningen mellan de tre populäraste klasserna består och för övriga huvudavdelningar är antalet annotationer mycket litet. Huvudavdelningarna *Medicin*, *Naturvetenskap* och *Lantbruk* som i tidigare tabeller har förefallit ganska populära ligger för denna aktivitet långt efter de populäraste.

#### 6.1.4. Att tagga en post från Franklin

40 av 208 användare har utnyttjat möjligheten att tagga poster i Franklin. Dessa användare har tillsammans taggat poster i Franklin vid 316 tillfällen. Hur taggningen fördelar sig över huvudavdelningarna i LCC kan ses i tabell 7.

**Tabell 7. Antal tillfällen som en användare har taggat en post från Franklin från respektive huvudavdelning i LCC samt antal användare som taggat poster från respektive huvudavdelning**

<u>LCC-kod</u>		<u>Antal tillfällen (antal användare)</u>
P	Språk och litteratur	112 (24)
H	Samhällsvetenskap	59 (15)
B	Filosofi. Psykologi. Religion	43 (7)
D	Världshistoria	12 (5)
R	Medicin	10 (3)
T	Teknik	10 (5)
E-F	De amerikanska kontinenternas historia	9 (6)
K	Juridik	9 (4)
G	Geografi. Antropologi. Fritid	6 (2)
M	Musik och böcker om musik	6 (4)
Z	Biografi. B & I. Allmänna informationsresurser	6 (4)
J	Statsvetenskap	4 (2)
C	De historiska hjälpvetenskaperna	4 (2)
N	Konst	3 (1)
Q	Naturvetenskap	3 (2)
A	Allmänna arbeten	2 (1)
L	Utbildning/pedagogik	2 (2)
S	Lantbruk	0 (0)
U	Militärväsen	0 (0)
V	Sjökrigsväsen	0 (0)
	Bortfall	17
<u>Totalt</u>		316

I tabell 7 visas att även när det gäller att tagga poster är *Språk och litteratur* nästan dubbelt så populär som den näst populäraste huvudavdelningen. Precis som för föregående aktiviteter är *Samhällsvetenskap* den näst populäraste huvudavdelningen.

När det gäller taggning återtar *Filosofi. Psykologi. Religion* platsen som tredje populäraste huvudavdelning. Vid sidan av dessa huvudavdelningar har endast *Världshistoria, Medicin* och *Teknik* taggats fler än 10 gånger.

Från huvudavdelningen *Lantbruk*, som låg högt upp i tabellen både när det gällde att ha många poster och många poster i projekt har överhuvudtaget inga poster taggats. I tabell 7 kan man också se att inga poster från *Lantbruk* har taggats av användarna. En likadan utveckling kan ses för huvudavdelningen *Naturvetenskap*.

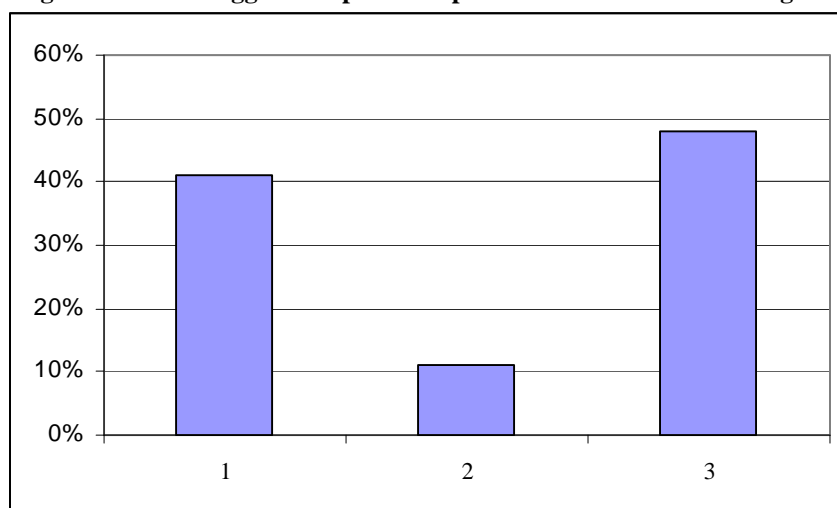
Tittar man på hur många användare som taggat poster inom varje huvudavdelning bekräftas i stora drag den trend som tidigare kunnat ses, med några undantag. Men eftersom endast de två stora huvudavdelningarna *Språk och litteratur* och *Samhällsvetenskap* taggats av fler än 10 användare känns det dock osäkert att uttala sig om övriga huvudavdelningars placeringar eftersom det rör sig om små skillnader.

## 6.2. Tillförandegrad och ämnesområde

De allra populäraste huvudavdelningarna när det gäller att lägga till taggar till poster i Franklin är *Språk och litteratur, Samhällsvetenskap* och *Filosofi. Psykologi. Religion*. För de 30 poster som analyserats för respektive ämnesområde har användarna lagt till 99 taggar till posterna inom *Språk och litteratur*, 74 taggar till *Samhällsvetenskap* och 42 taggar till *Filosofi. Psykologi. Religion*. Totalt har 215 taggar analyserats.

I diagram 2 visas hur i hur stor utsträckning de studerade taggar speglar tillförandegrad 1-3.

**Diagram 2. Andel taggar som passar in på kriterierna för tillförandegraderna (i %)**



Som man kan se i diagram 2 hör nästan hälften (48 %) till tillförandegrad 3 vilket innebär att de tillför de fördelar som taggning kan förväntas göra. Nästan lika många (41 %) av de studerade taggarna faller in på de kriterier som satts för tillförandegrad 1, det vill säga tillför de nackdelar som taggning förväntas innebära.

Till tillförandegrad 2 hör endast 11 % av de studerade taggarna. Tillförandegrad 2 innebär att tillföra termer som skulle kunna sägas vara användbara, exempelvis

synonymer till begrepp som redan finns i dokumentbeskrivningen, men på ett sätt som inte är i linje med de principer som vokabulärkontroll vilar på.

Diagram 3 visar hur fördelningen ser ut över tillförandegraderna inom respektive av de studerade huvudavdelningarna.

