

# Varat och webben

PETER BÖRJESON



**KTH Datavetenskap  
och kommunikation**

# Varat och webben

P E T E R   B Ö R J E S O N

Examensarbete i medieteknik om 15 högskolepoäng  
vid Programmet för medieteknik  
Kungliga Tekniska Högskolan år 2010  
Handledare på CSC var Leif Dahlberg  
Examinator var Nils Enlund

URL: [www.csc.kth.se/utbildning/kandidatexjobb/medieteknik/2010/  
borjeson\\_peter\\_K10090.pdf](http://www.csc.kth.se/utbildning/kandidatexjobb/medieteknik/2010/borjeson_peter_K10090.pdf)

Kungliga tekniska högskolan  
*Skolan för datavetenskap och kommunikation*

**KTH** CSC  
100 44 Stockholm

URL: [www.kth.se/csc](http://www.kth.se/csc)

## Sammanfattning

Tekniker att systematisera information har funnits ända sedan människan började försöka styra över sitt liv. Genom historien har olika systematiseringstekniker konstruerats i försök att systematisera information.

Denna uppsats undersöker hur kunskap om tidigare systematiseringstekniker i historien kan användas för att utveckla dagens sökverktyg i hur de systematiserar och skapar mening av information på webben.

Först ges en introduktion till de två systematiseringstekniker som dominerat modern tid, *Kortindex* skapat av Carl von Linné i mitten av 1700-talet och *Principle of Provenance* som introducerades 1881 vid Privy State Archive i Berlin. För att sedan få en tydlig bild av skillnaden mellan de två teknikerna används systemteorin.

För att ge en bakgrund till hur människor skapar mening av information introduceras Heidegger och Gadamer och deras tankar om hermeneutiken (läran om tolkning) och ontologin (läran om varat).

Sedan undersöks sökverktygen *Yahoo!*, *Google* och *The Wayback Machine* med fokus på vilken möjlighet sökverktygen ger att söka i tiden jämfört med att söka på ämne och vilken möjlighet sökverktygen ger att sätta sig in i en annan historisk kontext genom att kombinera de två typerna av sökning.

## Abstract

Techniques to systematize information have existed ever since man began to try to control his life. In attempt to systematize information humans have throughout history constructed different techniques. This paper examines how past techniques in history can be used to develop the current way search engines systematize and create meaning of information on the web.

The paper begins with an introduction to the two techniques that have dominated modern times: *index cards* created by Carl Linnaeus in the middle of 1700s and *the Principle of Provenance*, which was introduced in 1881 at the Privy State Archive, Berlin. To get a clear picture of the difference between the two techniques system theory is used.

To provide a background to how meaning is created, Heidegger and Gadamer are introduced and their thoughts on hermeneutics (Science of interpretation) and ontology (the study of being).

Then search engines *Yahoo!*, *Google* and *The Wayback Machine* are investigated focusing on the possibilities they provide in search in time compared to search on topic and to make possible to place oneself into a different historical context by combining the two types of search.



# Innehållsförteckning

1. Inledning .....	1
1.1 Mål och syfte och Problemformulering.....	1
1.2 Uppsatsens uppbyggnad och Metod .....	2
2. Systematisera information .....	3
2.1 Bakgrund .....	3
2.2 Kortindex.....	3
2.3 Principle of Provenance.....	3
2.4 Bakgrund .....	4
2.5 Systemteoretisk analys .....	4
2.6 Sammanfattning .....	5
3. Skapa mening av information.....	6
3.1 Bakgrund .....	6
3.2 Hermeneutik .....	7
3.3 Heidegger och varaförståelse .....	7
3.4 Sammanfattning .....	8
4. Information på webben .....	8
4.1 Om undersökningen.....	9
4.2 <i>Yahoo!</i> .....	9
4.2.1 Förstasidan .....	10
Figur 1. www.yahoo.com .....	10
4.2.2 Avancerad sökning .....	11
Figur 2. Avancerad sökning .....	11
4.2.3 "within a year" och "anytime" .....	12
Figur 3. Sökresultat av avancerad sökning på KTH, "Within a year" och "anytime" .....	12
4.3 <i>Google</i> .....	13
4.3.1 Förstasidan .....	13
Figur 4. www.google.com.....	13
4.3.2 Avancerad sökning .....	14
Figur 5. Avancerad sökning .....	14
4.3.3 Vanlig sökning, tidssökning och länksökning .....	15
Figur 6. Sökresultat för "kth" .....	15
Figur 7. Sökresultat för "kth" från perioden 2008-7-5 till 2009-7-5 .....	16
Figur 8. Sökresultat för "kth.se" från perioden 1970-1-1 till 1971-1-1.....	16

Figur 9. Sökresultat för länksökning på "kth.se" .....	17
Figur 10. Sökresultat för länksökning på "kth" 2008-7-5 till 2009-7-5 .....	17
4.4 <i>The Wayback Machine</i> .....	18
4.4.1 Förstasidan .....	18
Figur 11. Waybackmachine.org .....	18
4.4.2 Avancerad sökning .....	19
Figur 12. Avancerad sökning .....	19
4.4.3 Historia? .....	20
Figur 13. Sökresultat för "kth.se" .....	20
Figur 14. Sökresultat för "kth.se" från November 1996.....	20
4.5 Slutsats.....	21
5. Slutsats .....	22
Litteraturlista .....	25

---

# 1. Inledning

Tekniker att systematisera information har funnits ända sedan människan började försöka styra över sitt liv. För att hantera stora mängder information behövs ett sätt att systematisera informationen och i försök att systematisera har människan genom historien konstruerat olika systematiseringstekniker.

Sättet att systematisera information är inte något som finns inbyggt i själva informationen och i modern historia<sup>1</sup> har två skilda tekniker dominerat: *Kortindex*, utvecklat av Carl von Linné i mitten av 1700-talet och som en reaktion på den, *Principle of Provenance* som introducerades vid Privy State Archive i Berlin år 1881.

Frågan hur information bör systematiseras bygger på uppfattningen om hur människor skapar mening av information. Systematiseringsteknikerna *Kortindex* och *Principle of Provenance* skiljer sig i uppfattningen om vilken information som är meningsfull och vad som utgör informationens kontext.

Många texter om medier och samhället idag beskriver dagens samhälle som ett informationssamhälle där informationsmängden växer okontrollerat. Ofta nämns nödvändigheten av att systematisera informationen för att kunna kontrollera och skapa mening av denna nya informationsmängd. Att systematisera stora mängder information och att försöka definiera vad i informationen som är meningsfullt är ingenting nytt idag jämfört med tidigare informationssamhällen i historien. Teknikutvecklingen och möjligheten att sprida information via Internet<sup>2</sup> har lett till en aldrig tidigare skådad ström av information. Problemet att skapa mening av information på webben<sup>3</sup> är dock inte något nytt problem vilket samhället aldrig tidigare ställts inför. Istället för att se hantering av information på webben som ett nytt problem som kräver nya metoder har jag valt att jämföra de olika idéer om hur information bör systematiseras som funnits i historien med de sätt som information systematiseras idag på webben. De tekniker som används idag för att systematisera och skapa mening av information på webben är s.k. sökverktyg som brukar delas in i sökmotorer och webbkataloger. De första sökmotorerna introducerades i början av 1990-talet och även om det skapats nya sökverktyg sedan dess så liknar de varandra i många avseenden och bygger till stor del på tekniken hos de tidigaste söktjänsterna.

*Google* har som mål att ”organize the world's information and make it universally accessible and useful” (Google.com). Idén är inte ny utan har uppkommit tidigare i historien, men med olika förslag på lösningar.

## 1.1 Mål och Syfte och Problemformulering

Jag vill med hjälp av systemteorin undersöka skillnaden i hur de två systematiseringsprinciperna *Kortindex* och *Principle of Provenance* systematiserar information. Sen vill jag utifrån hermeneutiken (läran om tolkning) och ontologin (läran om varat) föra en diskussion om hur människor skapar mening av information. Sedan vill jag utifrån de två systematiseringsprincipernas olika idéer om hur information bör ordnas föra en diskussion om

---

<sup>1</sup> Tidsperiod från upplysningen (ca år 1750) till nutid. Räknas ibland också som tiden efter medeltiden

<sup>2</sup> Internet är nätverket som består av datorer och fiberoptiska kablar runt hela jorden.

<sup>3</sup> Webben består av webbsidor som finns på en viss dator och görs tillgängliga via Internet.

---

de olika metoder som söktjänsterna *Yahoo!* (år1994), *Google* (år1996) och *The Wayback Machine* (år1996) använder för att systematisera information. Utifrån de olika sätt som *Google*, *Yahoo!* och *The Wayback Machine* systematiserar information, och hur vi människor skapar mening av information, vill jag slutligen undersöka likheterna mellan informationssökning på webben och i tidigare informationssamhället i historien för att se hur vetenskapen om tidigare systematiseringstekniker kan användas för att utveckla dagens söktjänster?

Trots flera andra söktjänster har jag valt att titta på två av de vanligaste, *Google* och *Yahoo!*, och utöver dem söktjänsten *The Wayback Machine*.

## 1.2 Uppsatsens uppbyggnad och Metod

Uppsatsen är indelad i tre delar: systematisering av information, att skapa mening av information och information på webben.

Del 1. Systematisering av information:

Första delen beskriver systematiseringsteknikerna *Kortindex* och *Principle of Provenance* och ställer frågan om vad som utgör kontexten i en samling information och om det finns ett optimalt sätt att systematisera information. För att få en tydlig bild av skillnaden mellan de två teknikerna används systemteorin som beskriver hur delarna i ett system kopplas ihop och hur delarna tillsammans fungerar som en helhet. Genom att analysera hur de olika systematiseringsteknikerna definierar och kopplar ihop delarna beskrivs skillnaden i själva systematiseringsprocessen. Hur delarna skapar en helhet utgör bakgrunden till de två teknikernas olika förutsättningar att skapa mening av information.

