

## Innehållsförteckning:

<b>1 INLEDNING, BAKGRUND OCH ÄMNESVAL .....</b>	<b>1</b>
<b>2 SYFTE OCH PROBLEMSTÄLLNINGAR .....</b>	<b>3</b>
2.1 AVGRÄNSNINGAR .....	3
<b>3 TERMINOLOGI .....</b>	<b>5</b>
<b>4 METOD .....</b>	<b>8</b>
4.1 TIDIGARE FORSKNING .....	9
<b>5 WEBBEN.....</b>	<b>10</b>
5.1 VAD KARAKTERISERAR WEBBEN?.....	10
5.2 DIGITALA OBJEKT.....	12
5.2.1 <i>Intellektuellt innehåll</i> .....	14
5.2.2 <i>Åtkomst av digitala objekt</i> .....	15
5.2.2.1 <i>Teknisk åtkomst</i> .....	16
5.2.2.2 <i>Intellektuell åtkomst</i> .....	17
5.3 BROWSING OCH SÖKNING .....	19
5.4 SÖKTJÄNSTER.....	21
5.4.1 <i>Indexbaserade söktjänster</i> .....	22
5.4.2 <i>Ämneshierarkiska söktjänster</i> .....	23
5.4.3 <i>Platta listor</i> .....	23
5.5 SAMMANFATTNING.....	23
<b>6 KLASSIFIKATION.....</b>	<b>25</b>
6.1 VAD ÄR KLASSIFIKATION? .....	25
6.2 ÄMNESREPRESENTATION.....	26
6.2.1 <i>Kontrollerat och naturligt indexeringsspråk</i> .....	27
6.3 KLASSIFIKATIONSSYSTEM .....	28
6.3.1 <i>Bibliografisk klassifikation</i> .....	29
6.3.1.1 <i>Enumerativa och syntetiska system</i> .....	31
6.3.2 <i>Olika typer av klassifikationssystem</i> .....	32
6.4 ETABLERADE KLASSIFIKATIONSSYSTEM .....	33
6.4.1 <i>DDK</i> .....	33
6.4.1.1 <i>För- och nackdelar med DDK</i> .....	36
6.4.1.2 <i>DDK på webben - BUBL LINK</i> .....	38
6.4.2 <i>UDK</i> .....	41
6.4.2.1 <i>För- och nackdelar med UDK</i> .....	43
6.4.2.2 <i>UDK på webben - NISS Directory of Networked Resources</i> .....	43
6.5 ICKE-ETABLERADE KLASSIFIKATIONSSYSTEM.....	47
6.5.1 <i>Punkt se</i> .....	47
6.5.2 <i>Yahoo! Sverige</i> .....	51
6.6 SAMMANFATTNING.....	55
<b>7 DISKUSSION.....</b>	<b>57</b>
7.1 ÄR WEBBEN ETT INFORMATIONSSYSTEM? .....	57
7.2 INFORMATIONSSYSTEM PÅ WEBBEN .....	59
7.2.1 <i>Skillnader mellan webbaserade och traditionella informationssystem</i> .....	59
7.2.1.1 <i>Digitala objektens konstitution</i> .....	60
7.2.1.2 <i>Standardiserad metadata</i> .....	61
7.2.1.3 <i>Intellektuellt värde</i> .....	63
7.2.2 <i>Ämneshierarkiska söktjänster</i> .....	63
7.3 KLASSIFIKATIONSSYSTEM PÅ WEBBEN .....	66
7.3.2 <i>Etablerade och icke-etablerade klassifikationssystem</i> .....	67

7.3.2.1 Anpassning till användaren .....	67
7.3.2.2 Notation .....	70
7.3.2.3 Flexibilitet .....	71
7.3.2.3 Subjektiv klassificering .....	72
7.4 AVSLUTNING.....	72
7.4.1 Fortsatt forskning.....	74
<b>8 SAMMANFATTNING .....</b>	<b>75</b>
<b>9 KÄLL- OCH LITTERATURFÖRTECKNING.....</b>	<b>77</b>
9.1 TRYCKTA KÄLLOR OCH LITTERATUR.....	77
9.2 ELEKTRONISKA KÄLLOR.....	80

## **Bilagor**

BILAGA 1	FIGUR 1 OCH FIGUR 2
BILAGA 2.1	BUBL LINK, HUVUDKLASSER.
BILAGA 2.2	BUBL LINK, EXEMPEL PÅ POSTER.
BILAGA 2.3	BUBL LINK, EXEMPEL FRÅN ÄMNESHIERARKIN.
BILAGA 3.1	NISS DIRECTORY OF NETWORKED RESOURCES, HUVUDKLASSER.
BILAGA 3.2	NISS DIRECTORY OF NETWORKED RESOURCES, EXEMPEL PÅ POST.
BILAGA 3.3	NISS DIRECTORY OF NETWORKED RESOURCES, EXEMPEL FRÅN ÄMNESHIERARKIN.
BILAGA 3.4	NISS DIRECTORY OF NETWORKED RESOURCES, EXEMPEL FRÅN ÄMNESHIERARKIN.
BILAGA 4.1	PUNKT SE, HUVUDKLASSER I KATEGORIN BRANSCH.
BILAGA 4.2	PUNKT SE, EXEMPEL PÅ POST.
BILAGA 5.1	YAHOO! SVERIGE, HUVUDKLASSER.
BILAGA 5.2	YAHOO! SVERIGE, EXEMPEL FRÅN ÄMNESHIERARKIN OCH POSTER.

# 1 Inledning, bakgrund och ämnesval

Denna uppsats behandlar problem och möjligheter kring organisering av länkar till digitala objekt i ämnehierarkiska söktjänster på webben. Uppsatsen är skriven inom utbildningen på institutionen Bibliotekshögskolan vid Högskolan i Borås.

Vi har intresserat oss för Internet i en trepoängsuppsats tidigare i vår utbildning. Då undersökte vi hur ett antal bibliotekarier vid ett högskolebibliotek använde Internet för att täcka sina yrkesmässiga informationsbehov. Vid denna undersökning upptäckte vi att det var en vanlig åsikt hos professionella informationssökare att kunskapsåtervinning på Internet är onödigt tidskrävande på grund av att informationen inte är organiserad i enlighet med etablerade bibliografiska system. Att det inte alltid är det lättaste att hitta vad man söker på Internet har vi själva också erfarit. Våra personliga erfarenheter av Internet sträcker sig flera år tillbaka och vi använder dagligen nätet och dess applikationer. Utan att ha någon övertro på Internet som världsförbättrare har vi ändå en positiv grundinställning och anser att det är en mycket användbar teknik för informationsöverföring. Det finns dock utrymme för förbättringar på flera områden.

Internet och webben håller på att utvecklas till att innehålla stora mängder seriös information. Informationsåtervinningsaspekten blir därför allt viktigare. Egentligen behövs det ingen djupare förklaring till varför informationåtervinning på webben är ett problem idag. Alla som någon gång har försökt finna ett specifikt objekt eller information om ett visst ämne på webben vet att det kan vara mycket svårt. Birger Hjørland har uppmärksammat att utvecklingen av webben ställer nya krav på system för kunskapsorganisation:

*Traditional areas of IS [information science] include classification, indexing, knowledge representation, information seeking, search logic, and relevance judgements. All these traditional problems must of course be reconsidered in the new reality consisting of full-text retrieval, the World Wide Web, and so on.<sup>1</sup>*

Det anses alltså nödvändigt att utveckla verktyg och metoder för organisering och återvinning av information på webben. Det finns flera sätt att åstadkomma detta och att strukturera och klassificera kan eventuellt vara en lösning. Vi hoppas att vi med denna undersökning kan öka förståelsen för problemen med kunskapsorganisation och informationsåtervinning på webben.

Bibliotekens, bibliotekariernas och andra intermediärers roll har redan ändrats i och med Internets framväxt. Det finns även en risk att de traditionella institutionerna, som till exempel bibliotek, kommer att få minskad betydelse i framtiden. Samtidigt erbjuder

---

<sup>1</sup> Hjørland, Birger, (1997), Information seeking and subject representation: an activity-theoretical approach to information science. - Westport, Conn.: Greenwood Press. - s. 22.

svårigheterna med informationsåtervinning på webben nya möjligheter för bibliotekarier och informationsspecialister att visa på sin specialiserade yrkeskunskap.

## 2 Syfte och problemställningar

Syftet med denna uppsats är att genom kvalitativa textstudier beskriva och analysera hur digitala objekt som är tillgängliga via webben kan organiseras med hjälp av olika klassifikationssystem.<sup>2</sup> Vidare syftar vi till att visa på de problem som finns med kunskapsorganisation på webben. Utifrån detta söker vi svara på följande frågor:

- Vilka problem och möjligheter finns med klassificering av digitala objekt på webben?
- Vilka grundläggande karaktäristika har klassifikationssystem som används för ämneshierarkiska söktjänster i webbmiljö?
- Vilka skillnader finns mellan etablerade och icke-etablerade klassifikationssystem i en webbkontext?
- Vilka för- och nackdelar har de olika systemen?

Vi syftar även till att ge en lägesrapport om de aktuella förutsättningarna för kunskapsåtervinning på webben. Tonvikten ligger på ämneshierarkiska söktjänster och deras användning av klassifikationssystem.

### 2.1 Avgränsningar

Eftersom webben är en såpass ny företeelse analyserar vi endast delar av de senaste fyra årens diskussion. Studien omfattar texter publicerade såväl på webben som i tryckt form. Texterna som är publicerade på webben är alla från åren 1994 till 1998 men många av de texter som är i tryckt form är utgivna tidigare än 1994.

En av anledningarna till att vi har valt att göra en litteraturstudie är att vi anser att det redan finns många empiriska studier av sökverktyg på webben som fokuserat på återvinningseffektivitet och att det vore mindre angeläget att göra ytterligare en sådan undersökning. Ytterligare en anledning som motiverar en litteraturstudie är de begränsningar som uppsatsen ställer. Ett bra komplement till textstudier skulle vara att göra intervjuer med till exempel utvecklare och användare av söktjänster, vilket vi dock har valt bort eftersom det skulle bli allt för omfattande för att rymmas inom de uppsatta ramarna.

Vi avser endast behandla klassifikationssystem som syftar till att omfatta alla ämnesområden. Alla ämnesspecifika och nationella klassifikationssystem är därför uteslutna från närmare analys. De system som vi har valt att granska närmare är DDK och UDK. Anledningen till att vi har valt de universella klassifikationssystemen DDK och UDK är att de är väl etablerade och inarbetade system som används flitigt på bibliotek över hela världen och som även förekommer på webben. Vi har begränsat oss till att endast behandla organisering av digitala objekt på webben och inte på Internet i ett bredare perspektiv.

---

<sup>2</sup> Termen digitalt objekt definieras i stycke 5.2 Digitala objekt.

Ursprungligen var syftet med den empiriska avsnittet att jämföra svenska, danska och norska webbtjänster som använder sig av olika slags bibliografiska klassifikationssystem. Efter noggranna undersökningar kan vi dock konstatera att det i nuläget inte finns några webbtjänster inom de tre länderna som använder sig av *Universell Decimalklassifikation* (UDK) eller *Dewey Decimalklassifikation* (DDK), bortsett från bibliotekskataloger. Då de tre nordiska länderna inte erbjuder några söktjänster som använder dessa klassifikationssystem vidgade vi avgränsningen till det engelskspråkiga söktjänster i Europa. De söktjänster som vi valt att granska närmare är *NISS Directory of networked resources* och *BUBL LINK* som är ordnade enligt UDK respektive DDK. Dessutom granskas två svenska söktjänster som använder sig av icke-etablerade klassifikationssystem. Dessa är Punkt Se och Yahoo! Sverige.

Måhända förefaller det underligt att göra en geografisk avgränsning vid en undersökning av webben. Det bör påpekas att den geografiska lokaliseringen av söktjänster oftast är av underordnad betydelse. Till exempel inriktar sig svenska tjänster ofta på svenska användare och klassificerar således mest svenskt material.

### 3 Terminologi

Syftet med detta avsnitt är att visa på de svårigheter kring språkbruket som råder inom området. Vi definierar dessutom några av våra centrala termer.

Vid beskrivningar av webben används ofta maritima metaforer. Det talas om att *surfa* eller *navigera* på nätet och det finns rentav en del som *simmar*. Samtidigt varnas det för risken att *drunkna* i det överväldigande informationsflödet.<sup>3</sup> Dessa nautiska influenser kan vara funktionella, i synnerhet i det vardagliga språket men precis som när det gäller andra termer krävs en klar definition av vad som avses för att missförstånd inte ska uppstå.

Inom området Internet och webben förekommer en mängd termer som inte är klart definierade och därför ofta är mångtydiga. Användningen av dessa termer, till exempel förkortningen "IT", sker ofta oreflekterat och begreppsinnebörden blir därför vag. Flera termer är dessutom komplicerade för oss på grund av sin tekniska karaktär. Något som ofta uppmärksammas är det flitiga bruket av anglicismer.

Vidare måste nya upplevelser som människor har relateras till upplevelser som de redan har haft för att bli begripliga eller lättare att förstå.<sup>4</sup> När termer okritiskt lånas från andra områden och språk för att appliceras på nya företeelser förändras dock det semantiska innehållet. Samma term kan alltså användas i olika sammanhang och med olika innebörd. Det tar tid att etablera en gemensam vokabulär inom en disciplin, om det överhuvudtaget är möjligt med tanke på att inget ämne står stilla utan utvecklas hela tiden.

I vetenskapliga sammanhang ska termerna tjäna som otvetydiga arbetsredskap. Stringenta och tydliga definitioner är väsentligt för att undvika missförstånd om vad som avses. Vi är medvetna om att många av de termer vi använder kan komma att vara utbytt om några år och att vi kan ge termer annat innehåll än vad de tidigare har haft.

Svenska datatermgruppen definierar Internet kort och gott som "det internationella datornät som har den största utbredningen och som bygger på TCP/IP, en standard för datakommunikation".<sup>5</sup> När det gäller den standard på Internet som denna uppsats avser belysa rekommenderar Svenska datatermgruppen termen *webben* i stället för det engelska World Wide Web eller den vanliga förkortningen WWW.<sup>6</sup> Eventuellt skulle termen den världsvida väven, som en direkt översättning från det engelska World Wide Web, kunna användas. Trots att det engelska inflytandet är påtagligt i termen *webben* är det en idag

---

<sup>3</sup> Crawford, Walt och Gorman, Michael, (1995), *Future libraries: dreams, madness & reality*. - Chicago: American Library Association. - s. 80-84.

<sup>4</sup> Palmquist, Ruth A, (1996), *A qualitative study of Internet metaphors // Proceedings / National Online Meeting: New York, May 14-16, 1996 / edited by Martha E. Williams*. - Medford N.J.: Learned Information. - s. 293-298.

<sup>5</sup> Svenska datatermgruppen [URL: <http://www.nada.kth.se/dataterm/>] 98-04-27. TCP/IP betyder *Transmission Control Protocol/Internet Protocol* och är ett samlingsnamn för två protokoll som används för att överföra data via Internet.

<sup>6</sup> Ibid.

etablerad term som används och som även fortsättningsvis kommer att användas i denna uppsats.

Webben är *en* del av det globala nätverk som går under namnet Internet. Övriga funktioner är bland annat elektronisk post och diskussionsgrupper som dock inte ingår i vår definition av webben. Standarden för dataöverföring via webben är Hypertext Transfer Protocol (HTTP).<sup>7</sup> Eftersom tekniska standarder byts ut när nya bättre utvecklas finns det även på Internet standarder som är föregångare till HTTP, till exempel GOPHER och FTP.<sup>8</sup> Dessa standarder är nu inkluderade i de mest använda webbläsarprogrammen och ingår därför även i vår definition av webben. Alla slags slutna nätverk, till exempel ett intranät vid ett företag som bygger på samma teknik som webben men som inte finns globalt tillgängligt inkluderas inte i definitionen.

För att återvinna objekt kan man använda sig av olika sökverktyg. Ett sökverktyg är enligt vår definition alla typer av verktyg som erbjuder någon slags hjälp för att finna objekt på webben. En söktjänst är en kvalificerad form av sökverktyg och kan antingen vara av indexbaserad eller ämneshierarkisk typ. Detta är en bred definition och kvalitet och omfång skiftar mellan de olika söktjänsterna.

Klassifikationssystemen delar vi in i de två generella grupperna *etablerade* och *icke-etablerade* system. Denna indelning är inte helt problemfri eftersom gränsen mellan vad som är etablerat och vad som inte är det är svår att dra, något som bland annat forskaren Jeannette Woodward uppmärksammat i artikeln *Cataloging and Classifying Information Resources on the Internet*.<sup>9</sup> I denna uppsats använder vi termen etablerat system som synonymt med de system som tillkommit innan webben utvecklades. Exempel på dessa är de bibliografiska klassifikationssystemen UDK och DDK. Ett bibliografiskt klassifikationssystem är ett system som är utvecklat för att klassificera böcker och andra fysiska objekt i ett bibliotek. Ett icke-etablerat system däremot har utvecklats för att användas vid klassificering av digitala objekt i en webbkontext. Exempel på dessa är klassifikationssystemen som används av söktjänsterna Yahoo! och Punkt se.

Användning av vissa traditionella bibliotekstermer fungerar inte speciellt bra vid undersökningar av webben. Att okritiskt överföra redan etablerade termer från ett område till ett annat medför risk för språkförbistring. Oftast förknippas termer hämtade från biblioteksområdet med fysiska objekt vilket vi inte har för avsikt att studera. I flera fall har därför redan etablerade bibliotekstermer valts bort då deras betydelse är obestämd i denna nya kontext. Att förvirring råder är uppenbart. Det finns inget konsekvent bruk av termer mellan olika författare inom området. Då ingen författare har någon komplett terminologi som innefattar alla väsentliga delområden använder vi termer från olika författare och olika

---

<sup>7</sup> Ett protokoll är i detta sammanhang en teknisk standard för överföring av data mellan olika datorer. Ett exempel är HTTP.

<sup>8</sup> FTP är en förkortning av File Transfer Protocol, en standard som förefick GOPHER. FTP är en standard för filöverföring. GOPHER är en standard som hanterar textbaserad hypermedia.

<sup>9</sup> Woodward, Jeannette, (1996), *Cataloging and Classifying Information Resources on the Internet // Annual Review of Information Science and Technology (ARIST)*. - Vol. 31. - s. 189-220.

sammanhang. Vi har valt att definiera de övriga termer allteftersom de används i uppsatsen. Då det anses nödvändigt skrivs även den ursprungliga termen ut på det språk den översatts från.

## 4 Metod

Vår undersökning är en litteraturstudie som består av en kvalitativ analys av texter som är intressanta för ämnesområdena klassificering och klassifikationssystem för underlättandet av informationsåtervinning på webben. Utifrån textanalys redogörs för hur digitala objekt tillgängliga via webben kan organiseras med hjälp av olika klassifikationssystem. Vi har både använt oss av texter hämtade från webben och texter publicerade i tryckta böcker och tidskrifter. Urvalet av texter har baserats på vår uppfattning om vad som kan anses relevant. Facklitteratur inom det biblioteks- och informationsvetenskapliga området är en viktig utgångspunkt i uppsatsen, främst för det deskriptiva kapitlet om klassifikationssystem. Ett problem med flera av de äldre teoretiska verken är att de utgår från pappersbaserade eller andra fysiska objekt. Grundteorierna om klassifikationssystem är dock i de allra flesta fall applicerbara på organisation av digitala objekt och ämnesrepresentationer av innehållet i digitala objekt på webben.

Vi belyser vår problemställning främst utifrån textanalys men gör även kopplingar till nu existerande söktjänster. Genom att exemplifiera med detta empiriska material syftar vi till att belysa de problem som texterna behandlar samt att ge en bild av hur klassifikationssystem används på webben idag.

De utvalda söktjänsterna ger ett förtroendeingivande intryck avseende omfattning, uppdateringsfrekvens och huvudman. Uppfattningen av vad som är kvalitet är ett utslag av våra subjektiva förstahandsintryck. Detta är ett metodologiskt problem som vi är medvetna om. Anledningen till att vi använder dessa exempel är att vi i vår analys utgår ifrån hur webben och dess söktjänster ser ut idag och inte hur de kan komma att se ut i framtiden. Även när det gäller webbens tekniska och användarrelaterade förutsättningar utgår vi ifrån dagsläget.

Urvalet av söktjänster har även skett utifrån vilket klassifikationssystem de använder. Först och främst görs en avgränsning till svenskspråkiga tjänster. Yahoo! Sverige är vald för att dess internationella motsvarighet anses vara en av de främsta söktjänsterna på webben. För att anlägga ett svenskt perspektiv väljer vi att titta på den svenska Yahoo!. Punkt Se är intressant då det är en svensk tjänst som dessutom har en alternativ form av klassifikation. BUBL LINK och NISS är båda brittiska söktjänster som använder de etablerade klassifikationssystemen DDK och UDK. Dessa är utvalda dels på grund av deras geografiska placering, i detta fall inom Europa och dels på grund av att de använder etablerade klassifikationssystem.

Redovisning av resultatet sker i form av en diskussion kring ett antal problemområden. Uppsatsen är uppdelad i tre huvudsakliga kapitel. I kapitlet 5 ger vi en bakgrund till webben, beskriver digitala objekt, browsing och sökning samt olika typer av söktjänster. I kapitel 6 för vi ett teoretiskt resonemang kring klassifikation samt beskriver två typer av system. Dessutom exemplifierar vi med fyra webbaserade söktjänster. I det sista kapitlet problematiserar och diskuterar vi utifrån syftet och de tidigare kapitlen.

## 4.1 Tidigare forskning

Vi upplever att forskningen kring webben har fördjupats de senaste åren. Likaså har förhållningssättet till webben förändrats. Det är alltså inte bara den tekniska utvecklingen som går fort. På grund av ämnets aktualitet finns förhållandevis lite tidigare forskning genomförd. Traugott Koch, verksam vid Netlab i Lund, är en av de mest namnkunniga forskarna. Andra forskare inom området är Bella Hass Weinberg, Steve Steinberg, Angela Oehler, Carol A. Mandel, och Robert Wolven, Robert Kahn och Robert Wilensky, Willaim Y. Arms.

Ett stort intresseområde som sammanför bibliotek och Internet är det som kallas för digitala bibliotek. Det finns ingen klar definition av termen *digitalt bibliotek* men enligt Nürnberg kan forskningen inom detta område bland annat syfta till att beskriva ut hur de befintliga fysiska biblioteken kan anpassa och utöka sina tjänster med hjälp av bland annat Internet. Biblioteket står i fokus och man undersöker hur dess system, standarder och metoder kan användas i den digitala tidseran.<sup>10</sup> Vi försöker i denna uppsats att ställa oss litet vid sidan av denna diskussion och i stället ha ett bredare perspektiv på problemområdet. För även om denna uppsats skrivs inom ämnet biblioteks- och informationsvetenskap så fokuserar vi inte enbart på de fysiska bibliotekens tjänster utan även andra typer av tjänster framträder, däribland flera kommersiella. Gemensamt är att de alla försöker tillhandahålla informationstjänster, men med olika typer av organisationsprinciper.

---

<sup>10</sup> Nürnberg et al, Digital libraries: issues and architectures.  
[URL: <http://csdl.tamu.edu/csdl/DL95/papers/nuernberg/nuernberg.html>] 98-05-14.

## 5 Webben

Detta kapitel syftar till att beskriva webben och dess innehåll för att visa vilka förutsättningar som finns för klassificering och återvinning av digitala objekt. När det gäller beskrivning av Internet, dess bakgrund, beståndsdelar, historia och hur nätverket fungerar rent praktiskt hänvisar vi till någon av de många guider som finns inom området.<sup>11</sup> Vi diskuterar även termerna *browsing* och *sökning* utifrån webbaserade söktjänster. Vi definierar även relevanta termer för att lägga en grund för senare diskussion. I ett separat stycke definieras även olika typer av söktjänster.

Det kan ses som symptomatiskt att det mest är problem och inte möjligheter som diskuteras i de texter vi tagit del av. Webbens fördelar och möjligheter är ändå ganska uppenbara. Webben är ett globalt accepterat distributionssystem där åtkomst kan ske från alla anslutna persondatorer. Både publicering och återvinning kan ske utan några mellanhänder eller större tekniska hinder. Problemen är dock fortfarande stora och det är intressantare att beskriva dessa och lösa dem än att enbart visa på de potentiella möjligheter som finns. Många texter förefaller peka på problematiken som finns med webben men kommer sällan med något konstruktivt förslag om förbättring.

### 5.1 Vad karakteriserar webben?

Webben expanderar i mycket hög takt både när det gäller antalet användare, antalet värddatorer och antalet tillgängliga digitala objekt. Det faktum att mängden digitala objekt ökar i en så pass snabb takt är ett av de största problemen för den som på något sätt vill organisera dessa för att underlätta återvinning. Det är av många skäl i princip omöjligt att exakt bestämma webbens storlek och hur många objekt den hyser men det finns undersökningar som uppskattar storleken. Resultatet av dessa undersökningar skall alltså uppfattas som uppskattningar som antyder något om hur snabbt expansionen sker.

För att visa på tillväxten av webben exemplifierar vi med en undersökning, gjord av det amerikanska företaget Network Wizards, som visar antalet värddatorer (eng: hosts) uppkopplade till Internet.<sup>12</sup> Att mäta antalet värddatorer är *ett* sätt att undersöka hur snabb Internets tillväxt är. Det bör påpekas att en värddator dock inte behöver användas enbart som webbserver utan kan även användas för att utföra andra Internettjänster. Internet i detta sammanhang innefattar både Usenet News, elektronisk post, filöverföring med FTP samt GOPHER- och WWW-tjänster. En diagram över resultatet av undersökningen finns i bilaga 1, figur 1. Som framgår av undersökningen ökar antalet värddatorer snabbt i antal. Hur väl dessa siffror överensstämmer med det verkliga antalet uppkopplingar är som sagt svårt att bedöma. Network Wizards kommenterar själva resultatet så här:

---

<sup>11</sup> Boutell, Thomas, World Wide Web Frequently Asked Questions (With Answers, of Course!) [URL: <http://www.boutell.com/faq/oldfaq/index.html>] 98-03-05. ; Borg, Tage, (1994), Internet@Sverige. - Stockholm: Bonnier DataMedia.

<sup>12</sup> Network Wizards, Internet Domain Survey [URL: <http://www.nw.com/>] 98-04-27.

*We consider the numbers presented in the domain survey to be fairly good estimates of the minimum size of the Internet. [...] In summary, it is not possible to determine the exact size of the Internet, where hosts are located, or how many users there are.*<sup>13</sup>

Ett annat sätt att mäta expansionen är att mäta antalet registrerade domännamn. En värddator kan hysa många olika domäner med ett varierande antal objekt. Här redovisas hur många domännamn som registrerats i Sverige de senaste 11 åren och en bild över antalet registrerade domännamn i Sverige under perioden 1985 till 1996 finns i bilaga 1, figur 2.<sup>14</sup> Statistik över användningen av den svenska delen av Internet finns i rapporten *Svenska delen av Internet* utförd av statskontoret, Swedish Network Users Society (SNUS) på uppdrag av kommunikationsdepartementet.<sup>15</sup> Den visar bland annat att "andelen svenskar som har tillgång till Internet via arbetet, skolan, hemmet eller vänner och bekanta är 46 %, vilket motsvarar 2,9 miljoner personer. I oktober 1996 var motsvarande siffra 30 %. Detta innebär att på ett halvår har tillgången ökat med cirka 1 miljon personer."<sup>16</sup> Tillgången och användningen av Internet i Sverige har alltså ökat kraftigt det senaste året. Dessa tre undersökningar visar på Internets expansion vilket inte behöver vara samma som webben. Detta är ett problem som vi är medvetna om.

Bristen på standarder för både återvinning och beskrivning av innehållet i digitala objekt på webben är ytterligare ett problem. Webben saknar en central kontrollinstans i form av en organisation, ett företag, en myndighet eller någon annan slags grupp som sätter upp villkor för hur nätverkets objekt ska återvinnas utifrån innehåll. Det finns självklart organisationer och företag som påverkar hur webben utvecklas men ingen av dem har makt nog att själva bestämma exakt hur webben och dess objekt ska se ut eller organiseras för återvinning. De standarder som finns omfattar främst tekniska aspekter på åtkomst samt formatering av innehåll. Vi kommer att beskriva några av dessa standarder längre fram. De som har viss makt över utvecklingen är dels tillverkarna av de just nu mest använda webbläsarprogrammen, till exempel Netscape och Microsoft, och dels W3-konsortiet som definierar och utvecklar HTML-standarderna. Internet Engineering Task Force (IETF) är ytterligare en organisation som arbetar med att utveckla den tekniska standarderna.<sup>17</sup>

Sammanlänkningsen av objekt på webben gör webben till en hypertext. Hypertext kännetecknas av en icke-linjär uppbyggnad, det vill säga det finns ingen fastställd ordning hur en text skall läsas. I en bok, vars struktur oftast är linjär, är följden given och läsaren tar del av innehållet från början till slut, sida efter sida och kapitel efter kapitel. En användare

---

<sup>13</sup> Network Wizards, Domain Survey Notes [URL: <http://www.nw.com/zone/WWW/notes.html>] 98-04-27.

<sup>14</sup> Användningen av Internet i Sverige // Svenska delen av Internet [URL: <http://www.snus.se/internetutred/rapport/kap5.html>] 98-04-27.

<sup>15</sup> Svenska delen av Internet, [URL: <http://www.snus.se/internetutred/>] 98-05-01.

<sup>16</sup> Svenska delen av Internet: bilaga 10: användning av Internet i Sverige [URL: <http://www.snus.se/internetutred/bilaga/bil10.html>] 98-05-01.

<sup>17</sup> Netscape Inc. [URL: <http://www.netscape.com/>] 98-03-03. ; Microsoft [URL: <http://www.microsoft.com/>] 98-03-03. ; World Wide Web Consortium [URL: <http://www.w3.org/>] 98-03-03. ; Internet Engineering Task Force [URL: <http://www.ietf.org/>] 98-03-03.

av digitala objekt i hypertextformat kan dock förflytta sig fram och tillbaka inom objektet och vidare till andra objekt med hjälp av hyperlänkar. Dessa fungerar binärt och en länk förbinder två så kallade noder. Den ena noden utgör ett ankare, det vill säga det digitala objekt som länken leder ifrån, och den andra noden utgör destinationen.<sup>18</sup> Hypertext lämpar sig synnerligen väl just för browsing. ”With the development and applications of hypertext technology, browsing capabilities have been recognized as one of the central features of hypertext systems”.<sup>19</sup>

Sammanfattningsvis kännetecknas webben bland annat av en kraftig expansion gällande både antalet digitala objekt och antalet domäner. Det saknas även centrala kontrollinstanser och standardiserad beskrivning av innehåll och återvinning av detsamma. Webben bygger på hypertext som är icke-linjär till sin natur. Detta stödjer browsing på ett bra sätt.