**Diagram 3. Andel taggar som passar in på kriterierna för tillförandegraderna i huvudavdelningarna B, H, P (i %)**

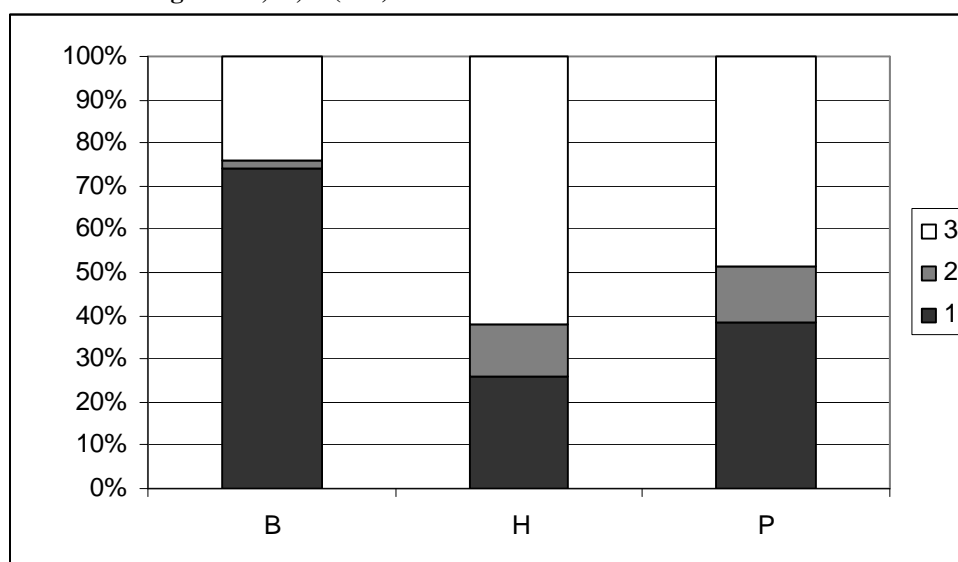


Diagram 3 visar att tillförandegraden i taggningen är något olika om de olika huvudavdelningarna. I huvudavdelningen *Filosofi. Psykologi. Religion* (B) ryms hela 74 % av de studerade taggarna inom kriterierna för tillförandegrad 1, det vill säga skapar ett mycket begränsat värde. Motsvarande siffra för *Språk och litteratur* (P) är 38 %. I huvudavdelningen *Samhällsvetenskap* (H) återfinns den minsta andelen av taggar som tillför ett begränsat värde, 26 %.

Tillförandegrad 2 är den kategorin som utgör den minsta andelen av taggarna inom samtliga huvudavdelningar. Inom *Filosofin, psykologi, religion* är denna andel knappt märkbar i diagrammet, så lite som 2 %. I huvudavdelningarna *Samhällsvetenskap* och *Språk och litteratur* utgör tillförandegrad 2 ungefär lika stor andel, 12 respektive 13 %.

De taggar som speglar tillförandegrad 3 och därigenom medför de fördelar som taggning kan förväntas göra utgör 62 % av de studerade taggarna inom *Samhällsvetenskap*. I huvudavdelningen *Språk och litteratur* faller 48 % in under de kriterier som gäller för tillförandegrad 3. Den huvudavdelning som har minst andel taggar inom denna tillförandegrad är *Filosofi. Psykologi. Religion* med 24 %.

## 7. SLUTSATSER

I detta avsnitt presenteras de svar som den empiriska undersökningen ger på de forskningsfrågor som ställts upp för studien. Forskningsfrågorna är:

- I hur stor utsträckning utnyttjas PennTags tekniska möjligheter för poster från Franklin inom olika ämnesområden?
- Vad tillför användarna genom taggning till dokumentbeskrivningen i Franklin inom olika ämnesområden?

När det gäller användarnas aktivitet på PennTags måste en första slutsats bli att det främsta användningsområdet för PennTags inte verkar vara att lägga till poster från Franklin då bara drygt en femtedel av användarna någon gång utnyttjat denna möjlighet. De användare som ändå väljer att lägga till poster från Franklin till PennTags är framförallt intresserade av att tagga och annotera poster. Att lägga Franklinposter i projekt är en relativt ovanlig aktivitet bland användarna.

Tittar man på hur användarnas aktivitet är spridd över ämnesområden framträder ett ämnesområde, *Språk och litteratur*, som det totalt dominerande oavsett vilken av de fyra aktiviteterna det är som avses. Inom denna huvudavdelning i LCC är det dessutom en enda underavdelning på den tredje hierarkiska nivån – *Film* – som står för över hälften av tilläggen från *Språk och litteratur*. Från denna underavdelning har dessutom mer än en tredjedel av användarna någon gång lagt till poster från Franklin. Det enda ämnesområden vid sidan av *Språk och litteratur* som håller en någorlunda jämn hög profil är *Samhällsvetenskap*.

En annan slutsats är att det inom ett mycket stort antal ämnesområden inte förekommer någon aktivitet att tala om på PennTags. *Allmänna arbeten, Biografi. B & I. Allmänna informationsresurser, De historiska hjälpvetenskaperna, Geografi. Antropologi. Fritid, Konst, Militärväsen, Musik och böcker om musik, Sjökrigsväsen, Statsvetenskap och Utbildning/pedagogik* är huvudavdelningar varifrån poster har lagts till PennTags vid färre än 10 tillfällen. Inom vissa ämnesområden som kanske särskilt *Lantbruk* men i viss mån även *Filosofi. Psykologi. Religion, Medicin* och *Naturvetenskap* gör enskilda användare ett stort jobb när det gäller att lägga till poster till PennTags.

Undersöker man vad taggningen tillför inom de mest frekvent taggade ämnesområdena *Språk och litteratur, Samhällsvetenskap* och *Filosofi. Psykologi. Religion* framträder sammantaget en bild av att nästan hälften av taggarna tillför de fördelar som taggning förväntas kunna tillföra och 59 % tillför nya meningsfulla sökingångar.

Tittar man på varje ämnesområde för sig framträder dock vissa skillnader i vad taggningen tillför. I *Samhällsvetenskap* tillför den största andelen taggar de fördelar som taggning förväntas kunna tillföra, 62 % jämfört med 48 % i *Språk och litteratur*. Endast 24 % taggar inom *Filosofi. Psykologi. Religion* är taggar som tillför dessa fördelar. Inom *Samhällsvetenskap* skapar 74 % av taggarna vad som betraktas som meningsfulla nya sökingångar och motsvarande siffra för *Filosofi. Psykologi. Religion* är 26 %.