Del 2. Skapa mening av information:

Andra delen ställer frågan om hur människor skapar mening av information och presenterar hermeneutiken, (läran om tolkning), och ontologin (läran om varandet). Idéerna om varat baseras på Martin Heideggers filosofi, och till viss del Hans-Georg Gadamers, och ger absolut inte en full bild av deras filosofi utan fungerar mer som inspiration till en diskussion om det som Heidegger kallar varaförståelse.

Del 3. Information på webben:

Tredje delen beskriver först hur sökverktyg generellt systematiserar information, sedan undersöks sökverktygen *Google*, *Yahoo!* och *The Wayback Machine*. De olika teknikerna som sökverktygen använder är bara delvis kända men det går att dra vissa slutsatser genom att se vilka sökmöjligheter som ges och genom att utföra olika sökningar. I undersökningen har jag fokuserat på att undersöka vilken möjlighet sökverktygen ger att söka i tiden jämfört med sökning på ämne och möjligheten att sätta sig in i en annan historisk kontext genom att kombinera de två typerna av sökning. Slutligen, baserat på Del 1 och Del 2, beskrivs hur dagens sökverktyg kan utvecklas.



---

## 2. Systematisera information

*Genom historien har människan konstruerat olika tekniker för att systematisera information. Hur en samling information ska systematiseras är inget som ges av själva informationen och i historien har information systematiserats olika beroende på olika uppfattningar om hur informationens kontext uppkommer.*

*Kapitlet ifrågasätter uppfattningen om att det skulle finnas ett optimalt sätt att systematisera information och tar upp hur uppfattningen av vad som väver samman information ändrats i historien och beskriver de två systematiseringsteknikerna som dominerat modern tid, Kortindex (ca år 1750), och Principle of Provenance (år 1881).*

*De två teknikerna står i motsatsförhållande till varandra och genom att studera de två teknikerna ur ett systemteoretiskt perspektiv fås en tydlig bild av skillnaden i hur de två teknikerna kopplar samman information och hur informationen interagerar som en helhet.*

### 2.1 Bakgrund

Systematiseringsprinciper har funnits ända sedan människan började försöka styra över sitt liv. För att hantera stora mängder information behövs ett sätt att systematisera informationen och genom historien har människan konstruerat olika systematiseringstekniker. Att betrakta systematisering av information som en teknik beror på att sättet att systematisera en samling information inte är något som finns inbyggt i själva informationen, "Classification systems are not discovered, but designed" (Colomb, 2002).

### 2.2 Kortindex

Systematiseringstekniken *Kortindex* utvecklades av Carl von Linné (1707-1778) och hade sitt ursprung i behovet av att föra in information i en bestämd ordning och samtidigt kunna integrera den ordningen med ny information. I ett kortindex delas informationen in på separata blad beroende på ämne, vilka sedan kan kompletteras och omfördelas. Kortindextekniken lagrar information i en viss fast ordning som samtidigt går att uppdatera genom att inom den ursprungliga ordningen lägga till och flytta information. Kortindextekniken delar in informationen beroende på ämne och skapade ett nytt sätt att betrakta information där informationen inte längre var fixerad vid en linjär tidsskala. Naturen var inte längre linjär utan kunde klassificeras in i ett övergripande system oberoende av tid, "Although a seemingly mundane and simple innovation, Linnaeus' use of index cards marks a major shift in how eighteenth-century naturalists thought about the order of nature" (Mueller-Wille, 2009).

### 2.3 Principle of Provenance

År 1881 vid Privy State Archive i Berlin introducerades den s.k. Provenienzprinzip, eller *Principle of Provenance*. Enligt den systematiseringstekniken ska information arrangeras efter dess ursprungliga ordning. Istället för att dela upp insamlad information i större samlingar så bygger *Principle of Provenance* på att bevara den ordning i vilken informationen uppkommit. Det relevanta är varifrån och framförallt när ett dokument uppkommit, eftersom det är i den ursprungliga ordningen som informationens kontext uppkommit.

*Principle of Provenance* bygger på uppfattningen att insamlade dokument inte ska arrangeras beroende på ämnesinnehåll eftersom, "a subject arrangement is alien to their nature"

---

(Schellenberg, 1965). Meningen med att systematisera information är istället att göra det möjligt att återskapa de förhållanden som skapat informationen: de medier som hjälpte till att producera den, den verksamhet den en gång var en del av och den teknologi som var avgörande för att den kunde uppstå (Spieker, 2008). Målet med att bevara den ursprungliga ordningen var att skapa ett objektivt förhållande till den ursprungliga källan för att sen kunna återskapa den.

## 2.4 Systemteori

För att få en tydlig bild av hur de två systemiseringsprinciperna skiljer sig i sättet att systematisera information använder jag systemteorin. Enligt systemteorin definieras ett system som "... a set of *objects* together with *relationships* between the *objects* and between their *attributes* related to each other and to their *environment* so as to form a *whole*" (Schoderbek, 1998).

Systemteorin bygger på att ett system är en helhet som inte kan tas isär utan att förlora väsentliga egenskaper. Ett system har en karakteristik som är mer än summan av delarna och på grund av samspelet mellan delarna försvinner ett systems egenskaper om en del tas bort och därför måste det studeras som en helhet. (Schoderbek, 1998).

## 2.5 Systemteoretisk analys av *Kortindex* och *Principle of Provenance*.

**Set:** Set är en väl definierad samling föremål, t.ex. en samling dokument eller webbsidor.

### Kortindex:

Det som avgör om ett dokument hör till en samling baseras på dokumentets ämnesinnehåll. En samling skapas genom att identifiera liknande egenskaper hos en grupp dokument och indela informationen i ämneshierarkier. Kortindextekniken bygger på att samlingar splittras och delas upp i en övergripande samling där kontexten utgörs av ämne.

### Principle of Provenance:

Samlingar sätts ihop baserat på tiden och platsen för dokumentens ursprung eftersom det är de tidshistoriska förhållandena som samlingen ska återskapa. Samlingar bevaras i sin ursprungliga ordning eftersom kontexten skapas av tiden då samlingen uppkom.

**Objects:** Objekten är de delar som systemet består av, t.ex. ett dokument eller en webbsida.

### Kortindex:

Objekten är frikopplade från tiden och delas in i större system baserat på deras ämnesinnehåll eftersom kontexten utgörs av objekten själva. De objekt som delats in i helheten kan omgrupperas fritt och fortfarande vara en del av helheten.

### Principle of Provenance:

Objekten beskrivs av tiden och platsen för deras uppkomst, dvs. det är inte objekten i sig själva som är det intressanta utan kontexten uppkommer av förhållandena runt objekten. Objekten ses som resultatet av en viss tid och betraktas som i tiden diskreta enheter som har en fast plats i historien.

**Relationships:** Relationer är länkarna mellan objekten.

### Kortindex:

Länkarna bygger på informationen inuti objekten och objekten i sig själva och baseras på liknande drag såsom ett dokument eller en webbsidas ämne.

---

Principle of Provenance:

Länkarna bygger på information utanför objekten och baseras på tiden och platsen för ett dokumentets uppkomst. Länkarna mellan två skilda samlingar bygger på kronologisk tid.

**Attributes:** Attributen beskriver egenskaper hos objekten.

Kortindex:

Attributen beskriver objektets innehåll.

Principle of Provenance:

Attributen beskriver tid och plats.

**Environment:** Samtidigt som omgivningen är något yttre så påverkar den systemet. Det som systemet kan kontrollera räknas till systemet medan det som inte kan kontrolleras är en del av omgivningen. För ett system är växelverkan med omgivningen en förutsättning för dess överlevnad.

Kortindex:

Omgivningen utgörs av det som ännu inte är känt och målet är att indela hela omgivningen i systemet. Informationen som redan delats in i systemet betraktas som universella sanningar oberoende av tid och insamling av ny information utgår från den redan indelade informationen och ser vad i omgivningen som kan delas in i systemet.

Principle of Provenance:

Omgivningen utgörs av kronologisk tid och den insamlade informationen utgörs av diskreta enheter i tiden vilka växelverkar med omgivningen. Gränsen mot omgivningen är bredare dragen än i kortindexsystemet då allt i historien är relevant och utövar ett inflytande på systemet.

**Whole:** Helheten är större än summan av delarna. Helheten är en ram inom vilken delarna spelar en distinkt roll.

Kortindex:

Målet är att identifiera informationsobjekt i tillvaron och sedan organisera den insamlade informationen. Kortindextekniken delar in världen i två system, ett som utgör tillvaron (det okända), och ett där all världens information är samlad (det kända).

Principle of Provenance:

Information som samlats in i nuet utgör en helhet i sig men får sin innebörd när den tolkas mot historien som är den totala helheten. Idén att allt kan förklaras gentemot historien får följden att all information kan komma att bli användbar i framtiden, allt blir intressant och sparas på hög.

## 2.6 Sammanfattning

Kapitlet beskriver systematiseringsteknikerna *Kortindex* och *Principle of Provenance* och ställer frågan om vad som utgör kontexten i en samling information och om det finns ett optimalt sätt att systematisera information.

Hur en samling information ska systematiseras är inget som ges av själva informationen utan beror på skiftande uppfattningar om hur informationens kontext uppkommer. De olika tekniker som konstruerats för att systematisera information har speglat det då rådande synsättet om vad som väver samman information.

---

### Kortindex:

Målet är att identifiera informationsobjekt i tillvaron och sedan organisera den insamlade informationen. Kontexten baseras på objekten i sig själva och är till skillnad från *Principle of Provenance* oberoende av tid. Omgivningen utgörs av det som ännu inte är känt och målet är att indela hela omgivningen i systemet. Tekniken utgår från den redan indelade informationen och ser vad i omgivningen som kan delas in i systemet med målet att skapa en övergripande samling.