## 5.2 Digitala objekt

Detta stycke handlar om digitala objekt på webben och vilka förutsättningar för åtkomst av dessa som finns. Vi skiljer i det här sammanhanget på teknisk- och intellektuell åtkomst. Intellektuell åtkomst, det vill säga återvinning av digitala objekt i huvudsak utifrån deras innehåll, är vad söktjänster försöker erbjuda. Man kan bland annat använda sig av klassifikationssystem för att uppnå detta. Anledningen till att vi tar upp den tekniska sidan av åtkomst är att den utgör en förutsättning för intellektuell åtkomst. För att förstå problematiken kring intellektuell åtkomst måste man även förstå grunderna för den tekniska åtkomsten.

Ett av grundbegreppen i sammanhanget är *digitalt objekt*. Ett digitalt objekt är bärare av intellektuellt innehåll och för att kunna ta del av det intellektuella innehållet måste man skapa teknisk åtkomst till det digitala objektet. Innan vi ger oss in på att försöka beskriva vad som karakteriserar digitala objekt på webben måste vi definiera termen *digitalt objekt*. De amerikanska forskarna Robert Kahn och Robert Wilensky, verksamma vid Corporation for National Research Initiatives (CNRI) samt Berkeley universitetet, försöker gemensamt reda ut några trådar i begreppshärvan kring digitala objekt.<sup>20</sup> Deras artikel *A Framework for Distributed Digital Object Services* syftar till att skapa en modell för en alternativ arkitektur av ett nätverksbaserat informationssystem. Eller som författarna själva väljer att benämna det, ”a universal, open, wide-area digital information infrastructure system”, kortfattat *systemet*.<sup>21</sup> Kahn och Wilenskys modell syftar till att skapa ett teoretiskt ramverk genom att definiera de grundläggande komponenter som ingår i nätverket. Det finns många exempel

---

<sup>18</sup> Nielsen, Jacob, (1995), *Multimedia and hypertext: the Internet and beyond*. - Boston: AP Professional. - s. 1-7.

<sup>19</sup> Chang, Shan-Ju och Rice, Roland E, (1993), *Browsing: a multidimensional framework* // *Annual review of information science and technology (ARIST)*, edited by Martha E. Williams. - Vol. 28. - s. 240.

<sup>20</sup> Kahn, Robert och Wilensky, Robert, (1995), *A framework for distributed digital object services* [URL: <http://www.cnri.reston.va.us/home/cstr/arch/k-w.html>] 98-04-28. ; För liknande modeller se även Van der Werf-Davelaar, Titia, (1995), *Organizing file servers on the Internet: role of the library* // *Publications of Essen University Library*. - Vol. 18. - s. 194-206.

<sup>21</sup> Kahn, Robert och Wilensky, Robert, (1995).

på tillämpningar av deras teoribygge då deras definitioner är medvetet generella. Webben är en möjlig infrastruktur men teorierna kan även appliceras på andra typer av nätverk. Det bör understrykas att Kahn och Wilensky lägger tonvikten på de tekniska aspekterna hos nätverket och befattar sig inte med de innehållsmässiga.

Enligt Kahn och Wilenskys definition är ett digitalt objekt ett av en upphovsman sammanställt objekt, vars beståndsdelar är digitalt material, det vill säga data. Digitala objekt i detta sammanhang omfattar dock inte alla typer av objekt i digital form utan avser endast objekt med en specifik datastruktur anpassad för nätverkshantering. Kanske vore det lämpligare att använda termen *digitala infrastrukturella objekt* för att undvika missförstånd, vilket Kahn och Wilensky själva föreslår. Termen förefaller dock en aning otymplig att hantera varför vi förkortar det till digitalt objekt.

Ett digitalt objekt kan vara antingen enkelt (eng:elemental) eller sammansatt (eng:composite). Ett sammansatt digitalt objekt består av olika typer av digitalt material i olika dataformat, till exempel en webbsida som kombinerar text, bild och ljud. Exempel på enkla digitala objekt är webbsidor i rent ASCII-format, FTP-filer eller GOPHER-sidor.<sup>22</sup> Författarna urskiljer två typer av data som utgör ett digitalt objekt. Dels *ordinär data*, dels *metadata* (eng: key-metadata). Data utgör objektets innehåll, medan metadata omfattar information *om* det digitala objektet.

Ett digitalt objekt är alltså bärare av intellektuellt innehåll. Arms, Blanchi och Overly, vid CNRI, som utgår från Kahn och Wilenskys modell, visar hur en webbsida är uppbyggd som digitalt objekt. Deras exempel utgår från att ett HTML-baserat objekt utgörs av text med tre tillhörande bilder. Texten utgör en enhet av data och bilderna utgör ytterligare tre enheter data som tillsammans utgör ett digitalt objekt.<sup>23</sup> Rent tekniskt består de flesta digitala objekt på webben av text i ASCII-format, formaterade med sidbeskrivningsspråket (eng: markup-language) HTML. Inbäddat i HTML-objektet finns ofta länkar till grafikfiler som visas på datorskärmen tillsammans med texten. Förutom textobjekt och grafiska objekt finns bland annat objekt med ljudinnehåll och objekt som innehåller datorkod, till exempel programvara.

En av grundtankarna i Kahn och Wilenskys modell är att globala unika namn, vad vi kan kalla handtag (eng: handle), skall tillskrivas alla digitala objekt. Dessa namn skall ingå i de digitala objektens metadata. Idén går ut på att ett digitalt objekt skall kunna återvinnas om dess handtag är känt. I en webbkontext kan ett handtag motsvaras av till exempel en

---

<sup>22</sup> ASCII-format står för *American Standard for Information Interchange* och är en standard för teckenuppsättning med, oftast, 256 tecken. En text skriven i ASCII-format innehåller ingen formatering såsom fet- och kursiv stil, utan bara tecken, mellanslag och tabbar.

<sup>23</sup> Arms, William Y., Blanchi, Christophe och Overly, Edward A., (1997), An architecture for information in digital libraries // D-Lib Magazine. - February. - [URL: <http://www.dlib.org/dlib/february97/cnri/02arms1.html>] 98-03-05. ; se även Arms, William Y., (1995), Key concepts in the architecture of the digital library // D-Lib Magazine. - [URL: <http://www.dlib.org/dlib/July95/07arms.html>] 98-04-28.

URN.<sup>24</sup> Möjligheten att kunna återvinna ett digitalt objekt enbart utifrån att dess globalt unika namn är känt är dock en sanning med modifikation. Handtaget innehåller nämligen inte någon nätverksadress. För att återvinna ett digitalt objekt måste handtaget först knytas till en nätverksadress. Detta sker med hjälp av en handtagsserver (eng: handleserver). Servern kontrollerar att alla handtag är unika samt håller reda på lagringsplatserna (eng: repository) för de digitala objekten. Vid sökning på ett handtag sänds förfrågan till handtagsservern som svarar med att returnera nätverksadresser där det digitala objektet med motsvarande handtag är lokaliserat. Flyttas ett objekt från en lagringsplats uppdateras handtagsservern med den nya nätverksadressen. Genom detta tillvägagångssätt kommer man förbi problemet med att inte kunna hitta filer för att de bytt lagringsplats.

Forskningen kring handtag befinner sig dock än så länge i försöksstadiet. Denna uppsats tar dessutom avstamp i de förhållanden som råder idag, inte hur ett framtida nätverkssystem kan konstrueras, varför handtag lämnas därhän i fortsättningen. Vi använder *digitalt objekt* enligt Kahn och Wilenskys definition med reservation för deras idé om handtag. Sammanfattningsvis beror en stor del av problemen kring åtkomst av digitala objekt dels på nätverkets tekniska arkitektur och dels på de digitala objektens karakteristika.

### 5.2.1 Intellectuellt innehåll

Detta avsnitt syftar till att visa på olika typer av intellektuellt innehåll i digitala objekt på webben. Arms påpekar att det trots allt är det intellektuella innehållet som användare efterfrågar, inte de digitala objekten som enbart är bärare av detta innehåll.<sup>25</sup>

Det är relativt svårt att finna statistik över innehållet i digitala objekt på webben. Frågor som var informationen kommer ifrån, vilka källor som använts, trovärdigheten och vilket språk texterna är skrivna på är viktiga men, om inte okänt, så i alla fall inte speciellt väldokumenterat i nuläget. De statistiska undersökningar som gjorts kostar dessutom ofta pengar att ta del av. De tekniska förutsättningarna är mer kända vilket säkert beror på att teknik alltid är lättare att hantera än mänskligt skapat innehåll. Några av anledningarna till att innehållet är relativt okänt är att webben är ett nytt fenomen som uppkommit utan några riktlinjer för vad som ska publiceras.

Webben består av flera olika typer av digitala objekt. Objekten har alla digital form men deras intellektuella innehåll är mycket skiftande. Webben är i princip världsomspännande och många varierande mänskliga uttryck finns representerade. En populär klyscha är att *allting* finns på webben. Så är det självklart inte men däremot finns en mångfald av information. Informationen har dessutom många olika syften och förekommer i många olika sammanhang. På dagens webb finns till exempel:

- Akademiska texter, rapporter och uppsatser.
- Nyhetstext, notiser, intervjuer.

---

<sup>24</sup> För förklaring av URN se stycke 5.2.2.1 Teknisk åtkomst.

<sup>25</sup> Arms, William Y., (1995).

- Krönikor, kåserier, romaner, noveller och annan fiktion.
- Reklam i både text och bildformat.
- Bilder, fotografier och grafik med ett stort antal motiv.
- Datorprogram.
- Multimediala presentationer.
- Information både om och från myndigheter, föreningar, organisationer, företag och privatpersoner.
- Elektroniska tidskrifter.
- Mer svårdefinierad information som till exempel sökresultat från olika söktjänster, listor med länkar till andra objekt och objekt som innehåller blandningar av flera olika typer av information.

Det är svårt att finna någon naturlig ordning bland alla dessa olika mediatyper och dess innehåll. Dessutom förvärras problemet av att kvaliteten på innehållet är mycket ojämn. Det finns inte heller någon kontroll av det intellektuella innehållet. Informationens ursprung och syfte skiftar med en blandning av privata, kommersiella och offentliga utgivare som alla har olika intressen, motiv och mål.

Vi kan dock med hjälp av Edward O'Neill dela upp innehållet i två de grupperna *statiska* och *interaktiva* webbsidor.<sup>26</sup> En statisk webbsida är den typ som mest liknar en bok eller en artikel och innehållet kan läsas från början till slut utan att man behöver lämna sidan. Om sidan inte har modifierats visar den alltid samma information för alla slags läsare, det vill säga den ändrar inte innehåll beroende på vem, när eller varifrån den läses.

En interaktiv webbsida är ett objekt som använder sig av en koppling till ett externt datorprogram för att utföra vissa funktioner. Sidan kan innehålla olika typer av möjligheter för användaren att påverka innehåll form och struktur på sidan. Exempel på dessa funktioner är möjligheten att fylla i ett formulär som kan ställa en fråga till en databas eller skicka någon annan typ av information till sidans administratör. Webbsidan kan innehålla fält för lösenord som i sin tur ger tillgång till vissa tjänster. Användaren kan ibland även påverka hur sidan presenteras på skärmen. Söksidan på en söktjänst är ett konkret exempel på en interaktiv webbsida. Där kan man dels fylla i ett sökformulär, begränsa sin sökning och ofta även välja hur resultatet ska presenteras. Vanliga tekniker för att göra interaktiva webbsidor idag är till exempel CGI-script och Java.

## 5.2.2 Åtkomst av digitala objekt

Utgår man från etablerade informationsvetenskapliga metoder och teorier och applicerar det på objekten i det system som beskrivits ovan kan det verka helt omöjligt att hitta någonting över huvud taget på webben. Vi ska nu beskriva de förutsättningar för teknisk och intellektuell åtkomst av digitala objekt som finns.

---

<sup>26</sup> O'Neill, Edward T. (1997), Characteristics of web accessible information // 63<sup>rd</sup> IFLA General conference - Conference programme and proceedings - August 31-September 5, 1997. - [URL: <http://www.nlc-bnc.ca/ifla/IV/ifla63/63onee.htm>] 98-04-28.

### 5.2.2.1 Teknisk åtkomst

För att skapa åtkomst till information måste man dels identifiera eller upptäcka och sedan lokalisera det informationsbärande objektet. För att återvinning av ett digitalt objekt på webben ska vara möjlig måste nätverksadressen till den eller de datorer som tjänar som lagringsplats för objektet vara känd. Den standard som är i bruk idag för lokalisering av digitala objekt på webben heter Uniform Resource Locator (URL). Funktionen hos en URL är, som namnet antyder, att utgöra en pekare (eng:locator) till den nätverksadress på webben där det aktuella objektet finns lagrat. En URL innehåller alltså ingen information om ett objekts intellektuella innehåll, till exempel ämne. Den innehåller ej heller någon information om objektets storlek. URL hänvisar enbart till var objektet är beläget. Detta faktum är dess främsta nackdel och är även orsaken till att en URL inte heller fungerar för att identifiera innehållet i det digitala objekt som den hänvisar till.

En URL är uppbyggd av flera delar och kan till exempel se ut så här: <http://www.hb.se/>. Den första delen talar om vilken slags nätverksprotokoll som används, till exempel HTTP, FTP eller GOPHER. Sedan kommer en avgränsning (://), och efter det en adress till en specifik dator, till exempel [www.hb.se](http://www.hb.se). Sist finns ofta även en hänvisning till en mapp eller en fil i datorns filhierarki, till exempel [/bhs/](http://www.hb.se/bhs/). URL har som sagt inte med objektens intellektuella innehåll att göra vilket påvisas av att URL:en inte är den minsta beståndsdel i systemet. När man klickar på en hyperlänk eller på något annat sätt söker åtkomst till ett visst objekt översätts bokstäverna i den del av URL:en som hänvisar till maskinadressen till siffror. Dessa siffror kallas IP-nummer och kan se ut så här, 193.10.166.34. Översättningen från URL till IP-nummer sköts automatisk av en så kallad *resolution service* eller *domain name server* (DNS).<sup>27</sup>

URL är inte speciellt bra som lokaliseringsverktyg eftersom både den dator dit den pekar mot och det protokoll som används kan ändras utan att adressen automatiskt uppdateras. Clifford Lynch påpekar att URL aldrig var tänkt som en identifiering av innehåll utan utvecklades för att vara ett snabbt och flexibelt adresseringssystem.<sup>28</sup> Det är dock viktigt att påpeka att det inte finns några tekniska hinder som gör att adresser till digitala objekt på webben måste förändras. Av olika skäl sker det dock alltför ofta vilket gör att länkar ibland är ”döda”, det vill säga länken hänvisar till ett objekt som har raderats eller flyttats. URL ger därför ingen garanti för åtkomst.

Det faktum att URL inte har det som krävs för tillförlitlig återvinning av webbobjekt uppmärksammades tidigt i webbens historia och nya standarder håller på att utvecklas.<sup>29</sup>

---

<sup>27</sup> För mer information om tekniska standarder för TCP/IP-baserade nätverkssystem se bland annat: The architecture domain [URL: <http://www.w3.org/Architecture/>] 98-01-11 ; Internet tools summary [URL: <http://www.december.com/net/tools/index.html>] 98-01-11.

<sup>28</sup> Lynch, Clifford, Identifiers and their role in networked information applications [URL: <http://www.arl.org/newsltr/194/identifier.html>] 97-11-03.

<sup>29</sup> Internet Engineering Task Force [URL: <http://www.ietf.org/>] 97-11-03.

Det finns redan flera olika standarder, bland annat URN, DOI, PURL.<sup>30</sup> Vi avser inte fördjupa oss i dessa tekniskt orienterade standarder utan beskriver bara den som vi tycker verkar mest intressant. Anledningen till detta är att visa på att det i alla fall förekommer försök att komma till rätta med problematiken kring intellektuell åtkomst av webbobjekt.

Arms skirver i artiklen *An Architecture for Information in Digital Libraries* att 'Handles are sometimes called "Uniform Resource Names (URN)" because they identify Internet resources by name, in contrast to the widely used Uniform Resource Location (URL) which identifies resources by location.'<sup>31</sup> URN kan till skillnad från URL även omfatta information om det digitala objektets innehåll. Sollins och Masinter skriver i artikeln *Functional Requirements of Uniform Resource Names* att URN är tänkt att fungera som en "globally unique persistent identifier used for recognition, for access to characteristics of the resource or for access to the resource itself".<sup>32</sup> URN ska alltså vara en globalt enhetlig standard där det digitala objektets handtag är unikt. Det innebär att en URN bara ska leda till ett objekt och leda till detta oberoende var det är lagrat. Det är även tänkt att URN ska vara för evigt bestående, i alla fall så länge Internet existerar.

Både etablerade standarder som *International Standard Book Number (ISBN)* och *International Standard Serial Number (ISSN)* och andra ännu inte färdigställda standarder för beskrivning av objekten och deras innehåll ska kunna ingå i URN. Primärt hanterar URN inte lokalisering av objekt utan måste konverteras till URL, vilket ska skötas av speciella *resolution services*. Detta är ett system som kommer att likna det system som idag konverterar URL till IP-nummer. URN liknar till viss del ISBN och ISSN som också är unika och bestående identifikatorer. Dessa fungerar globalt och man kan, med rätt förkunskaper, läsa ut fakta om objektet ur sifferkombinationen i ISBN. Ett ISBN kan till exempel se ut så här: 0-566-07547-4. Den första siffran står för utgivningsland, den andra sifferkombinationen för förlag, den tredje är ett objektnummer och det sista numret är en kontrollsiffra. URN-standarderna kommer med all sannolikhet att göra det lättare att återvinna digitala objekt på webben men vi fördjupar oss inte mer i detta nu. De texter som handlar om detta är dels mycket tekniskt inriktade och trots att URN standarderna redan är färdigutvecklade är de inte i bruk i praktiken.

Skillnaden mellan URN och URL kan i korthet sägas vara att den förstnämnda standarden identifierar det digitala objektet i enlighet med dess unika namn, det vill säga dess handtag, medan URL identifierar objektet i enlighet med dess lokalisering.<sup>33</sup>

### 5.2.2.2 Intellektuell åtkomst

---

<sup>30</sup> För information om PURL och DOI se Introduction to persistent Uniform Resource Locators [URL: <http://purl.oclc.org/OCLC/PURL/INET96>] 98-03-05 ; The digital object identifier (DOI) system [URL: <http://www.doi.org/>] 98-04-21.

<sup>31</sup> Arms, William Y., et al, (1997).

<sup>32</sup> Sollins, K och Masinter, L, (1994), Functional requirements of Uniform Resource Names. - December, RFC 1737. - [URL: <http://ds.internic.net/rfc/rfc1737.txt>] 98-04-27.

<sup>33</sup> Arms, William Y., et al., (1997).

Till skillnad från teknisk åtkomst syftar intellektuell åtkomst till att finna ett objekt utifrån dess intellektuella innehåll. Eftersom det saknas tekniska standarder för intellektuell åtkomst måste detta ske genom en söktjänst. Söktjänster använder sig av olika metoder för att indexera och att strukturera, det vill säga etablera relationer mellan objekt.

För att skapa intellektuell åtkomst till digitala objekt underlättas återvinningen av någon slags ämnesrepresentation. Mycket av de svårigheter med att finna det som eftersöks på webben beror på de indexeringsmetoder som används. Jon Wallis och Peter Burden, vid School of Computing & Information Technology vid Wolverhamptons universitet, urskiljer tre metoder för inhämtning av data som ligger till grund för indexering i söktjänster.<sup>34</sup>

- Aktivt förvärv - med indelning i två undergrupper:
  - Automatiskt förvärv - ett datorprogram, så kallad robot eller spindel samlar in samt indexerar data från de digitala objekten.
  - Manuellt förvärv - data insamlas på mänsklig väg.
- Passivt förvärv - data om ett digitalt objekt sänds in av en person som vill att detta skall bli funnet.
- Hybrid förvärv - en kombination av några eller alla av de ovannämnda förvärvstyperna.

Aktivt automatiskt förvärv möjliggör inhämtning av stora mängder data och därmed indexering av många digitala objekt. Vad som indexeras är i vissa fall hela texten eller utifrån algoritmer valda ord ur texten. Text behöver nödvändigtvis inte vara detsamma som det egentliga ämnesinnehållet. Vid sökning på termen *universitet* kan digitala objekt som på något sätt innehåller detta ord erhållas. Det innebär nödvändigtvis inte att innehållet handlar om universitet. Det kan likväl betyda att en text innehåller ordet universitet eftersom objektet är lokaliserat vid ett universitet och ordet nämns därmed i texten. Texten behöver för övrigt inte befatta sig innehållsmässigt med ämnet universitet. Indexering av ord är därmed nödvändigtvis inte passande för att beskriva innehåll. Ett problem med den automatiska indexering som sker i dag är att den vanligtvis enbart indexerar text. Ett digitalt objekt omfattar, som vi tidigare beskrivit, flera typer av data, till exempel bilder eller ljud, som oftast inte indexeras.

En form av ämnesrepresentation är klassificering. Vid klassificering av ett digitalt objekt avser man beskriva det intellektuella innehållet som objektet är bärare av varpå intellektuell åtkomst skapas. Det är vår avsikt att i denna uppsats ta fasta på problemen med ämnesrepresentation genom klassifikation i söktjänster på webben.

Manuell insamling är mer tidskrävande men kan medföra högre kvalitet på ämnesrepresentationen. Passivt förvärv är en mer osäker insamlingsmetod då datamängden som skall ligga till grund för indexering helt beror på bidragsgivarna. Insamling med

---

<sup>34</sup> Burden, Peter och Wallis, Jon, Towards a classification-based approach to resource discovery on the web [URL: <http://scitsc.wlv.ac.uk/wwlib/position.html>] 98-04-28.

hybridmetod kan utnyttja de fördelar som alla de övriga förvärvstyperna omfattar. Sammanfattningsvis finns det på dagens webb svårigheter med såväl teknisk som intellektuell åtkomst av digitala objekt.

### 5.3 Browsing och sökning

Syftet med detta stycke är att definiera *browsing* och *sökning* så som vi kommer att använda dem. Att definiera dessa termer är inte helt oproblematiskt då deras betydelse varierar mellan författare, sammanhang och epok. Exempelvis tar Chang och Rice i sin ofta citerade artikel upp fem olika ämnesområden inom vilka olika typer av browsing förekommer.<sup>35</sup> I biblioteksmiljö används ofta browsing som beteckning för att söka efter litteratur direkt från hyllan och att mer eller mindre ostrukturerat gå omkring och titta efter något intressant, ett söksätt som markant skiljer sig från det målmedvetna sökandet i en biblioteksdatabas. Under uppslagsordet *browsing* i *Encyclopedia of library and information science* dras slutsatsen att:

*browsing would seem to be not one but many things. It is sometimes a purely random, unstructured, and undirected activity. Other times it is closely directed and structured, where, although the final sources or media may not be known, the desired product or goal is clear.*<sup>36</sup>

Dessa ord skrevs i början av 1970-talet. Avseende browsing med hjälp av datorer uppfattades tekniken vid denna tid som allt för outvecklad. Problemet ansågs vara att dialogen mellan användare och datorsystem var undermålig. Försök till förbättringar pågick dock under denna tid och ett tidigt exempel är företaget DIALOGs insatser.<sup>37</sup> Sedan dess har dock stora förändringar skett av såväl datorer och programvara. I samband med utvecklingen av allt mer sofistikerade datorsystem under 1980-talets mitt och senare hälft sker en markant förändring i användningen av termen browsing, påpekar Kurth och Peters.<sup>38</sup> För det är just inom områden som knyter an till datorer som övertagit termen browsing och ytterligare komplicerat dess betydelse. Till dessa områden hör studier kring hypermediasystem, ljud- och bilddatabaser och inte minst Internet.

Nästan tjugofem år efter den citerade texten ovan skrevs definierar Keenan i *Concise dictionary of Library and information science* innebörden av termen browsing enligt följande: "looking through a store of documents at random, with no conscious search strategy; choosing from among a number of documents by examining each item."<sup>39</sup> Enligt

---

<sup>35</sup> Chang, Shan-Ju och Rice, Ronald E., (1993).

<sup>36</sup> Encyclopedia of library and information science: vol 3: bookmobiles to California/ ed. Allen Kent och Harold Lancour. - (Encyclopedia of library and information science). - New York: Dekker. - Cop 1970. - s. 414.

<sup>37</sup> Ibid., s. 408-415.

<sup>38</sup> Kurth, Martin och Peters, Thomas A., (1995), Browsing informations systems : an extensive annotated bibliography of the literature [URL: [http://ai.iit.nrc.ca/II\\_public/BrowsingReferences/](http://ai.iit.nrc.ca/II_public/BrowsingReferences/)] 97-11-03

<sup>39</sup> Keenan, Stella, (1996), Concise dictionary of library and information science. - London: Bowker Saur. - s.

denna definition är browsing att på ett slumpartat, godtyckligt sätt granska en samling av objekt utan någon som helst sökstrategi. Å andra sidan kan browsing dessutom uttydas som ett sätt att utifrån ett antal objekt välja ut något eller några objekt genom granskning av dem en och en. Även denna definition antyder att det finns två sätt att uppfatta browsing, antingen som en ostrukturerad metod där sökning sker på måfå eller som en grundlig genomgång av ett urval av objekt.

Genom att ställa olika termer mot varandra kan definitioner göras tydligare och därmed klargöra vad termen inte är. Vanligtvis brukar en tydlig skiljelinje dras mellan användningen av termerna sökning och browsing, där sökning uppfattas som grundligare och mer strukturerad metod som används då informationsbehovet är klart definierat. Traugott Koch definierar sökning på webben som den komplicerade process som sker i en databas speciellt skapad för detta ändamål. Sökning i dessa databaser kan ske med hjälp av ett sökspråk och till exempel boolesk logik.<sup>40</sup>

I denna uppsats görs en åtskillnad mellan termerna sökning och browsing för att skilja mellan två olika metoder för intellektuell åtkomst av digitala objekt på webben. För att skilja browsing från sökning används termen browsing i betydelsen navigering i ämnehierarkiskt strukturerade samlingar av information. Denna definition är hämtad från Angela Oehlers *Browsingdienste im Internet*.<sup>41</sup> I sin text hänvisar hon i sin tur till Traugott Koch som använder browsing i betydelsen systematisk bläddring i strukturerade samlingar av information. Användaren klickar sig fram i ett ämnehierarkiskt ordnat system istället för att mata in söksträngar för sökning i ett index. I jämförelse med Keenans vaga definition framstår Kochs och Oehlers som tydligare, vilket kan ha sin förklaring i att de sistnämnda författarna avgränsar användningen av browsing till söktjänster på webben som tillhandahåller strukturerade samlingar av information samt ställer termen browsing mot sökning. Man ska dock inte uppfatta browsing som en sämre metod för åtkomst av information än sökning utan som ett alternativ.

Avseende browsing på webben skall termen inte förväxlas med vad som populärt brukar benämnas surfning. Koch är noga att påpeka skillnaden mellan just termen browsing och surfning. Den sistnämnda termen kännetecknas enligt hans definition av osystematiskt följande av hyperlänkar.<sup>42</sup> Denna definition stämmer väl överens med Websters och Pauls beskrivning av surfning som browsing utan hjälpmedel.<sup>43</sup> Detta är en definition som även vi instämmer i.

---

21.

<sup>40</sup> Koch, Traugott, (1996a), Internet search services.

[URL: <http://www.ub2.lu.se/tk/demos/DO9603-meng.html>] 98-04-28.

<sup>41</sup> Oehler, Angela, (1996), Browsingdienste im Internet [URL: <http://userpage.fu-berlin.de/~angela/bond.htm/browsing>] 98-04-28.

<sup>42</sup> Koch, Traugott, (1996a).

<sup>43</sup> Webster, Kathleen och Paul, Kathryn, (1996), Beyond surfing: tools and techniques for searching the web [URL: <http://magi.com/~mmelick/it96jan.htm>] 98-04-27.

Sammanfattningsvis utgår vår definition av browsing från Koch och Oehler. De diskuterar dock inte användarnas informationsbehov och hur det påverkar browsing. En användare har alltid ett informationsbehov som kan befinna sig på olika medvetandenivåer, alltifrån klart uttalat till vagt formulerat. Browsing kan användas oberoende om användaren har ett vagt eller ett klart uttalat informationsbehov.

Termen *browsability*, det vill säga hur bra ett informationssystem stödjer browsing är också ett centralt begrepp i sammanhanget. Chang och Rice påpekar att strukturering och organiseringen av objekt kan underlätta browsing.<sup>44</sup> Ordnar man samlingen efter en ämnehierarkisk struktur ökar det samlingens browsability, det vill säga underlättar browsing.

Betydelsen av termen browsing är komplex och det är svårt att finna ett svenskt ord som på ett korrekt sätt motsvarar dess betydelse. Man skulle kunna använda termen *bläddring* men den överensstämmer inte helt och hållet med browsing. Fortsättningsvis kommer browsing att användas trots att det är en term som är hämtad från det engelska språket.

## 5.4 Söktjänster

Syftet med detta stycke är att definiera termerna söktjänst, indexbaserad söktjänst och ämnehierarkisk söktjänst. Vi ska även kort beskriva vad som kännetecknar de olika tjänsterna.

Terminologin kring söktjänster på webben är minst sagt inkonsekvent. Man talar om bland annat sökmaskiner, sökindex, sökrobotar, ämneskataloger, ämnesindex, ämnesguider, ämnestråd, blädderkataloger, länklister, webbkataloger, webbindex, gateways, clearinghouses och istografier. För vårt fortsatta resonemang är det viktigt att försöka bringa någon slags klarhet i detta terminologiska kaos.