## 8. ANALYS & DISKUSSION

När det gäller användarnas aktivitet på PennTags måste en första slutsats bli att det främsta användningsområdet för PennTags inte verkar vara att lägga till poster från Franklin då bara drygt en femtedel av användarna någon gång utnyttjat denna möjlighet. Men PennTags används – åtminstone i min mening – i relativt liten utsträckning överhuvudtaget. Av de minst 24 000 potentiella användare som studenterna utgör – eller 48 000 om man räknar in hela Penn Community – har endast drygt 800 någon gång skapat en post på PennTags under de 2 år som PennTags har funnits.

Guy och Tonkin diskuterar vikten av att ha riktlinjer för hur taggar ska sättas, men samtidigt vikten av att dessa riktlinjer inte ska vara för stränga eftersom folksonomins naturliga charm då kan gå förlorad. Också Bar-Ilan et al finner att en tydlig struktur skapar kvalitativt ”bättre” metadata men att riktlinjer samtidigt skapar högre barriärer för deltagande. På PennTags finns riktlinjer och det skulle naturligtvis kunna vara så att dessa skapar en miljö som användarna uppfattar som för strikt. Att PennTags används i så liten utsträckning skulle också kunna förklaras mot bakgrund av Shirkeys resonemang om folksonomier som en framtvungad rörelse i brist på bättre alternativ. I PennTags kontext finns ett alternativ. Posterna i Franklin är redan beskrivna med ämnesord av indexerare och det kanske gör att behovet att själv tagga dem inte känns lika stort. Samtidigt finns en rad andra tänkbara förklaringar till förhållandet. Kanske tycker användarna att de dokumentbeskrivningar som finns i Franklin är perfekta, eller kanske bryr de sig helt enkelt inte om att de inte speglar deras egen vokabulär. Svaret på denna fråga ligger naturligtvis utanför vad min empiriska undersökning kan ge svar på, men det är ett förslag till vidare forskning att med hjälp av mer direkta metoder undersöka hur användarna uppfattar read/write OPAC.

När användarna ändå är aktiva på PennTags är det framförallt att tagga och att skriva annotationer de är intresserade av. Att lägga till Franklinposter i projekt är en relativt ovanlig aktivitet bland användarna. Att taggning är så vanligt skulle kunna bero på att det är en förutsättning för att övriga tillägg ska synas i Franklin. Det är inte otänkbart att användarna känner till att PennTags funkar så och taggar av just den anledningen. Det skulle också kunna bero på att de använder taggarna för att organisera sina egna samlingar av favoritresurser på PennTags. Men det skulle faktiskt också kunna vara ett tecken på ett intresse från användarnas sida av att påverka hur dokument beskrivs i Franklin.

Genom taggningen tillför användarna generellt en stor andel nya sökingångar vilket ligger i linje med hur Lin et al fann en överlappning på endast 11 % mellan de taggar och de ämnesord ett antal dokument getts. Att överlappningen är så liten betyder att en stor del av termerna är nya och alltså tillför något, vilket var fallet även i min studie. I likhet med Trant försökte jag i denna studie även ta reda på i vilken utsträckning de nya sökingångarna var meningsfulla och fann att de generellt var det. Denna slutsats ligger alltså även i linje med Spiteri förhoppning om att taggningen ska kunna fungera som ett komplement till den kontrollerade vokabulären, åtminstone inom huvudavdelningarna *Språk och litteratur* och *Samhälle*. Även om det är som Tennis säger, att taggning har andra syften och fyller andra funktioner än indexering, så lyckas PennTags användare genom taggning tillföra specificitet och

uttömmandegrad till dokumentbeskrivningar i PennTags. Men å andra sidan består en stor andel av taggarna av språkvarianter och upprepningar och därigenom tillför taggningen nästan lika ofta de nackdelar som taggningen förväntas kunna tillföra, så som Spiteri förutspår.

Den museikontext som Trants taggningsförsök genomfördes i liknar på många vis PennTags. I båda fallen finns redan en beskrivning av dokumentet, det finns redan en relation mellan användare och institution och i denna relation har man iakttagit ett glapp som man genom introduktionen av taggning hoppas kunna överbrygga. Detta kanske kan vara anledningen till att både min och Trants studie hamnar i slutsatser om att användarna har mycket meningsfullt att bidra med, till skillnad från exempelvis Guy och Tonkin som i studien av Flickr finner att nästan hälften av taggarna är felstavade eller felkonstruerade. Kanske tar man sig i kragen på ett annat vis en biblioteks- eller museikontext på grund av att man antar att saker och ting ska vara på ett visst vis och anpassar sig. Eller också har de regler som Guy och Tonkin föreslår för taggning slagit igenom på PennTags. Riktlinjer finns ju på PennTags och kanske tycker användarna av att de är meningsfulla att följa.

En annan tänkbar anledning till att både min och Trants studie finner att användarnas taggning generellt tillför någonting meningsfullt till skillnad från Guy och Tonkin är att de analyserat sitt material maskinellt och utan att ta hänsyn till att taggars uppbyggnad ofta skiljer sig ifrån ämnesords. Flera forskare, bland annat Kipp och Campbell och Golder och Huberman pekar på att det med tiden uppstår överenskommelser mellan delar av användarskaran omkring exempelvis hur termer ska sättas samman i fraser precis som det på PennTags är allmänt vedertaget att film noir skrivs *film\_noir*. Om jag liksom Guy och Tonkin ignorerat detta och betraktat en tagg som *film\_noir* som felstavad hade jag onekligen fått ett annorlunda resultat.

En del skillnader finns dock även mellan min och Trants studie. Trants studie rör beskrivning av bilder, något som ofta sägs vara annorlunda än beskrivning av textuellt material som jag studerat. Ett större tolkningsutrymme, som man brukar säga finns med bilder, kanske skulle kunna förklara varför de användare Trant studerade förmådde skapa en större andel nya och användbara sökingångar. Att Trant fann att hela 77 % av de nya sökingångarna var meningsfulla i jämförelse med 59 % i min studie skulle också kunna bero på att de kriterier Trant använder för att bedöma användbarhet skiljer sig från hur jag har bedömt meningsfullhet. I Trants studie gavs dessutom användarna ett antal bilder som de ombads beskriva medan de i min studie redan hade beskrivit dokumenten på eget initiativ vilket kanske påverkar resultaten. Även om Trants användargrupp hade möjlighet att avstå från att beskriva någon bild gjorde kanske vetskapen om att man deltog i ett experiment att man handlade annorlunda än vad man gjort i ett verkligt scenario. Denna risk finns hur som helst inte i min studie, men det skulle kunna vara en förklaring till att Trants användare skapar fler användbara sökingångar än de jag studerat.