### Principle of Provenance:

Samlingar sätts ihop baserat på tiden och platsen för informationens ursprung med målet att återskapa de tidshistoriska förhållandena. Samlingar bevaras därför i sin ursprungliga ordning till skillnad från kortindextekniken som delar upp samlingar i en övergripande samling. Objekten beskrivs av tiden och platsen för deras uppkomst, dvs. det är inte objekten i sig själva som är intressanta och till skillnad från kortindex så uppkommer kontexten av förhållandena runt objekten. Objekten ses som resultatet av en viss tid och betraktas som i tiden diskreta enheter som har en fast plats i historien.

Länkarna mellan två skilda samlingar bygger på kronologisk tid medan kortindextekniken delar in informationen i kategorier i en övergripande samling oberoende av tid. Delarna i systemet består av diskreta enheter i tiden som växelverkar med omgivningen. Information som samlats in i nuet utgör en helhet i sig men får sin innebörd när den tolkas mot historien som är den totala helheten.

## 3. Skapa mening av information

*För att förstå hur människor systematiserar information behöver man undersöka hur människor skapar mening av information. Filosoferna Martin Heidegger (f. 1889) och Hans-Georg Gadamer (f. 1900) sysselsatte sig med frågan om varat och hur vi skapar vår tillvaro genom tolkning. Med utgångspunkt från deras filosofi ställer kapitlet frågan hur människor förstår och skapar mening av information. Först presenteras hermeneutiken, (läran om tolkning) och sedan Heideggers studier av varat, främst baserat på boken *Varat och Tiden* från 1927.*

### 3.1 Bakgrund

En vanlig uppfattning om hur mening skapas av information är att den uppstår när information samlats in och organiserats, ett antagande som kommer från det systematiska och metodiska tillvägagångssätt som kännetecknar vetenskapen. Heidegger skiljer på det arbete som utförs inom vetenskapen och tänkandet inom filosofin. Ett av Heideggers mer kända uttryck är, "Science does not think" (Heidegger, 1968), med vilket han menade att vetenskapen endast utför ett arbete inom vissa givna parametrar utan att ifrågasätta bakgrunden till deras uppkomst. Istället för att använda vetenskapens fokus på mätbarhet och effektivitet ansåg Heidegger att man för att förstå hur mening uppkommer behöver ställa de mer grundläggande frågorna om varat.

---

## 3.2 Hermeneutik

Hermeneutiken är en teori som beskriver hur vi skapar mening genom tolkning. Ursprungligen användes hermeneutiken som en metod för att förstå religiösa texter vilka trots att de skrevs i en annan tid fortsätter att skapa mening än idag. En text som funnits över flera århundraden är beroende både av det sammanhang i vilken den skrevs och i det sammanhang i vilken den existerar i nuet. Hermeneutiken motsäger inte att betydelsen finns i själva texten, men att om en viss person vid ett visst tillfälle hade en tolkning, så får texten en annan tolkning i en annan tid. En text får olika betydelser genom historien och dess mening bygger på en process där både textens historiska läge och historien själv spelar en avgörande roll. Vi tolkar ett historiskt sammanhang (delen) mot historien själv (helheten), där vår uppfattning av helheten utgår från nuet vilket ständigt ändras. En texts innebörd går inte att definiera i någon absolut mening utan är beroende av både dess ursprung och nutiden, dvs. det handlar inte om att försöka övervinna sin egen kultur för att återfå den ”rätta” tolkningen (Winograd, 1987).

Denna cirkulära process mellan ett visst ögonblick och historien beskrivs av den s.k. hermeneutiska cirkeln. För att skapa mening rör vi oss fram och tillbaka mellan delen och helheten. Med cirkulär menas inte att det är en statisk rörelse ”fram och tillbaka” mellan delen och helheten utan förhållandet dem emellan ändras hela tiden genom själva tolkningsprocessen. Heidegger skriver, ”Den begreppslighet, som hör till det varande som ska utläggas, kan utläggningen antingen hämta från detta varande självt, eller också kan den tvinga in detsamma i sådana begrepp som det varande enligt sin varaart motsätter sig” (Heidegger, 1992), med vilket menas att det går att komma någonstans genom tolkning. Genom att sätta in ett visst begrepp i ett annat ursprung ändras innebörden och tolkningsprocessen leds framåt.

Heidegger och Gadamer menade att det som tolkningen sker emot är våra förutfattade meningar, ”Utläggningen är aldrig någonsin ett förutsättningslöst uppfattande av något redan föregivet. När den specifika konkretion, som hör till utläggningen i bemärkelsen exakt textuttolkning (Textinterpretation), så gärna återropar sig på det som »står där», är det som »står där» ingenting annat än utläggarens egen självklara, Odiskuterade föremening” (Heidegger, 1992). Gadamer använder uttrycket horisont<sup>4</sup> för att beskriva denna förförståelse och menar att våra preliminära uppfattningar är något vi inte kan komma ifrån utan att de alltid finns där. Vi har alla vår historiska horisont som bestäms av vår kulturella bakgrund och historiska plats, och Gadamer menar att tolkning och meningsskapande alltid sker inom en horisont där vi genom att röra oss mellan nuet och det förflutna skapa förståelse genom att på så sätt sammansmälta våra historiska horisonter, “Understanding is not to be thought of so much as an action of one’s subjectivity, but as the placing of oneself within a process of tradition, in which past and present are constantly fused” (Gadamer, 1975).

Gadamer, och före honom Heidegger, tog hermeneutiken bortom analysering av texter och satte tolkning som grunden för mänsklig kunskap. På samma sätt som tolkning spelar en roll för en persons förståelse av en text så skapar den vår förståelse av världen som helhet. (Winograd, 1987)

## 3.3 Heidegger och varaförståelse

Heidegger skriver att syftet med *Varat och Tiden* är att undersöka varats mening och att förstå tiden som grunden för varat, ”när helst tillvaron Outtryckligt förstår och utlägger sådant som varat, gör den det med utgångspunkt från *tiden*. Tiden måste bringas fram i ljuset och i begreppsliga termer genuint fattas såsom horisonten för all varaförståelse och för varje sätt att utlägga varat” (Heidegger, 1992).

Heidegger menar att när vi tänker på varat och tiden så tänker vi på tiden som befintlighet och

---

<sup>4</sup> Uttrycket kommer från Nietzsche och har även använts av Husserl. Innebörden har skiftat något men Heidegger och Gadamer använder det i likhet med Husserl som en plats eller position som utgångspunkt för den hermeneutiska tolkningsprocessen.

---

närvaro i nuet då vi istället bör förstå tiden som horisonten för förståelse av varat. Att förstå varat i tiden istället för i nuet är precis som en text inte har en exakt innebörd utan tolkas mot historien så tolkar vi i varje ögonblick vår existens mot tiden. Att förstå varat i tiden betyder inte nödvändigtvis att se tiden som kalendariskt datum: "Även utan dylika "datumangivelser", är "nu", "sedan" och "den gången" mer eller mindre bestämt daterade. Om dateringens bestämdhet uteblir, innebär detta dock inte att daterbarhetens struktur skulle saknas eller vara godtycklig" (Heidegger, 1992). Istället för "klocktid" där tiden fungerar som numerisk bestämning så kommer tid först från den "naturliga" klockan: "först denna 'naturliga' klocka motiverar och möjliggör tillika framställning och användning av mer hanterliga klockor, närmare bestämt på så vis att de "konstgjorda" klockorna måste "ställas" efter de "naturliga" (Heidegger, 1992). Med "naturlig" tid menar Heidegger den tid som utgör vår tillvaro, något vi delar med de andra vilka vi delar världen med, att tillvaron existerar i form av *medvaro* med andra, "att väsendet i ombesörjandet av tiden *inte* ligger i användning av numeriska bestämningar vid datering" (Heidegger, 1992).

### 3.4 Sammanfattning:

För att förstå hur människor systematiserar information behöver man undersöka hur människor skapar mening av information. Heidegger och Gadamer tog den hermeneutiska idén om tolkning bortom analysering av texter och placerade tolkning som grunden för vår varaförståelse. De menade att varje person för att förstå världen ständigt deltar i tolkning. Möjligheten att skapa mening av vår vardag genom tolkning utgörs av tiden, där tiden är horisonten för vår varaförståelse. Heidegger skiljer på tid såsom den uppfattas inom vetenskapen respektive filosofin och menar tid som i kultur och vår kulturella bakgrund och inte tid som en linje vars intervall kan mätas med en klocka och som prioriterar tiden i nuet vilket präglat det vetenskapliga tänkandet.

## 4. Information på webben

*Hur fungerar sökverktyg på webben och hur liknar de tidigare tekniker att systematisera och skapa mening av information? Först ges en generell beskrivning av komponenterna (hårdvaran) som samlar in och systematiserar information på webben, något som i stort sett är lika för de flesta sökverktyg. Sedan introduceras och undersöks sökverktygen Yahoo!, Google och The Wayback Machine och de olika tekniker de använder för att systematisera informationen på webben och för att tolka vilken information som är meningsfull för den som söker.*

*De olika teknikerna som sökverktygen använder är bara delvis kända men det går att dra slutsatser genom att utföra olika sökningar. Sökresultaten skiftar något från dag till dag men på ett sett som inte påverkar själva undersökningen. Sökningarna är gjorda juli och augusti 2010. Jag har fokuserat på att undersöka vilken möjlighet sökverktygen ger att söka i tiden jämfört med sökning på ämne, och möjligheten att sätta sig in i en annan historisk kontext.*

---

## 4.1 Sökmotorer och Webbkataloger

Det finns två huvudtyper av sökverktyg på Internet, sökmotorer och webbkataloger. Sökmotorer hämtar sidor från webben och ordnar in dem maskinellt i index medan webbkataloger är indelade efter ämne och innehåller utvalda länkar till webbsidor inom ett specifikt ämnesområde som en redaktör har bedömt och granskat.