Den instans som definierar termer inom datorområdet heter Svenska Datatermgruppen. Gruppen är en sammanslutning med representanter från både den akademiska världen, näringslivet och olika mediaföretag. Som samordnare fungerar Svenska språknämnden och Tekniska nomenklaturcentralen.<sup>45</sup> Datatermgruppen rekommenderar bruket av termen *söktjänst* för att beteckna en tjänst som ”erbjuds på en webbplats och som är inriktad på att tillhandahålla sökmöjligheter i text på webbsidor och i meddelanden i diskussionsgrupper”.<sup>46</sup> Exempel på söktjänster är enligt datatermgruppen: AltaVista, Euroseek, Lycos och Yahoo.<sup>47</sup> Vi utgår från denna definition trots att den inte innehåller någon tydlig distinktion mellan de olika typer av tjänster som står till buds för användaren

---

<sup>44</sup> Chang, Shan-Ju och Rice, Ronald E., (1993), s. 237.

<sup>45</sup> Tekniska nomenklaturcentralen [URL: <http://www.tnc.se/>] 98-04-18 ; Svenska språknämnden [URL: <http://www.spraknamnden.se/>] 98-04-18.

<sup>46</sup> Svenska datatermgruppens rekommendationer, version 9, 4 augusti 1997 [URL: <http://www.nada.kth.se/dataterm/rek/v9.html#a67>] 98-04-27.

<sup>47</sup> AltaVista [URL: <http://altavista.telia.com/>] ; EuroSeek [URL: <http://www.euroseek.net/>] ; Lycos: [URL: <http://www.lycos.com/>] ; Yahoo! [URL: <http://www.yahoo.com/>] 98-03-09.

eller hur informationen är strukturerad. Kurt Munson väljer att skilja mellan två typer av hjälpmedel för åtkomst av information på webben.<sup>48</sup> Dels finns det *indexbaserade söktjänster*, som erbjuder sökning i automatiskt genererade index, och dels finns det *ämnessierarkiska söktjänster* som möjliggör browsing i hierarkiska ämnesstrukturer. Det är den sistnämnda kategorin som är av största intresse för denna uppsats.

Tilläggs bör att många av dagens söktjänster erbjuder både sökning och browsing eller en kombination av båda. Den svenska söktjänsten Kvasir erbjuder till exempel både en ämnessierarkisk lista baserad på passivt förvärv med resurser godkända av en redaktion och ett automatiskt genererat index med sökfunktion. En annan variant finns hos Yahoo! Sverige, som erbjuder både en ämnessierarki för browsing men man kan även söka i densamma.<sup>49</sup>

#### 5.4.1 Indexbaserade söktjänster

Termen indexbaserad söktjänst definieras här som en söktjänst som är baserad på ett index och som möjliggör formulering av en söksträng. Den indexbaserade söktjänsten består av tre huvudsakliga delar, ett sökbart index, en sökmotor samt ett sökgränssnitt. Det sökbara indexet består av hela eller delar av digitala objekt insamlade från webben, till exempel URL-adress, titel, de 20 första orden i objektet eller vissa metadatakoder. Sökmotorn är ett specialskrivet datorprogram som söker efter digitala objekt på webben och utifrån specifika urvalsregler indexerar dem i en databas.<sup>50</sup> Hur detta går till i praktiken varierar från program till program. Svenska Datatermsgruppen har definierat en sökmotor som ett ”program för indexering av och sökning i stora textmassor, till exempel samtliga webbsidor på Internet”.<sup>51</sup> Andra namn som förekommer i stället för sökmotor är sökrobotar, spindlar och krypare (eng: robots, spiders, crawlers). Sökgränssnittet är den del av tjänsten som dels omvandlar användarens sökformulering till en söksträng som kan appliceras på indexet och sedan presenterar resultat för användaren. Dessutom kan man kalla själva det grafiska gränssnittet för sökgränssnitt.<sup>52</sup>

Viktigt att påpeka i sammanhanget är att den engelska termen *search engine*, vilket ordagrant kan översättas med sökmotor, används både i betydelsen söktjänst och sökmotor. Även den del av söktjänsten som applicerar användarens sökformulering på databasens index betecknas ibland *search engine*. Dessa språkliga problem har uppmärksammats vid läsandet av engelska texter.

---

<sup>48</sup> Munson, Kurt I., (1996), World Wide Web indexes and hierarcical lists: finding tools for the Internet // Computers in Libraries. - Medford, N.J.: Information Today. - Vol. 16(6), June. - s. 54-57.

<sup>49</sup> Kvasir Sverige [URL: <http://kvasir.se/>] 98-04-05. ; Yahoo! Sverige [URL: <http://www.yahoo.se/>] 98-04-05.

<sup>50</sup> De Bra, P.M.E., Finding information on the web [URL: <http://www.wis.win.tue.nl/~debra/cwi-qw/article.html>] 97-11-04.

<sup>51</sup> Svenska datatermgruppen.

<sup>52</sup> För mer information om söktjänster se bland annat: De Bra, P.M.E., Finding information on the web [URL: <http://www.wis.win.tue.nl/~debra/cwi-qw/article.html>] 97-11-04. ; Dong, Xiaoying och Su, Louise S., (1997), Search engines on the world wide web and information retrieval from the Internet: a review and evaluation // Online & CD-rom review. - Oxford: Learned Information. - Vol. 21, nr 2. - s. 67-81.

### 5.4.2 Ämneshierarkiska söktjänster

I vår definition av en ämneshierarkisk söktjänst utgår vi från Angela Oehler som använder termen *browsingtjänst* (ty: browsingdiensten) för att beteckna en hierarkisk ämneslista som möjliggör browsing. Enligt hennes definition består browsingtjänster av länkar till digitala objekt och tjänstens hierarkiska struktur möjliggör browsing i överordnade, koordinerade och underordnade kategorier.<sup>53</sup> Länkarna är indelade i kategorier enligt etablerade eller icke-etablerade klassifikationssystem. Vi vill så långt det är möjligt använda svenskt språkbruk varför vi introducerar termen *ämneshierarkisk söktjänst* vars definition till fullo översensstämmer med Oehlers term browsingtjänst. En ämneshierarkisk söktjänst skiljer sig från en indexbaserad genom att den har en ämneshierarki som möjliggör browsing till skillnad från sökning i ett index.

De ämneshierarkiska söktjänster vi exemplifierar med tillhandahåller länkar till digitala objekt vilka är direkt tillgängliga via webben. De är alltså inte frågan om så kallade *online public access catalogues* (OPAC:s) med webbgränssnitt. Rent tekniskt kan även ämneshierarkiska söktjänster bestå av ett index men med den stora skillnaden att det presenteras i form av en ämneshierarki i stället för att enbart tillåta sökningar med sökspråk.

### 5.4.3 Platta listor

Förutom indexbaserade och ämneshierarkiska söktjänster finns även olika slags platta listor, det vill säga länkarna radas upp i en lista utan hierarkisk uppdelning. Wallis och Burden kallar dessa listor för *flat lists*.<sup>54</sup> Dessa listor befinner sig på en annan nivå än de professionellt skapade tjänsterna men trots detta kan de självklart fylla vissa informationsbehov i vissa sammanhang.

## 5.5 Sammanfattning

I inledningen av detta kapitel visar vi på webbens stora expansionen som en av de faktorer som gör det problemematiskt att skapa åtkomst och återvinna information på webben. Tillväxten avser såväl antalet anslutna datorer, registrerade domäner, digitala objekt som användare. Ytterligare ett problem är den stora mängden objekt med mycket skiftande innehåll. Termen digitalt objekt definieras som ett objekt i digital form som är sammanställt för nätverkshantering. I samband med definition av digitalt objekt och beskrivning av dess karaktäristika diskuteras förutsättningarna för teknisk och intellektuell åtkomst. Dagens teknik och frånvaron av en central kontrollinstans försvårar kontrollen över objekten. Den tekniska åtkomsten sker med hjälp av URL som inte ger någon garanti för åtkomst och ej

---

<sup>53</sup> Oehler, Angela, (1996).

<sup>54</sup> Burden, Peter och Wallis, Jon, Towards a classification-based approach to resource discovery on the web.

omfattar uppgifter om intellektuellt innehåll. När det gäller intellektuell åtkomst saknas etablerade standarder för återvinning eller beskrivning av digitala objekt. indexbaserade och ämneshierarkiska söktjänster är de huvudsakliga medel som erbjuder intellektuell åtkomst. En ämneshierarkiskt ordnad söktjänst möjliggör browsing i en ämneshierarki och en indexbaserad sökning i ett index. Detta index är ofta insamlat med hjälp av aktivt automatiskt förvärv medan de ämneshierarkiska söktjänsterna ofta bygger på passivt eller hybrid förvärv. Vi skiljer på browsing och sökning som är två sätt att skapa intellektuell åtkomst av ett objekt. Browsing definieras som navigering i en ämneshierarkiskt strukturerad samling information och sökning är formulering av en söksträng och applicering av denna på ett index.

## 6 Klassifikation

Detta kapitel syftar till att ge en teoretisk bakgrund till vad klassifikationssystem är och hur de fungerar. Vi beskriver de två befintliga bibliografiska klassifikationssystemen DDK och UDK. Vi kommer i korthet även att behandla ämnesrepresentation samt kontrollerat och naturligt indexeringspråk. Stycket syftar även till att ge exempel på hur olika klassifikationssystem används på webben genom att kort beskriva fyra webbaserade söktjänster som alla använder olika system. Eftersom litteraturen inom detta område till största del utgår från traditionella förhållanden där man klassificerar fysiska objekt, används termen *dokument* ofta i texterna. Vi använder i detta kapitel termen dokument i den generella betydelsen *en fysisk bärare av intellektuellt innehåll*.

### 6.1 Vad är klassifikation?

Klassifikation innebär en systematisk indelning av objekt i grupper eller kategorier. Enligt Ana och Donald Cleveland är klassificering, det vill säga systematisering av omvärlden, en allmänmänsklig företeelse som ständigt försigår.<sup>55</sup>

Birger Hjørland urskiljer tre nivåer för organisering av kunskap, med olika ambitionsnivå beroende på den praktiska tillämpningen.<sup>56</sup> Den första och lägsta ambitionsnivån är *ad hoc-klassifikation*, även kallat *kategorisering*. Hjørland ger som exempel på ad hoc-klassifikation när man arrangerar blommor och växter i hemmet. Man utgår då från sin godtyckliga uppfattning om vad som är smakfullt för stunden, vilken omgivningen är och vilka andra föremål som kan passa till blommorna. Mycket av den klassifikation och kategorisering som görs utgår i grunden från ad hoc-klassifikation. Det är en vardaglig form av organiserande av kunskap där utgångspunkten är vad som är lämpligt för tillfället. Till denna kategori hör de icke-etablerade klassifikationsystem som är utvecklade för användning i ämneshierarkiska söktjänster på webben. *Pragmatisk klassifikation* är den mellersta ambitionsnivån och reflekterar olika kunskapsintressen. Varierande intressen genererar olika varianter av kunskapsorganisation. Hjørlands exempel på pragmatisk klassifikation är även det förbundet till blommor. Till exempel en trädgårdsmästare på amatörstadiet och en professionell biolog klassificerar rhododendron på olika sätt beroende på att de har olika förhållningssätt till ämnet. De använder olika språk och terminologin är anpassad för att motsvara den praktiska användningen inom respektive intresseområde. Bibliografiska klassifikationssystem, som till exempel UDK och DDK, räknas till gruppen pragmatisk klassifikation. De är utvecklade för praktisk tillämpning i bibliotek av bibliotekarier och informationsspecialister. Den översta klassifikationsnivån kallas *vetenskaplig klassifikation* och är den mesta ambitiösa formen av kunskapsorganisation. Den syftar till att beskriva verkligheten så som den är och ska inte blandas ihop med bibliografisk klassifikation. Vetenskaplig klassifikation är abstrakt och generell till sin

---

<sup>55</sup> Cleveland, Donald och Cleveland, Ana, (1990), Introduction to indexing and abstracting. - Englewood, Colo.: Libraries Unlimited. - s. 17.

<sup>56</sup> Hjørland, Birger, (1997), s. 46f.

karaktär. Till exempel klassificeras flora och fauna på denna nivå med biologisk taxonomi, alltså vetenskaplig systematik. Exempel på vetenskaplig klassifikation är det periodiska systemet som klassificerar alla kända grundämnen eller Carl von Linnés sexualsystem. I denna uppsats beaktas endast ad hoc och pragmatisk klassifikation.

Inom biblioteks- och informationsvetenskap är grundsyftet med all klassificering att skapa ordning i en samling för att göra det möjligt att återvinna samlingens objekt, såväl fysiska som digitala. Indexering av objekt, det vill säga när en utvald indexeringsterm tillskrivs ett objekt, kan betraktas som en form av klassificering skriver Cleveland och Cleveland.<sup>57</sup> Metoderna uppfattas åtminstone vara nära besläktade med varandra. När ett objekt tillskrivs en indexterm infogas objektet indirekt i en klass som omfattar den tillskrivna index termen. Birger Hjørland ansluter sig till denna uppfattning då han hävdar att varje form av klassificering kan betraktas som indexering och varje form av indexering kan betraktas som klassificering. Enligt hans åsikt styrks denna uppfattning av det faktum att klassifikationssystem emellanåt görs om till thesauri och tvärtom. UDK är ett exempel på ett klassifikationssystem som genomgått en förvandling till thesaurus.<sup>58</sup> Detta innebär dock inte att ordindexering och klassificering är samma sak, påpekar Hjørland. Traditionell klassificering utgår från en analys av det generella mot det specifika inom ett ämne eller kunskapsområde, så kallad top-down analys. Vid ordindexering sker analysen i omvänd ordning, det vill säga bottom-up analys. Analysen utgår från det specifika för att fortsätta mot det mer generella.<sup>59</sup> En annan skillnad är att ett klassifikationssystem påvisar relationer mellan ämnen och kunskapsområden, såväl inbördes relationer som interna.

## 6.2 Ämnesrepresentation

Många användare av informationssystem söker ofta efter information om ett specifikt ämne, varför det är viktigt att tillhandahålla verktyg som möjliggör just ämnesvis återvinning. Klassifikationssystem är ett medel för att återvinna objekt utifrån just ämne. Syftet med en sökning på eller browsing efter ett ämne är dels att påvisa vad som finns om ett specifikt ämne och dels att visa på relaterade ämnen. Utan att gå djupare in på vad ett ämne är bör det ändå påpekas att det råder många olika uppfattningar om hur man ska definiera ett ämne. Jennifer Rowley skriver att det antingen kan vara ett område av särskilt intresse, ett yrkes- eller forskningsområde, ett område som någon skriver om eller ett studerat kunskapsområde.<sup>60</sup> Flera av hennes definitioner förefaller vara ganska lika och överlappar sannolikt varandra i viss mån.

Indexeringstermer betraktas ofta som surrogat eller kondensationer av ämnesinnehållet. Hjørland hävdar att ämnesrepresentationer i vissa fall, i synnerhet i manuella system, kan fylla denna funktion, men han anser samtidigt att detta inte är den huvudsakliga funktionen.

---

<sup>57</sup> Cleveland, Donald och Cleveland, Ana, (1990), s. 17.

<sup>58</sup> Hjørland, Birger, (1997), s. 29.

<sup>59</sup> Ibid., s. 29f.

<sup>60</sup> Rowley, Jennifer, (1996), Organizing knowledge: an introduction to information retrieval. – Aldershot: Ashgate. - s. 159-162.

Först och främst skall ämnesrepresentationer fungera som en hjälp för återvinning. Detta innebär inte att indexeringstermerna måste vara hämtade från det representerade objektet, även om det ofta förhåller sig så.<sup>61</sup> Vid all slags indexering måste man även ta hänsyn till användaren. Det är trots allt användarens informationsbehov som uttrycks genom språket, vilket inte får förbises vid ämnesrepresentation.<sup>62</sup> Ett objekt skall därför inte bara analyseras utifrån sitt intellektuella innehåll utan även utifrån vilka frågor det kan besvara. Under indexeringen gäller det helt enkelt att försöka förutse på vilket sätt användarna kommer att söka efter objektet. En avvägning mellan en användar- och en innehållsmässigt orienterad indexering bör alltså göras.

### 6.2.1 Kontrollerat och naturligt indexeringsspråk

Harter skiljer mellan två huvudgrupper för hur ämnesrepresentation kan ske, antingen med *kontrollerat indexeringsspråk* eller *naturligt indexeringsspråk*. Rowley nämner ytterligare en typ, så kallat *fritt indexeringsspråk*.<sup>63</sup> Kontrollerat indexeringsspråk innebär att det finns utvalda termer som är godkända att användas för indexering. Urvalet av godkända termer görs av människor och sammanställs i en så kallad auktoritetslista. Med listan som referens under indexeringsarbetet förbättras möjligheterna till konsekvent ämnesrepresentation. Det finns två typer av kontrollerade indexeringsspråk enligt Rowley, nämligen *klassifikationssystem* och *alfabetiskt indexeringsspråk*.<sup>64</sup> I det här sammanhanget använder vi alltså termen indexering i en bred bemärkelse som omfattar både klassifikationssystem och ordindexering.

Användning av naturligt indexeringsspråk vid automatisk indexering ställer som enda krav att indextermerna hämtas från det digitala objektet. Alla ord som förekommer i ett objekt utgör därmed potentiella indexeringstermer. Dessa termer kan hämtas från valda delar av ett objekt, till exempel titel och abstract, men även från hela objektet. Indexering av fulltextformat kan bli mycket omfattande och detaljerad, såvida inte någon form av urvalsmekanism av indexeringstermer fastställts. Vid automatiserad indexering med hjälp av datorer kan detta innebära en statistisk analys av termers frekvens för att urskilja betydelsebärande ord. Datorns indexeringsprogram kan använda en sammanställning av användbara termer för att kunna identifiera lämpliga termer för indexering. Datorbaserad indexering av denna typ används bland annat av söktjänster på webben, exempelvis AltaVista. Om indexering med naturligt indexeringsspråk sker av en person är det upp till hans eller hennes omdöme att göra ett urval. Begränsning av om var urvalet får ske underlättar verksamheten.<sup>65</sup>

---

<sup>61</sup> Hjørland, Birger, (1997), s. 25-34.

<sup>62</sup> Harter, S. P., (1986), Languages for information retrieval // Online information retrieval. - Orlando: Academic Press. - s. 22-63. - (Library and information science (New York)).

<sup>63</sup> Ibid., s. 22-63 ; Rowley, Jennifer, (1996), s. 160-162.

<sup>64</sup> Rowley, Jennifer, (1996). – s. 160-162.

<sup>65</sup> Ibid., s. 160-162.

Fritt indexeringspråk ställer inte några som helst krav på vilka ord som får användas som indextermer. Indexering med fritt språk kan utföras på såväl automatisk väg som med mänsklig kraft. Automatiserad indexering i detta fallet innebär i praktiken detsamma som automatisk indexering med naturligt indexeringspråk, det vill säga termerna hämtas från det digitala objektet såvida inte en lista över termer tjänar som grund för urvalet. Intellectuell indexering ställer höga krav på indexerarens förmåga. Om denne är skicklig kan indexeringen bli konsekvent och samtidigt ta fasta på användarnas perspektiv.<sup>66</sup>

Vare sig naturligt eller artificiellt så karaktäriseras varje indexeringspråk av vad som brukar benämnas *vokabulär*, *syntax*, *logisk struktur* samt *domän*. De termer som används för indexering utgör den så kallade vokabulären. Det är de enskilda termerna i vokabulären som utgör de minsta byggstenarna utifrån vilka komplexa sammansättningar kan ske. Syntaxen beskriver hur dessa enstaka element i vokabulären får kombineras. Den logiska strukturen hänför sig till relationen mellan de enskilda elementen i vokabulären, vilket är viktigt i synnerhet för kontrollerat indexeringspråk. Avseende domän är det viktigt att skilja mellan begrepp, vilket är abstrakta idéer, och de symboler som används för att representera dessa begrepp. Ta som exempel satsen ”1+1=2”. Siffrorna uttrycks i exemplet med arabiska symboler, men kan likväl representeras av romerska siffersymboler eller någon annan form av symboler. Det semantiska innehållet, begreppen, förblir dock samma även om olika typer av symboler används för att representera dessa. Men syntaxen i en sats kan vara felaktig beroende på kontext. Den ovanstående satsen är sann såvida den tolkas utifrån ett decimalbaserat nummersystem. Men betraktad med ett binärt nummersystem som bas är den felaktig. En sann sats ur detta perspektiv är till exempel ”1+1=10”. Det semantiska innehållet är med andra ord kontextberoende. Ett indexeringspråk måste alltså ta fasta på objekts kontext, det vill säga domän, för att kunna genomföra en indexering. Detta måste uppmärksammas med tanke på de frågor som användare ställer. Förfrågningar som förefaller vara likartade inledningsvis kan vid närmare betraktelse handla om olika saker beroende på kontext.

### 6.3 Klassifikationssystem

Ett klassifikationssystem syftar till att ordna ett eller flera kunskapsfält på ett organiserat sätt genom att gruppera angränsande områden tillsammans och relaterade områden nära varandra. Detta skall genomföras på ett sådant sätt att det övergripande föregår det specifika så att en hierarki skapas.<sup>67</sup> Grupper, som omfattar ett ämne eller en viss aspekt av ett ämne, benämns vanligtvis *klasser*. Dessa klasser kan vara *överordnade*, *koordinerade* eller *underordnade* i förhållande till varandra. En överordnad klass representerar ett ämne ur ett mer generellt perspektiv än en underordnad klass. I den underordnade klassen representeras ämnet på en mer specifik nivå än den överordnade. Klasser som befinner sig på samma hierarkiska nivå klass benämns koordinerade.<sup>68</sup> Denna struktur brukar liknas vid ett träd,

---

<sup>66</sup> Ibid., s. 160-162.

<sup>67</sup> Aluri, Rao, Kemp, D. Alasdair och Boll, John J., (1991), Subject analysis in online catalogs. – Englewood, Colo.: Libraries Unlimited. - s. 113.

<sup>68</sup> Ibid., s. 113.

där de mest generella huvudklasserna bildar stammen, för att förgrena sig i mindre, mer specifika klasser.

### 6.3.1 Bibliografisk klassifikation

Regler för hur ett bibliografiskt klassifikationssystem skall struktureras och omsättas i praktiken, så kallade klassifikationskonventioner, har funnits i över etthundra år. Dessa regler har med åren förfinats och formaliserats till att utgöra vad man idag kan kalla en klassifikationsteori. Det finns två grundläggande klassifikationsprinciper, enumerativ klassifikation och syntetisk klassifikation vilka kommer att behandlas senare.<sup>69</sup>

Vid skapande av klassifikationssystem, eller revideringar av dessa, kan man utgå från de objekt som ska ingå i systemet. Hänsyn tas då till de perspektiv på ämnet eller ämnena så som de framställs i objekten. Klassifikationssystem speglar och påverkas av de ämnen och ämnesrelationer som finns i de objekt som skall organiseras. Därigenom kringgås några av de svårigheter som kan uppstå vid definition av ett ämne. Detta tillvägagångssätt kan uppfattas som pragmatiskt, det vill säga med den omedelbara nyttan i fokus. Det bör dock understrykas att kunskap är dynamisk till sin natur och att uppfattningen av innehållet i objekt förändras med tiden. Ett klassifikationssystem reflekterar därmed ämnen och deras relationer så som de uppfattas och definieras vid tillfället för systemets tillkomst. För att hålla jämna steg med tillväxten av objekt och nya eller förändrade ämnesdefinitioner måste därför kontinuerlig uppdatering ske. Samtidigt som plats måste beredas för nytillkomna ämnen måste utrymme fortfarande ges för tidigare ämnesdefinitioner. En alternativ metod vid konstruktion av klassifikationssystem är att istället för att anpassa verktyget efter objektens innehåll utgår man från ett teoretiskt perspektiv på kunskapens natur och struktur.<sup>70</sup> Den innehållsrelaterade indexeringsformen är dock den vanligast förekommande metoden.

Aluri, Kemp och Boll ger exempel på vanliga karaktäristiska som används vid indelning av grupper i hierarkiska system.<sup>71</sup>

- Från klass till klassmedlem; primater är överordnat apor.
- Från helhet till delar; nervsystemet i förhållande till ryggmärgen.
- Processers händelseförlopp; för textilier finns förloppet karda, spinna och väva.
- Från mindre till ökad komplexitet; musik för duett, trio, kvartett.
- Kronologi; ett lands historia.
- Enligt rymd och territorium; en kontinentets geografi och geografien i ett land på denna kontinent.

---

<sup>69</sup> Ibid., s. 115.

<sup>70</sup> Rowley, Jennifer, (1996), s. 159f.

<sup>71</sup> Aluri, Rao, Kemp, D. Alasdair och Boll, John J., (1991). - s. 119.

Till dessa typer av indelning skall fogas de godtyckliga grupperingar som förekommer trots alla försök att åstadkomma en logiskt korrekt hierarki. Det bör noteras att samma typ av indelning kan användas mer än vid ett tillfälle vid skapandet av klasser. Exempelvis så kan ett klassifikationsystem med inriktning mot fåglars anatomi indelas efter kroppsdelar, med klasser som huvud, vingar och ben. Klassen huvud kan i sin tur även den delas in utifrån kroppsdelar, till exempel näbb och ögon.

Bibliografisk klassifikation utgörs av en sekvens av klasser som är systematiskt ordnade. Att de är systematiskt ordnade innebär att relaterade klasser hålls samman i största möjliga mån. Det finns två sätt att upprätta en dylik sekvens. Antingen genom induktiv metod, då utgångspunkten är från det specifika till det generella, eller deduktiv metod, då man utgår från det generella mot det specifika.<sup>72</sup> Om möjligt skall detta genomföras på basis av särskilda utvalda skillnader och likheter mellan ämnena. Aluri, Kemp och Boll urskiljer två olika utgångspunkter vid klassifikation. Den ena typen är naturlig klassifikation, där organisationsstrukturen grundas på de inneboende egenskaper som de klassificerade objekten besitter. Den andra typen av utgångspunkt vid skapandet av klassifikationsstrukturer är artificiell klassifikation då klassificerarens omdöme och uppfattning påverkar indelningen.<sup>73</sup>

Vid bibliografisk klassifikation, vars syfte är att underlätta informationsåtervinning utifrån ämne, är det viktigt att inte enbart utnyttja en logisk indelning utan även använda dubbelklassificering. Dubbelklassificering innebär att ett objekt representeras av flera klasser. Klasser kan vara av antingen enkla eller sammansatta ämnen. Enkla ämnen reflekterar enbart en typ av indelning medan sammansatta ämnen reflekterar mer än en typ av indelning. Till exempel en bok om 1700-talet reflekterar enbart en typ av indelning under *historia* och därmed ett enkelt ämne. En bok om Sverige under 1700-talet är däremot ett sammansatt ämne.<sup>74</sup> Ett objekt som behandlar flera ämnen, det vill säga sammansatta ämnen, dubbelklassificeras lämpligen. Vid sökning i flera klasser kan detta underlätta återvinningen för användaren.<sup>75</sup> Dubbelklassificering är mycket användbart och går utmärkt att implementera i hypermediala nätverk som till exempel webben.

De klasser som faller under en överordnad klass, efter det att en viss typ av indelning använts, benämns på engelska *array*. Det finns ingen etablerad svensk term som motsvarar det engelska ordet, så vi använder fortsättningsvis ordet *array*. För att undvika missförstånd och förvirring hos användaren är det lämpligast att använda endast en typ av indelning inom varje *array*. En vanlig indelning vid klassificering av och om människor är kön och ålder. I till exempel DDK användes båda dessa kategorier för indelning i undergrupper av klassen *Education*. Inom samma *array* återfanns tidigare klasser som till exempel *Education for women* och *Education for adults*. Detta kunde åstadkomma svårigheter vid återvinning och

---

<sup>72</sup> Mills, Jack, (1977), The structure of a bibliographic classification // Bliss bibliographic classification. - 2<sup>nd</sup> ed. - London: Butterworths. - 35-47.

<sup>73</sup> Aluri, Rao, Kemp, D. Alasdair och Boll, John J., (1991). - s. 119f

<sup>74</sup> Mills, Jack, (1977). - 35-47.

<sup>75</sup> Aluri, Rao, Kemp, D. Alasdair och Boll, John J., (1991). - s. 120.

klassificering. Den motsägelsefulla indelningen påstås ha sin förklaring i de varierande aspekter som objekten avhandlar inom olika ämnen.<sup>76</sup> Man utgår från ett innehållsorienterat perspektiv vid skapande och revidering av klasser inom DDK. Användarens behov förefaller inte vara prioriterat.

Ett begrepp som är nära förbundet med array är *facett*. Termen definieras ofta på ett liknande sätt som array: "as the product of a characteristic of division."<sup>77</sup> Men facett och array skiljer sig åt på så vis att en facett är resultatet av enbart en typ av indelning. För att fortsätta med exemplet ovan om indelningen av klassen *Education* så får enbart en av dessa kategorier, antingen kön eller ålder, användas för att räknas som en facett.

I klassifikationssystem används ibland förkortade ersättningar för att representera klasser, vilka kallas för *klassymboler*. Ibland förekommer termerna *klassnummer* eller *klassifikationsnummer*. Som framgår av benämningarna utgörs klassymbolerna i detta fall av siffror. *Notation* är det regelverk som styr hur klassymboler kan skapas.<sup>78</sup> Enligt Aluri med flera anses en av notationens fördelar vara att den dels påvisar och upprätthåller gruppernas inbördes ordning dels att den är kortare än ord och därmed fungerar som minnesstöd.<sup>79</sup> Vi anser att frågan om huruvida klassymboler är lättare att minnas än ämnesordet för en klass kan diskuteras. Faktorer som påverkar är bland annat användarens vana vid klassifikationssystemet samt vilken grupp av användare informationssystemet riktar sig till.

### 6.3.1.1 Enumerativa och syntetiska system

Enumerativa system kännetecknas av att varje ämne som kan klassificeras inom systemet är representerat av en klass. Ämnen som inte är tilldelade en egen klass klassificeras under närmast överordnade grupp. För att underlätta konsekvent klassificering med enumerativa klassifikationssystem är de konstruerade med fokus just på det huvudsakliga, övergripande ämnet.<sup>80</sup> Vid användning av enumerativa system för hylluppställning uppstår dock problem om en bok är klassificerad under flera klasser. Det finns enbart en bok och den kan enbart ställas upp vid en hyllplats även om det är klassificerat under flera klasser. I en hypermedial miljö, som till exempel på webben, infinner sig inte samma problem då åtkomst till ett digitalt objekt kan åstadkommas oberoende under vilken klass det förekommer.