Golder och Huberman samt Kipp och Campbell iakttog att ett stort antal taggar fyllde andra funktioner än att beskriva ett dokumentens ämne och så skulle kunna vara fallet även på PennTags eftersom så mycket som 41 % av taggarna inom de mest frekvent taggade områdena speglar tillförandegrad 1, dit sådana taggar skulle räknas. Men hit räknas också taggar som är identiska med, eller språkvarianter av, termer som

förekommer i ett annat sökbart fält. Lin et al resonerar i sin studie om att användarna i stor utsträckning hämtar termer från exempelvis titeln enligt en teori om att användarna eftersträvar minsta möjliga ansträngning och de 41 % av taggarna som speglar tillförandegrad 1 skulle också kunna vara ett tecken på att detta stämmer.

Tittar man på hur användarnas aktivitet är spridd över ämnesområden framträder ett ämnesområde, *Språk och litteratur*, som det totalt dominerande oavsett vilken av de fyra aktiviteterna det är som avses. Det enda ämnesområdet vid sidan av *Språk och litteratur* som håller en någorlunda jämn hög profil är *Samhällsvetenskap*. En annan slutsats blir därför att det inom ett mycket stort antal ämnesområden inte förekommer någon aktivitet att tala om på PennTags. *Allmänna arbeten, Biografi. B & I. Allmänna informationsresurser, De historiska hjälpsvetenskaperna, Geografi. Antropologi. Fritid, Konst, Militärväsen, Musik och böcker om musik, Sjökrigsväsen, Statsvetenskap och Utbildning/pedagogik* är huvudavdelningar varifrån poster har lagts till PennTags vid färre än 10 tillfällen. Inom vissa ämnesområden som kanske särskilt *Lantbruk* men i viss mån även *Filosofi. Psykologi. Religion, Medicin* och *Naturvetenskap* gör enskilda användare ett stort jobb när det gäller att lägga till poster till PennTags.

Anledningen till att *Språk och litteratur* dominerar beror på den mycket populära underavdelningen *Film*. PennTags förefaller onekligen fylla en större funktion för en grupp människor som är intresserade av film än för användare i allmänhet. Nästan en tredjedel av de användarna som någon gång har lagt till en post från Franklin till PennTags har faktiskt lagt till en post från underavdelningen *Film*. Denna grupp verkar använda PennTags för att tagga Franklinposter och tipsa varandra om intressant litteratur. En del av dessa användare skapar minst sagt imponerande annoterade bibliografier inom ett visst filmområde som absolut måste sägas generera ett värde. Detta värde kommer även Franklin till nytta eftersom man via ett enkelt klick på en länk kan ta sig från posten i Franklin till posten i PennTags om exempelvis ett projekt som en post ligger i skulle verka intressant. I projektet kanske man sedan hittar fler intressanta dokument inom samma ämne.

Men vad beror denna filmdominans på PennTags på? Kanske är dessa användare klasskamrater eller rentav goda vänner som tipsar varandra om PennTags. Kanske har någon gjort reklam för PennTags på föreläsningar inom filmvetenskap. Anledningarna till detta kan jag på basis av min empiriska undersökning egentligen bara spekulera i. Habib menar att en read/write OPAC skulle kunna fungera som ett sätt för studenter att tipsa om material till en viss kurs eller ett visst projekt, och kanske är det just det som visar sig i mitt material, att det har hänt bland filmvetarna.

Men film – som konstform – kan också ses som en del i det lager som Brand kallar *Mode och konst*. Detta är det lager som enligt Brand har den allra snabbaste förändringshastigheten. Detta ämnesområdes popularitet på PennTags och det faktum att taggningen inom film generellt skapar många nya och användbara sökingångar kan därför sägas bekräfta att ett ämnes förändringshastighet påverkar användarnas incitament till aktivitet och möjlighet att tillföra. Ämnen inom *Mode och konst* går snabbt framåt och den kontrollerade vokabulären kanske inte hänger med i samma tempo som ämnet och användarna som intresserar sig för det.

Eftersom en ganska stor andel av de poster som har lagts till inom *Filosofi. Psykologi. Religion* lagts till från just underavdelningar för religion tycker jag att detta kan sägas spegla det lager i samhället som Brand kallar *Kultur*. Religion är enligt Brand (1999) det allra trögaste området inom lagret *Kultur* (s. 42). Inom ämnesområdet religion verkar heller inget större allmänt intresse för taggning finnas. När taggning ändå sker förmår den inte tillföra särskilt mycket utan medför främst de nackdelar som taggning kan förväntas tillföra. Än en gång verkar förändringshastigheten hos ämnet påverka användarnas incitament till aktivitet och möjlighet till vettiga tillägg genom taggning. Kanske finns det helt enkelt inget nytt att tillägga utöver det som täcks in i den kontrollerade vokabulären eftersom denna i god tid kunnat uppfatta i princip alla nya aspekter.

Aktiviteten och taggningen inom *Språk och litteratur* och *Filosofi. Psykologi. Religion* kan på detta vis ganska enkelt förklaras med hjälp av Brands lagermodell, mycket på grund av att enskilda snävare ämnesområden som lätt kan härledas till ett enda lager i modellen är dominerande. När det gäller *Samhällsvetenskap* är nöten något svårare att knäcka. *Samhällsvetenskap* är en bred huvudavdelning som dock är något snävare än vad namnet ger sken av. Ett antal områden som vi traditionellt förknippar med samhällsvetenskap finns nämligen inte i denna huvudavdelning. De discipliner som lätt skulle kunna härledas till det tröga lagret *Statsskick* – statsvetenskap och juridik – har fått egna huvudavdelningar och från ingen av dessa har poster lagts till PennTags i någon större utsträckning. Historia, geografi och antropologi har också fått egna huvudavdelningar och i ingen av dem förekommer någon större aktivitet. *Samhällsvetenskap* består istället av allmän samhällsvetenskap, ekonomi och sociologi. Att *Samhällsvetenskap* genom att det innehåller ekonomi kan härledas till lagret *Handel* som nästan rör sig lika fort framåt som *Mode och konst* känns som en frestande slutsats. Men i själva verket hämtar den sortens ekonomi som återfinns i denna huvudavdelning sitt studieobjekt från flera av lagren i Brands modell för här finns plats för både handel och offentlig ekonomi. Det är heller inte så att någon övervägande majoritet av de posterna kommer från underavdelningarna för ekonomi. En av de totalt sett största underavdelningarna – *Industrier. Markanvändning. Arbetskraft* – tillhör visserligen ekonomi men dessa områden brukar sällan anses tillhöra de snabbaste inom området, kanske snarare tvärt om.