### Sökmotor

En sökmotor ett dataprogram som söker information i ett index som det byggt upp av webbsidor som det hämtat från webben och sparar i ett index. Söker alltså inte på webben utan i sitt index. En sökmotor har tre grundläggande komponenter, en sidhämtare eller sökspindel, en indexerare och ett sökgränssnitt.

Sökspindeln är den komponent i sökmotorn som hämtar sidor från webben. Den går igenom webbsidorna och lagrar deras innehåll i en databas. Efter att sökspindeln gått igenom en webbsida följer den länkarna på webbsidan för att hitta nya webbsidor. Spindeln söker ständigt igenom webben efter nya webbsidor och om en webbsida redan har besökts kontrollerar spindeln i så fall för hur länge sedan och om innehållet har uppdaterats.

Indexeringsprogrammet analyserar de webbsidor som hämtats av spindelprogrammet. indexerar ord och fraser från dokumenten, liksom metadata (information om dokumentets källa och dess innehåll) samt information om filtyp, filstorlek och tidpunkten då dokumentet hämtades ned. Är den del som systematiserar informationen.

Sökgränssnittet är sökmotorns ansikte utåt och är den komponent som söker i indexet. Sökgränssnittet har ett sökformulär med sökfunktioner och sökmöjligheter.

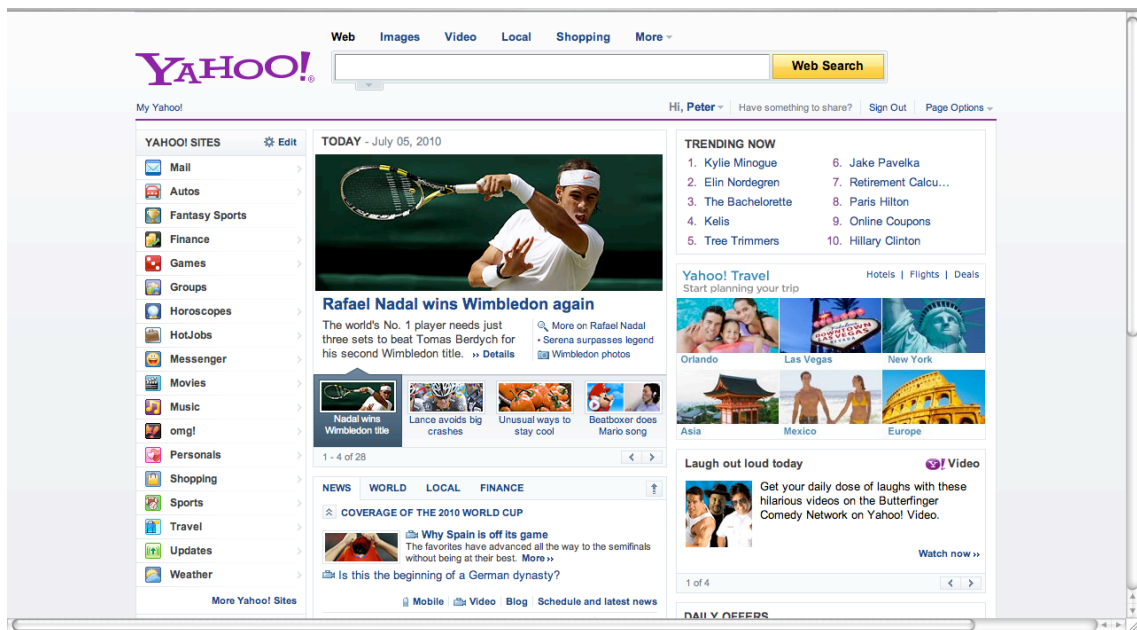
### Webbkatalog

En webbkatalog är en samling länkar inom ett eller flera bestämda ämnesområden. Vanligtvis skapas dessa av människor som valt ut länkarna efter vissa kriterier. Webbkataloger fokuserar på att ha ett bra urval av noggranna utvalda resurser med hög kvalitet men med problemet att redaktörerna inte hinner kvalitetsgranska och ämneskategorisera mer än en bråkdel av webben, eftersom den helt enkelt blivit för stor. En tydlig trend är att sökmotorn mer och mer har kommit att slå ut webbkatalogen som användarnas förstahandsval vad gäller söktjänst. Webben har helt enkelt blivit för stor för redaktörerna på webbkatalogerna ska hinna kvalitetsgranska och ämneskategorisera ens bråkdel av webben.

## 4.2 Yahoo!

Grunden till det som kom att bli söktjänsten *Yahoo!* lades 1994 av David Filo and Jerry Yang, då doktorander i elektroteknik vid Stanford University. Under namnet "Jerry and David's Guide to the World Wide Web" började de lista sina favoritplaster på webben. Antalet användare ökade snabbt och i ett försök att kontrollera alla nya tips om intressanta webbsidor började de dela in dem i olika kategorier, "Eventually, Jerry and David's lists became too long and unwieldy, and they broke them out into categories, then subcategories...and thus the core concept behind *Yahoo!* was born" (Yahoo.com, 2010-08-25). I slutet av 1994 hade *Yahoo!* fått en miljon sökningar och idag har de mer än 600 miljoner användare (Yahoo.com, 2010-08-25). *Yahoo!* har visionen att vara "the center of people's online lives by delivering personally relevant, meaningful Internet experiences" (Yahoo.com, 2010-08-25).

## 4.2.1 Förstasidan



(Figur 1. [www.yahoo.com](http://www.yahoo.com))

*Yahoo!* har både en sökmotor och en webbkatalog, och utöver det visar hemsidan nyheter, senaste nytt inom underhållning och kändisskvaller och de har en egen reseplanerare, och *Yahoo.com* kan därför kallas Webbportal. Till vänster på sidan finns *Yahoo!'s* webbkatalog som innehåller ett tjugotal kategorier. Det finns fler kategorier att välja mellan än vad som visas på hemsidan och sammanlagt är de cirka 100 kategorier som i sin tur innehåller subkategorier. De olika kategoriernas innehåll har samlats in och redigerats av *Yahoo!'s* personal, "Instead of relying on automated spider programs to build its database it maintains a staff of some 150 editors" (Glossbrenner, 1998). Sökmotorns index är uppbyggt både av *Yahoo!'s* och av *Google's* databas. Först söks relevanta webbsidor ur *Yahoo!'s* webbkatalog och sedan kompletteras sökningen med sidor ur *Google's* index. I de sökningar som görs i *Yahoo!'s* databas baseras en webbsidas popularitet framförallt på om den tillhör någon av kategorierna i *Yahoo!'s* webbkatalog, vilket främst baseras på webbsidans titel och utöver det på en kort sammanfattning av sidans innehåll. Sökningarna som görs i *Google's* databas använder bl.a. *Google's* teknik PageRank som licenserats av *Yahoo!*. PageRank bedömer en sidas relevans genom att se vilka sidor den länkar till och vilka sidor som i sin tur länkar till den. *Yahoo!* använder även tekniken "click-through tracking", som baserar en webbsidas popularitet på antalet användare som valt just den sidan bland alla sökresultat.



## 4.2.2 Avancerad sökning

YAHOO! SEARCH [Yahoo!](#) - [Search Home](#) - [Help](#)

### Advanced Web Search

You can use the options on this page to create a very specific search. Just fill in the fields you need for your current search.

**Show results with**

all of these words

the exact phrase

any of these words

none of these words


Tip: Use these options to look for an exact phrase or to exclude pages containing certain words. You can also limit your search to certain parts of pages.

**Updated**

**Site/Domain**

Any domain  
 Only .com domains  Only .edu domains  
 Only .gov domains  Only .org domains  
 only search in this domain/site:

Tip: You can search for results in a specific website (e.g. yahoo.com) or top-level domains (e.g. .com, .org, .gov).

**Creative Commons Search** 

Search only for Creative Commons licensed content  
 Find content I can use for commercial purposes  
 Find content I can modify, adapt, or build upon

Tip: This special Yahoo! Search finds pages that have content with a Creative Commons license. [Learn more...](#)

**File Format** Only find results that are:

**SafeSearch Filter** Applies when I'm signed in:

Filter out adult Web search results - SafeSearch On  
 Do not filter Web results (results may include adult content) - SafeSearch Off