Med den andra typen av klassifikationssystem, syntetiska system, konstrueras klasser genom att klassymboler från flera olika klasser kombineras. De första att utveckla detta var skaparna av UDK, Paul Otlet och Henri La Fontaine. Syntetisk klassifikation i sin mest utvecklade form benämns facetterad klassifikation. Facetterad klassifikation kännetecknas av att endast en typ av egenskap, så kallade facetter, används åt gången vid skapande av

---

<sup>76</sup> Ibid., s. 121.

<sup>77</sup> Ibid., s. 121.

<sup>78</sup> Ibid., s. 113.

<sup>79</sup> Ibid., s. 114.

<sup>80</sup> Ibid., s. 122-128.

klasser.<sup>81</sup> Till exempel vid skapandet av klasser för kläder kan egenskapen funktion vara en första utgångspunkt för skapandet av en array. Först när alla upptänkliga funktioner är urskiljda kan nya underordnade klasser infogas, även de indelade utifrån en typ av egenskap.<sup>82</sup> Syntetiska klassifikationssystem används först och främst i klassifikationssystem som är specialiserade på vissa ämnen. Det finns inga klara gränser mellan enumerativa och syntetiska klassifikationssystem, då enumerativa klassifikationssystem ofta har syntetiska inslag.<sup>83</sup>

### 6.3.2 Olika typer av klassifikationssystem

Klassifikationssystem kan indelas i flera grupper. Traugott Koch urskiljer fyra typer, som alla är exempel på bibliografiska klassifikationssystem.<sup>84</sup> Den första gruppen är de *universella systemen*. I denna grupp finner man kända klassifikationssystem som *Universell Decimalklassifikation* (UDK), *Dewey Decimalklassifikation* (DDK) och *Library of Congress Classification* (LCC). Dessa klassifikationssystem syftar till att omfatta all mänsklig kunskap. För att möjliggöra detta skall de tillåta och möjliggöra expansion och kontraktion av systemet. Därmed är det tänkt att universella klassifikationssystem ständigt skall kunna hålla sig ajour med kunskapsutvecklingen. Gemensamt för dessa system är att de utger sig för att vara anpassade för en internationell tillämpning. En annan grupp är de *nationella generella klassifikationssystemen*, till vilka bland annat SABs *Klassifikationssystem för svenska bibliotek* och det nederländska *Nederlandse Basisclassificatie* (BC) hör. Liksom sina internationellt universella motsvarigheter försöker nationella generella system täcka in all kunskap. Det finns ofta en tydlig kulturell och etnocentrisk prägel på dess system, med klassindelning anpassad till det geografiska och språkliga ursprungsområdets fördel. God kännedom om nationella generella system är dessutom ofta begränsad till deras ursprungsplats. Förutom dessa finns diverse ämnesspecifika system, vilka utgör en tredje grupp. *Engineering Information* (EI), *National Library of Medicine* (NLM), *1991 Mathematics Subject Classification* (MSC 1991), *Association for Computing Machinery's Computing Classification System* (ACM CCS) och *ICONCLASS* är alla exempel på ämnesspecifika klassifikationssystem. Ofta är de internationellt anpassade.

Som vi tidigare nämnt delar vi in de typer av klassifikationssystem vi studerar i de två grupperna etablerade och icke-etablerade. De vi kallar för etablerade klassifikationssystem är vanligt förekommande på bibliotek idag och dessa system är ursprungligen även utvecklade för användning just i en bibliotekskontext. Utöver dessa etablerade system finns specialsystem utvecklade för organisering av ämnesspecifika samlingar, vilket vi inte avser

---

<sup>81</sup> Ibid., s. 128-133.

<sup>82</sup> Mills, Jack, (1977), s. 35-47.

<sup>83</sup> Aluri, Rao, Kemp, D. Alasdair och Boll, John J., (1991), s. 128-133.

<sup>84</sup> Koch, Traugott, (1996b), Specification for resource description methods Part 3. The role of classification schemes in Internet resource description and discovery.

[URL: [http://www.ub2.lu.se/desire/radar/reports/D3.2.3/class\\_v10.html](http://www.ub2.lu.se/desire/radar/reports/D3.2.3/class_v10.html)] 98-04-28.

behandla. De system vi benämner icke-etablerade är utvecklats för att användas vid klassificering av digitala objekt i en webbkontext.

## 6.4 Etablerade klassifikationssystem

Detta kapitel ger en bakgrund till de två olika klassifikationssystemen DDK och UDK. Vi beskriver systemens utveckling och grundläggande karakteristika samt några för- och nackdelar med respektive system. Dessutom behandlas två webbaserade söktjänster som använder olika icke-etablerade system.

### 6.4.1 DDK

Dewey Decimal Klassifikation publicerades anonymt 1876 under titeln *A Classification and Subject Index for Cataloging and Arranging the Books and Pamphlets of a Library*. I förordet skriver Melvil Dewey att klassifikationssystemet utvecklades redan 1873 efter enträgna studier av hundratalet litterära alster och femtioalet biblioteksinstitutioner.<sup>85</sup> Inspirationen till DDK:s blivande struktur och indelningen av klasser hämtade han från ett flertal ställen.<sup>86</sup> Den kanske viktigaste förebilden var ett tidigare klassifikationssystem sammanställt av William Torrey Harris. Notapparaten däremot, med 0 som generellt nummer för var klass och siffrorna 1-9 för underordnade klasser, hämtade Dewey sannolikt från Jacob Schwartz, hävdar Comaromi.<sup>87</sup>

Både Chan och Barry anser det sannolikt att Harris vid klassindelningen av sitt klassifikationssystem har utgått från Francis Bacons, 1551-1626, modell för kunskapsklassifikation.<sup>88</sup> John Comaromi påstår dock att det snarare var Hegel som var den huvudsakliga inspirationskällan för Harris. Till den hegelianska ämnesstrukturen fogade Harris en generell klass som omfattade verk som behandlar flera olika ämnen, till exempel uppslagsverk, samlingsverk och periodika. Comaromi anser alltså att Harris främsta inspirationskälla var Hegels modell, vilket alltså även skall reflekteras i DDK:s klassindelning.<sup>89</sup> Tydligt är att DDK bygger på gamla idéer om hur världen skall klassificeras. Varifrån den faktiska inspirationen än kom ifrån, den baconska eller hegelianska traditionen, så mynnade Deweys indelning ut i följande huvudklasser.<sup>90</sup>

---

<sup>85</sup> Chan, Lois Mai, (1994), *Cataloging and classification: an introduction*. - 2<sup>nd</sup> ed. - New York, McGraw-Hill. - (Library Science Series). - s. 269f. ; Comaromi, John, (1976a), *The historical development of Dewey classification system // Major classification systems: the Dewey centennial: / ed by Kathryn Luther Henderson*. - Urbana-Champaign, Ill. Forest Press. - s. 22.

<sup>86</sup> För utförligare beskrivning se Comaromi, John, (1976b) *The eighteen editions of the Dewey decimal classification*. - Albany, N.Y. - s. 1-29.

<sup>87</sup> Chan, Lois Mai, (1994), s. 269f. ; Comaromi, John, (1976a), s. 20.

<sup>88</sup> Chan, Lois Mai, (1994), s. 269f. ; för utförligare beskrivning se Comaromi, John, (1976b), Appendix II, s. 615-618.

<sup>89</sup> Comaromi, John, (1976a), s. 21f. ; Chan, Lois Mai, fotnot 2, s. 269.

<sup>90</sup> OCLC / About the DDC / Dewey Decimal Classification (DDC 21) Summaries  
[URL: <http://www.oclc.org/oclc/fp/about/ddc21sm1.htm>] 98-04-27.

- 000 Allmänt och blandat
- 100 Filosofi och psykologi
- 200 Religion
- 300 Samhällsvetenskap
- 400 Språkvetenskap
- 500 Naturvetenskap och matematik
- 600 Teknologi (tillämpad vetenskap)
- 700 Konst
- 800 Litteraturvetenskap och retorik
- 900 Geografi och historia

Att påstå att klassindelningen sker ämnesvis i DDK är en sanning med modifikation då huvudklasser och i viss utsträckning även underordnade klasser uppenbarligen delas in i akademiska discipliner, åtminstone uppfattades de som sådana vid DDKs tillkomst. Generellt är DDKs hierarkiska struktur först indelad efter disciplin och underdiscipliner. De disciplinära klasserna följs av en ämnesvis indelning med tillhörande undergrupper, varpå en geografisk eller tidsmässig specificering kan göras. Den sista typen av indelning är enligt form.<sup>91</sup>

Notationshierarkin indikerar relationerna mellan klasserna. Till exempel så är klassen *500 Naturvetenskap och matematik* överordnad klasserna med klasssymbol *510* till och med *590*. Av detta följer att klasssymbol *511* till och med *519* är underordnade *510*. Ingen av klasserna kan ha mindre än tresiffriga klasssymboler, varför huvudklasserna och deras direkt underordnade indelningar fylls ut med nollor. Krävs mer än tre siffror för exaktare indelning placeras ett decimaltecken efter den tredje siffran, till exempel *554.5 Italiens geologi*.<sup>92</sup>

- 500 Naturvetenskap och matematik
  - 510 Matematik
    - 516 Geometri
      - 516.3 Analytisk geometri
        - 516.37 Metrisk differential geometri
          - 516.372 Euklidisk geometri
- 520 Astronomi
- 530 Fysik
- 540 Kemi
  - 541 Fysikalisk och teoretisk kemi

De banbrytande och nyskapande förtjänster som Dewey införde med DDK, i jämförelse med tidigare klassifikationssystem, var *relativ lokalisering* och *relativt index*. Vid relativ lokalisering tillskrivs varje bok en eller flera klasssymboler efter det ämne som den avhandlar. En boks innehåll är statiskt och på så vis samlas material efter deras ämnesrelationer. Införandet av relativt index innebär att det under en ämnesterm finns

<sup>91</sup> Chan, Lois Mai, (1994), s. 275-277.

<sup>92</sup> Ibid., s. 278-80.

hänvisningar till klassymbolerna för de olika aspekter och discipliner som ämnet kan studeras utifrån, det vill säga ämnets domäner.<sup>93</sup> Vad Dewey åstadkom var en metod för upprättandet av en ämneskatalog där hylluppställningen av böcker överensstämmer med ordningen i denna.<sup>94</sup> Det kan även tilläggas att bara tanken på ett universellt klassifikationssystem på Deweys tid måste ha betraktats som banbrytande. Dåtidens bibliotekarier skapade vanligtvis sina egna klassifikationssystem efter lokala preferenser eller anpassade det till det fysiska rum där böckerna skulle stå uppställda.<sup>95</sup>

Nya upplagor ges ut så snart den tidigare sålt slut eller revideringar ansågs berättigade.<sup>96</sup> Den upplaga som nu föreligger är den tjugoförsta i ordningen.<sup>97</sup> Sedan 1894 ges dessutom en förkortad version av DDK ut och det är den trettonde upplagan av denna version som finns tillgänglig för närvarande.<sup>98</sup> Det finns nu även en elektronisk version av det kompletta systemet, *Dewey for Windows*.<sup>99</sup> Idag används DDK i uppskattningsvis 135 länder och finns i flera olika översättningar, såväl det kompletta systemet som den förkortade versionen.<sup>100</sup> Någondera av de senaste versionerna är översatt eller håller på att översättas till arabiska, franska, grekiska, hebreiska, italienska, persiska, ryska, spanska samt turkiska.<sup>101</sup>

Från att ha varit ett utpräglat enumerativt system är DDK idag i allt större omfattning ett syntetiskt klassifikationsystem. Ett flertal klassymboler finns ej som enskilda klasser utan hämtas från särskilda tilläggstabeller. Till den tjugonde upplagan hör sju huvudsakliga tilläggstabeller. Den första tabellen omfattar tilläggsnummer för standardindelningar och kan användas varhelst det är möjligt inom klassifikationsystemet. Tabell två omfattar indelningar enligt geografi, historisk period och personer. Denna typ av indelning får enbart användas när det uttryckligen står så. Detta gäller även för tilläggstabellerna fem, sex och sju. Med dessa tabeller kan indelningar göras enligt etnisk, nationell tillhörighet, språk samt grupper av personer.

Trots att DDK är ett väl beprövat klassifikationssystem med gamla anor krävs regelbunden uppdatering och revidering för att det inte skall bli otidsenligt. Organisationen som står bakom administrationen av DDK är välutvecklad, vilket borgar för att systemet skall vara någorlunda aktuellt. För närvarande publiceras nya utgåvor av DDK i intervall på sju till tio år. Sedan det nittonde DDK sker dessutom en fortlöpande uppdatering och revidering mellan de olika utgåvorna. Klassifikationsystemet ses över, befintlig notation och indextermer granskas och nya ämnen tillfogas. Med denna service kan systemet hållas

---

<sup>93</sup> Ibid., s. 269f.

<sup>94</sup> Comaromi, John, (1976a), s. 18.

<sup>95</sup> Batty, David, (1976), *Library classification: one hundred years after Dewey // Major classification systems: the Dewey centennial* / ed. by Kathryn Luther Henderson. - Urbana-Champaign: Univ. Of Illinois, Forest Press. - s. 2f. ; Comaromi, John, (1976a), s. 18.

<sup>96</sup> Comaromi, John, (1976a), s. 24.

<sup>97</sup> Dewey Decimal System Home Page [URL: <http://www.oclc.org/fp/>] 98-04-27.

<sup>98</sup> Chan, Lois Mai, (1994), s. 272.

<sup>99</sup> OCLC / About DDC / Expanded intro [URL: <http://www.oclc.org/oclc/fp/about/expand.htm>] 98-04-27.

<sup>100</sup> Chan, Lois Mai, (1994), s. 269.

<sup>101</sup> DDC translations [URL: [http://www.oclc.org/oclc/fp/about/trans\\_frameset.htm](http://www.oclc.org/oclc/fp/about/trans_frameset.htm)] 98-04-27.

någorlunda aktuellt.<sup>102</sup> Den administrativa servicen bakom klassifikationsystem är en viktig aspekt. Ett klassifikationsystem som inte uppdateras eller revideras blir snart otidsenligt och därmed mindre användbart.

De vanligast förekommande revideringarna av DDK som görs omfattar expansion, det vill säga införandet av nya ämnen och mer precisa underavdelningar i befintliga klasser. Till de mer ovanliga företeelserna hör reduktion, det vill säga motsatsen till expansion, som eliminerar sällan nyttjade underavdelningar. Eventuella underklasser som kan tänkas existera flyttas högre upp i det hierarkiska systemet för att inordnas under en mer generell klass. Relokalisering är ytterligare en form av revidering som återkommer i var upplaga av DDK. Genom ändringar i notationen flyttas ämnen inom klassifikationsschemat. Syftet är bland annat att undvika överlappande ämnesavdelningar. Innan den tjugonde upplagan omfattades till exempel ämnet avfallsteknik dels av klasssymbolen *604.6 Avfallsteknik*, dels av *628.4 Offentlig sanitet*. I den tjugonde upplagan återfinns avfallsteknik under *628.4 Avfallsteknik, offentliga toaletter, gatusopning*.

#### **6.4.1.1 För- och nackdelar med DDK**

Det finns en rad för- och nackdelar med att använda DDK i en bibliotekskontext. Det är viktigt att skilja på för- och nackdelar vid klassificering av dels fysiska och dels digitala objekt. Frågorna om DDK:s användning på webben kommer att behandlas i ett avsnitt längre fram.

Lois Mai Chan, professor vid institutionen för biblioteks- och informationsvetenskap vid universitetet i Kentucky, visar på några styrkor och svagheter hos DDK. Något som talar till DDK:s fördel är att det är ett praktiskt system som är etablerat och inarbetat sedan åtminstone etthundratjugo år tillbaka.<sup>103</sup> Chan får medhåll av Christopher Turner som skriver att bara det faktum att DDK finns tillgängligt, redo att snabbt tas i bruk är en väsentlig fördel. Dessutom är många biblioteksanvändare samt flertalet biblioteks- och informationsvetare bekanta med DDK:s notation. Till dags dato är det ju dessutom det mest använda och utbredda klassifikationssystemet i världen. Det finns med andra ord en utbredd kunskap om och förtrogenhet med systemet som talar för DDK.

Att DDK bestått tidens tand anser Chan vittna om just dess praktiska betydelse och värde.<sup>104</sup> Det är också praktiskt att hålla kvar vid ett system som använts länge och som fyller sin funktion. Med åtanke på de kostnader, omställningar och arbetsinsatser som skulle föreligga med att ersätta DDK med ett annat klassifikationssystem är det kanske inte förvånande att DDK som ett fullt funktiondugligt klassifikationssystem fortfarande begagnas.

---

<sup>102</sup> Chan, Lois Mai, (1994), s. 272-275.

<sup>103</sup> Ibid., s. 280.

<sup>104</sup> Ibid., s. 280.

Flera av de fördelar som Chan pekar på rör notationen och dess decimala system. Just bruket av arabiska numeriska tecken i notationen är en av DDK:s förtjänster, då dessa tecken är allmänt kända i hela världen, påstår Chan. På så vis anser Chan att alla människor, oberoende av språklig eller kulturell tillhörighet, med lätthet kan omfatta systemet. Huruvida dessa tecken verkligen är universellt kända och nyttjade ger hon dock inga belägg för. Likaså anser hon att den numeriska sekvensen i ett decimalt system är självklar, vilket underlättar organisationen av information. En annan betydelsefull aspekt med notationen är att den tydliggör den hierarkiska strukturen och tjänar som stöd för att upprätthålla förhållandet mellan och inom klasserna. Notationens hierarkiska uppdelning kan dessutom underlätta vid sökning i en OPAC eller annan online-databas. Genom att ta bort eller lägga till en siffra kan användaren antingen göra en sökning bredare eller smalare. Notapparaten främjar dessutom användares sökningar med sin mnemoniska, det vill säga minnesunderstödjande karaktär. Med framtiden i fokus garanterar ett decimalt system dessutom obegränsad utvidgning och indelning av klasser. Med kontinuerliga revideringar och nya publikationer kan klassifikationssystemets aktualitet upprätthållas.<sup>105</sup>

Chan kritiserar även DDK och pekar till exempel på den tydliga och uppenbara anglo-amerikanska dominans som råder. Hon nämner att denna partiskhet blir synnerligen uppenbar i huvudklasserna *800 Litteratur* och *900 Geografi och historia*, men framförallt i huvudklassen som omfattar religion, *200 Religion*. I den sistnämnda pekas på den tydliga snedvridning som förekommer med en markant betoning på protestantism i Amerika.<sup>106</sup> Chan är inte ensam i sin kritik på denna punkt. Jean Aitchison uppmärksammar även detta problem i sin genomgång av DDK och skriver att en förhärskande västerländsk dominans, utifrån ett anglo-amerikanskt perspektiv, återspeglas i DDK. En del huvudklasser uppfattar Aitchison rentav vara otidsenliga. Precis som sin kollega finner Aitchison huvudklass *200 Religion* vara förlegad, men ger även *700 Konst* som typexempel.<sup>107</sup> Utav tio klasser finner författarna alltså tydlig snedvridning i representationen av ämnen till västerländsk fördel, nämligen *200 Religion*, *700 Konst*, *800 Litteratur* och *900 Geografi och historia*.

Ett annat föremål för kritik är att relaterade klasser i flera fall är skiljda åt. Så är fallet med huvudklasserna *300 Samhällsvetenskap* som skiljs från *900 Geografi* samt *400 Språk* som skiljs från *800 Litteratur* påpekar Chan. Dessutom har ämnesindelningen av huvudklasserna ifrågasatts, med till exempel biblioteksvetenskap under *000 Allmänt*, psykologi under *100 Filosofi* och idrott och underhållning under *700 Konst*. Indelningen speglar dåtidens akademiska discipliner. Ämnesindelningen förvrängs dessutom av det faktum att antalet huvudklasser begränsas till enbart tio stycken. Klasserna blir därmed inte tillräckligt uttömmande. Att DDK i huvudsak är ett enumerativt system bidrar till att begränsa antalet ämneskombinationer som kan tas upp samt inskränker möjligheten att inkorporera tämligen enkla ämneskategorier i systemet. I en bibliotekskontext är det lätt att korrigera detta genom att anpassa hylluppställningen.

---

<sup>105</sup> Ibid., s. 280f.

<sup>106</sup> Ibid., s. 281.

<sup>107</sup> Aitchison, Jean, (1992), *Indexing languages and indexing // Handbook of special librarianship and information work* / ed. Patti Dossett. - 6<sup>th</sup> ed. - London: Aslib. - s. 191-233.

Samtidigt som ett decimalt notationssystem anses erbjuda möjlighet till oändlig utvidgning och indelning av klasser, sätter det gränser. Som påpekades ovan kan antalet likvärdiga klasser, till exempel huvudklasserna, aldrig bli fler än tio till antalet. Skulle det finnas behov av att expandera klassifikationssystemet med en ny koordinerad ämneskategori, till exempel mellan 610 och 620, finns det inte utrymme för denna operation. Istället får det nya ämnet inhysas under en befintlig ämneskategori, alltså bli en underkategori till denna.<sup>108</sup> Detta problem blir påtagligt vid betraktande av den varierande utvecklingen av huvudklasserna, då begränsningen gett upphov till att ämnesstrukturen blivit ojämn i huvudklasser som *300 Samhällsvetenskap*, *500 Naturvetenskap* och *600 Teknologi*. Decimalbaserade klassymboler har även nackdelen att de blir väldigt långa för specifika ämnen då facetter begagnas. I bibliotekssammanhang uppmärksammas detta problem framförallt vid hylluppställning.<sup>109</sup>

#### 6.4.1.2 DDK på webben - BUBL LINK

BUBL är en förkortning för *BULLETIN BOARD FOR LIBRARIES*. Åtminstone var det så vid tjänstens tillkomst år 1990 men numera är den dock mest känd under namnet *BUBL INFORMATION SERVICE* eller i korthet bara *BUBL*.<sup>110</sup> Tjänstens har sitt ursprung i ett projekt vars mål var att fungera som en anslagstavla (eng: bulletin board) för forskningsbibliotekarier. Inom kort blev det dock uppenbart att den information som fanns tillgänglig via BUBL även var intressant för användare inom andra akademiska områden. BUBL beskrivs som en nationell informationstjänst med inriktning mot akademisk utbildning och forskning eller som det uttrycks i BUBL:s målsättning, "[t]o provide value-added access to Internet resources and services of academic, research, and professional significance to the UK Higher Education community".<sup>111</sup>

Målsättningen skall förverkligas genom att erbjuda användare olika medel för intellektuell åtkomst till utvalda högkvalitativa digitala objekt som anses vara akademiskt relevanta. Mellan åren 1993 fram till mars 1997 tillhandahöll BUBL enbart *en* ämneshierarki för att möjliggöra just ämnesvis åtkomst av webbobjekt men denna tjänst ansågs aldrig som fullt tillräcklig i sig själv. För att förbättra och utveckla möjligheterna till ämnesvis åtkomst till digitala objekt på webben initierade BUBL projektet *LIBRARIES OF NETWORKED KNOWLEDGE*, förkortat LINK. Hur många digitala objekt som klassificerats och som söktjänsten omfattar framgår dock inte.

Alla digitala objekt som är tillgängliga via BUBL LINK är i grunden katalogiserade och organiserade enligt DDK men det finns olika hjälpmedel för intellektuell åtkomst av dessa. Idag tillhandahåller BUBL LINK en indexbaserad söktjänst, två ämneshierarkiska

---

<sup>108</sup> Chan, Lois Mai, (1994), s. 281f.

<sup>109</sup> Ibid., s. 281f. ; Turner, Christopher, (1987). *Organizing information: principles and practice*. - London: Bingley. - s. 76ff ; Aitchison, Jean, (1992), 203f.

<sup>110</sup> BUBL Admin: Frequently Asked Questions about BUBL [URL: <http://bubl.ac.uk/admin/faq.htm>] 98-04-07.

<sup>111</sup> BUBL Admin: BUBL Funding and Purpose [URL: <http://www.bubl.ac.uk/admin/purpose.htm>] 98-04-09.

söktjänster samt ett alfabetiskt ämnesordlista.<sup>112</sup> De båda ämneshierarkiska söktjänsterna skiljer sig åt genom att de är organiserade på olika sätt. Den senast tillkomna tjänsten är *BUBL 5:15*. Dess ämneshierarkiska struktur är i sig ett icke-etablerat klassifikationsystem vars struktur grundar sig på en kombination av termer från olika ämnesordlistor (eng: subject headings) och klassifikationsystem.<sup>113</sup> Den andra söktjänstens ämneshierarki är organiserad enligt DDK och det är den som kommer att beskrivas närmare. Intressant att notera är att denna ämneshierarki tidigare var organiserad enligt UDK. Varför de övergått till DDK finns det ingen dokumentation om.

I bilaga 2.1 visas hur söktjänstens inledande sida, med huvudklasserna från DDK, är upplagd. Bredvid BUBL:s symbol står rubriken BUBL LINK, följt av ett antal länkar. Härifrån kan användaren via BUBL:s symbol förflytta sig till BUBL:s hemsida men framför allt orientera sig med hjälp av länkar mellan BUBL LINK:s utbud av sök-, browsing- och hjälpmöjligheter. Denna meny finns alltid tillgänglig för användaren vid browsing i ämneshierarkin. Så oberoende av var i hierarkin man befinner sig kan man förflytta sig till någon av de andra sök- och browsing metoderna. Ytterligare en meny med länkar till BUBL:s olika avdelningar finns alltid tillgänglig nederst på sidan.

Under den översta menyn finns rubriken *Browse LINK by DDC* följt av huvudklasserna från DDK, med såväl klasssymbol som klassnamn utsatt. Efter varje huvudklass finns dessutom en kortfattad beskrivning om vad för typ av innehåll de omfattar. Med huvudklasserna som utgångspunkt kan användaren förflytta sig djupare ned i ämneshierarkin. Vid browsing i klassifikationssystemet indikerar rubriken, alldeles under den översta menyn, den aktuella ingången. I bilaga 2.2 illustreras detta med rubriken *551.5 Meteorology: national centres*, där användaren utifrån klasssymbol och klassnamn kan avgöra var i hierarkin han eller hon befinner sig och vad det är för ämne som avses. Även på den andra nivån i ämneshierarkin förekommer kortfattade beskrivningar av innehållet för en del klasser, som ett komplement till DDK:s klassnamn. Så är fallet för klasser under *000 Generalities*, *300 Social sciences* samt *600 Technology (applied sciences)*. För klasser underordnade övriga huvudklasser finns dock inte någon kommentar om innehållet mer än de ämnesord som klassifikationssystemet ger. Måhända anses klassnamnen i dessa fall ge tillräckligt med information om objektet.

BUBL LINK:s ämneshierarki överensstämmer i hög grad med DDK:s struktur. Avvikelser förekommer dock, till exempel klass *351 Model League of Arab States* som illustreras i bilaga 2.3. *Model League of Arab States* är en organisation som utgör en alldeles egen klass. Observera att det finns ytterligare klasser med samma klasssymbol, klasserna *351 Public administration: general resources* och *351 Governments worldwide*. Enligt den tjugoförsta upplagan av DDK är *351 Public administration* den korrekta klassen.<sup>114</sup>

---

<sup>112</sup> BUBL LINK: about LINK [URL: <http://link.bubl.ac.uk:80/service.htm>] 98-04-09 ; DDK beskrivs som "the underlying filing system" vid About BUBL 5:15 [URL: <http://link.bubl.ac.uk:80/about515.htm>] 98-04-16.

<sup>113</sup> About BUBL 5:15 [URL: <http://link.bubl.ac.uk:80/about515.htm>] 98-04-16.

<sup>114</sup> OCLC / About the DDC: DDC 21 Summaries (level 3) [URL: <http://www.oclc.org/oclc/fp/about/ddc21sm3.htm#350>] 98-04-27.

För att fortsätta exemplifiera med bilaga 2.3 så påvisas inte relationerna mellan klasser i ämnehierarkin mer än med hjälp av klassymboler. En layout med till exempel indrag av underordnade klasser i förhållande till de hierarkiskt överordnade klasserna är att föredra, då detta skulle förtydliga de hierarkiska relationerna. Den typografiska utformningen är inte alltid konsekvent mellan klasserna. Samtliga huvudklasser är skrivna med fetstil text, vilket visas i bilaga 2.1, men på den andra nivån i ämnehierarkin är typografin inte lika konsekvent. Klasser på andra nivån som är underordnade *000 Generalities*, *300 Social sciences*, *600 Technology (applied sciences)* och *700 The arts* är skrivna med fetstil men klasser på motsvarande nivå under övriga huvudklasser är skrivna i normal stil.

Hur posterna kan presenteras visas i bilaga 2.2. Just i detta exempel visas titlarna på samtliga digitala objekt inom klassen *551.5 Meteorology: national centres* överst på sidan. Genom att klicka på symbolerna intill titlarna förflyttar man sig till posten över respektive klassificerat digitalt objekt något längre ned på samma sida. Denna form av presentation är inte speciellt vanlig inom BUBL LINK utan vanligtvis presenteras posterna rakt upp och ner, en efter en.

Posterna inleds med titeln på det klassificerade digitala objektet, vilken även fungerar som länk till detsamma. Den första posten i bilaga 2.2 bär exempelvis titeln *Atmospheric Research and Information Center*. Till varje post hör en kort beskrivning om det klassificerade digitala objektets innehåll. Under fältet *Author* finns uppgifter om upphovsman registrerat, vilket kan vara såväl organisationer, institutioner eller personer. Härpå följer så det klassnummer som tillskrivits det digitala objektet. Det finns även en kommentar om vad det är för typ av digitalt objekt. Den sista upplysningen som ges är om var det digitala objektet är beläget. För det första exemplet i bilaga 2.2 så är det frågan om Storbritannien. För att återigen knyta an till klass *351 Model League of Arab States* som nämnts ovan så presenteras posterna i denna klass enbart med titel, vilket därmed avviker från övriga poster. Orsaken till detta kan bero på att samtliga klassificerade digitala objekt under denna klass är lokaliserade på samma server. Det finns enbart en kort, sammanfattande beskrivning om organisationen överst på sidan. Generellt samsas dock inte klasser och poster på samma uppslag. Poster redovisas vanligtvis på en egen sida, men även här finns det undantag. I klassen *616.8 Diseases of nervous system and mental disorders* uppenbarar sig plötsligt en post följt av en rad underordnade klasser.