Från *Samhällsvetenskap* kommer två till av de totalt sett största underklasserna, och dessa – *Sociologi (allmänna verk)* och *Familjen. Äktenskapet. Kvinnor* – hör till den sociologiska delen av huvudavdelningen. Sociologi beskrivs av Nationalencyklopedin (1995) som vetenskapen om samhället och man beskriver hur disciplinen tar sin utgångspunkt i att mänskligt handlande och tänkande – liksom skillnader mellan individer eller grupper – måste ses som ett resultat sociala förhållanden och historiskt nedärvda skillnader. ”Historiskt nedärvda skillnader” för åtminstone mina tankar till något som rör sig framåt i mycket långsam hastighet. Och om sociologi är studiet av samhället så borde disciplinens studieobjekt finnas inom flera av Brands lager eftersom denna modell är en modell över just samhället. Min slutsats blir därför att anledningen till att så många användare väljer att lägga till poster till PennTags från *Samhällsvetenskap* och dessutom förmår skapa så många nya och användbara sökingångar när de taggar poster från *Samhällsvetenskap* helt enkelt måste sökas någon annanstans än i Brands lagermodell.

I bakgrundskapitlet tog jag upp att en hel del debatt har förekommit omkring vad som kallas en kulturell eller ideologisk vridning i LCSH och den frekventa taggningen inom *Samhällsvetenskap* skulle kunna vara försök att visa en annan sida. Fischer (2005) menar att ett av de områden som kritiken om ideologisk vridning riktats tydligast mot under de senare åren är hur mannen betraktas som norm i systemet och hur litteratur om feminism endast kan indelas i tre olika typer, trots att många menar att det existerar långt fler än så (s. 93-94). Att en underavdelning som *Familjen. Äktenskapet. Kvinnor* tillhör de populäraste skulle kunna vara ett tecken på hur användare inte håller med om hur dessa ämnen beskrivs i LCSH. Samtidigt finns en rad andra tänkbara förklaringar till förhållandet – kanske ett stort intresse för ämnet bland användarna – men det skulle kunna vara intressant för framtida forskning att utreda om det finns något samband mellan taggning och ämnesord som på detta vis anses något kontroversiella.

Samhällsvetenskap är bara ett av flera exempel på huvudavdelningar som är svåra att härleda till ett enda lager i Brands modell. Av denna anledning valde jag även att undersöka hur användarnas aktivitet på underavdelningarna såg ut. Men hur fungerar modellen för att förklara varför de största underavdelningarna på en andra hierarkisk nivå utgörs av just *Litteratur (allmänna verk)*, *Praktiskt teologi*, *Djuruppfödning*, *Invärtes medicin*, *Industrier. Markanvändning. Arbetskraft*, *Sociologi (allmänna verk)*, *Familjen. Äktenskapet. Kvinnor*, *Samhällsmedicin*, *Asien (historia)*, *Amerikansk litteratur*, *USA (historia)*? Tre av dessa, *Djuruppfödning*, *Invärtes medicin* och *Samhällsmedicin* är naturvetenskapliga ämnen. Naturvetenskapen tar generellt sitt studieobjekt i vad som kan räknas till lagret *Natur* och just dessa ämnen är dessutom närliggande ämnet biologi vilket kanske är det naturvetenskapliga ämnet som enklast kan kopplas till just naturen. Att dessa tre underavdelningar överhuvudtaget förekommer bland de populäraste visar sig emellertid vara ett resultat av enskilda användares stora intresse för ämnet. Få användare väljer alltså att vara aktiva på PennTags när det gäller poster från Franklin inom naturvetenskapliga ämnen och det skulle kunna förklaras med att det studieobjekt som naturvetenskapen berör rör tillhör de trögare lagren.

Underavdelningen *Samhällsmedicin* skulle också kunna kopplas till ett mer samhällsorienterat fält som flera av de populära underavdelningarna skulle kunna räknas till. *Industrier. Markanvändning. Arbetskraft. Sociologi (allmänna verk)*, *Familjen. Äktenskapet. Kvinnor* tillhör huvudavdelningen *Samhällsvetenskap* men hit skulle också de två populära underavdelningarna med Asiens respektive USA: s historia kunna räknas. Räknar man på detta vis tillhör hela sex av de elva populäraste underavdelningarna samhällsvetenskapligt orienterade ämnen. Återigen framstår *Samhällsvetenskap* som ett område där användarna vill och kan tillföra en hel del.

Utgår man från Brands lagermodell och mina antaganden om incitament till aktivitet finns det huvudavdelningar i LCC som borde vara minst lika populära som *Språk och litteratur*, och det borde inte vara *Samhällsvetenskap*. Jag tänker närmast på huvudavdelningen *Konst* men även på *Musik och böcker om musik*, huvudavdelningar som uppenbart kan härledas till lagret *Mode och konst* i Brands modell. *Musik och böcker om musik* är något populärare än *Konst* men från ingen av avdelningarna har fler än 10 användare lagt till poster vilket måste anses som ett mycket begränsat intresse för att påverka hur dokument inom dessa ämnesområden beskrivs. Samma problematik stöter man naturligtvis på om man tittar på de populäraste

underavdelningarna. Det går generellt ganska bra att förklara varför just dessa underavdelningar tillhör de populäraste, men samtidigt fattas en rad av underavdelningar som enligt teorin borde vara populärare än vad de är på PennTags

Det faktum att det i de allra flesta huvudavdelningar förekommer mycket låg aktivitet innebär dessutom att det blir svårt att argumentera för att skillnaden i popularitet skulle vara av betydelse, att det exempelvis skulle visa på någonting av vikt att 4 av 208 användare har valt att lägga poster till PennTags från *Musik och böcker om musik* medan endast 2 har gjort detsamma från *Geografi. Antropologi. Fritid*. Det blir närmast fånigt att försöka härleda detta till att *Musik och böcker om musik* kommer från ett snabbare lager än *Geografi. Antropologi. Fritid* – för den enda uppenbara slutsatsen är att användarna generellt är i princip ointresserade av PennTags, förutom när det gäller *Språk och litteratur* och i någon mån *Samhällsvetenskap*.