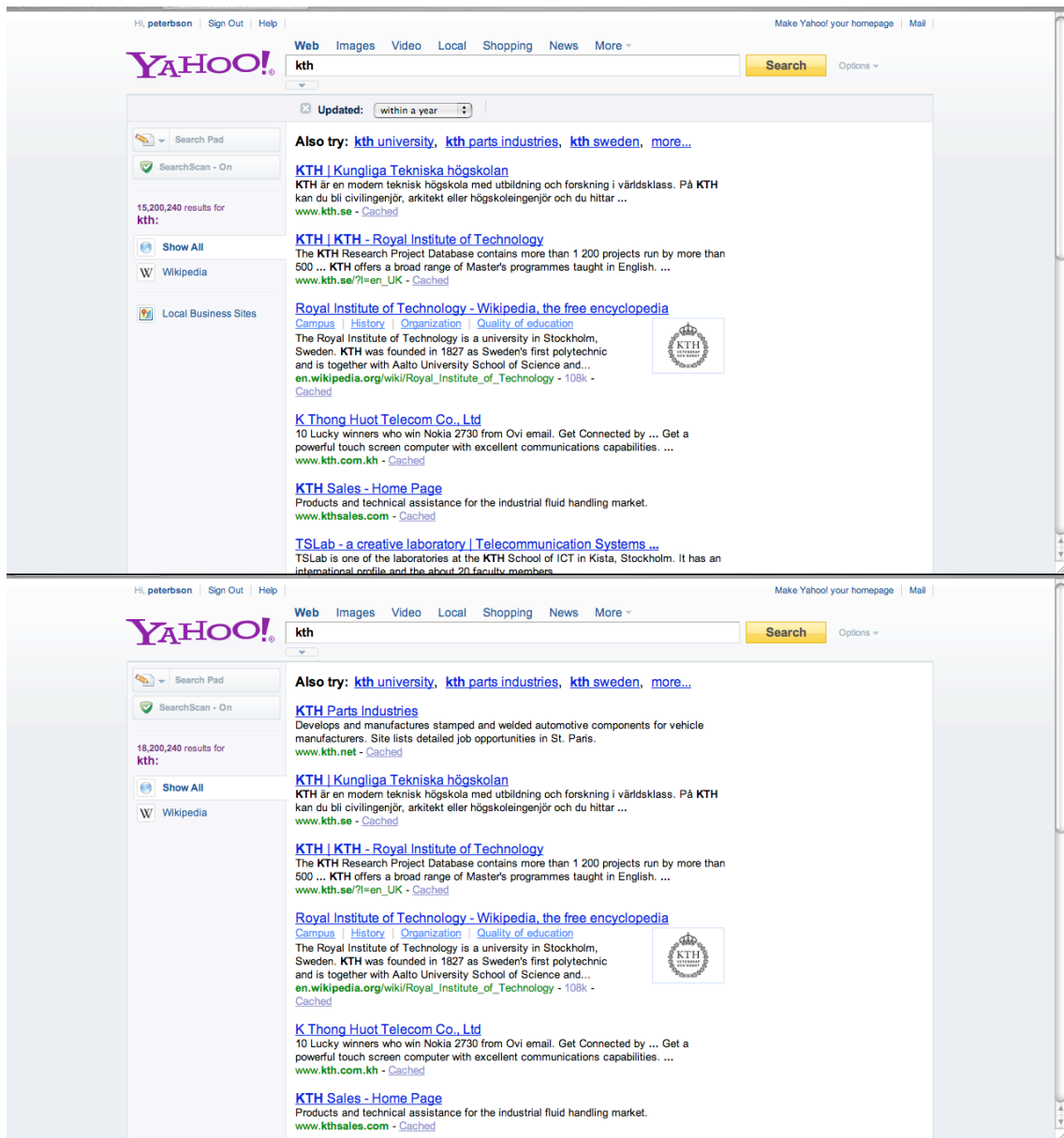
Note: Any user signed in on your computer as 18 or older can change this setting. We recommend periodically

(Figur 2. Avancerad sökning <http://search.yahoo.com/web/advanced>)

Gränssnittet för avancerad sökning visar att *Yahoo!*'s fokus ligger på att analysera webbsidor efter deras textinnehåll. Sökningen görs genom att söka på ord eller fraser, som det står längst upp på sidan, show results with: "all of these words", "the exact phrase", "any of these words" eller "none of these words".

*Yahoo!*'s sökverktyg ger ingen egentlig möjlighet att söka information i tiden då det endast går att välja tidsintervall för när webbsidor uppdaterats, dvs. när de besökts av sökspindeln, där det går att välja att söka på webbsidor som uppdaterats: "anytime", "within the past 3 months", "within the past 6 months" och "within a year". Om en sida uppdaterats de senaste 3 månaderna så visas samma version vid sökning på webbsidor som uppdaterats de senaste 6 månaderna, dvs. *Yahoo!* sparar inte äldre versioner.

## 4.2.3 "within a year" och "anytime"



(Figur 3. Sökresultat av avancerad sökning på KTH uppdaterad: "Within a year" och "anytime")

Sökning på ordet "kth" (Kungliga Tekniska Högskolan) bland de webbsidor i databasen som uppdaterats "Anytime" ger 18 200 240 resultat, vilket är 3 miljoner fler än sökning på de webbsidor som uppdaterats "within a year". De sex första sökningarna skiljer sig bara åt med ett resultat där sökning på webbsidor som uppdaterats "anytime" har kth.net som första resultat av vilket man kan dra slutsatsen att relevansen för kth.net rankas högt men att den inte uppdateras så ofta. Anledningen till att den rankas högt är antagligen att net är en större domän än se, 1 450 000 000 webbsidor jämfört med 253 000 000. På kth.net står det också att hemsidan inte uppdaterats på länge: "Revised: April 23, 2008" (www.kth.net, 2010-08-05).

Det går inte att veta vilka som är de 3 miljoner träffarna som skiljer de två olika sökningarna eller hur långt bak i tiden resultaten för "anytime" är, annat än att de inte uppdaterats det senaste året. Funktionen med de olika tidsintervallen för "updated" är inte att söka äldre information utan att begränsa urvalet: "Updated Limit your results to sites that have been updated within the

past 3, 6, or 12 months” (Yahoo.com, 2010-08-25).

Webbsidor som uppdaterats ”anytime” ger både webbsidor som nyligen uppdaterats och sådana som uppdateras mer sällan. De ger inte äldre information i form av äldre sparade kopior utan ”anytime” är själva defaultinställningen. Så även om det finns webbsidor från längre bak i tiden än ett år så går de inte att jämföra med yngre versioner för att undersöka hur en sida ändrats över tiden, eftersom det bara är den senaste versionen som finns sparad, de äldre försvinner när en webbsida uppdateras. Uppdatering av webbsidor kan vara missvisande eftersom många webbsidor uppdateras när de besöks, t.ex. en sida som inte uppdaterats de senaste 3 månaderna uppdateras automatiskt när den besöks, men det är den äldre icke uppdaterade versionen som står som uppdaterad, vilket är för att bara de webbsidor som används ska uppdateras.

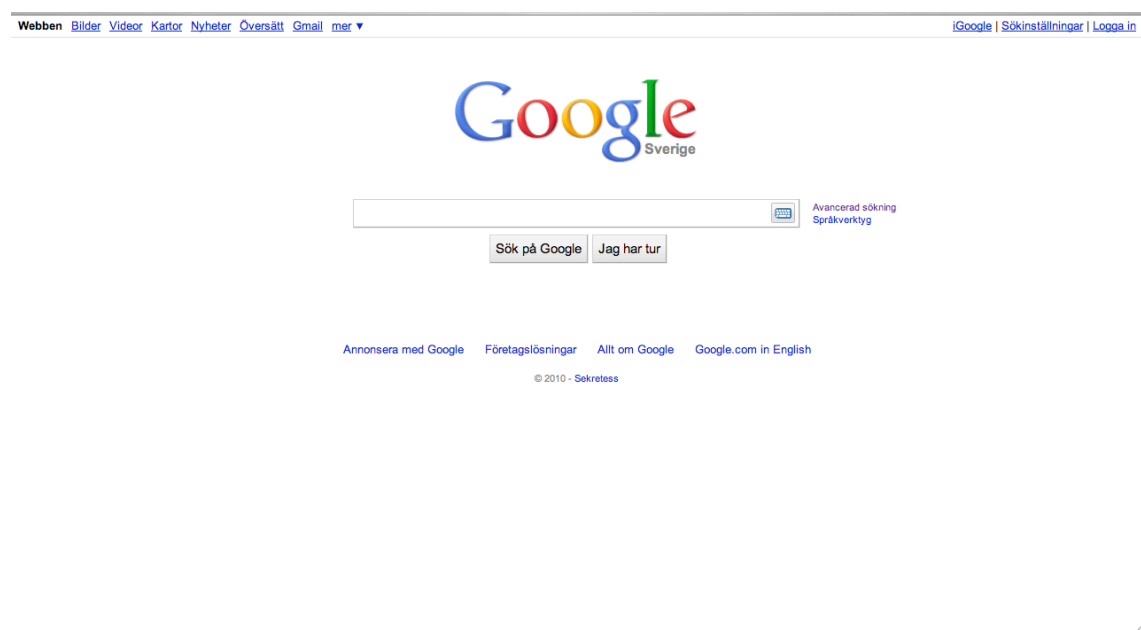
## 4.3 Google

*Google* började 1996 som ett forskningsprojekt av datateknikdoktoranderna Larry Page och Sergey Brin vid Stanford University. De sökverktyg som då fanns rankade resultaten beroende på hur många gånger ett visst sökord återfanns på en sida och målet med projektet var att skapa ett sökverktyg med ett nytt sätt att tolka relevansen hos webbsidor. Tekniken de skapade, PageRank (SidBetyg) analyserar länkarna mellan olika webbsidor och tolkar en webbsidas relevans beroende på antalet ut- och inlänkar.

Namnet *Google* kommer från en felstavning av den matematiska termen googol som introducerades 1938 och står för en etta följd av hundra nollor ( $10^{100}$ ). *Google* menar att “The name reflects the immense volume of information that exists, and the scope of *Google’s* mission: to organize the world's information and make it universally accessible and useful” (Google.com, 2010-08-25).

Idag innehåller *Google’s* index fler än en miljard webbsidor. (Google.com, 2010-06-25)

### 4.3.1 Förstasidan



(Figur 4. [www.google.com](http://www.google.com))

*Google* är framförallt en sökmotor även om de också har en webbkatalog. Meningen med den avskalade hemsidan är att visa att fokus ligger på informationssökning vilket de skriver om på

hemsidan med en uppenbar passning till *Yahoo!*: “*Google* ger ordning till nätet. *Google* skapades för att tillföra ordning i informationskaoset. Det är precis vad en sökmotor borde vara; inte en redigerad, begränsad katalog eller en lista med resultat som blivit utauktionerade till högstbjudande, utan en effektiv metod för att organisera Internet i gemenskap med dess egen struktur” (Google.com, 2010-06-25).