Ämnehierarkin är generellt grund, ofta används endast två, tre nivåer och i vissa fall fyra. Att till exempel klasserna *551.552 Hurricanes*, *796.358 Cricket* och *004.678 The Internet* finns i ämnehierarkin antyder att djupet skulle vara större men ibland så hoppas vissa klassnivåer över och för att använda exemplet ovan så återfinns *551.552 Hurricanes* under *551.5 Meteorology*. Under *796 Sport and outdoor activities* finns underordnade klasser på så skilda nivåer som *796 Other sports*, *796.48 Olympic Games* och *796.358 Cricket*. Då tomma kategorier inte finns med i ämnehierarkin är djupet på ämnehierarkin grund, grundare än vad befintliga klasser ibland antyder.

Det är dock värt att notera att det klassnummer som är tillskrivet det digitala objektet nödvändigtvis inte behöver överensstämma med den klass i ämnehierarkin vari det placerats. Klassificeringen av ett digitalt objekt är ibland mer detaljerad än den klass vari objektet befinner sig. Det finns även exempel där man gjort tvärtom, exempelvis under 551.48 *Hydrology* finns poster som är klassificerade 551.4. Klassifikationen av objekten är i detta fall mer generell än den klass som de återfinns i eller klassen är mer specifik än klassifikationen av det digitala objektet. Liknande felaktigheter finns bland annat under 551.552 *Hurricanes*.

Trots att dubbelklassificering lämpar sig väl i hypertextsystem förekommer det inte i BUBL LINK. Under vissa klasser ges istället hänvisningar till angränsande klasser. Till exempel under klassen 150 *Psychology* så hänvisas användaren till angränsande ämnesområden med hänvisningar av typen *See also: artificial intelligence, psychiatry and clinical psychology, social psychology, physiology and neuroscience*. Genom att klicka på någon av de uppräknade ämnesområdena förflyttas användaren dit.

En del poster hänvisar till söktjänsten Yahoo!s ämnehierarki. Vi undrar om inte BUBL LINK anser sig kunna konkurrera med Yahoo!, i synnerhet som det är professionella bibliotekarier och informationsspecialister som står bakom tjänsten? Till exempel samtliga sex poster under klassen 027.7 *University and college libraries worldwide* hänvisar till Yahoo!. Exempel på andra klasser med poster som hänvisar till Yahoo! är 296 *Judaism*, 820 *English and old English literature: general resources* och 616.89 *Psychiatry: general resources*. Gemensamt för dessa klasser är att de ofta hänvisar till digitala objekt vars innehåll fungerar som index över länkar.

Vid browsing i ämnehierarkin kan användaren följa huvudklasserna djupare ned i hierarkin. Men browsing försvåras av det faktum att det inte finns några länkar som leder uppåt i ämnehierarkin. För att återgå till huvudklasserna är användaren utlämnad till webbläsarens bakåt-knapp. Förbättrad browsability skulle uppnås med länkar till överordnade klasser.

## 6.4.2 UDK

Den första utgåvan av Universal Decimal Klassifikation (UDK) utvecklades av Paul Otlet och Henri La Fontaine och publicerades 1905.<sup>115</sup> Detta nya klassifikationssystem byggde till stor del på DDK, som dock byggdes ut och utvecklades på ett sätt som kommer att beskrivas närmare. Från 1905 till 1991 ägdes och administrerades UDK av en organisation

---

<sup>115</sup> Marcella, Rita och Newton, Robert, (1994), *A new manual of classification*. - Hampshire: Gower Publishing. - s. 89-99.

som på engelska heter International Federation for Documentation (FID). Det var även FID som ansvarade för de ständigt pågående uppdateringarna. Sedan 1991 ägs UDK av UDK-konsortiet som består av FID och fem nationella standardiseringsorganisationer från bland annat Storbritannien och Japan. För de UDK-versioner som publicerats nationellt ansvarar dock inte FID för utan detta ansvar ligger på nationella organisationer.

UDK finns i 22 olika språkversioner. Versionerna är antingen fullständiga, medium, vilket innebär att de består av 30% av det ursprungliga systemet, eller förkortade, vilket motsvarar cirka 10% av det. Systemet har blivit mest populärt i Europa och Ryssland.<sup>116</sup> UDK är ett universellt klassifikationssystem som till skillnad från DDK är mer inriktat på att fungera för ämnesvis organisation och återvinning av information än som hylluppställningssystem. Det är detaljerat nog för att kunna användas för klassificering av inte bara böcker utan även pamfletter, rapporter och periodika. Systemet används främst på specialbibliotek, dokumentationscentraler och informationsbyråer. Det används även vid indexerings- och abstract-tjänster inom ämnena naturvetenskap och teknik. Trots att det är menat som ett universellt system som omfattar all mänsklig kunskap har det inte blivit särskilt flitigt använt inom folk- och forskningsbiblioteken.<sup>117</sup>

UDK utvecklades som vi tidigare nämnt med DDK som förebild. På så sätt kunde DDK:s notation och huvudklasser användas. UDK har som DDK decimalbaserad notation, vilken påvisar på vilken nivå i ämnehierarkin som ämnet befinner sig på. Som framgår av följande lista över UDK:s huvudgrupper skiljer de sig inte nämnvärt från de som används i DDK:

- 0 Allmänt och blandat. Vetenskap. Information.
- 1 Filosofi. Psykologi.
- 2 Religion. Teologi.
- 3 Juridik. Förvaltning.
- 4 Vakant.
- 5 Matematik. Naturvetenskap.
- 6 Tillämpade vetenskaper. Medicin. Teknik.
- 7 Konst. Fritid. Sport.
- 8 Språk. Lingvistik. Litteratur.
- 9 Geografi. Biografi. Historia.<sup>118</sup>

Som framgår av uppställningen ovan så är huvudklass 4 ledig. Notationen bygger på det arabiska decimalsystemet, det vill säga det som vi använder i västvärlden, och man kan alltså se den hierarkiska ordningen utifrån hur högt nummer en viss post i systemet har. Vidare använder man punkter bland annat för att göra det lättare att läsa långa

---

<sup>116</sup> Ibid., 89-99.

<sup>117</sup> Ibid., 89-99.

<sup>118</sup> Universella decimalklassifikationen, (1977). -3., revid. utg. - Stockholm: Tekniska litteratursällskapet. - s. 10.

nummerserier. UDK har i sin officiella version cirka 60 000 huvudklasser och mer än 220 000 underklasser men det finns även utgåvor med färre grupper för olika behov.<sup>119</sup>

Den stora skillnaden mellan UDK och dess föregångare DDK är att UDK är ett i princip syntetiskt system. Systemet erbjuder stora möjligheter till detaljerad beskrivning och utvecklas även kontinuerligt mot än större facetteringsmöjligheter. Det finns flera olika typer av facetter i UDK. Bland annat kan man utgå från geografiska eller formmässiga aspekter och beskriva objekt utifrån till exempel språk, plats, tid, material och person.

#### **6.4.2.1 För- och nackdelar med UDK**

Precis som i kapitlet om DDK diskuteras här endast de för och nackdelar med UDK som gäller en bibliotekskontext. Användningen i en webbkontext behandlas i diskussionskapitlet.

Den stora fördelen med att använda UDK i stället för DDK är att systemet är mer syntetiskt. Det är utvecklat främst för informationsåtervinning och inte för hylluppställning. UDK är också ett mycket detaljerat system vilket bland annat gör att man kan uppnå hög detaljnivå vid ämnessökningar.

De olika klasserna är skapade av ämnesspecialister och därmed kan man bryta ut underavdelningar och använda dessa för klassificering av specialsamlingar. Det finns också stora möjligheter för den enskilde katalogisatören att få just sina behov tillfredställda, då UDK ger stora möjligheter till att uttrycka specifika ämnesrelationer och hierarkier genom vilken ordning de olika klassymbolerna placeras i. Detta kan dock vändas till en nackdel, eftersom standardisering och igenkännande inte gynnas. UDK är som sagt språkoberoende och kan därför användas som komplement till indexering för att visa på hierarkiska strukturer eller representera olika ämneskoncept. Att notationen består av siffor gör systemet dessutom lätt att hantera med datorer.

En nackdel med UDK är att systemet är byggt på en enumerativ grund, till skillnad från till exempel Ranganathans *Kolonklassifikation* som från grunden är helt syntetiskt. UDK har dock fördelen att ha en liknande notation som DDK och är därmed lättare att använda för den som redan känner till DDK-systemet.

Precis som föregångaren DDK ligger UDK:s tyngdpunkt vid västerländska förhållanden, något som är särskilt tydligt i avdelningen Religion.

#### **6.4.2.2 UDK på webben - NISS Directory of Networked Resources**

Förkortningen NISS betyder *National Information Services and Systems* och är en brittisk organisation med syfte att fungera som en central informationsinstans för utbildnings- och forskningssektorn i Storbritannien. NISS förser sin målgrupp med tillgång till olika slags

---

<sup>119</sup> Marcella, Rita och Newton, Robert, (1994), 89-99.

informationskällor både på webben och via andra nätverk. Bland sina många tjänster har NISS en webbaserad söktjänst kallade *Directory of Networked Resources*, vilken vi har valt att titta närmare på. NISS Directory organiserar länkar till webbtjänster inom många områden och enligt NISS består söktjänsten av information från både Storbritannien och andra länder. Söktjänsten innehåller länkar till både kommersiella tjänster och tjänster inom utbildningssektorn och dessutom finns alla NISS-sidor med. Förutom browsingmöjligheter så finns även en indexbaserad sökfunktion som medger sökning i både NISS Directory och på alla andra NISS-sidor. Förutom den indexbaserade och ämneshierarkiska söktjänsten finns även en lång alfabetisk ämnesordslista och en lista ordnad efter UDK. Den förstnämnda listan är alfabetiskt gruppindelad medan den andra listan består av en enda lång sida med UDK:s klasser ordnade enligt notationen.

Tjänsten syftar inte till att vara en söktjänst för hela webben, men den försöker täcka in alla ämnesområden. Huvudsidan har nio huvudkategorier som till största delen överensstämmer med UDK, vilket även visas i bilaga 3.1.

- 0 Generalities
- 1 Philosophy
- 2 Religion, Theology
- 3 Social Sciences, Law, Government
- 5 Mathematics and Natural Sciences
- 6 Applied Sciences, Medicine, Technology
- 7 The Arts, Recreation, Entertainment, Sport
- 8 Languages, Linguistics, Literature
- 9 Geography, Biography, History.<sup>120</sup>

Bredvid huvudklasserna, bortsett från religion och litteratur, finns länkar till andra söktjänster inom respektive ämnesområde. Exempel på dessa är *SOSIG* och *Biz/ed* i anslutning till klass 3 *Social Sciences, Law, Government* samt *ADAM* och *RUDI* i anslutning till 7 *The Arts, Recreation, Entertainment, Sport*.<sup>121</sup>

Går man vidare från förstasidan ner i ämneshierarkin hittar man på den andra nivån dels ett antal länkar på denna nivå och sedan många av UDK:s underordnade klasser som i sin tur leder ned till nivå tre. På den tredje nivån finns ytterligare länkar, underordnade klasser och så vidare. Som exempel på UDKs ämneshierarki använder vi huvudklass 7 varifrån man på nivå två finner följande nio underordnade klasser:

- 71 Physical planning
- 72 Architecture
- 73 Sculpture
- 74 Drawing

---

<sup>120</sup> NISS Directory of Networked Resources [URL: <http://www.niss.ac.uk/subject/>] 98-04-02.

<sup>121</sup> SOSIG [URL: <http://sosig.ac.uk/>] ; Biz/Ed [URL: <http://bized.ac.uk/>] ; ADAM [URL: <http://adam.ac.uk/>] ; RUDI [URL: <http://rudi.herts.ac.uk/>] 98-03-31.

- 75 Painting
- 76 Graphic arts
- 77 Photography
- 78 Music
- 79 Recreation

Söktjänsten NISS:s indelning överensstämmer till största delen med UDK:s kategorier i denna klass då den ser ut på följande vis:

- 71 Physical planning. Regional, town and country planning.  
Landscaping
- 72 Architecture
- 73 Plastic arts. Sculpture
- 74 Drawing. Design. Applied arts and crafts
- 75 Painting
- 76 Graphic arts. Graphics
- 77 Photography and similar processes
- 78 Music
- 79 Recreation. Entertainment. Games. Sport<sup>122</sup>

På nivå fyra finns under exempelkategorin *72 Architecture* 16 kortfattade poster till olika digitala objekt inom ämnet. Posterna innehåller dels en titel, till exempel *Cathedrals in Britain and Ireland*, och dels två hyperlänkar. Den ena hyperlänken heter *http* och går direkt till själva resursen och den andra heter *info* och leder till en sida med metadata om det aktuella objektet. Den beskrivning som finns under länken *info* liknar till stor del en post i en bibliografisk databas. Eftersom UDK används finns det anledning att tro att de som utför klassificeringen utgår från etablerade klassificeringsmetoder. Den metadata som finns tillgänglig är titel, eventuella övriga titlar, copyright, informationstyp, en abstractliknande beskrivning, upphov, URL, kontaktperson, UDK klass, Library of Congress klass, uppgift om vem som klassificerat det digitala objektet och datum när detta har skett. Exempel på hur posters utförliga metadata presenteras finns i bilaga 3.2.

Även om söktjänsten använder UDK så utnyttjar den inte klassifikationsystemet in i minsta detalj. För att ge exempel är det värt att beakta huvudklass *1 Philosophy*, som i NISS Directory innehåller följande underordnade klasser:

- 14 Philosophical systems
- 15 On to Psychology
- 16 Logic
- 17 Ethics

---

<sup>122</sup> NISS Directory of networked resources: UDC in Brief [URL: <http://www.niss.ac.uk/resource-description/udcbrief.html>] 98-04-28.

Trots att NISS Directory är organiserat utifrån UDK-klassifikation finns inte alltid alla UDK:s klasser representerade på alla nivåer. Att *alla* kategorier skulle finnas med är att kräva för mycket eftersom UDK är ett stort och omfattande system. Det är däremot en klar begränsning att man redan på nivå två inte har med alla underkategorier. Fler exempel på detta finns under klasserna 2 *Religion* och 8 *Languages, Linguistics, Literature*. Anledningen till att man har uteslutit vissa kategorier kan vara att man helt enkelt saknar objekt som kan passa in i dessa kategorier.

Vidare använder sig NISS inte av de möjligheter till facettering som finns i UDK. Dubbelklassificering används dock till viss del men i vilken omfattningen är svårt att säga bland annat på grund av att dokumentation även här är bristfällig. Som exempel kan dock nämnas med att det under 72 *Architecture* bland annat finns en länk till *The Shakespeare Birthplace Trust* som även återfinns under 821.111 *English literature*. Ytterligare ett exempel är *National Art Library* som finns under både 72 *Architecture* och 02 *Librarianship*. Detta samband återspeglas dock inte i ämneshierarkin utan man måste gå till beskrivningen av de enskilda objekten för att se dubbelklassificeringen.

NISS saknar information om hur många länkar tjänsten omfattar. Inte heller bredvid de olika huvudklasserna finns det någon notering om hur många länkar de innehåller. En granskning visar att många kategorier innehåller mindre än tio länkar samtidigt som vissa är förhållandevis omfattande med över hundra länkar inom samma kategori. Till exempel i klass 05 *Periodicals*, underordnade 0 *Generalities*, finns en lång lista över 98 olika elektroniska tidskrifter.

De olika klasserna inom NISS Directory har ett ganska varierande djup. I exemplet ovan med elektroniska tidskrifter är djupet enbart två nivåer. Inom till exempel huvudklass 5 *Science* finns dock hela fem undernivåer ända ned till exempelvis 518.110 *Accountancy systems: Application software*. Det är svårt att bedöma vad detta beror på eftersom tjänstens urvalsprinciper är höjda i dunkel. På en av NISS hjälpsidor står det att "[t]he Directory also contains hundreds of resources which are found nowhere else on NISS, which we think may be useful."<sup>123</sup> Utifrån detta korta uttalande skulle man kunna anta att urvalet sker efter personalens omdöme.

Det är förhållandevis svårt med browsing i NISS Directory och det finns flera anledningar till detta. Browsing mellan de olika klasserna sker med hjälp av grafiska pilar eller klick på hypertextlänkar på vanligt sätt. Pilarnas betydelse är dock inte tydligt utformade utan deras betydelse kan vara till förmån för tolkning, vilket visas i bilaga 3.3. Vidare har man konsekvent skrivit ut URL-adresser utan att göra dem till länkar. Användningen av hypertext är undermåligt genomförd, till exempel så saknas det hypertext kopplingar mellan posternas klassymboler och klassen med motsvarande klassymbol i ämneshierarkin. Dessa två sista aspekterna gör browsing svårare för användaren och strider mot tanken med hypertext.

---

<sup>123</sup> NISS Directory of Networked Resources: help [URL: [http://www.niss.ac.uk/help/subject\\_help.html](http://www.niss.ac.uk/help/subject_help.html)] 98-04-05.

Grafisk utformning kan tyckas vara mindre viktigt i detta sammanhang men flera studier visar att webbsidors utseende spelar roll för användningen. De flesta användare är till exempel inte benägna att bläddra på en sida och de vill inte heller ha för mycket information presenterad på en gång.<sup>124</sup> Ett delmål med en söktjänst är förmodligen att användarna ska ha lätt att hitta och att tjänsten ska ha hög browsability. Det finns hos NISS många sidor som är väldigt långa, i vissa fall upp över tio sidor med webbläsaren Netscape Navigators standardinställningarna, det vill säga typsnittet Times New Roman i 12 punkters storlek.<sup>125</sup> Varje sida innehåller också en stor logotyp som gör att man alltid måste bläddra för att komma åt sidans innehåll.

## 6.5 Icke-etablerade klassifikationssystem

Att försöka beskriva ad-hoc klassifikation utifrån generella principer förefaller i viss mån vara motsägelsefullt, vilket även Hjørland påpekar. Icke-etablerade klassifikationssystem, som är exempel på ad-hoc klassifikation, kategoriserar i enlighet med det givna tillfallets förutsättningar. Det finns med andra ord inga regler som kan generaliseras från det ena tillfället till det andra. Då det inte går att urskilja några klassifikationskonventioner ges istället i detta avsnitt exempel på olika icke-etablerade klassifikationssystem som används i två olika ämneshierarkiska söktjänster.

### 6.5.1 Punkt se

Punkt se är en kommersiell svensk söktjänst som startade under hösten 1996. Tjänsten har utvecklats av företaget Spray Network och administreras numera av medieföretaget Circus, som är en del av Spray Network.<sup>126</sup> Tjänsten innehåller tre olika system för kategorisering av länkar. Dels två ämneshierarkier som av Punkt se kallas för *blädderkataloger*, med huvudkategorierna *Branch*, *Intresse*, dels en förteckning över personliga så kallade hemsidor under kategorin *Person*. Förutom söktjänsten finns även ett antal kringtjänster med bland annat nyhetsbevakning, redaktionellt editerade guider och möjligheter att bygga egna webbsidor.

Kategorin *Branch* innehåller enligt Punkt se företag, organisationer och andra offentliga webbtjänster som ordnas efter formell branchtillhörighet. Kategorin *Intresse* är tänkt att fungera som en mindre formell indelning som utgår från användares intresse då de digitala objekten har klassificerats utifrån de intresseområden deras verksamhet kan tänkas beröra. De båda kategorierna är en blandning av länkar till både kommersiella och icke-kommersiella webbtjänster. Kategorin *Person* innehåller personliga hemsidor och är indelad

---

<sup>124</sup> Nielsen, Jacob, How Users Read on the Web [URL: <http://www.useit.com/alertbox/9710a.html>] 98-03-30. ; Nielsen, Jacob, Report From a 1994 Web Usability Study [URL: [http://www.useit.com/papers/1994\\_web\\_usability\\_report.html](http://www.useit.com/papers/1994_web_usability_report.html)] 98-03-30.

<sup>125</sup> Exempel på långa sidor: [URL: <http://www.niss.ac.uk/cgi-bin/GetUdc.pl?37>] 98-04-05 och [URL: <http://www.niss.ac.uk/subject2/new95udc.html>] 98-04-05.

<sup>126</sup> Spray Network [URL: <http://www.spraynetwork.com/>] 98-04-05. ; Circus [URL: <http://www.circus.net/>] 98-04-05.

alfabetiskt från a till ö. För att visa hur Punkt se ser ut finns en bild på ämneshierarkin i kategorin *Branch* som bilaga 4.1. I kategorin *Bransch* finns följande huvudklasser.

- Datorer, telekom & Internet
- Finansiell verksamhet
- Föreningar & organisationer
- Förvaltning & hälsovård
- Handel
- Kultur & nöje
- Media
- Resor & turism
- Tillverkning & industri
- Tjänster
- Utbildning & forskning

I kategorin *Intresse* finns följande huvudklasser:

- Fritid
- Hem
- Hälsa
- Information
- Internet
- Jobb
- Kultur & nöje
- Mat
- Pengar
- Politik
- Resande
- Spel
- Sport
- Stil
- Utbildning

Under huvudklasserna följer efter varje kategori antalet poster som dessa omfattar. Detta åskådliggörs till exempel på den andra nivån i ämneshierarkin, som innehåller följande kategorier:

- Bil- & båtuthyrning (25)
- Busstrafik (37)
- Båt (43)
- Flyg (47)
- Hotell & boende (529)
- Järnväg (7)
- Lokaltrafik (8)
- Resebyråer & arrangörer (205)
- Taxi (23)

- Turistattraktioner (171)
- Turistservice (173)
- Övrigt (49)

Siffrorna inom parentes visar alltså hur många länkar varje kategori innehåller och som synes är det stor spridning på mängden poster i de olika kategorierna. Anledningen till att man har stor variation i antalet poster i varje kategori är att man praktiserar ad-hoc klassifikation och alltså lägger till kategorier vartefter man anser att de behövs. Kategorin *Hotell & boende* innehåller i sin tur följande kategorier:

- Bed & Breakfast (40)
- Campingplatser (32)
- Hotell & Motell (197)
- Kursgård & konferens (183)
- Studios (4)
- Stugor (51)
- Övrigt (22)

I detta exempel är ämnesshierarkin relativt grund med endast två underordnade nivåer. Andra kategorier är djupare som till exempel *Intresse: Information: Nyheter: Nättidningar: Näringsliv & samhälle*, som är fem nivåer djup.

I likhet med flera av de andra söktjänsterna har även Punkt se vissa sidor som är mycket långa, i vissa fall upp till 20 sidor enligt Netscape Navigators standardkonfiguration. Detta är som vi tidigare nämnt inte speciellt bra vid browsing. Ett av många exempel på detta finns under kategorin *Bransch: Datorer, telekom & Internet: Försäljning av hård- & mjukvara: Företagsförsäljning* som innehåller hela 494 länkar, vilket motsvarar cirka 20 sidor. Förutom denna mindre lyckade aspekt stödjer Punkt se browsing på ett tillfredsställande sätt. Man vet hela tiden var i ämnesshierarkin man befinner sig och det är enkelt att förflytta sig med hjälp av browsing inom tjänsten.

Till varje klassificerat digitalt objekt hör en separat informationsida med uppgifter om titel, e-post, postadress, URL, datum för registrering samt vilken kategori det förekommer i. Denna sida nås via informationssymbolen framför titeln på det klassificerade objektet. Ett digitalt objekt kan förekomma inom flera kategorier utan att det framgår av informationssidan. Exempel på hur en informationsida kan se ut ges i bilaga 4.2.

Punkt se saknar detaljerade uppgifter om hur man har indexerat sina länkar men förmodligen använder man passivt förvärv, det vill säga man baserar till största delen sin input på att användarnas själva indexerar sina bidrag. Bidragen läggs dock in utan att de går igenom någon redaktionell kontrollfunktion. Man hävdar dock att "[f]ör att försäkra oss om att kategoriseringen är konsekvent och i enlighet med databasens tänkta struktur, förbehåller vi oss rätten att ändra kategorival som är felaktiga eller alltför långsökta."<sup>127</sup> Detta är något

<sup>127</sup> Om registrering i Punkt se [URL: <http://www.punkt.se/registrera/>] 98-04-13. ; Om redaktionell kontroll

som enligt vår undersökning inte praktiseras med någon större konsekvens eller ens överhuvudtaget. Av detta följer att tjänsten innehåller mängder av helt uppenbara felindexeringar. Exempel på dessa fel finns bland annat under kategorin *Bransch: Datorer, telekom&Internet: Internetresurser: Sökverktyg & index* där man finner *Dentistry Online HB*, ett företag i dentalbranchen, och Fenixgymnasiet i Vaggeryd. Under kategorin *Spel: På nätet* i ämneshierarkin *Intresse* hittar man bland annat företaget *Newton direct marketing agency* som sysslar med marknadsföring. Samma företag dyker även upp under ett stort antal olika kategorier där det helt uppenbarligen inte hör hemma, bland annat *Intresse: Utbildning: Val av: Vägledning och Bransch: Kultur & nöje: Bibliotek & arkiv: Allmänna bibliotek*. Detta är bara ett fåtal exempel på de mängder av felindexeringar som finns i Punkt.se och helt klart är att denna extremt låga kontroll av input gör att tjänsten som helhet håller låg kvalite. Att det finns så påtagligt mycket fel påverkar vår uppfattning av söktjänsten som helhet. Passivt förvärv av länkar kan vara föremål för diskussion även om detta inte är speciellt relevant för hur man har klassificerat. Varför innehåller till exempel ämneskategorin *Intresse: Internet: Lära sig om* mest företag som till exempel Klickdata AB, verksamt inom datorutbildning medan Skolverkets grundläggande guide till webben, *Ett smultronställe för Internetblåbär*, inte finns med. Detta är ett inputproblem.

Punkt.se använder sig av en extrem ad-hoc klassifikation och det finns många exempel på detta. Man har många fantasifulla kategorier som till exempel under ämneshierarkin *Intresse* i kategorin *Stil* finns till exempel *Gå ut & visa upp sig* och *Piffa upp sig*. Många kategorier har även en avdelning kallad *Övrigt* som innehåller allt möjligt. Många av dessa fantaifulla kategorier fungerar naturligtvis endast i sin speciella kontext. Till exempel kategorin *Internet* i ämneshierarkin under *Intresse* innehåller följande underkategorier:

- Etablera sig på (1961)
- Koppla upp sig mot (155)
- Lära sig om (402)
- Näthandel (483)
- Nöje på (883)
- Skaffa dator för (712)
- Söka på (92)
- Övrigt (237)

En stor brist med denna modell är att dessa kategoriers namn är domänspecifika och därmed endast fungerar i förhållande till sin överordnade kategori. Detta är inga problem att förstå om man använder ämneshierarkin. Om man däremot använder tjänstens sökfunktion är det minst sagt osannolikt att man söker på just frasen *lära sig om* när man till exempel letar efter en grundkurs i hur Internet fungerar.

Punkt.se har som sagt även en sökfunktion som medger enkel fritextsökning i de olika kategorierna och med begränsningar i tid. Dessutom kan användaren välja mellan de olika grupperna *Företag & organisationer*, *Privata webbtjänster*, *Personliga hemsidor* och

---

[URL: <http://www.punkt.se/om/hjalp/>] 98-04-13.

*Kategorier* för vilken information som skall visas. Man kan även matcha hela eller delar av ord eller söka i enbart den kategori där man befinner sig just nu. Den sistnämnda funktionen gör det lite lättare att hitta i de kategorier som innehåller flera hundra poster. Som tidigare påpekats gör de fantasifulla namnen på kategorierna att sökning med begränsning till kategori i de flesta fall förmodligen inte ger relevanta resultat.

Punkt se nämner i sin hjälptext att ”sökverktyget är uppbyggt utifrån användarens behov och erbjuder ett antal olika sökingångar beroende på vilken typ av information användaren letar efter och vilket perspektiv han eller hon vill anlägga.”<sup>128</sup> Detta tyder på att de har utgått från en användarcentrerad modell när tjänsten utvecklades.

Att diskutera dubbelklassificering är intressant i fallet Punkt se. Det ambitiösa företaget *Newton direct marketing agency* förekommer som sagt inom ett stort antal kategorier men detta kan inte kallas för dubbelklassificering utan är snarast felindexeringar som beror på bristande redaktionell kontroll av input.

Punkt se kallar sig själva för ”en definitiv informationstjänst för svenska Internet”, något som vi definitivt inte kan hålla med om.<sup>129</sup>

## 6.5.2 Yahoo! Sverige

Yahoo! är en av de äldsta söktjänsterna och även en av de mest använda på webben i dag. Förutom den ursprungliga internationella varianten finns det även nationella eller regionala anpassningar av Yahoo! som fokuserar på webbsidor från det egna täckningsområdet och erbjuder sökning på det aktuella språket. Söktjänsten Yahoo! Sverige är ett exempel på en dylik anpassning för svenska förhållanden och användare. Övriga europeiska länder med regionala anpassningar av Yahoo! är just nu Danmark, Frankrike, Norge, Storbritannien och Irland samt Tyskland. På världsbasis återfinns liknande tjänster i Australien och Nya Zeeland, Japan, Kanada och Korea. Dessutom finns en Yahoo! som täcker in Sydostasien. Till dessa kommer en rad *Yahoo! Metro*-tjänster som omfattar en rad storstadsmetropoler i USA.<sup>130</sup> Vad ordet Yahoo! egentligen betyder är oklart men en tolkning är att det är en förkortning av *Yet Another Hierarchical Officious Oracle*. Denna förklaring antyder att det är frågan om en hierarkisk organiserade tjänst.

Historien bakom utvecklingen av söktjänsten är inte särskilt lång. Det hela började 1994 som ett försök av de två studenterna David Filo och Jerry Yang vid Stanforduniversitet att organisera sina personliga länkar. Ganska snart var resultatet långa otympliga listor och allt mer tid togs i anspråk för att organisera Yahoo!. Samma år som dess tillkomst gjordes Yahoo! om till en allmänt tillgänglig söktjänst för Internet-användare.<sup>131</sup> Vid tillkomsten av Yahoo! utgick konstruktörerna från de fördelar som finns med hierarkiska strukturer i

---

<sup>128</sup> Punkt se hjälpsida [URL: <http://www.punkt.se/hjalp/>] 98-03-12.

<sup>129</sup> Vad är Punkt se? [<http://www.punkt.se/om/>] 98-03-12.

<sup>130</sup> Yahoo! [URL: <http://www.yahoo.com/>] 98-03-12.