Min slutsats blir att teorin om pace layering inte kan betraktas som en tillfredställande förklaringsmodell om man vill förstå användarnas aktivitet och möjlighet att tillföra användbara sökingångar på PennTags. Vissa ämnesområdets relativa popularitet går att härledas till modellen, men alltför många fall avviker från det mönster som man borde kunna iakta. Det står därför klart att fler teorier måste lyftas in för att på ett meningsfullt sätt kunna förklara användarnas aktivitet på PennTags.

Under arbetets gång har jag vid flera tillfällen kommit på mig själv med en önskan att komma i kontakt med användarna – för jag har så många frågor som jag skulle vilja att de svarade på. Varför använder de inte PennTags mer? Känner de inte till att möjligheten finns? Är det något fel på PennTags – och i så fall vad? För PennTags förefaller inte vara någon revolution av delaktighet från användarnas sida och någon demokratisering av dokumentbeskrivningarna i 2.0-anda tycks inte ske. Mot bakgrund av mitt empiriska material kan jag omöjligt kalla PennTags för biblioteksvetenskap när den är som allra bäst, så som Maness beskriver read/write OPAC. Men om man däremot sänker de uppskrivade förväntningarna en aning och betraktar PennTags för vad det egentligen är förefaller det vara ett system som fyller en funktion för en grupp användare som ges möjlighet att lagra och organisera länkar till sina favoritresurser och dela dem med andra. I vissa fall har detta dessutom tillfört meningsfulla sökingångar och andra användbara element till dokumentbeskrivningarna i Franklin som kommer till nytta för en större användargrupp.

## 9. SAMMANFATTNING

*Read/write OPAC* är en av flera komponenter i begreppet Biblioteket 2.0 och den brukar beskrivas som en OPAC där användarna ges möjlighet att exempelvis kommentera poster eller genom taggning sätta kompletterande ämnesord, så kallade taggar. Fenomenet diskuteras ofta i mycket positiva ordalag men inga vetenskapliga undersökningar har ännu gjorts på vad resultatet av en implementering av en *read/write OPAC* kan innebära.

Denna studies syfte är att undersöka hur en tillämpning av *read/write OPAC* kan fungera i praktiken genom att studera hur användarna vid universitetsbiblioteket vid Pennsylvania University i USA använder systemet *PennTags*. Detta för att ta reda på hur användarna utnyttjar möjligheten att påverka innehållet i OPAC: en Franklin och därigenom skapa sökingångar som inte täcks in i den befintliga dokumentbeskrivningen.

För att uppnå detta syfte har jag arbeta med forskningsfrågorna:

- I hur stor utsträckning utnyttjas *PennTags* tekniska möjligheter för poster från Franklin inom olika ämnesområden?
- Vad tillför användarna genom taggning till dokumentbeskrivningar i Franklin inom olika ämnesområden?

Materialet som används för att besvara forskningsfrågorna utgörs av de poster en sampl av användarna skapat på *PennTags* för poster från Franklin. Dessa poster undersöks utifrån en kvantitativ ansats i syfte att studera om det finns något samband mellan användarnas aktivitet och ämnesområde och/eller mellan vad jag har valt att kalla tillförandegrad i taggningen och ämnesområde. *Pace layering* är en teori som menar att samhället är indelat i olika lager med olika hög förändringshastighet och jag undersöker i denna studie om det finns anledning att tro att incitamenten för aktivitet på *PennTags* och/eller möjligheten till ett meningsfullt tillförande genom taggning är beroende av förändringshastigheten hos ämnesområdet som dokumentet som beskrivs handlar om.

För att definiera ämnesområde har jag använt mig av de *call number* i *Library of Congress Classification* som posterna getts i Franklin. De aktiviteter varje användare kan utföra när en post från Franklin läggs till *PennTags* är att lägga posten i ett projekt, att skriva en annotation till posten och/eller att ge posten ett valfritt antal taggar. Även att lägga till en post till *PennTags* betraktas som en aktivitet.

Tillförandegraderna är tre stycken och är konstruerade utifrån de för- och nackdelar som taggning förväntas kunna tillföra. Tillförandegrad 3 och 2 betraktas som tillförande av meningsfulla sökingångar. Tillförandegraden i taggningen har mätts för de 3 ämnesområden där taggningen är mest frekvent – *Språk och litteratur*, *Samhällsvetenskap* och *Filosofi. Psykologi. Religion*.

Studien finner att intresset för att genom *PennTags* påverka OPAC: en generellt är ganska lågt förutom för poster från huvudavdelningen *Språk och litteratur* som

dominerar mitt empiriska material när det gäller samtliga aktiviteter. Vid sidan av denna förekommer endast aktivitet för poster från huvudavdelningen *Samhällsvetenskap* i någorlunda hög nivå på PennTags. För poster från många huvudavdelningar förekommer liten eller ingen aktivitet och i några huvudavdelningar gör enskilda användare ett stort arbete genom annotationer och taggande.

När användarna taggar tillför de i 48 % av fallen de fördelar som taggning förväntas kunna tillföra och i 59 % av fallen skapas nya sökingångar som betraktas som meningsfulla. Mellan ämnesområdena skiljer sig dock graden av tillförande åt och i huvudavdelningen *Filosofi. Psykologi. Religion* är tillförandet av meningsfulla tillägg endast 26 % medan det i *Samhällsvetenskap* är 74 %.

Jag finner att populariteten hos *Språk och litteratur* skulle kunna förklaras med hjälp av teorin om *pace layering*, likaså att användarna inte förmår tillföra särskilt många meningsfulla sökingångar genom taggning inom *Filosofi. Psykologi. Religion*. Samtidigt är de fall som avviker från teorin för många för att denna ska anses vara en tillräcklig förklaringsmodell och jag föreslår därför att andra teorier måste användas för att till fullo förstå användarbeteende i system som Penntags

## KÄLL- OCH LITTERATURFÖRTECKNING

Allen, Laurie & Winkler, Michael (2006). *PennTags: Social bookmarking in a University Library*.