*Google* använder flera olika tekniker för att tolka vilken information som är meningsfull för den som söker. De olika teknikerna är en affärshemlighet men själva principerna bakom beskrivs på *Google*'s hemsida. Om deras mest omtalade teknik, länkanalystekniken PageRank står det: “Sidbetyg förlitar sig till Internets unika demokratiska natur genom att använda sig av dess väldiga länknät för att bedöma en enskild sidas värde. I grunden fungerar det genom att *Google* tolkar en länk från sida A till sida B som en röst, från sida A, på sida B. Men, *Google* tittar på mer än bara antalet röster, eller länkar till en sida, utan analyserar också sidan som lägger rösten. Röster lagda av sidor som själva är "värdefulla" väger tyngre och hjälper andra sidor att bli mer "värdefulla". Viktiga sidor av hög kvalitet erhåller en högre sidrankning” (Google.com, 2010-06-25).

Utöver PageRank använder *Google* andra tekniker: ”Givetvis så saknar viktiga sidor värde om de inte motsvarar din förfrågan. Därför kombinerar *Google* sidrankning med avancerade textjämförelsetekniker för att hitta sidor som är såväl viktiga som relevanta för din sökning. *Google* sträcker sig långt bortom hur många gånger en term förekommer på en sida och undersöker alla aspekter av sidans innehåll, (och innehållet på sidor som är länkade till den), för att avgöra om den motsvarar din förfrågan” (Google.com, 2010-06-25).

Utöver det analyserar *Google* vilka av sökningarnas svarsalternativ som användarna klickar och stannar på. Sökresultaten innehåller även resultat från *Google*'s webbkatalog vars innehåll kommer från ”the open directory project” som är en webbkatalog skapad av frivilliga redaktörer.

## 4.3.2 Avancerad sökning

Google Avancerad sökning [Sök tips](#) | [Allt om Google](#)

Hitta sidor med alla dessa ord  10 resultat

med den exakta frasen

med något av dessa ord

utan dessa ord

Språk Visa endast sidor på valfritt språk

Region Sök efter webbsidor på: alla regioner

Filtyp  Endast  visa resultaten av filtyp

Datum Returera webbsidor som dykt upp först under de/det

Placering Hitta sidor där mina sökord finns

Domän  Endast  sidor från denna domän

Användningsrättigheter Returera resultat som är

Säker sökning:  Ingen filtrering  Filtrera med [SafeSearch](#)

Exempel: .org, google.com [Mer information](#)

ej är filtrerade på grund av licens

Sidspecifik sökning

Liknande Sök efter sidor som liknar sidan

Exempel: www.google.com/help.html

Länkar Sök efter sidor som är länkade till sidan

©2010 Google

(Figur 5. Avancerad sökning [https://www.google.com/advanced\\_search?hl=sv](https://www.google.com/advanced_search?hl=sv))

Gränssnittet för avancerad sökning visar att *Google* främst söker webbsidor i sin databas

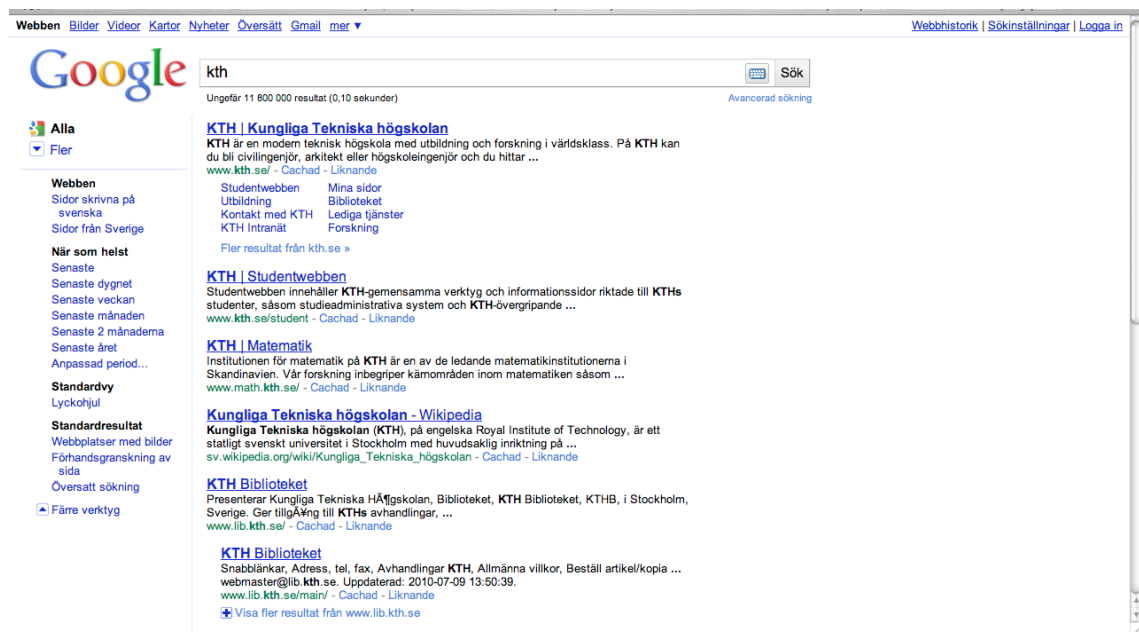
beroende på innehållet av ord. Längst upp på sidan markerat med blått går det att välja att söka på alla ord, den exakta frasen eller på ett visst ord.

Möjligheten att söka information beroende på tid finns under ”datum”, där det går att välja att returnera webbsidor som ”dykt upp”, antingen ”när som helst” eller ”först under det senaste dygnet, veckan, månaden eller året”. Sökningen påminner om *Yahoo!’s* ”updated” men med *Google* går det att se vilken information som dykt upp det senaste dygnet till skillnad från de senaste 3 månaderna. Möjligheten att söka på information som dykt det senaste dygnet är praktiskt om man vill söka på ett aktuellt ämne eftersom man då inte behöver gå via t.ex. nyhetstidningar på nätet eller redan kända nyhetsbloggar.

Vid sökning på datum visas svarsalternativen daterade, vilket gör att det går att se när en webbsida dykt upp eller uppdaterats. Första dagen visas antalet timmar informationen funnits, därefter antalet dagar och efter det datumet den spindlades. Det går inte att följa en webbsidas förändring över tiden eftersom efter att en webbsida dykt upp eller uppdaterats visas bara den senaste versionen.

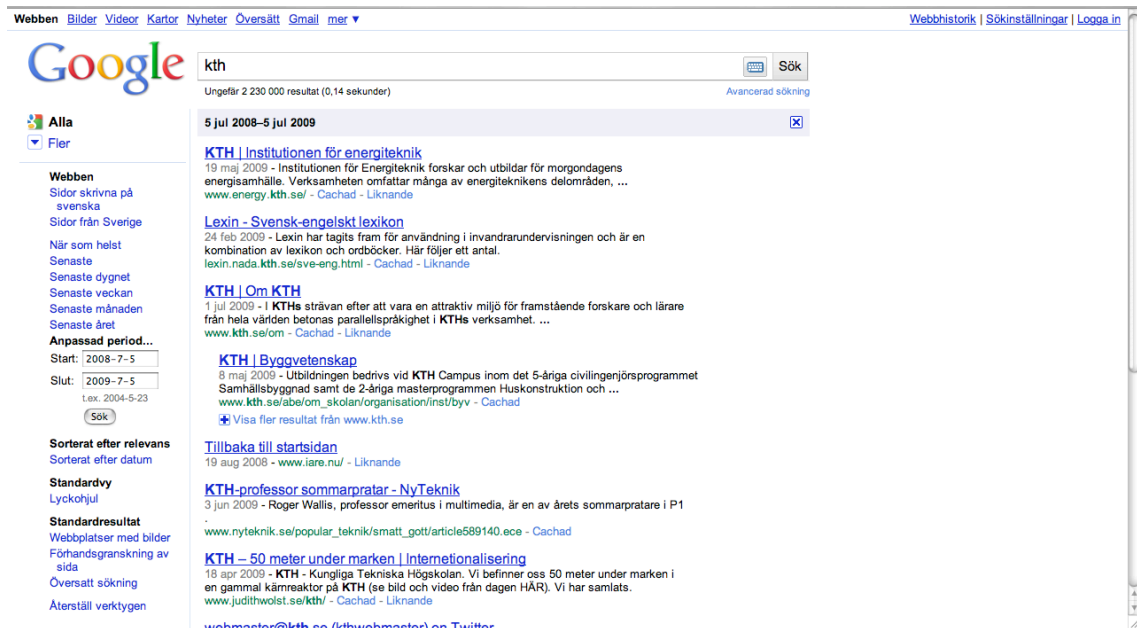
Längst ner finns på sidan finns länksökning, där man genom att skriva in en webbsidas adress får fram de sidor som är länkade till den sökta webbsidan.