<sup>131</sup> Yahoo!s historia [URL: <http://www.yahoo.ce/docs/pr/history.html>] 98-04-16.

hypertext miljö, för att förbättra browsing.<sup>132</sup> Yahoo! har växt i omfång sedan dess och betraktas idag som en av de största söktjänsterna tillgängliga på webben. I september 1997 ombesörjde Yahoo! dagligen ca 50 miljoner förfrågningar, så kallade *page views*. Den insats som Yahoo! gör har av somliga lite pretentiöst jämförts med Carl von Linnés prestation.<sup>133</sup> Företaget Yahoo! har dessutom växt till den grad att det idag är noterat vid börsen i New York.<sup>134</sup>

Yahoo! Sverige lanserades den 3 november 1997. Tjänsten påstås vara anpassad för svenska förhållande och framställer sig som ”ett navigationsverktyg som från ett svenskt perspektiv presenterar den mängd information som finns på Internet”.<sup>135</sup> Söktjänsten hänvisar enbart till digitala objekt med svenskspråkigt innehåll och målsättningen är att vara den mest heltäckande söktjänsten inom detta område. Längst ned på sidan under respektive kategori finns det dock en hänvisning till motsvarande kategorier i internationella Yahoo!.<sup>136</sup>

I inledningsskedet omfattade Yahoo! Sverige mer än 12 000 länkar.<sup>137</sup> Det är användare av Yahoo! Sverige som får ge förslag på digitala objekt som skall inkluderas i söktjänsten men det är redaktionen som i slutet avgör vilka objekt som hamnar i hierarkin. Objekten kategoriseras i fjorton huvudklasser, vilka motsvarar de huvudklasser som förekommer i internationella Yahoo!.<sup>138</sup> Som komplement till browsingsmöjligheterna i ämneshierarkin finns det en indexbaserad sökfunktion som möjliggör sökning i antingen den valda klassen eller i hela Yahoo! Sverige.<sup>139</sup> Ämneshierarkin för browsing är uppdelad i fjorton huvudklasser, se även bilaga 5.1.

- Datorer och Internet
- Hälsa och sjukvård
- Konst och kultur
- Naturvetenskap och teknik
- Nyheter och media [Extra!]
- Näringsliv och ekonomi
- Nöje
- Politik och förvaltning
- Referensverk
- Regionalt
- Samhälle och kulturer

---

<sup>132</sup> Dodd, David G., (1996), Grass-roots cataloging and classification: food for thought from World Wide Web subject-oriented hierarchical lists // *Library Resources and Technical Services*. - Vol. 40, nr 3, July 1996. - s. 276.

<sup>133</sup> Yahoo!s historia [URL: <http://www.yahoo.se/docs/pr/history.html>] 98-04-16.

<sup>134</sup> Yahoo! erövrar Sverige [URL: <http://www.yahoo.se/docs/pr/release1.html>] 98-02-05.

<sup>135</sup> Välkommen till Yahoo! Sverige [URL: <http://www.yahoo.se/docs/info/welcome.html>] 98-02-05.

<sup>136</sup> Ibid.

<sup>137</sup> Yahoo! erövrar Sverige [URL: <http://www.yahoo.se/docs/pr/release1.html>] 98-02-05 ; Now Scandinavia can Yahoo! too: Yahoo! introduces localized Internet guides in Denmark, Norway and Sweden [URL: <http://www.yahoo.com/docs/pr/release129.html>] 98-02-05.

<sup>138</sup> Yahoo! Sverige [URL: <http://www.yahoo.se/>] 98-02-05.

<sup>139</sup> Yahoo! Sverige nyheter [URL: <http://www.yahoo.se/nyheter/>] 98-04-16.

- Samhällsvetenskap och humaniora
- Sport och fritid
- Utbildning

Samtliga av dessa huvudklasser befinner sig på så kallad metanivå. Det innebär att klasserna av Yahoo! Sverige anses vara alltför generella för att kunna omfatta några objekt med detta ämnesinnehåll. Huvudklasserna innehåller därför enbart underordnade klasser. Det är alltså först på den andra nivån i den ämneshierarkiska strukturen som länkar till olika digitala objekt återfinns. Även på den andra nivån förekommer en klass som räknas som metanivå, nämligen klassen *Företag* som faller under huvudklassen *Näringsliv och ekonomi*.<sup>140</sup>

Enligt Srinija Srinivasan, ansvarig för ämneshierarkin i Yahoo!, är strävan att tillhandahålla så precisa och kortfattade kategorier som möjligt.<sup>141</sup> Klasserna skall utgöras av ord eller fraser som antas vara vanliga vid sökning eller browsing. En annan målsättning är att använda termer konsekvent för att underlätta användningen av ämneshierarkin. Till exempel så har kategorin *Index* i stort sett samma betydelse oberoende av i vilken kategori den förekommer, enligt Srinivasan.

I bilaga 5.2 ges ett exempel på hur det kan se ut vid browsing i Yahoo! Sverige. Överst på sidan finns en företagslogo. Intill på höger hand finns tre klickbara bilder. Bilden med ett frågetecken på leder till Yahoo! Sveriges Informationscentrum där användaren bland annat kan få hjälp med och svar på frågor om användning. Vill man ge förslag på ett digitalt objekt som skall inkluderas i Yahoo! Sverige är det bara att trycka på bilden bredvid med undertexten *Lägg till URL*. Användaren uppmanas att uppge titel, URL och skriva en kort beskrivning av innehållet i ett digitalt objekt. Det är främst med passivt förvärv som nya digitala objekt inkluderas i söktjänsten, men även aktivt automatiskt förvärv sker med hjälp av en sökmotor. Alla digitala objekt granskas dock av katalogisatörer innan de inkorporeras i söktjänsten. Bilden med Yahoo!-figuren leder, precis som det står under bilden, användaren till en rad länkar till andra Yahoo!-tjänster. Reklamen som infinner sig härunder bär vittnesbörd om att det är en reklamfinansierad tjänst.

För att skilja mellan vad som är klasser och vad som är poster vid browsing visas alla klasser i fetstilt medan posterna med länkar till digitala objekt är skrivna med normalstil.<sup>142</sup> För att fortsätta exemplifiera med bilaga 5.2 så visas nedanför reklamslaget alltid den aktuella ingången i ämneshierarkin samt dess överordnade klasser åtskiljda av kolontecken, i detta fall *Topp:Samhällsvetenskap och humaniora:Lingvistik och språk*. Därmed är det lätt för användaren att överblicka var i hierarkin han eller hon befinner sig. Genom att klicka på någondera av de överordnade klasserna kan användaren genom browsing snabbt förflytta sig mellan de olika nivåerna. Klassen *Topp* leder till huvudklasserna och *Samhällsvetenskap och humaniora* är i detta exempel den närmast överordnade klassen som användaren kan

<sup>140</sup> Yahoo!s reserverade kategorier [URL: <http://www.yahoo.se/docs/info/reserved.html>] 98-03-24.

<sup>141</sup> Dodd, David G., (1996), s. 276f.

<sup>142</sup> Yahoo! Hjälp [URL: <http://www.yahoo.se/docs/info/help.html>] 98-03-25.

förflytta sig till. Därefter följer den valda nivån, *Lingvistik och språk*. Denna tillämpning borgar för god browsability, vilket även konstruktörerna av Yahoo! var medvetna om vid skapandet av söktjänsten. Nedanför den indexbaserade sökfunktionen återfinns den första klassen, *Index*. Under denna kategori samlas objekt som har ett översiktligt innehåll eller en samling av länkar inom den berörda klassen. Liknande kategorier är återkommande under flera klasser i söktjänstens ämnesshierarki.

Siffrorna inom parenteser invid klasserna representerar antalet poster som finns under respektive klass samt eventuellt underordnade klasser. I dessa tal inkluderas dock inte dubbelklassificerade poster. I bilaga 5.2 så omfattar klassen *Index* i detta fall enbart en post. Övriga underordnade klasserna, *Datalogvistik*, *Institut*, *Ordböcker@*, *Organisationer*, *Språk*, *Teckenspråk*, *Tidskrifter*, *Tolkar och översättare* samt *Utbildning*, omfattar något fler poster. Klassen *Ordböcker@* är ett exempel på dubbelklassifikation, vilket indikeras av det avslutande tecknet @. Användaren hänvisas vid ingången *Ordböcker@* till hierarkin *Topp:Referensverk:Ordböcker:Språk*. Med detta förfarande kan användaren länkas till ämnen som är relaterade och därmed kan tänkas vara av intresse. Dubbelklassificering är dessutom synnerligen användbart då ett digitalt objekt behandlar flera ämnen och i Yahoo! Sverige är denna funktion väl omsatt. Klassnamnen överensstämmer dock inte alltid med de klasser som användaren hänvisas till. Ibland används synonymer, till exempel så hänvisar *Pacifism@* under *Topp:Samhällsvetenskap och humaniora:Freds- och konfliktstudier* till *Topp:Samhälle och kulturer:Icke-våld*. En annan svårighet är påvisandet av domän för ett ämne. För till exempel klassen *Storleksfrågor* under *Topp:Samhälle och kulturer* så framgår inte vilken domän som avses. Om användaren ser efter vad den nämnda klassen innehåller så finns enbart en hänvisning till klassen *Topp:Hälsa och sjukvård:Viktproblem*.

Nedanför de underordnade klasserna återfinns så de poster som faller under den aktuella klassen *Lingvistik och språk*. Precis som påpekades tidigare är posterna skrivna i normalstil för att skilja dem från klasser. Utifrån de uppgifter som användaren och Yahoo! Sveriges katalogisatörer lagt in består posten av en titel, som fungerar som länk till det avsedda digitala objektet, samt en kort en kort beskrivning av dess innehåll. Antalet poster i respektive klass, alltså ej inkluderat de som finns i underordnade klasser, varierar. Ofta är antalet poster mellan 1 och 10 stycken. I enstaka fall förekommer väldigt många poster i samma klass, exempelvis under *Topp:Näringsliv och ekonomi:Företag:Resor och turism:Resebyråer* finns 81 resebyråer förtecknade. Under *Topp:Samhälle och kulturer:Folk:Personliga hemsidor* finns hela ca 1500 poster. Finner användaren att det svenska ubudet inte förslår finns det möjlighet att ta del av internationella Yahoo!s utbud i motsvarande klass.

Ämnesshierarkin uppvisar i somliga fall logiska felaktigheter. Till exempel under klassen *Inte riktigt det bästa* som återfinns under *Topp:Datorer och Internet:Internet:World Wide Web:Det bästa på nätet*. Under kategorin *Det bästa på nätet* förväntar man sig förmodligen inte att finna klassen *Inte riktigt det bästa*. En del kategorier som kan förefalla annorlunda är till exempel *Topp:Datorer och Internet:Internet:Nöje:Intressanta saker sammankopplade med nätet* och *Topp:Nyheter och media:Bisarra nyheter* eller *Topp:Samhälle och kulturer:Storleksfrågor*.

Då Yahoo! Sverige använder ett icke-etablerat klassifikationsystem för sin ämnehierarki finns goda möjligheter att infoga klasser så fort behov uppstår. Till exempel så finns under *Datorer och Internet* klassen *År 2000-problemet*, vilket är ett nyligen uppmärksammat problem. Likaså finns det möjlighet att ta bort klasser som är tomma utan att det blir lika tydligt som vid användning av klassifikationsystem med klassymboler.

Emellanåt förekommer olika grafiska symboler invid klasser och poster. Symbolen i form av ett par solglasögon betyder att innehållet i en post bedömts som bra eller har en bra presentation. Denna symbol är alltså en form av betygssystem. För att indikera nya tillägg av klasser och poster finns det en symbol som heter *Nytt*. Om *Nytt*-symbolen följer en post är denna tillagd den senaste veckan. Följer samma symbol en klass så innebär det att posterna i denna katalog modifierats någon gång under de tre senaste dagarna.<sup>143</sup>

Att Yahoo! Sverige försöker ha ett så allmänt system som många användare kan förstå är sannolikt deras mål, i synnerhet som det är en kommersiell tjänst som är beroende av sina användare för att kunna ha några intäkter. Därför skall användarvänligheten i söktjänsten betraktas ett exempel på kommersialismens positiva inverknings. Omfattande dubbelklassificering och goda browsingsmöjligheter underlättar för användaren att orientera sig i ämnehierarkin.

## 6.6 Sammanfattning

Vi har i detta stycke beskrivit klassifikation utifrån Birger Hjørlands tre ambitionsnivåer för organisering av kunskap. Dessa är ad-hoc, pragmatisk och vetenskaplig klassifikation. Exempel på ad-hoc klassificering är de klassifikationssystem som används i Yahoo! Sverige och Punkt se och exempel på användning av pragmatiska system finns hos BUBL LINK och NISS Directory. Vi har diskuterat vad klassifikation är och konstaterat att huvudsyften är att på ett systematiskt sätt skapa ordning och påvisa relationer mellan ämnen. Klassifikation är ett bra sätt att underlätta åtkomst till objekt utifrån ämne. Vid ämnesanalys skall ett objekt inte bara analyseras utifrån sitt intellektuella innehåll utan även utifrån vilka frågor det kan besvara. Man skall även ta hänsyn till vilka användarna är. Ämnesrepresentation kan ske antingen med kontrollerat indexeringsspråk eller naturligt indexeringsspråk. Ett indexeringsspråk måste bland annat ta fasta på objekts kontext.

Vi har även behandlat klassifikationssystem och konstaterat att det finns olika typer av system som fyller olika syften. Gemensamt för alla system är att dess klasser ska ordnas så att det övergripande föregår det specifika vilket skapar en hierarki mellan olika klasser. När man skapar ett klassifikationssystem kan man bland annat utgå från vilken typ av objekt som ska ingå i systemet men även ta hänsyn till vilka ämnen som systemet ska omfatta. Det finns olika tillvägagångssätt när man skapar en hierarki, till exempel från klass till klassmedlem, från helhet till delar, spegling av en process förlopp, från mindre till ökad komplexitet, i kronologisk ordning eller enligt rymd och territorium. Vidare underlättas

---

<sup>143</sup> Yahoo! Specialare [URL: <http://www.yahoo.se/docs/info/features.html>] 98-04-16.

återvinning av att man använder dubbelklassificering, vilket innebär att ett objekt representeras i flera klasser. I klassifikationssystem används ibland förkortade ersättningar för att representera klasser, vilka kallas för klassymboler.

Man delar in klassifikationssystem huvudsakligen i enumerativa och syntetiska system. Enumerativa system karakteriseras av att varje ämne som kan klassificeras inom systemet är representerat av en klass. I ett syntetiskt system konstrueras klasser genom att klassymboler för olika klasser sammanfogas. Olika sammanhang kräver olika klassifikationssystem och de delas in i grupperna universella, nationellt anpassade och ämnesspecifika system. Ytterligare en gruppindelning av klassifikationssystem är i etablerade och icke-etablerade system.

Vi ger en historisk bakgrund till de etablerade klassifikationssystemen DDK och UDK och beskriver även några av de mer erkända för- och nackelar som finns med de båda systemen. Slutligen ger vi exempel på användningen av klassifikationssystem på webben genom en beskrivning av fyra söktjänster. Dessa är BUBL LINK, NISS Directory, Yahoo! Sverige och Punkt.se. Vi har beskrivit tjänsterna utifrån hur de har klassificerat sina objekt och hur deras ämnehierarkier ser ut. Resultatet visar att söktjänsterna är relativt olika både beroende på vilket klassifikationssystem de använder och på vilket sätt de har utformat sin tjänst.

## 7 Diskussion

Detta stycke syftar till att med utgångspunkt i de tidigare kapitlen diskutera, fördjupa och problematisera frågorna kring informationsåtervinning med hjälp av klassifikationssystem på webben. Dessutom redogör vi för mer självständiga tankar kring de ämnen som vi berört i tidigare kapitel.

### 7.1 Är webben ett informationssystem?

I den litteratur om informationsåtervinning och klassificering på webben som vi har tagit del av utgår författarna från att klassifikationssystemen ska användas i vad som kallas ett informationssystem. Det har hos oss väckt frågan om huruvida webben är att anses som ett sådant.

För att få svar på detta måste vi definiera vad ett informationssystem är. Birger Hjørland diskuterar i boken *Informationsvidenskabelige grundbegreber* frågan om vad ett informationssystem är.<sup>144</sup> Han påpekar att definitionen är avhängig av hur man definierar *information* vilket är komplicerat, omständigt och näst intill omöjligt. I boken *information seeking and subject representation: an activity-theoretical approach to information science* citerar han Bogdan som i sin tur skriver att information bland mycket kan anses vara ett kommunikationsmönster mellan en avsändare och en mottagare, ett slags kontroll och återkoppling av det samma, innehållet i ett meddelande som sänds via någon slags kanal, innehållet i ett kognitivt tillstånd, betydelsen av en språklig sats eller utplånandet av osäkerhet. Inom olika discipliner definieras dessutom information på olika sätt varför den ovanstående listan säkert kan göras ännu längre. Hjørland hänvisar även till Buckland som beskriver information som en term som kan användas på objekt, processer och kunskap. Information kan alltså vara kunskap, någonting som gör att man blir informerad, ett fysiskt eller digitalt objekt, data eller annan noterad kunskap samt en process som på något sätt innehåller behandling av data. Buckland menar att saker kan vara *informativa* samt att det är omöjligt att hitta någonting som inte är informativt och alltså är allting information. Hjørland påpekar även att inom informationsvetenskapen måste termen information definieras utifrån ett specifikt teoretiskt perspektiv. Ramarna för denna uppsats räcker dock inte till för att fördjupa diskussionen om hur man ska definiera information.<sup>145</sup>

Dagobert Soergel talar redan på sidan 3 i boken *Organizing information* om olika slags informationssystem. Några av de många exemplen på sådana är ett bibliotek, en databas innehållande lagtexter, en arbetsförmedling och ett företags lagerkataloger. Soergel anser att ett informationssystem ska kunna identifiera en sökares problem, ge användaren möjlighet att lösa problemet och leverera den eftersökta informationen.<sup>146</sup> Den komponent

---

<sup>144</sup> Hjørland, Birger, (1995), *Informationsvidenskabelige grundbegreber*: Bd 2: Informationssystem - Zoom. - København: Danmarks Biblioteksskole. - s. 193-194.

<sup>145</sup> Hjørland, Birger, (1997), s. 110.

<sup>146</sup> Soergel, Dagobert, (1985), *Organizing information: principles of data base and retrieval systems*. - New York: Academic Press. - s. 3.

i informationssystemet som benämns *Information Storage And Retrieval* (ISAR) syftar till att utifrån ett informationsbehov finna de objekt i systemet som kan täcka detta behov. Hur man bäst går tillväga för att återvinna objekten kallas för *Information Retrieval* (IR).

Diann Rusch-Feja nämner i sin text *Structuring Subject Information Sources in the Internet* att "widely dispersed information systems are being made available to the end user" och senare i texten där det handlar det om ämneshierarkier talar hon om "the benefits of linking distributed information systems using the relatively unstructured Internet".<sup>147</sup> Hon använder alltså termen informationssystem som synonymt med de olika söktjänster som erbjuder sökning och browsing.

Utifrån ovanstående definitioner anser vi inte att webben kan likställas med ett informationssystem. Webben i sig kan inte identifiera en användares problem och ej heller automatiskt leverera någon information, vilket däremot är möjligt för specifika webbtjänster. Yahoos ämneshierarki eller Alta Vistas indexbaserade söktjänst kan alltså anses utgöra separata informationssystem. Själva webben är snarare att betraktas som ett *distributionssystem*. Med termen distributionssystem menar vi ett sätt att överföra data. Andra sådana metoder är till exempel telefoni och telefaxöverföring. Om vi tar exemplet med telefonin så är telenätet ett distributionssystem och nummerupplysningens databas och telefonkatalogen är informationssystem som är kopplade till distributionssystemet.

Karen Coyle anser att webben är ett modernt och moget telekommunikationssystem med väl utvecklade tekniska standarder men är helt amatörmässigt som informationssystem för återvinning.<sup>148</sup> Som vi nämnt ovan anser vi inte att webben i sig är ett informationssystem, utan att dessa system snarare utgörs av vissa webbaserade tjänster. Coyle gör dock inte någon sådan åtskillnad.

Webben kritiserar ofta för att det är ett medium där det är svårt att hitta den information man söker. Användarna, bland annat de med biblioteks- och informationsvetenskaplig bakgrund, uppfattar av mer eller mindre medvetna skäl att webben är ett informationssystem. Eftersom webben genom sin struktur gör det svårt att hitta eftersökt information kritiserar den följaktligen och anses vara ett dåligt informationssystem. Webben är ett *öppet nätverkssystem för distribution* som inte kan jämföras med slutna informationssystem som till exempel en databas i ett bibliotek. Om man som vi i stället ser webben som ett distributionssystem så kan man konstatera att kritiken är grundad i en felaktig uppfattning om vad webben egentligen är. Därmed ställs även felaktiga krav på vad man skall kunna uppnå med nätverket. I stället är det de enskilda webbtjänsterna som bör betraktas som informationssystem och värderas som mer eller mindre bra sådana.

---

<sup>147</sup> Rusch-Feja, (1995), *Structuring subject information sources on the Internet // Wissen in elektronischen Netzwerken: strukturierung, erschliessung und retrieval von Informationsressourcen im Internet: eine Auswahl von Vorträgen der 19. Jahrestagung der Gesellschaft für Klassifikation Basel 1995 / red. Hans-Christoph Hobohm, Hans-Joachim Wätjen*. - Oldenburg: Bibliotheks- und Informationssystem. - s. 100. ; Ibid, s. 109.

<sup>148</sup> Coyle, Karen, (1995), Access: Not Just Wires // LIBRES, Library and Information Science Research Electronic Journal. - [URL: [gopher://gopher.fis.utoronto.ca:70/00/library/ejournal/libres/V5/N1/coyle.5n1](http://gopher://gopher.fis.utoronto.ca:70/00/library/ejournal/libres/V5/N1/coyle.5n1)] 98-03-18.

Vidare syftar inte till exempel en OPAC att skapa åtkomst till all världens information. Jämför man med bibliotekskataloger och andra databaser som endast syftar till att kontrollera en viss begränsad mängd objekt kan man inte ställa krav på att de webbaserade söktjänsterna ska täcka in all information på webben.

Kahn och Wilensky talar om en alternativ arkitektur för ett universellt öppet digitalt nätverkssystem där webben är en av flera möjliga applikationer. De talar alltså inte om ett informationssystem utan har en teknisk utgångspunkt där webben ses som en infrastruktur för distribution av information. Deras modell utgår från att det finns kontroll av de digitala objekten med hjälp av till exempel URN. Så länge URL används finns det ingen form av kontroll av de digitala objekten vilket bidrar till att webben inte kan kallas ett informationssystem. Tekniken sätter alltså för närvarande hinder för att webben skall fungera som ett informationssystem och frågan är om det överhuvudtaget kommer att bli ett informationssystem och blir webben i så fall ett informationssystem över distribuerade informationssystem?

Sammanfattningsvis anser vi att termen informationssystem inte bör appliceras på webben utan på de webbaserade söktjänsterna. Webben är snarare att anse som ett distributionssystem.

## **7.2 Informationssystem på webben**

Tanken på söktjänster som distribuerade informationssystem på webben har dykt under arbetets gång. Detta uppfattar vi så pass intressant att tanken är värd att utveckla inom ramen för uppsatsen. Syftet med detta stycke är att beskriva några av skillnaderna mellan traditionella informationssystem och webbaserade dito. Det är främst nackdelarna med de webbaserade systemen som blir uppenbara här, vilket beror på att många av de texter vi analyserat utgår från de etablerade systemen när de beskriver de icke-etablerade. Denna utgångspunkt är lätt att förstå eftersom de etablerade systemen är just etablerade och har haft många år på sig att utvecklas och det finns en inarbetad förtrogenhet bland användarna. Det finns även ett väl utvecklat teoribygge kring bibliografisk klassifikation. Det är därför naturligt att många jämförelser görs med utgångspunkt från etablerade system. Längre fram i detta avsnitt skall vi dessutom redogöra för några av skillnaderna mellan ämneshierarkiska och indexbaserade söktjänster.

### **7.2.1 Skillnader mellan webbaserade och traditionella informationssystem**

Vi kan konstatera att etablerade metoder för informationsåtervinning i flera fall inte går att applicera på webben. Detta beror på att de webbaserade informationssystemen och de digitala objekten påverkas av arkitekturen och de tekniska förutsättningarna hos webben, distributionsystemet. Söktjänsterna saknar flera av de mer traditionella bibliotekskatalogernas och onlinedatabasernas karaktäristika.

För att undersöka skillnaderna mellan informationssystem baserade på etablerade principer och de nyare webbaserade systemen kan man ta hjälp av den i katalogiseringsammanhang legendariske Charles A. Cutter. Cutter beskriver i sin bok *Rules for a Dictionary Catalog* ett antal grundkrav som katalogen, ett dåtida informationssystem, ska uppfylla:

- Katalogen ska göra det möjligt för en användare att hitta ett dokument om författare, titel eller ämne är känt.
- Användaren ska kunna se vad biblioteket har av en specifik författare, i ett visst ämne eller av en viss slags dokument.
- Katalogen ska hjälpa till i valet av andra dokument inom samma ämne eller av samma slags dokument.<sup>149</sup>

Eftersom detta skrevs 1876 talar man självklart om dokument, kataloger och bibliotek. Frågan är om man kan applicera Cutters krav på moderna verktyg för kunskapsorganisation och då även datorbaserade informationssystem som söktjänster på webben? Mandel och Wolven utgår i sin text, *Intellectual access to digital documents*, från att Cutters grundprinciper är användbara även för elektroniskt lagrad information.<sup>150</sup> Användaren ska även i detta sammanhang ha möjlighet att hitta ett objekt utifrån ämne eller titel och man ska även kunna skilja på olika versioner av samma objekt. Författarna menar vidare att man kan använda sig av de etablerade principer för kunskapsåtervinning som finns men att man behöver nya metoder för att nå upp till Cutters mål. Dessa metoder kan till exempel vara nya objektformat och protokoll samt en standardisering av metadata. De menar också att den omedelbara tillgången till digitala objekt som webben ger minskar behovet av traditionella katalogposter även om man gör avkall på den bibliografiska beskrivningen. Exempel på detta finns hos Yahoo! som inte använder fullständiga beskrivningar av objekten. Motiveringen till detta förfarande kan dels vara att objekt på webben ofta ändrar innehåll och dels på grund av att det är enkelt för användaren att klicka sig fram till originalkällan och kontrollera om innehållet är relevant. Enligt Mandel och Wolven kan man även med hjälp av automatisk indexering dra ut tillräckligt med metadata ur det befintliga innehållet från digitala objekt för att möjliggöra återvinning.<sup>151</sup> Vi ska dock inte fördjupa oss i automatisk indexering utan bara påpeka att det är viktigt att skilja mellan själva texten och dess intellektuella innehåll. Vid automatisk indexering är det vanligtvis enbart texten som ligger till grund för urvalet av indexeringstermer.

### 7.2.1.1 Digitala objektens konstitution

---

<sup>149</sup> Mandel, Carol A. och Wolven, Robert, (1996), *Intellectual access to digital documents: joining proven principles with new technologies* // *Cataloging & Classification Quarterly*. - New York: Haworth P. - Vol. 22, nr 34. - s. 29-31.

<sup>150</sup> Ibid., s. 25-42.

<sup>151</sup> Ibid., s. 25-42.

I ett traditionellt informationssystem kan användaren vara förhållandevis säker på att ett objekt som funnits där tidigare finns kvar vid ett senare tillfälle. Ofta finns det en central instans som lägger till nya objekt och dessutom förs en tillräcklig mängd metadata in om varje nytt objekt för att möjliggöra återvinning. Så är inte fallet på webben. På webben finns inte samma kontroll över vilka objekt som läggs till eller tas bort. Därför förekommer både i indexbaserade söktjänster, där indexeringen sker automatiskt och i ämnesshierarkiska söktjänster där indexeringen sker manuellt att länkar är ”döda”. Det vill säga objektet, som länken refererar till, finns inte kvar på sin ursprungliga plats.

Ytterligare en viktig skillnad mellan fysiska objekt och digitala objekt på webben är att innehållet i de sistnämnda aldrig är statiskt på samma sätt som i en bok. Om en bok ges ut i ny upplaga klassificeras den ofta igen så att man kan se vilka olika upplagor som finns. Vill man ändra i ett digitalt objekt på webben uppdaterar man ofta innehållet och den äldre versionen skrivs över med en ny. Både digitala objekt i mer kontrollerade system och på webben kan dupliceras och finnas i flera utgåvor men just på webben är det vanligare att man skriver över en äldre version med en ny. Detta kan rent praktiskt gå till så att man har en webbsida som till exempel innehåller nyheter. I stället för att radera eller duplicera hela objektet fyller man bara på med fler nyheter i det befintliga objektet.

Ett annat problem i detta sammanhang är att etablerade metoder för katalogisering bygger på fysiska objekt som har en klar början och ett slut. Det är svårt att fastställa en början och ett slut på ett digitalt objekt på webben eftersom hyperlänkar skapar en möjlig kontinuerlig fortsättning på objektet.

De digitala objektens karaktär påverkar även förhållningen till bevarandenaspekter och organiseringen. I traditionella bibliotek där det finns kontroll över beståndet finns det både intresse och möjlighet att bevara ett bestånd. På webben handlar det snarast om åtkomst. Åtkomst av digitala objekt som kan vara temporära och föränderliga avseende på såväl lokalisering som innehåll. Webben ställer inte bara krav på nya metoder för kunskapsorganisation utan kräver ett helt nytt förhållningssätt till den kunskap som skall organiseras.

### **7.2.1.2 Standardiserad metadata**

En stor skillnad mellan traditionella och webbaserade informationssystem är metoderna för beskrivning av fysiska respektive digitala objekt. För till exempel böcker formateras metadata på ett välbekant sätt utifrån de etablerade standarder som finns i bland annat KRS.<sup>152</sup> Enligt dessa standarder definieras vad som är titel, upphovsman etcetera. Information som finns i standardiserade format möjliggör en standardiserad beskrivning.