<http://www.slideshare.net/laurie.allen/penntags-presentation-at-educause-2006>  
[2007-05-08]

*Amazon*. [www.amazon.com](http://www.amazon.com) [2007-06-10]

Benito, Miguel (2001). *Kunskapsorganisation: En introduktion till katalogisering, klassifikation och indexering*. Borås: Taranco.

Brand, Stewart (1997). *How buildings learn: What happens after they're built*. London: Phoenix.

Brand, Stewart (1999). *The clock of the long now: Time and responsibility*. New York: Basic Books.

Brante, Thomas (1995). Sociologi. Ingår i Engström, Christer, ed. *Nationalencyklopedin*, band 17, s. 35-36. Höganäs: Bra Böcker.

Bryman, Alan (2002). *Samhällsvetenskapliga metoder*. Malmö: Liber ekonomi.

Casey, Michael E. & Savastinuk, Laura C. (2006). Library 2.0: Service for the next-generation library, *Library Journal*, issue 14 (september 1), 2006.  
<http://www.libraryjournal.com/article/CA6365200.html> [2007-05-14]

Chan, Lois Mai (1999). *Guide to the Library of Congress Classification*. 5 ed. Englewood, Colo.: Libraries Unlimited.

Chan, Lois Mai & Hodges, Theodora (2000). Entering the millenium: A new century for LCSH. *Cataloging & Classification Quarterly*, vol. 29, no. 1-2, s. 225-234.

*Delicious*. <http://del.icio.us/> [2007-06-10]

Exploring the characteristics of social classification (2006). Av Lin, Xia, Beaudoin, Joan E., Bui, Yen & Desai, Kaushal. Ingår i *Proceedings of the 17th SIG Classification Research Workshop, 2006*.  
<http://www.slais.ubc.ca/users/sigcr/sigcr-06lin.pdf> [2007-05-02]

Feinberg, Melanie (2006). An examination of authority in social classification systems. Ingår i *Proceedings of the 17th SIG Classification Research Workshop, 2006*.  
<http://www.slais.ubc.ca/users/sigcr/sigcr-06feinberg.pdf> [2007-05-06]

Fischer, Karen S. (2005). Critical views of LCSH, 1990-2001: The third bibliographic essay. *Cataloging & Classification Quarterly*, vol. 41, no. 1, s. 63-105.

Flickr. <http://www.flickr.com/> [2007-06-10]

Giger, Peter (2006). Constructing the Web 2.0 concept. Del 1 av 2 i *Participation literacy*. Lic. Karlskrona: Blekinge Tekniska högskola. (Dissertation series no 2006:07).

Golder, Scott A. & Huberman, Bernardo A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, vol. 32, no. 2, s. 198 - 208. <http://jis.sagepub.com/cgi/reprint/32/2/198> [2007-05-11]

Google. [www.google.com](http://www.google.com) [2007-06-10]

Guy, Marieke & Tonkin, Emma (2006). Folksonomies: Tidying up tags?. *D-Lib Magazine*, vol. 12, no. 1. <http://www.dlib.org/dlib/january06/guy/01guy.html> [2007-01-30]

Habib, Michael C. (2006). *Toward Academic Library 2.0: Development and Application of a Library 2.0 Methodology*. Chapel Hill, N.C.: School of Information and Library Science. (SILS Master's Papers). <http://hdl.handle.net/1901/356> [2007-01-25]

Kipp, Margaret E. I. & Campbell, Grant D. (2006). Patterns and inconsistencies in collaborative tagging systems: An examination of tagging practices. Ingår i *Proceedings of the ASIST Annual Meeting, 2006*. <http://eprints.rclis.org/archive/00008315> [2007-03-02]

Kroski, Ellyssa (2005). The hive mind: Folksonomies and user-based tagging. *Infotangle*. <http://infotangle.blogspot.com/2005/12/> [2007-03-02]

Lancaster, Frederick Wilfrid (1998). *Indexing and abstracting in theory and practice*. London: Library Association.

Lawley, Liz (2005). It's the social network, stupid! *Many2Many: A group weblog on social software*, January 20, 2005. [http://many.corante.com/archives/2005/01/20/its\\_the\\_social\\_network\\_stupid.php](http://many.corante.com/archives/2005/01/20/its_the_social_network_stupid.php) [2007-05-14]

Lawrence, K. F. & Schraefel, M. C. (2006) Freedom and Restraint: Tags, Vocabularies and Ontologies. Ingår i *Proceedings of 2nd IEEE International Conference on Information & Communication Technologies: From Theory to Applications, 24-28 April 2006, Damascus, Syria*. [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?isnumber=35470&arnumber=1684649&count=375&index=322](http://ieeexplore.ieee.org/xpls/abs_all.jsp?isnumber=35470&arnumber=1684649&count=375&index=322) [2007-05-11]

Library of Congress. *Library of Congress Classification Outline*. <http://www.loc.gov/catdir/cpsolcco/> [2007-05-11]

Macaskill, Wendy & Owen, Dylan (2006). Web 2.0 to go. Ingår i *Proceedings LIANZA Conference 2006, Wellington (New Zealand)*.  
[http://www.lianza.org.nz/library/files/store\\_013/Web2ToGo\\_WMacaskill.pdf](http://www.lianza.org.nz/library/files/store_013/Web2ToGo_WMacaskill.pdf)  
[2007-04-30]

Macgregor, George & McCulloch, Emma (2006). Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review*, vol. 55, issue 5, s. 291-300.

Maness, Jack M. (2006). Library 2.0 Theory: Web 2.0 and Its Implications for Libraries. *Webology*, vol. 3, no. 2.  
<http://www.webology.ir/2006/v3n2/a25.html> [2007-01-25]

Mejias, Ulises Ali (2005). Tag literacy. *IDEANT*, April 22, 2005.  
[http://ideant.typepad.com/ideant/2005/04/tag\\_literacy.html](http://ideant.typepad.com/ideant/2005/04/tag_literacy.html) [2007-04-20]

Merholz, Peter (2005). Clay Shirky's viewpoints are overrated. *Peterme.com*, August 07, 2005.  
<http://www.peterme.com/archives/000558.html> [2007-04-20]

Morville, Peter (2005). *Ambient findability*. Sebastopol, Calif.: O'Reilly.