### 4.3.3 Vanlig sökning, tidssökning och länksökning



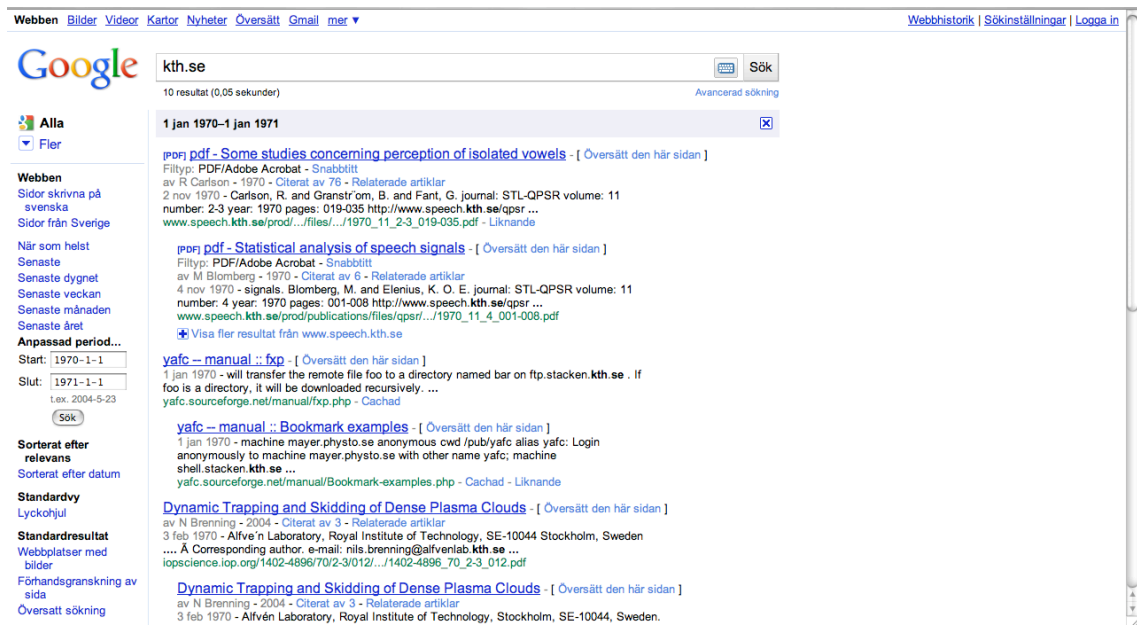
(Figur 6. Sökresultat för "kth")

Sökning på ordet "kth" ger 11 800 000 resultat. Förutom Wikipedia är de första svaren kopplade till adressen kth.se. I marginalen under "anpassad period" går det att söka på ett exakt tidsintervall tillskillnad från avancerad sökning där sökning i tiden alltid utgår från nuet.



(Figur 7. Sökresultat för "kth" från perioden 2008-7-5 till 2009-7-5)

Sökning på perioden 2008-7-5 till 2009-7-5 ger 2 230 000 resultat. Resultatens sidor skiljer sig beroende på om det är sidor som uppdaterats sedan dess eller inte. Även om sökresultaten visar ett annat datum så visar de sidor som fortfarande uppdateras samma webbsida som om man skulle söka på dagens datum. De resultat som visar webbsidor som inte uppdateras längre visar den information som webbsidan innehöll det sökta datumet, alltså hur den såg ut senast den uppdaterades. Även om det går att söka på ett specifikt tidsintervall visas inte äldre kopior av webbsidor utan senast uppdaterad version och därför går det inte att se äldre kopior för att jämföra hur information ändrats över tiden.



(Figur 8. Sökresultat för "kth.se" från perioden 1970-1-1 till 1971-1-1)

Det tidigaste datumet som det går att söka från är 1 januari 1970. Sökningen på kth.se, perioden 1970-1-1 till 1971-1-1, ger 10 resultat. De webbsidor som går att söka är de som inte uppdaterats sedan dess.



The screenshot shows a Google search interface with the search term 'link:kth.se'. The search results are displayed on the right side of the page, listing several links related to KTH (Kungliga Tekniska högskolan). The links include:

- KTH Executive School**: KTH Executive School utvecklar strategiskt ledarskap och affärsförmåga i teknikbaserade företag. Kunskapsbildningen sker främst i kvalificerade ...
- KTH | NORDITA**: 10 mar 2010 ... Grupper, Adress. NORDITA, ROSLAGSTULLSBACKEN 23, 10691 STOCKHOLM. Alla anställda inom NORDITA. Förmann, Eftermann, E-post, Telefon, Titel ...
- KTH | Vicerektorer**: Vicerektorer utses av rektor för speciella uppdrag.
- KTH | Institutionen för Mark- och vattenteknik**: Vi vill medverka till att göra verklighet av samhällets krav på ett utållt och effektivt utnyttjande av naturresurser. Vår vetenskapliga bas är en ...
- PARTNERS - SARnet**: [ Översätt den här sidan ] ARC Seibersdorf Research GmbH Austria (ARCS) Association Vincotte Nucleaire Belgium (AVN) Atomic Energy of Canada Limited Canada (AECL) ...
- Meetings and events - Sustainablewaste.info**: [ Översätt den här sidan ] Up coming, December 10, 2008. Third programme meeting of Towards Sustainable Waste Management, Göteborg, September 2009 ...
- TKK - Informaatio- ja luonnontieteiden tiedekunta - Vaihto**: Etusivu » Opinnot » Kansainvälinen opiskelu » Vaihto-opiskelu » Löydä oma kohteesi » Vaihto-opiskelukohdeet. Vaihto-opiskelukohdeet ...
- Kungliga Tekniska högskolan. KTH images - MyNewsdesk**: KTH i Stockholm svarar för en tredjedel av Sveriges kapacitet av teknisk forskning och ...

(Figur 9. Sökresultat för länksökning på "kth.se")

Genom att skriva in en webbsidas adress får fram de sidor som är länkade till den sökta webbsidan. Länksökning på kth.se ger 1490 sökresultat.

The screenshot shows a Google search interface with the search term 'link:kth.se'. The search results are displayed on the right side of the page, listing various links related to KTH. The results are filtered by date, showing results from 5 Jul 2008 to 5 Jul 2009. The links include:

- KTH Executive School**: 1 jul 2010 - KTH Executive School utvecklar strategiskt ledarskap och affärsförmåga i teknikbaserade företag. Kunskapsbildningen sker främst i kvalificerade ...
- Kungliga Tekniska högskolan. KTH images - MyNewsdesk**: 3 jul 2010 - KTH i Stockholm svarar för en tredjedel av Sveriges kapacitet av teknisk forskning och ingenjörsutbildning på högskolenivå. Utbildningen och forskningen ...
- KTH | NORDITA**: 3 jul 2010 - Förmann, Eftermann, E-post, Telefon, Titel. Eddy - Ardonne, ardonne@nordita.org, 0855378804, Gästprofessor. Iouri - Belokopytov, iouri.belokopytov@albanova. ...
- KTH | Vicerektorer**: 30 jun 2010 - Vicerektorer utses av rektor för speciella uppdrag.
- KTH | KTH Datavetenskap och kommunikation**: 2 jul 2010 - CIAM medarbetare från KTH Datavetenskap och kommunikation.
- KTH | KAFI**: 2 jul 2010 - Det är självklart att universitet och högskolor bedriver forskning och utbildning. Detta sker i den dagliga verksamheten på KTH genom tillämpad forskning.
- PARTNERS - SARnet**: [ Översätt den här sidan ] 4 jul 2010 - ARC Seibersdorf Research GmbH Austria (ARCS) Association Vincotte Nucleaire Belgium (AVN) Atomic Energy of Canada Limited Canada (AECL) ...
- Nordiska Radiosamfundet (NRS)**: 18 jun 2010 - Svifte och ändringar. Nordiska Radiosamfundet (NRS) är en stiftelse verksam i ...

(Figur 10. Sökresultat för länksökning på "kth" från perioden 2008-7-5 till 2009-7-5)

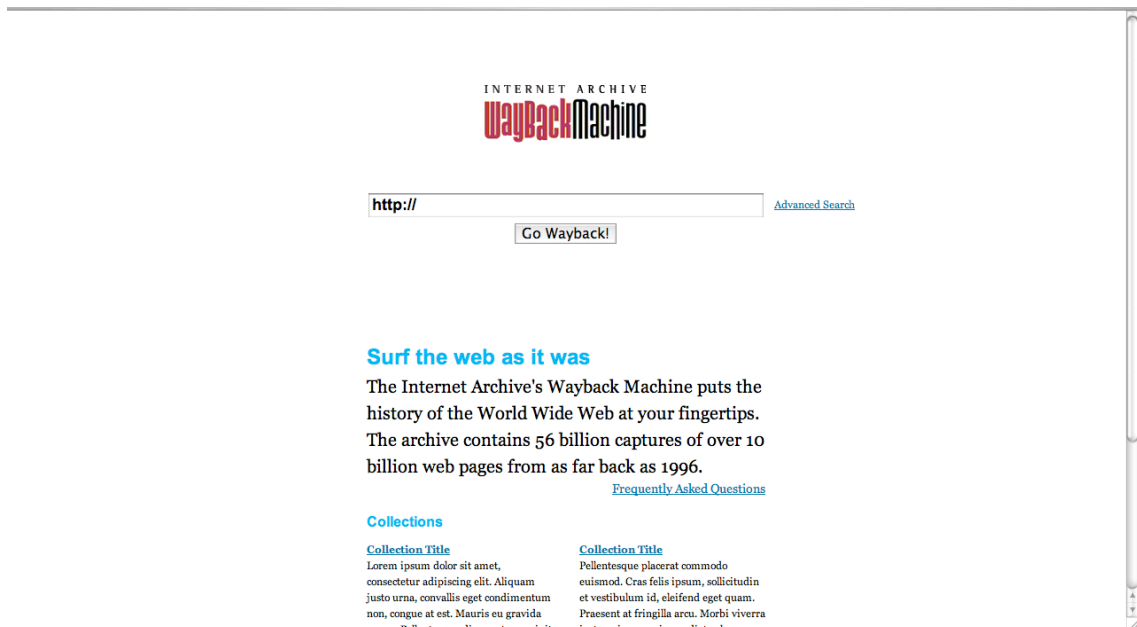
Kombinerad länksökning och tidssökning från samma period, 7/5 - 2008 till 7/5 - 2009, ger 1460 resultat men sökresultatens datummärkning visar att sökresultaten är från år 2010 (gäller alla resultat). Kombinerad länksökning och tidssökning ger ingen möjlighet att se hur länkar till en viss webbadress förändrats över tiden.