Problemet är att det på webben inte finns inte några standardiserade format för beskrivning av digitala objekt. Enligt HTML-standarden finns det koder som är tänkta att beskriva

---

<sup>152</sup> Katalogiseringsregler för svenska bibliotek, (1990). - 2. uppl / red. Bodil Gustavsson. - Lund: Bibliotekstjänst.

innehåll utifrån titel och rubriker. Återvinning av digitala objekt på webben försvåras dock av att det råder en bristande konsekvens i användningen av den logiska struktur som HTML innebär. Många HTML-kodare är okunniga om eller struntar helt enkelt i att använda till exempel rubriknivåer i enlighet med HTML-standarderna. Alla rubriker kodas inte med de speciella rubrikkoderna som `<H1>` och `<H2>`. Om rubrikkoderna användes på *alla* webbsidor skulle den automatiska indexeringen underlättas genom att man skulle kunna indexera rubrikinnehållet separat och därmed få bättre information om textens innehåll än vid ren fulltextindexering. Wallis och Burden påpekar att HTML är ett språk som beskriver innehåll och att vissa HTML-koder då borde säga mer om innehållet än andra. Titel borde väga tyngre vid indexering än brödtext. Deras undersökning visar dock att sökning på både rubriknivå 1 och 2 gav bättre resultat än sökning på titel. Detta beror på att de flesta författare inte förser sina texter med lämplig titel. Att så inte är fallet kan bero på att HTML-taggen `<TITLE>` visas just i titellisten i webbläsarprogrammet och därför inte uppfattas som en del av texten. Jämför man med mer etablerade informationssystem som till exempel en OPAC så kan man där återvinna objekt utifrån titel och titeln fungerar som en viktig ingång för att identifiera ett objekt. Det finns som vi tidigare påpekat även etablerade standarder för vad som verkligen är titel. I etablerade system följs reglerna för titelkodning medan det på webben ofta inte används på ett enligt HTML-standarderna korrekt sätt.

De ovanstående HTML-koderna tar enbart fasta på titel och rubrik. För att kunna göra mer fullständiga beskrivningar av objekt finns särskilda metadata-koder framtagna. Dessa definieras av till exempel Dublin Core.<sup>153</sup> Metadatakoderna kan beskriva bland annat upphovsman, lokalisering, publiceringsdatum, abstract men inte minst ämnesdata i form av nyckelord eller klassifikationskoder från bibliografiska klassifikationssystem. Metakoderna syftar bland annat till att underlätta intellektuell åtkomst. De webbaserade informationssystemen påverkas av att det inte råder någon konsensus kring användningen av dessa metakoder. Dels finns det flera olika standarder utvecklade och dels är användningen av dessa inte speciellt utbredd.

En utgångspunkt för hur ett informationssystem ska utformas är vilka *objekttyper* systemet ska innehålla. Mandel och Wolven menar att typen av objekt som systemet skall omfatta påverkar både beskrivningen av dessa samt på vilket sätt de skall återvinnas.<sup>154</sup> Som påpekats tidigare så har böcker till exempel en viss fysisk form och ett antal standardiserade metadata, som författare, titel, förlag och ISBN. Böcker ges ut av ett förvisso stort men ändå begränsat antal förlag. Detta gör det *relativt* enkelt att skapa en katalog över bokutgivning. Självklart finns det mängder av problem med att skapa kataloger även över standardiserade informationsbärare men exemplet med böcker visar dock att information som finns i standardiserade format möjliggör en standardiserad beskrivning. Finns det en standardiserad beskrivning kan objekten även ordnas i ett system för återvinning. Icke standardiserade format kräver mer flexibla beskrivnings- och återvinningsystem. På

---

<sup>153</sup> The Nordic Metadata Project [URL: <http://linnea.helsinki.fi/meta/>] 98-03-18.

<sup>154</sup> Mandel, Carol A. och Wolven, Robert, (1996), s. 26-29.

webben finns inga etablerade standarder för beskrivning av digitala objekt, varför det krävs alternativa metoder för kunskapsorganisation och återvinning.

### 7.2.1.3 Intellectuellt värde

Den sista stora skillnaden som vi kunnat urskilja är frågan om det intellektuella värdet på digitala objekt på webben. Det är tveksamt om det intellektuella värdet i ett webbaserat digitalt objekt generellt kan anses vara lika högt som i till exempel en bok eller en fulltextpost hos databasvärden Dialog. Åtminstone under webbens första tre år har vi uppfattat att mycket av innehållet i de digitala objekten på webben ofta inte varit lika genomarbetat och angeläget som innehållet i tryckta tidningar, tidskrifter och böcker. Detta är dock något som vi tror kan komma att ändras i och med att användningen av webben ökar och mediet etableras. Undantag från detta finns inom framför allt inom forskarvärlden där Internet använts professionellt i många år. De webbaserade informationssystemen kan påverkas av webbobjektens förhållandevis låga värde. Eftersom söktjänsternas innehåll inte anses lika betydelsefulla som till exempel en bibliotekskatalog, anses de inte vara värda lika mycket omsorg i form av tid, pengar och arbete. Ett exempel på detta kan vi se hos söktjänsten Punkt se som uppenbarligen inte varit speciellt noggranna när de indexerade sina objekt. Det förekommer många felindexeringar som kan bero på att man inte värderat sina objekt speciellt högt.

Sammanfattningsvis skiljer sig de webbaserade informationssystemen på flera punkter från de mer traditionella. De påverkas av arkitekturen och de tekniska förutsättningarna hos webben som den fungerar idag. Några av dessa förutsättningar är bristen på kontroll av vilka objekt som läggs till och tas bort, bristen på konsekvent användning av metadata och den ibland felaktiga användningen av HTML. Vidare är objekten på webben dynamiska och kan ändras mycket snabbt. Slutligen har värdet på innehållet i webbobjekten hittills varit lägre än i mer etablerade system vilket påverkar söktjänsternas kvalitet.

### 7.2.2 Ämneshierarkiska söktjänster

Det finns som vi tidigare nämnt två huvudtyper av söktjänster på dagens webb, ämneshierarkiska och indexbaserade söktjänster. Syftet med detta stycke är att visa på några för- och nackdelar med just ämneshierarkiska söktjänster. För att synliggöra problem och möjligheter med denna typ av söktjänst ställer vi dem i förhållande till andra söktjänster, det vill säga indexbaserade sådana.

Det är lätt att konstatera att ett av problemen med de indexbaserade tjänsterna är att de ofta genererar mycket brus både i jämförelse med browsing i ämneshierarkiska system och sökningar i traditionella bibliografiska databaser. Informationsvetenskapliga begrepp som *precision* och *recall* framstår rentav som meningslösa i samband med de indexbaserade söktjänsterna och det finns flera anledningar till detta. Koch påpekar att det dels inte fogas några ämnesord till objekten genom mänskligt arbete och dels är den automatiska indexeringen i dessa tjänster för närvarande såpass dålig att den inte kan mäta sig med de

bibliografiska databasernas manuellt inmatade ämnesdata.<sup>155</sup> Bella Weinberg påpekar att det i och med de senaste årens ökade användning av fritextsökning utvecklats en tro på att mänsklig ämnesanalys inte spelar lika stor roll längre eftersom alla ord är sökbara.<sup>156</sup> Detta kan göra att utvecklare av söktjänster ofta tror att det räcker med att indexera *allting*, det vill säga så mycket som möjligt av webbens objekt för att skapa god återvinning.

En ämneshierarkiskt ordnad söktjänst innehåller oftast ett mindre antal objekt än en indexbaserad tjänst som genererats automatiskt. En ämneshierarkisk söktjänst kan därför inte erbjuda samma täckningsgrad och dessutom expanderar webben alldeles för snabbt för att någon söktjänst ska hinna med att manuellt strukturera alla nya objekt. Oftast är det dock inte de ämneshierarkiska tjänsternas syfte att ens försöka indexera *hela* webben utan de flesta har som syfte att erbjuda ett mindre antal kvalitetskontrollerade objekt och då kan den begränsade omfattningen rentav vara en fördel. För indexbaserade söktjänster är det dock önskvärt att ha en så stor täckning som möjligt och vissa tjänster påstår i sin marknadsföring till och med ha en total täckning av hela webben. Detta är i dagens läge naturligtvis omöjligt men det tycks finnas prestige bland söktjänsterna i att kunna erbjuda kvantitet.

Browsing och sökning på ämnesord är vanliga i alla typer av söktjänster men ämnesordssökningar fungerar bäst i informationssystem som använder en begränsad och kontrollerad vokabulär. De ämneshierarkiska tjänsterna fungerar i detta sammanhang bättre än de indexbaserade söktjänsterna eftersom de har större kontroll över sin input beroende på grundligare ämnesanalys. Talesättet ”skräp in - skräp ut” gäller även vid ämnesrepresentation, eftersom en representation av ett objekt kan aldrig vara bättre än den analys av innehållet som görs.

Anne Callery, som arbetar som katalogisator på Yahoo!, menar i sin artikel *Yahoo! Cataloging the Web* att sökresultaten generellt får högre precision om informationssystemet har ett ämneshierarkiskt index.<sup>157</sup> Hon exemplifierar med att när man i en indexbaserad söktjänst efterfrågar information om *surfning* kan man få träffar rörande både sporten surfning och på metaforen för ostrukturerad navigering på webben. Ett ämneskontext, vad som i klassifikations-sammanhang även kallas domän, är lättare att bestämma i en ämneshierarkisk söktjänst. Ytterligare en av fördelarna med klassificering som används i de ämneshierarkiska söktjänsterna blir uppenbar vid sökningar efter objekt inom ett brett

---

<sup>155</sup> Koch, T., Ardö, A., Brümmer, A. och Lundberg, S., (1996), The building and maintenance of robot based internet search services: a review of current indexing and data collection methods. [URL: <http://www.ub2.lu.se/desire/radar/reports/D3.11/tot.html>] 98-04-28.

<sup>156</sup> Weinberg, Bella Hass, (1996), Complexity in indexing systems – abandonment and failure: implications for organizing the Internet // ASIS '96 : proceedings of the 59th ASIS Annual Meeting, October 19-24. - [URL: <http://www.asis.org/annual-96/ElectronicProceedings/weinberg.html>] 98-04-27. ;

Se även Weinberg, Bella Hass, (1996), Complexity in indexing systems – abandonment and failure: implications for organizing the Internet // ASIS '96 : proceedings of the 59th ASIS Annual Meeting, Baltimore, Maryland, October 21-24, 1996 / ed. Steve Hardin. - Medford, N.J.: Information Today. - Cop. 1996. - (Proceedings of the ASIS annual meeting ; 33). - S. 84-90. - ISBN: 1-57387-037-4

<sup>157</sup> Callery, Anne, (1996), Yahoo! Cataloging the Web [URL: <http://www.library.ucsb.edu/untangle/callery.html>] 98-03-19.

ämne. En sökning efter översiktlig information om till exempel *Danmark* ger oerhört många träffar i en indexbaserad söktjänst. Flera av dessa handlar troligtvis om Danmark ur mer detaljerade perspektiv än vad som är avsett med sökningen. Med hjälp av ett klassifikationssystem kan objekt av generell karaktär placeras i den översta generella klassen. Därigenom kan man särskilja de generella och de mer specifika objekten från varandra och återvinningen underlättas.<sup>158</sup> Denna fördel gäller även i icke webbaserade system men kan vara speciellt tydlig på webben.

Marcella och Newton påpekar i *A New Manual of Classification* att ett klassifikationssystem inte bara kan hjälpa användaren att hitta ett objekt inom ett ämne, utan även kan vara användbart genom att peka på fler intressanta objekt *inom* samma ämne. Den minsta gemensamma nämnaren för objekt ordnade enligt ett klassifikationssystem är inte författare eller titel utan just ämne.<sup>159</sup> De ämneshierarkiska söktjänsterna på webben är alla hypertextbaserade system där man kan bläddra sig fram till den nivå i hierarkin som hyser det sökta objektet. På vägen dit passerar man flera närliggande ämneskategorier som man då kan använda sig av. Detta är speciellt värdefullt när man inte vet exakt vad man letar efter. Även Callery uppmärksammar det som Marcella och Newton påpekat, det vill säga att den browsande användaren får tips om relaterade objekt som kan vara av intresse. Om han eller hon inte hittar den eftersökta informationen direkt kan förmodligen några av de närliggande grupperna vara till hjälp. Användaren av en ämneshierarki inte behöver känna till lika många specialtermer och synonymer inom det aktuella ämnesområde som den som ska formulera en söksträng till ett index. Användaren kan välja en bred kategori och sedan browsa vidare bland de mer specifika ämnesrubrikerna.<sup>160</sup> Ytterligare en fördel med ämneshierarkiska söktjänster är att användaren inte behöver kunna avancerade sökspråk med till exempel trunkeringsregler och boolesk logik, vilket ofta används vid sökning i indexbaserade söktjänster.

Traugott Koch anser dock att det över huvud taget finns få fördelar med intellektuell indexering eller klassificering i en webbkontext.<sup>161</sup> Det som talar för en intellektuell indexeringsmetod är att kvaliteten på indexeringen *kan* bli bättre, men nödvändigtvis inte blir det. Ett litet urval objekt med relevant innehåll eller som uppfyller vissa kvalitetskriterier kan väljas ut. Detta tillvägagångssätt gör att sammanfattningar av innehållet kan skrivas för att underlätta återvinning. Manuell indexering lämpar sig därför enligt Koch bäst för specialiserade ämnesspecifika söktjänster, där det även är möjligt att genomföra uttömmande sökningar.

En annan viktig skillnad mellan indexbaserade- och ämneshierarkiska söktjänster är att de senare vanligtvis är skapade manuellt. Manuell klassifikation är alltid förenad med stora kostnader. Detta kan antagligen förklara varför det på dagens webb finns fler indexbaserade än ämneshierarkiska söktjänster. Det anses förmodligen vara både enklare

---

<sup>158</sup> Hjørland, Birger, (1997), s. 30f.

<sup>159</sup> Marcella, Rita och Newton, Robert, (1994), s. 4f.

<sup>160</sup> Callery, Anne, (1996).

<sup>161</sup> Koch, Traugott, Ardö, A., Brümmer, A. och Lundberg, S., (1996).

och billigare att köpa eller utveckla ett datorprogram som automatiskt indexerar på webben än att som Yahoo! ha en hel stab av katalogisatorer anställda.

Sammanfattningvis kan sägas att ämneshierarkiska tjänster kan fungera bättre än indexbaserade vid sökning på ämne. Detta beror på att de har större kontroll över sin input, att klassificeringen hjälper till att ge förslag på andra objekt inom samma ämne samt att man undviker problem med homonymer. Om man ska använda sig av manuell indexering passar det bäst i specialiserade ämnesspecifika söktjänster. Ämneshierarkiska söktjänster innehåller färre antal objekt men oftast med mer metadata om varje objekt. Tjänsterna är ofta skapade manuellt vilket förenat med stora kostnader.

### 7.3 Klassifikationssystem på webben

I föregående avsnitt pekade vi på några av de för- och nackdelar som finns med ämneshierarkiska söktjänster på webben. Syftet med detta stycke är att visa på för- och nackdelar med etablerade respektive icke-etablerade klassifikationssystem i dessa söktjänster. Vi inleder med några allmänt hållna funderingar kring klassifikationssystem på webben.

Både Steinberg och Weinberg hävdar att webben främst används för att finna den allra färskaste informationen som ännu inte hunnit ut i tryck.<sup>162</sup> Att det går snabbt att publicera information talar till viss del mot all form av mänsklig klassifikation av den enkla anledningen att det tar för lång tid. När väl en klassificering av ett digitalt objekt är genomförd kan beskrivningen vara föråldrad.

Katalogisering och klassificering med etablerade metoder innebär att man gör en mer eller mindre noggrann och ofta tidsödande analys av det intellektuella innehållet i ett digitalt eller fysiskt objekt och därefter tilldelar det en mängd metadata och en plats i en ämneshierarki. Traditionellt utförs detta arbete för till exempel böcker och tidskriftsartiklar i en bibliotekskatalog. I ett informationssystem vars bestånd består av fysiska objekt fyller beskrivningen ett viktigt syfte för identifiering och återvinning. Försök till klassificering med etablerade metoder förekommer även i begränsad skala på webben. Exempelvis så har NISS Directory en relativt hög detaljnivå på sina poster vilket tyder på att de har använt sig av etablerade klassifikationsmetoder. Ett problem med detta enligt Dong Zhao är att katalogiseringsregler, som till exempel KRS, först och främst tar fasta på den fysiska beskrivningen. Han fortsätter med "[t]he importance of physical description no longer exist in the world of of electronic information on the WWW, where retrieval of the intellectual content take a higher priority over the retrieval of physical items."<sup>163</sup>

---

<sup>162</sup> Weinberg, Bella Hass, (1996). ; Steinberg, Steve G., (1996), Seek and ye shall find (maybe) [URL: <http://www.wired.com/wired/4.05/features/indexweb.html>] 98-04-28.

<sup>163</sup> Zhao, Dong, (1997), Managing electronic information: web page, subject folder, and onweb catalog // Proceedings / National Online Meeting: New York, May 13-15, 1997 / edited by Martha E. Williams. Medford, N.J.: Learned Information. - Cop. 1997. - s. 387.

Problemen med etablerade klassifikationsmetoderna är att de är tidsödande, kräver mycket personal med lång utbildning och ekonomiskt krävande. Även de icke-etablerade systemen kräver tid, pengar och personal men eftersom de vanligtvis har lägre kvalitetskrav och mindre detaljnivå kan de vara mindre krävande än de etablerade metoderna. Använder man sig av denna tidsödande och detaljerade klassificering på digitala objekt på webben är risken stor att man har gjort arbetet i onödan eftersom åtkomst till digitala objekt eventuellt inte går att åstadkomma i framtiden. Obeständigheten hos digitala objekt på webben i kombination med höga ekonomiska och tidsmässiga kostnader för klassificering talar mot användning av detaljerade bibliografiska metoder på webben.

Vidare måste alla informationssystem anpassas till sina användare vilket även gäller i webbsammanhang. De ämneshierarkiska söktjänster vi intresserar oss för har som syfte att täcka in alla slags ämnen och företeelser. Detta oberoende av om de använder etablerade eller icke-etablerade klassifikationssystem. Man kan dock fråga sig vilken målgrupp de olika tjänsterna riktar sig till. Tjänsterna saknar i många fall dokumentation kring både omfattning, syfte och målgrupp. För både NISS Directory och BUBL är den akademiska målgruppen tydligt uttalad. De kommersiella tjänsterna riktar sig förmodligen till den breda allmänheten. Då dessa söktjänster är inriktad mot Sverige är det främst den svenskspråkiga allmänheten som är den primära målgruppen.

### **7.3.2 Etablerade och icke-etablerade klassifikationssystem**

De etablerade klassifikationssystemen har varit föremål för kritik redan innan webben kom in i bilden. Jesse Shera kritiserade i mitten av 1960-talet bruket av bibliografisk klassifikation då denna metod enligt honom misslyckats på två sätt. För det första har bibliografiska klassifikationssystem i allt för hög grad grundat sig på boken som fysiskt objekt istället för att ta fasta på dess intellektuella innehåll.<sup>164</sup> Detta är något som även Bella Weinberg tar upp.<sup>165</sup> Hennes kritik riktar sig framför allt mot DDK som är utvecklat för att organisera böcker, vilket talar till dess nackdel vid användningen på webben. För det andra kritiserade Shera de begränsningar som hierarkiska klassifikationssystem omfattar. Till exempel en bok som behandlar flera ämnen ur flera olika perspektiv måste infogas i ett system under en klass. Steinberg framför en liknande invändning mot de etablerade systemen då han påstår att de är mer begränsade eftersom en bok endast kan placeras på en hylla medan digitala objekt kan dubbelklassificeras i flera kategorier.<sup>166</sup> Detta är lite som att jämföra äpplen och päron eftersom både etablerade och icke-etablerade klassifikationssystem kan använda sig av dubbelklassificering. Exempel på detta finns både hos Yahoo! Sverige och hos NISS Directory.

#### **7.3.2.1 Anpassning till användaren**

---

<sup>164</sup> Shera, Jesse H., (1965), *Libraries and the organization of knowledge*. - London: Crosby Lockwood & Son. - (New Librarianship Series). - s. 84-86.

<sup>165</sup> Weinberg, Bella Hass, (1996).

<sup>166</sup> Steinberg, Steve G., (1996).

Detta stycke beskriver ett antal argument för- och emot etablerade och icke-etablerade klassifiaktionssystem främst utifrån ett användarperspektiv.

Tidigare i uppsatsen berördes Hjørlands indelning av de varierande ambitionsnivåer som förekommer vid klassificering. Bibliografiska etablerade klassifikationssystem har en högre ambitionsnivå än icke-etablerade klassifikationssystem. De olika ambitionsnivåerna säger dock inget om att det ena systemet är bättre än det andra. Vid användning av till exempel bibliografiska klassifikationssystem i den miljö de är konstruerade för, det vill säga i en bibliotekskontext, är de utmärkta som en metod för både hylluppställning och återvinning. Detta behöver nödvändigtvis inte betyda att de är lämpade i webbsammanhang.

Hjørland hävdar, som vi tidigare nämnt, att all kunskapsorganisering syftar till att fylla en praktisk funktion. Klassificering syftar alltid till att underlätta mänsklig verksamhet och avseende klassifikationssystem på webben är syftet att möjliggöra återvinning av digitala objekt utifrån deras intellektuella innehåll, det vill säga deras ämne. Hjørland utgår från ett pragmatiskt förhållningssätt där den praktiska användningen sätts i fokus. Det är utifrån detta synsätt viktigt att fråga sig vem det är som skall använda söktjänsten, vilket informationsbehov och vilka förkunskaper de har och att anpassa söktjänsten till detta.

Marcella och Newton hävdar att de etablerade klassifikationssystemens användbarhet i en webbkontext framför allt beror på hur väl de svarar mot de krav majoriteten av användarna på webben har.<sup>167</sup> Hänsyn måste då tas till på vilket sätt användaren kan tänkas gå till väga för att finna det eftersökta. Optimalt skall användarens beteende på något sätt kunna förutses och systemet ska kunna anpassas efter detta. Hjørland nämner att det finns många exempel på när bibliotekarier förespråkat DDK i sammanhang där det inte alls passat. Dessa bibliotekarier utgår då ifrån sina egna behov och förespråkar användandet av ett etablerat system då de själva ofta är förtrogna med detta. Analys av användarens behov och krav förbises.

Det finns även användare som inte är utbildade bibliotekarier eller informationsspecialister och som ändå är bekanta med etablerade klassifikationssystem. En av fördelarna med ett etablerat klassifikationssystem är just att användarna kanske känner till hur ett system är uppbyggt och därför enkelt kan hitta rätt. Används samma klassifikationsschema i mer än en söktjänst kan browsing då ske med liknande tillvägagångssätt i flera informationssystem. Traugott Koch har frågat några ansvariga vid ett antal söktjänster som använder UDK om vad de anser vara systemets fördelar.<sup>168</sup> De bekräftar att UDK är ett universellt standardssystem som används av många och därmed finns det en utbredd förtrogenhet med systemet och att det kan användas i många olika sammanhang. Det är dessutom, i de allra flesta fall, gratis att använda, vilket naturligtvis inte på något sätt förbättrar användarvänligheten av systemet.

---

<sup>167</sup> Marcella, Rita och Newton, Robert, (1994), s. 5.

<sup>168</sup> Koch, Traugott, (1996b).

Problemen med tomma klasser eller klasser som saknas talar mot användningen av etablerade klassifikationsystem på webben. Detta visar sig tydligt när man studerar till exempel NISS Directory. Att klasser helt saknas kan vara speciellt besvärande för de användare som har god kunskap om systemet och således förväntar sig att finna vissa kategorier på vissa platser i hierarkin. Om det däremot som i NISS Directory eller BUBL LINK saknas klasser där de borde finnas faller lite av fördelarna med att användarna känner till systemet. Detta är även något som Kock påpekar och som han tror beror på att innehållet i digitala objekt *kan* skilja sig markant från innehållet i tryckt form. Måhända är det så. Ett exempel på detta är att BUBL LINK valt att upprätta *BUBL 5:15*, som alltså använder ett icke-etablerat system som komplement till den DDK baserade söktjänsten. Anledningen till detta är att ”many Internet resources cover a big subject area and are more like libraries than books.”<sup>169</sup> Precis som Koch påpekar så är uppfattningen vid BUBL LINK att de digitala objektens innehåll skiljer sig från innehållet i tryckta objekt.

Ytterligare ett problem med etablerade klassifikationssystem är att de är mycket stora och komplicerade. Det kan diskuteras hur många av webbans användare som egentligen kan använda komplicerade system som DDK och UDK. Hur representativa är de som kan använda etablerade klassifikationsystem för den genomsnittlige användaren av söktjänster på webben? Det bör påpekas att DDK och UDK i första hand är konstruerade av och för bibliotekarierna och informationsspecialisterna. Weinberg påpekar att de etablerade systemens komplexitet kräver mänskliga intermediärer som kan hjälpa användarna, vilket dessutom innebär ökade kostnader.<sup>170</sup> De etablerade klassifikationssystemen kräver att användaren sätter sig in i hur de är uppbyggda för att kunna använda dem optimalt men så är fallet även för icke-etablerade system, vilket Weinberg dock inte berör. Ytterligare ett av hennes argument mot etablerade klassifikationssystem är att användarna inte vill lära sig komplicerade system utan de vill komma åt informationen på snabbast möjliga sätt. Weinberg refererar dessutom till undersökningar som hävdar att användare ofta är motvilliga att ta hjälp av intermediärer och att de hellre vill finna det eftersökta på egen hand. En av grundtankarna med webben är just oberoendet av tid och rum, varför det ter sig underligt om användaren måste befinna sig vid en viss plats där det finns utbildade intermediärer, till exempel bibliotekarierna, för att kunna utnyttja ett webbaserat informationssystem fullt ut.

Frågan om vilka användarna egentligen är har vi varit inne på tidigare. Att olika söktjänster har olika målgrupper är självklart liksom de har olika typer av innehåll. Söktjänster vars målgrupp och innehåll är av akademisk karaktär vinner enligt Koch på att använda ett etablerat klassifikationsystem. DDK och UDK är utvecklade för användning i en akademisk kontext.<sup>171</sup> De är dock inte konstruerade för användning specifikt för webben, något som Koch inte tar upp. Weinberg hävdar att DDK är mindre lämpligt för användning just i en webbkontext och motiverar detta bland annat med att DDK:s primära indelning är disciplin

---

<sup>169</sup> About BUBL 5:15 [URL: <http://link.bubl.ac.uk:80/about515.htm>] 98-04-16

<sup>170</sup> Weinberg, Bella Hass, (1996).

<sup>171</sup> Koch, Traugott, (1996b).

och ej konkret ämne. Detta är ytterligare en aspekt som talar för att DDK är ett föråldrat system.<sup>172</sup>

Presentationen av en ämnehierarkiskt system är betydelsefull för att underlätta användarvänligheten. Hypertext är den grund som webben vilar på och om hypertext används på ett genomtänkt sätt kan ämnehierarkier i hypertext underlätta ämnesåtkomst. Traugott Koch pekar i texten *The role of classification schemes in Internet resource description and discovery* på att användningen av en hierarkisk struktur på webben stödjer browsing.<sup>173</sup> Detta är oberoende av om man använder ett etablerat eller ett icke-etablerat klassifikationssystem. Ett lyckat exempel på detta anser vi vara Yahoo! Sverige medan NISS Directory är ett exempel på en mindre lyckad användning av hypertext.

Koch visar slutligen på att det även finns ekonomiska nackdelar med icke-etablerade system och att de förmodligen aldrig kommer att bli lika spridda som till exempel UDK och DDK. Om ingen större spridning av systemen sker är de organisationer eller de företag som har utvecklat systemen själva också ansvariga för eventuella revidering och uppdatering som kan komma att behövas, vilket kan innebära stora kostnader. Koch refererar här i sin tur till Steinberg.<sup>174</sup>

### 7.3.2.2 Notation

Till de bibliografiska klassifikationssystemens fördelar brukar nämnas deras användning av klassymboler. Som påpekats ovan anses en av klassymbolernas fördelar enligt Aluri, Kemp och Boll vara att de är kortare än ord och därmed fungerar som minnesstödjande. Lois Mai Chan ansluter sig till samma uppfattning och tillägger att den numeriska sekvensen är självklar. Vi frågar oss dock om numeriska och andra klassymboler är lika välkända bland användare utanför det biblioteks- och informationsvetenskapliga området? För dessa användare kan det vara komplicerat att utan speciell träning i informationssökning att ta till sig den notation som UDK och DDK använder. Exempelvis kan ett objekt som omfattar en statistisk sammanställning av olyckor i brittiska kolgruvor från andra världskriget och framåt beskrivas i UDK med klassymbolen 622.33:622.86(410)''1945/98'':31. En lång rad siffror som inte är speciellt enkel att komma ihåg! Koch anser dessutom att notationen närmast kan vara störande och förespråkar att denna inte skall vara synlig för att inte störa användaren. Däremot kan den användas som ett underliggande organisationssystem.<sup>175</sup>

Att notationen kan tjäna som en språklig gränsöverskridare är ett argument som brukar föras fram till de etablerade systemens försvar.<sup>176</sup> Till detta kommer att exempelvis DDK och UDK är översatta till många språk och samma söktjänst kan därmed erbjudas på olika språk. Notationen kan underlätta för användaren eftersom den är konsekvent genomförd

---

<sup>172</sup> Weinberg, Bella Hass, (1996).

<sup>173</sup> Koch, Traugott, (1996b).

<sup>174</sup> Steinberg, Steve G, (1996).

<sup>175</sup> Koch, Traugott, (1996b).

<sup>176</sup> Ibid.

oberoende av språk. Om notation används behöver användaren rent teoretiskt inte vara bekant med det språk som ämneshierarkin är skriven på. Frågan är dock om det är tillämpbart i praktiken på webben. Om en fransk motsvarighet till BUBL LINK upprättas skulle innehållet i de digitala objekt som original versionen omfattar förmodligen inte vara lika intressanta för franska användare. Anledningen till detta är att det huvudsakliga materialet i BUBL LINK är engelskspråkigt.

En annan fördel med notation är att om den används på ett bra sätt kan den påvisa hierarkiska relationer i ämneshierarkin. Men om man som till exempel i NISS Directory och BUBL LINK använder notation men trots det inte tydliggör den hierarkiska strukturen så utnyttjar man inte systemets möjligheter till fullo.

### 7.3.2.3 Flexibilitet

En fördel med icke-etablerade klassifikationssystem är att man snabbt kan åstadkomma förändringar i ämneshierarkin om så behövs, vilket bekräftas av Traugott Koch. Man kan lägga till och ta bort klasser efter behov utan att behöva vänta på nästa uppdatering av systemet. Allteftersom nya ämnen och delämnena dyker upp kan man även ändra i systemet och därmed spegla snabba förändringar.<sup>177</sup> Exempel på denna snabba anpassning finns hos Yahoo!, som bland annat infogat en särskild klass för *År 2000-problemet* under huvudklassen *Datorer och Internet*.