O'Reilly, Tim (2005). *What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software*.  
<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>  
[2007-04-20]

Penn Libraries (2007-04-19) *Visitor information*.  
<http://www.library.upenn.edu/access/visitors/visitors.html> [2007-05-11]

*PennTags*. <http://tags.library.upenn.edu/> [2007-06-10]

Peterson, Elaine (2006). Beneath the metadata: Some philosophical problems with folksonomy. *D-Lib Magazine*, vol. 12, no. 11.  
<http://www.dlib.org/dlib/november06/peterson/11peterson.html> [2007-04-30]

Rowley, Jennifer (2000). *Organizing knowledge: An introduction to managing access to information*. 3. ed. Aldershot: Gower.

Shirky, Clay (2005a). Folksonomies + controlled vocabularies. *Many2Many : A group weblog on social software*, January 7, 2005.  
[http://many.corante.com/archives/2005/01/07/folksonomies\\_controlled\\_vocabularies.php](http://many.corante.com/archives/2005/01/07/folksonomies_controlled_vocabularies.php) [2007-05-11]

Shirky, Clay (2005b). Folksonomies are a forced move: A response to Liz. *Many2Many : A group weblog on social software*, January 22, 2005.  
[http://many.corante.com/archives/2005/01/22/folksonomies\\_are\\_a\\_forced\\_move\\_a\\_response\\_to\\_liz.php](http://many.corante.com/archives/2005/01/22/folksonomies_are_a_forced_move_a_response_to_liz.php) [2007-05-05]

Shirky, Clay (2005c). *Ontology is overrated: Categories, links and tags*.  
[http://shirky.com/writings/ontology\\_overrated.html](http://shirky.com/writings/ontology_overrated.html) [2007-04-20]

Spiteri, Louise F. (2006). The Use of Folksonomies in Public Library Catalogues.  
*The Serials Librarian*, vol. 51, no. 2, s. 75-89.

Strout, Ruth French (1956). The development of the catalog and cataloging codes,  
*The Library quarterly*, vol. 26, no. 4, s. 254-275.

Stockholms stadsbibliotek. *Biblioteket.se*, <http://www.biblioteket.se/>

Stove, Alva T. (2000). The LCSH Century: A brief history of the Library of Congress Subject Headings, and introduction to the centennial essays. *Cataloging & Classification Quarterly*, vol. 29, no. 1-2, s. 1-15.

Structured vs. unstructured tagging: A case study (2006). By Judit Bar-Ilan, Snunith Shoham, Asher Idan, Yitzchak Miller & Aviv Shachak. Ingår i *Proceedings of the 15<sup>th</sup> International World Wide Web Conference, Edinburgh Scotland 2006*.

Taylor, Arlene G. (2004). *The organization of information*. 2 ed. Westport, Conn.: Libraries Unlimited.

Tennis, Joseph T. (2006). Social tagging and the next steps for indexing. Ingår i *Proceedings of the 17th SIG Classification Research Workshop, 2006*.  
<http://www.slais.ubc.ca/users/sigcr/sigcr-06tennis.pdf> [2007-05-15]

Trant, Jennifer (2006). Exploring the potential for social tagging and folksonomy in art museums: proof of concept, *New Review of Hypermedia and Multimedia*, vol. 12, no. 1, s. 83-105.  
<http://www.informaworld.com/smpp/content?content=10.1080/13614560600802940> [2007-05-15]

Vander Wal, Thomas (2007). Folksonomy. *Vanderwal.net*, Februari 2, 2007.  
<http://vanderwal.net/folksonomy.html> [2007-03-02]

Weinberger, David (2005). *Tagging and why it matters*.  
<http://cyber.law.harvard.edu/home/2005-07> [2007-05-02]

*Wikipedia*. [www.wikipedia.org](http://www.wikipedia.org) [2007-06-10]

## Bilaga 1 – Tillförandegraderna

### Tillförandegrad 1

Denna variabel mäter antalet taggar som tillför någon av de nackdelar som taggning förväntas kunna tillföra.

- Taggar som är identiska med en subject heading som redan återfinns i posten
- Taggar som är språkvarianter av en subject heading som återfinns i posten, exempelvis stavnings- och böjningsvarianter.
- Taggar som är identiska med en textsträng som återfinns i ett fält som är sökbar genom keywordsökning (titel, författare, tidskriftsnamn)
- Taggar som är språkvarianter av en textsträng som återfinns i ett fält som är sökbar genom keywordsökning (titel, författare, tidskriftsnamn), exempelvis stavnings- och böjningsvarianter.
- Taggar som identifiera upphov för dokumentet
- Taggar som förfinar eller vidareindelar existerande kategorier
- Taggar som identifierar olika kvaliteter eller karaktäristika hos dokumentet eller som gör detsamma för ämnet som behandlas
- Taggar som innehåller en referens till personen som taggar
- Taggar som organiserar en uppgift

Bruket av understreck för att markera mellanrum när man skapar fraser anses så vedertaget på PennTags att detta inte noteras som skillnad vid jämförelsen. Samma sak gäller bruk av versaler och gemener. *Film noir*, *Film Noir* och *film\_noir* betraktas alltså som identiska textsträngar.

### Tillförandegrad 2

Denna variabel mäter antalet taggar som utgör synonymer för ett begrepp som redan är närvarande i indexeringen och därigenom kan sägas skapa nya sökingångar men inte på ett sätt som ligger i linje med de principer som vokabulärkontroll vilar på.

Hit räknas:

- Taggar som är synonymer till en subject heading som återfinns i posten
- Taggar som lyfter ut en enda aspekt av en subject heading, och därigenom är en mer generell term än den som återfinns i posten.

- Taggar som är synonyma med en enda aspekt av en subject heading, och därigenom är en mer generell term än den som återfinns i posten.

### Tillförandegrad 3

Denna variabel mäter antalet taggar som tillför någon av de fördelar som taggning förväntas kunna tillföra.

Hit räknas:

- Taggar som ökar uttömmandegraden i beskrivningen, det vill säga rör en aspekt av dokumentet som inte är närvarande i subject headings dokumentet getts.
- Taggar som ökar specificiteten i beskrivningen, det vill säga rör en aspekt av dokumentet som är närvarande i indexerarens indexering men som är en mer specifik term för att beteckna aspekten än den indexeraren använt.