Steve Steinberg argumenterar i sin text *Seek and Ye shall find (maybe)* för att de etablerade klassifikationssystemen är föråldrade och inte räcker till i den snabbt växande floran av ämnen och ämneskombinationer på webben.<sup>178</sup> Till exempel så publiceras nya upplagor av DDK vart tionde år och mellan dessa utkommer häften med uppdateringar och revideringar. Liknande åsikter förs fram av Weinberg som hävdar att de närmast rigida klassifikationssystemen inte hinner med i den snabba utvecklingen inom modern vetenskap. Lösningen på dessa problem är att uppdatera och omstrukturera systemen med jämna mellanrum, vilket också sker men inte tillräckligt snabbt. Dessa åsikter bekräftas ju även av de intervjuer Koch utfört med ansvarig för söktjänster som använder UDK.<sup>179</sup> Det bör ju påpekas att icke-etablerade system troligtvis inte heller skulle hinna med i de snabba vetenskapliga framstegen, men det är vanligtvis inte heller deras avsikt. Koch hävdar att de icke-etablerade klassifikationssystemen vanligtvis är utvecklade för att klassificera bredare innehåll, till exempel underhållning, kommersiell och offentlig information.<sup>180</sup>

För ett bibliotek med OPAC, vari objekten klassificeras enligt ett etablerat system, kan det dock finnas fördelar med att använda ett etablerat klassifikationssystem även för digitala objekt på webben. Bibliotekets OPAC och webbaserade söktjänst kan då integreras med varandra. Anne Callery beskriver i sin text *Yahoo! Cataloging the Web* fördelarna med det

---

<sup>177</sup> Ibid.

<sup>178</sup> Steinberg, Steve G., (1996).

<sup>179</sup> Koch, Traugott, (1996b).

<sup>180</sup> Ibid.

icke-etablerade klassifikationssystem som Yahoo! använder.<sup>181</sup> En av anledningarna till att Yahoo! inte använder något etablerat klassifikationssystem är att de till skillnad från de flesta fysiska bibliotek inte har någon existerande samling som objekten måste integreras i.

### 7.3.2.3 Subjektiv klassificering

De icke-etablerade klassifikationssystemens flexibilitet kan ha både för och nackdelar. En nackdel enligt Koch är bristen på konsekvens vilket leder till subjektiv klassificering. Utgår man, som till exempel i Yahoo! Sverige, inte från något nedtecknat system utan från de åsikter som de enskilda klassifikatörerna har så finns det inget som säger att en användare uppfattar ämnen på samma sätt. Ytterligare ett exempel på detta finns hos Punkt.se som använder ett mycket subjektivt klassifikationssystem, vilket förmodligen passar vissa användare men kanske inte alla. Detta resonemang gäller även för etablerade system vars upphov är resultatet av ett fåtal personers insatser. Etablerade system som DDK och UDK utgår från ett akademiskt synsätt där klasser delas in enligt erkända discipliner. Det finns med andra ord en tydlig ordning vid infogande av klasser med utgångspunkt i *disciplin* snarare än *innehåll*.

Angela Oehler hävdar i *Browsingdienste im Internet* att det faktum att ämnehierarkin i till exempel Yahoo! tilltar efterhand som söktjänsten växer leder till inkonsekvens i klassifikationsstrukturen. Våra studier bekräftar att det förekommer logiska brister i till exempel Yahoo! Sverige och Punkt.se. Dessa består av kategorier som utger sig för att innehålla länkar med ett visst innehåll men som sedan innehåller något annat. Exempel på detta i Yahoo! Sverige är kategorin *Topp:Datorer och Internet:Internet:World Wide Web:Det bästa på nätet* som innehåller kategorin *Inte riktigt det bästa*. En klass som säger sig innehålla *det bästa* kan logiskt sätt inte innehålla *inte riktigt det bästa*.

Subjektiv klassificering får dessutom konsekvenser vid detaljerade precisionssökningar på klasser. Exempel på detta finns hos Punkt.se som med sin fantasifulla namngivning av klasser inte erbjuder så bra sökning på klasser medan däremot browsing fungerar bra. Etablerade klassifikationssystem har här en fördel enligt Koch.<sup>182</sup> Användningen av notation sätter söktermerna i en kontext och möjliggör filtrering och sökning med hög precision.

## 7.4 Avslutning

Det finns många åsikter kring olika aspekter på klassificering av digitala objekt på webben. Vi anser att det inte finns något direkt rätt eller fel och vi syftar inte heller till att försöka bestämma vilken typ av system som är *bäst*. Däremot är det relevant att diskutera vilka system som kan fungera i en specifik kontext och vilka för- och nackdelar varje system har.

Vid utveckling av en webbaserad söktjänst måste man göra en grundläggande analys av vilka förutsättningar, målsättningar och resurser man har. Vi menar att vilken typ av klassifikationssystem som är lämpligast beror mycket på i vilken kontext systemet ska

---

<sup>181</sup> Callery, Anne, (1996).

<sup>182</sup> Koch, Traugott, (1996b).

användas, vad tjänstens syfte är och vilken målgrupp den riktar sig till. Dessa aspekter kan anses vara självklara men de söktjänster vi undersökt och de texter vi läst tyder på att det behöver upprepas.

Hänsyn måste även tas till vilka resurser i form av personal, tid och pengar som finns, bland annat eftersom klassificering är mycket personalkrävande och personalen måste ha god utbildning. De tekniska förutsättningarna och vilka krav på kvalitet man har är andra faktorer som måste beaktas. Man måste även se till att man har resurser för uppdatering och underhåll. Olika användargrupper har olika informationsbehov och skiftande kunskaper i ämnessökning vilket måste påverka valet och utformningen av klassifikationssystem.

Vidare beror valet av system till stor del på hur webben, de digitala objekten och deras innehåll uppfattas. Som vi ser det är webben bara ett nytt sätt att förmedla information på, det vill säga ett distributionssystem. Webben som system ställer nya krav på kunskapsorganisation på grund av bristen på kontroll av digitala objekt. Det intellektuella innehållet i objekten skiftar mycket och en fråga som följer av detta är huruvida innehållet i webbobjekten kan jämföras med innehållet i pappersburen information. Om så inte är fallet är det då lämpligt att använda klassifikationssystem som är utvecklade för att klassificera böcker? Att nöja sig med de etablerade bibliografiska klassifikationsystemen kan kanske vara en bekväm inställning eftersom systemen redan finns färdiga att användas. Väljer man däremot att nyttja ett icke-etablerat system bör man inte bortse från all den kunskap som finns i de etablerade klassifikationsystemen. Bella Weinberg uttrycker det på följande sätt:

*It is recognized that the decentralized nature of the Internet requires original solutions to the problems of access: no single organization is providing all the requisite cataloging and indexing data for the wide variety of media available electronically. It is hoped that the systems for organizing information which were developed in the last century will not be ignored, that their design flaws will not be replicated, and that our increasing knowledge of human factors will be incorporated into systems for indexing the Internet.*<sup>183</sup>

Även Aluri, Kemp och Boll har liknande åsikter och menar att man inte får förbise de möjligheter och den kunskap som finns om beprövade klassifikationssystem: ”Existing bibliographical classification schemes provide a mine of information and ideas that should not be overlooked by the designers of online catalogs or of IR systems in general.”<sup>184</sup> Weinberg påpekar också att icke-etablerade system förr eller senare kommer att stöta på många av de problem som användarna och utvecklarna av de etablerade bibliografiska klassifikationssystemen redan känner till.

En av flera förklaringar till varför man använder olika typer av klassifikationssystem vid upprättandet av ämneshierarkiska söktjänster kan vara de kulturella skillnader som råder mellan institutioner och företag som utvecklar söktjänster på webben. Denna fråga ligger

---

<sup>183</sup> Weinberg, Bella Hass, (1996).

<sup>184</sup> Aluri, Rao, Kemp, D. Alasdair och Boll, John J., (1991), s. 115.

lite utanför syftet med denna uppsats men den är ändå intressant att fundera kring. Vi tror att anledningen till att man väljer ett visst klassifikationssystem kan bero på vilken syn man har på kunskapsorganisation och vilken organisationskultur man är del utav. Det finns en tendens bland dagens söktjänster att de som är kommersiella använder ett icke-etablerat klassifikationssystem medan de som har en biblioteksorganisation som huvudman använder etablerade system. Tjänsternas olikheter kan delvis bero på de olika kulturella skillnaderna som finns hos institutioner inom till exempel utbildningsvärlden och hos vinstdrivande företag. Vid bibliotek finns ett förtroget sätt att organisera kunskap på och det kan då vara naturligt att fortsätta använda detta. Det kan underlätta för användaren om hela biblioteket som informationssystem är strukturerat utifrån samma system så att hylluppställning, OPAC, andra databaser samt webbaserad söktjänst är integrerade med varandra. Vänder sig en söktjänst dessutom till den akademiska världen förefaller det naturligt att använda ett klassifikationsystem som delar in sina klasser efter disciplin. För ett icke-etablerat system med kommersiella syften kan ett icke-etablerat system vara fördelaktigast. Kommersiella tjänster är beroende av sina användare och inriktar sig därför på att locka så många användare som möjligt. Därför kan det vara av intresse att snabbt kunna inkludera nya populära ämnen samt försöka tillhandahålla en ämneshierarki vars kategorier motsvarar användarnas språkbruk och uppfattningar.

#### **7.4.1 Fortsatt forskning**

Ett uppsatsarbete kräver alltid att man begränsar sitt ämne och går på djupet i några få problemställningar vilket gör att flera idéer inte kan utvecklas. Som avslutning ger vi här några förslag på fortsatt forskning inom området.

Webben utvecklas hela tiden och om utvecklingstakten fortgår är det intressant att skriva en liknande uppsats inom samma område redan inom ett par år. Vidare skulle det vara spännande att göra en uppsats baserade på kvalitativa intervjuer med några projektledare ansvariga för utveckling av webbaserade söktjänster. Man skulle då kunna studera deras idéer kring användare, klassifikationssystem och kunskapsorganisation i en webbkontext. Att på ett mer uttalat sätt än vad vi har gjort här fokusera på skillnader mellan kommersiella söktjänster och tjänster utan vinstsyfte skulle kunna vara föremål för en fördjupad undersökning. Även om de tekniska frågorna inte ingår i de huvudsakliga yrkesuppgifterna för bibliotekarier och informationsspecialister är det dock viktigt att inte bortse från detta område. Genom att vara medveten om de tekniska förutsättningarna kan möjligheterna tas till vara och hindren i bästa fall undvikas.

Användning av metadata är ett annat område som är värt att studera. Lyckas man genomföra en standardisering av metadata och dessutom implementera denna på något intelligent sätt i HTML-standarden har man nått långt. Vi ser framtiden an och hoppas att det kommer att skrivas mer om webben på Bibliotekshögskolan.

## 8 Sammanfattning

Denna uppsats handlar om kunskapsorganisation på webben. Uppsatsens syfte är att undersöka de problem och möjligheter som finns med organisering av digitala objekt på webben med hjälp av klassifikationssystem. I uppsatsen belyser vi följande frågeställningar:

- Vilka problem och möjligheter finns med klassificering av digitala objekt på webben?
- Vilka grundläggande karaktäristika har klassifikationssystem som används för ämneshierarkiska söktjänster i webbmiljö?
- Vilka skillnader finns mellan etablerade och icke-etablerade klassifikationssystem i en webbkontext?
- Vilka för- och nackdelar har de olika systemen?

Vi ger även en lägesrapport om de aktuella förutsättningar för kunskapsåtervinning på webben med utgångspunkt från användning av klassifikationssystem i ämneshierarkiska söktjänster. Som exempel på hur klassifikationssystem används i söktjänster idag görs en kort beskrivning av fyra ämneshierarkiska söktjänster. Metoden för att belysa frågeställningarna är kvalitativa litteraturstudier och texterna som ligger till grund för uppsatsen är hämtade både från webben och tryckta tidskrifter och böcker. I anslutning till metodavsnittet diskuteras ett källkritiskt förhållningsätt till källmaterialet och i synnerhet för artiklar hämtade från webben.

Mycket av svårigheterna kring webben består i att det inte finns några klara definitioner. Vi resonerar och problematiserar därför kring termer som är relevanta för uppsatsen. Termer som webben, digitala objekt, browsing och sökning definieras. Vi gör även en indelning av söktjänster i ämneshierarkiska respektive indexbaserade. Vi pekar på svårigheter med bibliografisk kontroll som beror på webbens och de digitala objektens karaktäristika samt bristen på standarder. I samma kapitel görs även en distinktion mellan teknisk och intellektuell åtkomst. Det är den sistnämnda formen av åtkomst, med tyngdpunkten på klassifikationssystem, som huvudsakligen diskuteras vidare.

Kapitlet om klassifikation ger en bakgrund till klassifikation och klassifikationssystem. Utifrån Birger Hjørland delas klassifikationssystem in i de två huvudsakliga kategorierna pragmatisk och adhoc-klassifikation. Till den första gruppen räknas etablerade klassifikationssystem som Dewey Decimal Klassifikation (DDK) samt Universell Decimal Klassifikation (UDK), vilka även beskrivs i separata avsnitt. I samband med denna beskrivning ges även exempel på hur två ämneshierarkiska söktjänster använder respektive klassifikationssystem. Förutom exempel på etablerade klassifikationssystem beskrivs även två söktjänster vars ämneshierarkier grundar sig på icke-etablerade klassifikationssystem.

I diskussionskapitlet fördjupas diskussionen kring användningen av etablerade och icke-etablerade klassifikationssystem. För- och nackdelar med användning av de båda

systemtyperna i en webbkontext tas upp. En annan tanke som utvecklas är uppfattningen om webben som ett distributionsystem med distribuerade informationsystem.

Resultatet av studien visar att den typ av klassifikationssystem som är lämpligast beror på i vilken kontext systemet skall användas i, vad tjänstens huvudsyfte är och vilken målgruppen är. Ekonomiska resurser, tekniska förutsättningar och krav på kvalitet är andra faktorer som måste beaktas. Webbens och de digitala objektens konstitution är ytterligare en aspekt att ta fasta på.

## 9 Käll- och litteraturförteckning

### 9.1 Tryckta källor och litteratur

**Aitchison**, Jean, (1992), Indexing languages and indexing // Handbook of special librarianship and information work / ed. Patti Dossett. - 6th ed. - London: Aslib. - S. 191-233. - ISBN: 0-85142-269-1

**Aluri**, Rao, Kemp, D. Alasdair, och Boll, John J., (1991), Subject analysis in online catalogs. – Englewood, Colo.: Libraries Unlimited. - ISBN: 0-87287-670-5

**Batty**, David, (1976), Library classification: one hundred years after Dewey // Major classification systems: the Dewey centennial: papers presented at the Allerton Park institute / ed. by Kathryn Luther Henderson. - Urbana-Champaign, Ill.: Forest Press. - S. 1-16. - ISBN: 0-87845-044-0

**Borg**, Tage, (1994), Internet@Sverige. - Stockholm: Bonnier datamedia. - ISBN: 91-644-0010-7

**Chan**, Lois Mai, (1994), Cataloging & Classification: an introduction. - 2nd ed. - New York: McGraw-Hill. - (Library Science Series). - ISBN: 0-07-113253-8

**Chang**, Shan-Ju och Rice, Roland E, (1993), Browsing: a multidimensional framework // Annual Review of information science and technology (ARIST), edited by Martha E. Williams. - Vol. 28. - S. 231-276

**Cleveland**, Donald och Cleveland, Ana, (1990), Introduction to indexing and abstracting. - Englewood, Colo.: Libraries Unlimited. - ISBN: 0-87287-677-2.

**Comaromi**, John, (1976a), The historical development of Dewey classification system // Major classification systems: the Dewey centennial: papers presented at the Allerton Park institute / ed. by Kathryn Luther Henderson. - Urbana-Champaign, Ill.: Forest Press - S. 17-31. - ISBN: 0-87845-044-0

**Comaromi**, John, (1976b), The eighteen editions of the Dewey decimal classification. - Albany, NY: Forest Press Division. - ISBN: 0-910608-17-2

**Crawford**, Walt och Gorman, Michel, (1995), Future libraries: dreams, madness & reality. – Chicago: American Library Association. - ISBN: 0-8389-0647-8

**Dodd, David G.**, (1996), Grass-roots cataloging and classification: food for thought from World Wide Web subject-oriented hierarchical lists // *Library Resources and Technical Services*. - Vol. 40, Nr 3, July 1996 - s. 275-286. - ISSN: 0024-2527

**Dong, Xiaoying och Su, Lousie S.**, (1997), Search engines on the world wide web and information retrieval from the Internet : a review and evaluation // *Online & CD-rom review*. - Oxford: Learned Information. - Vol. 21, nr 2. - S. 67-81. - ISSN: 1353-2642

**Encyclopedia** of library and information science: vol. 3: bookmobiles to California / ed. Allen Kent och Harold Lancour. - (Encyclopedia of library and information science). - New York: Dekker. - Cop. 1970

**Harter, S. P.**, (1986), Languages for information retrieval // *Online information retrieval*. - Orlando: Academic Press. - S. 22-63. - (Library and information science (New York)). - ISBN: 0-12-328455-4

**Hjørland, Birger**, (1997), Information seeking and subject representation: an activity-theoretical approach to information science. - Westport, Conn.: Greenwood Press. - ISBN: 0-313-29893-9

**Hjørland, Birger**, (1995), Informationsvidenskabelige grundbegreber: Bd 2: Informationssystem - Zoom. - København: Danmarks Biblioteksskole. - ISBN: 87-7415-252-1.

**Katalogiseringsregler** för svenska bibliotek, (1990). - 2. uppl / red. Bodil Gustavsson. - Lund: Bibliotekstjänst. - ISBN 91-7018-324-4.

**Keenan, Stella**, (1996), Concise dictionary of library and informations sciences. - London: Bowker Saur. - ISBN: 1-85739-022-9

**Mandel, Carol A. och Wolven, Robert**, (1996), Intellectual access to digital documents: joining proven principles with new technologies // *Cataloging & Classification Quarterly*. - New York: Haworth P. - Vol. 22, nr 3/4. - S. 25-42. - ISSN: 0163-9374

**Marcella, Rita och Newton, Robert**, (1994), A new manual of classification. - Hampshire: Gower Publishing Limited. - ISBN 0-566-07547-4

**Mills, Jack**, (1977), The structure of a bibliographic classification // *Bliss bibliographic classification*. - 2nd ed. - London: Butterworths. - S. 35-47. - ISBN: 0-408-70835-2

**Munson, Kurt I.**, (1996), World Wide Web indexes and hierarchical lists: Finding tools for the Internet // *Computers in Libraries*. - Medford, N.J.: Information today. - Vol. 16(6), June. - S. 54-57. - ISSN: 1041-7915

**Nielsen, J.**, (1995), *Multimedia and hypertext: the Internet and beyond*. - Boston: AP Professional. - ISBN 0-12-518408-5.

**Palmquist, Ruth A.**, (1996), *A qualitative study of Internet metaphors // Proceedings / National Online Meeting: New York, May 14-16, 1996 / edited by Martha E. Williams*. - Medford, N.J.: Learned Information. - ISBN: 1-57387-026-9

**Rowley, Jennifer**, (1996), *Organizing knowledge: an introduction to information retrieval*. - Aldershot: Ashgate. - ISBN: 1-85742-005-5

**Rusch-Feja**, (1995), *Structuring subject information sources on the Internet // Wissen in elektronischen Netzwerken: strukturierung, erschliessung und retrieval von Informationsressourcen im Internet: eine Auswahl von Vorträgen der 19. Jahrestagung der Gesellschaft für Klassifikation Basel 1995 / red. Hans-Christoph Hobohm, Hans-Joachim Wätjen*. - Oldenburg: Bibliotheks- und Informationssystem. - S. 99-132. - ISBN: 3-8142-0534-0

**Shera, Jesse H.**, (1965), *Libraries and the organization of knowledge*. - London: Crosby Lockwood & Son. - (New librarianship series)

**Soergel, Dagobert**, (1985), *Organizing information: principles of data base and retrieval systems*. - New York: Academic Press. - ISBN 0-12-654261-9.

**Turner, Christopher**, (1987), *Organizing information: principles and practice*. - London: Bingley. - ISBN: 0-85157-379-7

**Universella** decimalklassifikationen, (1977). - 3., revid. utg. - Stockholm: Tekniska litteratursällskapet.

**Van der Werf-Davelaar, Titia**, (1995), *Organizing fileservers on the Internet: role of the library // Publications of Essen University Library*. - Vol. 18. - S. 194-206. - ISSN: 0931-7503

**Zhao, Dong**, (1997), *Managing electronic information: web page, subject folder, and onweb catalog // Proceedings / National Online Meeting: New York, May 13-15, 1997 / edited by Martha E. Williams*. - Medford, N.J.: Learned Information. - Cop. 1997. - S. 381-392. - ISBN: 1-57387-043-9

**Woodward, Jeannette**, (1996), *Cataloging and Classifying Information Resources on the Internet // Annual Review of Information Science and Technology (ARIST)*. - Vol. 31. - ISSN: 0066-4200.

## 9.2 Elektroniska källor

**Användningen** av Internet i Sverige // Svenska delen av Internet

[URL: <http://www.snus.se/internetutred/rapport/kap5.html>] 98-04-27.

**Arms**, William Y., (1995), Key concepts in the architecture of the digital library // D-Lib Magazine. - July. - ISSN: 1082-9873.

[URL:<http://www.dlib.org/dlib/July95/07arms.html>] 98-04-28

**Arms**, William Y., Blanchi, Christophe och Overly, Edward A., (1997), An architecture for information in digital libraries // D-Lib Magazine. - February. - ISSN: 1082-9873.

[URL:<http://www.dlib.org/dlib/february97/cnri/02arms1.html>] 98-03-05

**Boutell**, Thomas, World Wide Web Frequently Asked Questions (With Answers, of Course!)

[URL: <http://www.boutell.com/faq/oldfaq/index.html>] 98-03-05.

### **BUBL**

About BUBL 5:15

[URL: <http://link.bubl.ac.uk:80/about515.htm>] 98-04-16

BUBL Admin: Frequently Asked Questions about BUBL

[URL: <http://bubl.ac.uk/admin/faq.htm>] 98-04-07.

BUBL Admin: BUBL Funding and Purpose

[URL: <http://www.bubl.ac.uk/admin/purpose.htm>] 98-04-09.

BUBL LINK: about LINK

[URL: <http://link.bubl.ac.uk:80/service.htm>] 98-04-09

**Burden**, Peter och Wallis, Jon, Towards a classification-based approach to resource discovery on the web

[URL: <http://scitsc.wlv.ac.uk/wwlib/position.html>] 98-04-28.

**Callery**, Anne, (1996), Yahoo! Cataloging the Web

[URL: <http://www.library.ucsb.edu/untangle/callery.html>] 98-03-19.

**Coyle**, Karen, (1995), Access: not just wires // LIBRES, Library and Information Science Research Electronic Journal. - ISSN 1058-6768.

[URL: [gopher://gopher.fis.utoronto.ca:70/00/library/ejournal/libres/V5/N1/coyle.5n1](mailto:gopher://gopher.fis.utoronto.ca:70/00/library/ejournal/libres/V5/N1/coyle.5n1)] 98-03-18.

**De Bra**, P.M.E., Finding information on the web

[URL: <http://wwwis.win.tue.nl/~debra/cwi-qw/article.html>] 97-11-04.

**Kahn**, Robert och Wilensky, Robert, (1995), A framework for distributed digital object

services

[URL: <http://www.cnri.reston.va.us/home/cstr/arch/k-w.html>] 98-08-28.

**Koch**, Traugott, (1996a), Internet search services.

[URL: <http://www.ub2.lu.se/tk/demos/DO9603-meng.html>] 98-04-28.

**Koch**, Traugott, (1996b), Specification for resource description methods Part 3. The role of classification schemes in

[URL: [http://www.ub2.lu.se/desire/radar/reports/D3.2.3/class\\_v10.html](http://www.ub2.lu.se/desire/radar/reports/D3.2.3/class_v10.html)] 98-04-28.

**Koch**, T., Ardö, A., Brümmer, A. och Lundberg, S., (1996), The building and maintenance of robot based internet search services: A review of current indexing and data collection methods

[URL: <http://www.ub2.lu.se/desire/radar/reports/D3.11/tot.html>] 98-04-28.

**Kurth**, Martin och Peters, Thomas A., (1995), Browsing informations systems: an extensive annotated bibliography of the literature

[URL: [http://ai.iit.nrc.ca/II\\_public/BrowsingReferences/](http://ai.iit.nrc.ca/II_public/BrowsingReferences/)] 97-11-03

**Lynch**, Clifford, Identifiers and their role in networked information applications

[URL: <http://www.arl.org/newsltr/194/identifier.html>] 97-11-03.

### **Network Wizards**

[URL: <http://www.nw.com>]. 98-04-27.

Domain Survey Notes

[URL: <http://www.nw.com/zone/WWW/notes.html>] 98-04-27.

**Nielsen**, Jacob, How Users Read on the Web

[<http://www.useit.com/alertbox/9710a.html>] 98-03-30.

**Nielsen**, Jacob, Report From a 1994 Web Usability Study

[URL: [http://www.useit.com/papers/1994\\_web\\_usability\\_report.html](http://www.useit.com/papers/1994_web_usability_report.html)] 98-03-30.

**NISS** Directory of Networked Resources

NISS Directory of Networked Resources

[URL: <http://www.niss.ac.uk/subject/>] 98-04-02.

NISS Directory of Networked Resources: Help

[URL: [http://www.niss.ac.uk/help/subject\\_help.html](http://www.niss.ac.uk/help/subject_help.html)] 98-04-05.

NISS Directory of Networked Resources: UDC in brief

[URL: <http://www.niss.ac.uk/resource-description/udcbrief.html>] 98-04-28

**Nürnberg**, Peter, J, Digital libraries: issues and architectures

[URL: <http://csdl.tamu.edu/csdl/DL95/papers/nuernberg/nuernberg.html>] 98-05-14.

**OCLC**

**DDC translations**

[URL: [http://www.oclc.org/oclc/fp/about/trans\\_frameset.htm](http://www.oclc.org/oclc/fp/about/trans_frameset.htm)] 98-04-27

**Dewey Decimal System Home Page**

[URL: <http://www.oclc.org/fp/>] 98-01-09

**OCLC / About the DDC / DDC 21 Summaries (level 3)**

[URL: <http://www.oclc.org/oclc/fp/about/ddc21sm3.htm#350>] 98-04-27.

**OCLC / About the DDC / Dewey Decimal Classification (DDC 21) Summaries**

[URL: <http://www.oclc.org/oclc/fp/about/ddc21sm1.htm>] 98-04-27.

**OCLC / About DDC / Expanded intro**

[URL: <http://www.oclc.org/oclc/fp/about/expand.htm>] 98-04-27

**Oehler, Angela, (1996), Browsingdienste im Internet.**

[URL: <http://userpage.fu-berlin.de/~angela/bond/browsing.htm>] 98-04-28

**O'Neill, Edward T., (1997), Characteristics of web accessible information // 63<sup>rd</sup> IFLA General Conference - Conference programme and proceedings - August 31- September 5, 1997.**

[URL: <http://www.nlc-bnc.ca/ifla/IV/ifla63/63onee.htm>] 98-04-28.

**Nordic Metadata Project**

[URL: <http://linnea.helsinki.fi/meta/>] 98-03-18.

**Punkt se**

**Om redaktionell kontroll**

[URL: <http://www.punkt.se/om/hjalp/>] 98-04-13.

**Om registrering i Punkt se**

[URL: <http://www.punkt.se/registrera/>] 98-04-13.

**Punkt se hjälpsida**

[URL: <http://www.punkt.se/hjalp/>] 98-03-12

**Steinberg, Steve G., (1996), Seek and ye shall find (maybe)**

[URL: <http://www.wired.com/wired/4.05/features/indexweb.html>] 98-04-28.

**Sollins, K. och Masinter, L., (1994), Functional requirements of Uniform Resource Names. - December, RFC 1737**

[URL: <http://ds.internic.net/rfc/rfc1737.txt>] 98-04-27.

**Svenska datatermgruppen**

[URL: <http://www.nada.kth.se/dataterm/>] 98-04-27.

**Svenska delen av Internet**

[URL: <http://www.snus.se/internetutred/>] 98-05-01.

**Weinberg, Bella Hass, (1996), Complexity In Indexing Systems – Abandonment And**

Failiure: Implications For Organizing The Internet // ASIS '96 : proceedings of the 59th ASIS Annual Meeting, October 19-24.

[URL: <http://www.asis.org/annual-96/ElectronicProceedings/weinberg.html>] 98-04-27.

Se även Weinberg, Bella Hass, (1996), Complexity In Indexing Systems – Abandonment And Failiure: Implications For Organizing The Internet // ASIS '96 : proceedings of the 59th ASIS Annual Meeting, Baltimore, Maryland, October 21-24, 1996 / ed. Steve Hardin. - Medford, N.J.: Information Today. - Cop. 1996. - (Proceedings of the ASIS annual meeting ; 33). - S. 84-90. - ISBN: 1-57387-037-4.

### **Yahoo!**

Yahoo!

[URL: <http://www.yahoo.com>] 98-03-12.

Now Scandinavia can Yahoo! too: Yahoo! introduces localized Internet guides in Denmark, Norway and Sweden.

[URL: <http://www.yahoo.com/docs/pr/release129.html>] 98-02-05.

### **Yahoo! Sverige**

Yahoo! Sverige

[URL: <http://www.yahoo.se/>] 98-04-27.

Yahoo!s historia

[URL: <http://www.yahoo.se/docs/pr/history.html>] 98-04-16.

Yahoo! erövrar Sverige

[URL: <http://www.yahoo.se/docs/pr/release1.html>] 98-02-05.

Välkommen till Yahoo! Sverige

[URL: <http://www.yahoo.se/docs/info/welcome.html>] 98-02-05.

Yahoo! Sverige Nyheter

[URL: <http://www.yahoo.se/nyheter/>] 98-04-16.

Yahoo!s reserverade kategorier

[URL: <http://www.yahoo.se/docs/info/reserved.html>] 98-03-24.

Yahoo! Hjälp

[URL: <http://www.yahoo.se/docs/info/help.html>] 98-03-25.

Yahoo! Specialare

[URL: <http://www.yahoo.se/docs/info/features.html>] 98-04-16.

**Webster**, Kathleen och Paul, Kathryn, (1996), Beyond surfing: tools and techniques for searching the web

[URL: <http://magi.com/~mmelick/it96jan.htm>] 98-04-27.

### **World Wide Web Consortium**

[URL: <http://www.w3.org>] 98-03-03.