## 4.4 The Wayback Machine

The Internet Archive är en ideell organisation som 1996 började spindla webben med målet att bygga ett digitalt bibliotek av webbsidor. The Internet Archives söktjänst heter *The Wayback Machine* och genom att söka på en sidas URL (webbadress) får man fram tidigare versioner av en webbsida. Förebilden för The Internet Archive är alexandri´nska biblioteket<sup>5</sup> som innehöll en kopia av varje bok i hela världen. (Archive.org 2010-08-25)

Idag innehåller deras arkiv 10 miljarder webbsidor (Archive.org, 2010-08-25) jämfört med *Google's* 1 miljard. The Internet Archive har ett uttalat mål att spara digital information från webben, "Most societies place importance on preserving artifacts of their culture and heritage. Without such artifacts, civilization has no memory and no mechanism to learn from its successes and failures. Our culture now produces more and more artifacts in digital form. The Archive's mission is to help preserve those artifacts and create an Internet library for researchers, historians, and scholars" (Archive.org, 2010-08-25).

### 4.4.1 Förstasidan



(Figur 11. <http://www.waybackmachine.org/>)

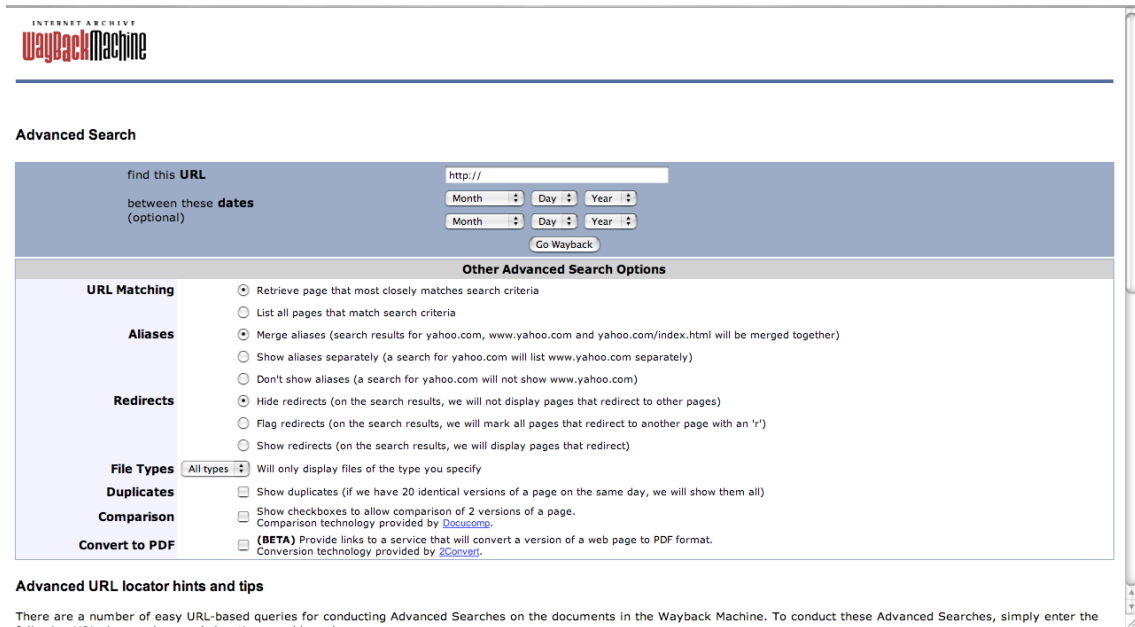
Internet Archive skriver att sökverktyget *The Wayback Machine* "puts the history of the World Wide Web at your fingertips". Genom att spara en webbsida varje gång den besöks av sökspindeln ges möjligheten att söka information från webben som den såg ut tidigare. Internet

<sup>5</sup> alexandri´nska biblioteket är det tidigast kända offentliga biblioteket i antiken och grundades i början av 200-talet f.Kr. av de ptolemeiska kungarna av Egypten. Ovisshet råder om när biblioteket förstördes: uppgiften att det brändes 48 f.Kr. av Caesar har ofta anförts, men nämns inte av samtida källor. (<http://www.ne.se.focus.lib.kth.se/kort/alexandrinska-biblioteket/1162218>)



Archive spindlar all information som finns i Open Directory Project, som är en webbkatalog där vem som helst kan lägga till en sida.

## 4.4.2 Avancerad sökning



The screenshot shows the 'Advanced Search' page of the Internet Archive's Wayback Machine. At the top left is the 'WaybackMachine' logo. The main search area is titled 'Advanced Search' and contains a form with the following elements:

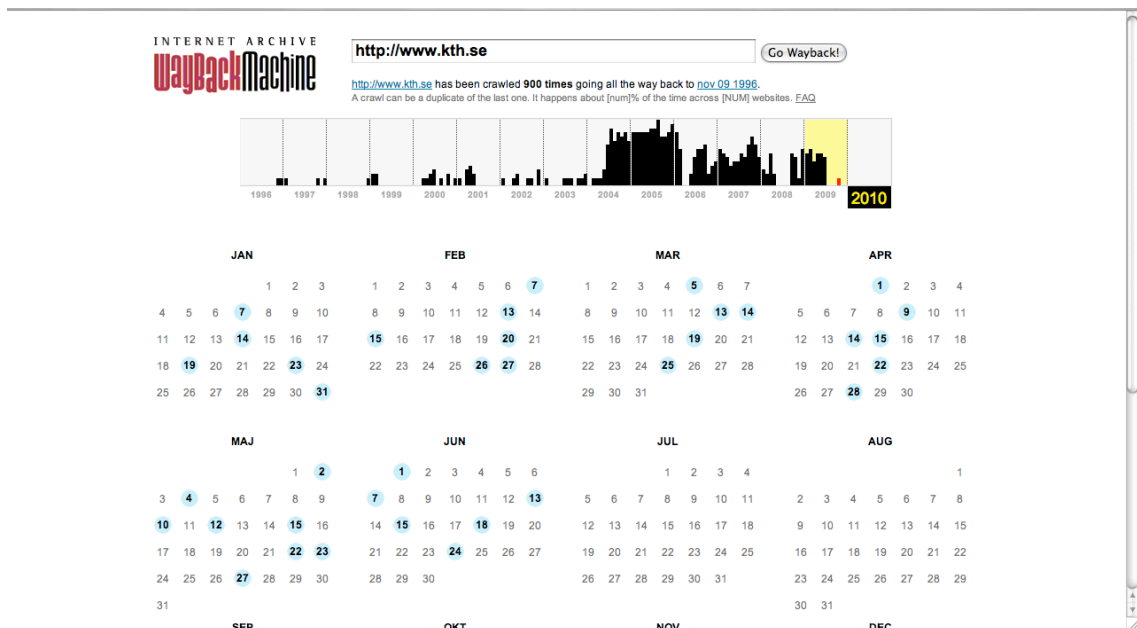
- find this URL**: A text input field containing 'http://'.
- between these dates (optional)**: Two sets of date pickers, each with 'Month', 'Day', and 'Year' dropdown menus.
- Go Wayback**: A button.
- Other Advanced Search Options**: A section with several radio button and checkbox options:
  - URL Matching**:
    - Retrieve page that most closely matches search criteria
    - List all pages that match search criteria
  - Aliases**:
    - Merge aliases (search results for yahoo.com, www.yahoo.com and yahoo.com/index.html will be merged together)
    - Show aliases separately (a search for yahoo.com will list www.yahoo.com separately)
    - Don't show aliases (a search for yahoo.com will not show www.yahoo.com)
  - Redirects**:
    - Hide redirects (on the search results, we will not display pages that redirect to other pages)
    - Flag redirects (on the search results, we will mark all pages that redirect to another page with an '!')
    - Show redirects (on the search results, we will display pages that redirect)
  - File Types**: A dropdown menu set to 'All types' with the text 'Will only display files of the type you specify'.
  - Duplicates**:
    - Show duplicates (if we have 20 identical versions of a page on the same day, we will show them all)
  - Comparison**:
    - Show checkboxes to allow comparison of 2 versions of a page. Comparison technology provided by [Docucomp](#).
  - Convert to PDF**:
    - (BETA)** Provide links to a service that will convert a version of a web page to PDF format. Conversion technology provided by [2Convert](#).

Below the search options is a section titled 'Advanced URL locator hints and tips' with the text: 'There are a number of easy URL-based queries for conducting Advanced Searches on the documents in the Wayback Machine. To conduct these Advanced Searches, simply enter the'.

(Figur 12. Avancerad sökning <http://web.archive.org/collections/web/advanced.html>)

Enligt gränssnittet för avancerad sökning går det att hämta den senast arkiverade kopian av en specifik webbadress, eller en arkiverad kopia från viss tidpunkt. Det går även att söka efter alla kopior av en specifik URL från en viss tidsperiod eller alla webbadresser som börjar med t.ex. adressen <http://www.kth.se>.

## 4.4.3 Historia?



(Figur 13. Sökresultat för webbadressen "kth.se")

Sökningen visar från vilka datum det finns arkiverade kopior. Sökresultaten visar att den senaste kopian av kth.se är från slutet av 2009, på hemsidan skriver de att det kan ta upp till 6 månader innan en sida kommer upp. Diagrammet visar att den tidigaste kopian är från 1996 då Internet Archive började sin verksamhet och att kth.se spindlades som mest 2005. Sökresultaten visar också antal kopior som sparats från varje månad och att det t.ex. det var uppehåll i juli och augusti.

VETENSKAP OCH KONST  
KUNGL. TEKNISKA HÖGSKOLAN

[In English, please!](#)

Välkommen till KTH

- [Aktuellt](#)
- [Allmän information](#)
- [Utbildning](#)
- [Forskning](#)
- [Institutioner](#)
- [Centrala funktioner](#)
- [KTHB - Biblioteket](#)
- [Dator- och nätverksinformation](#)
- [Att hitta rätt person vid KTH](#)
- [SUNET - det Svenska universitetsdatornätet](#)

Kungl. Tekniska Högskolan, KTH, är Sveriges största och äldsta tekniska högskola och svarar för en tredjedel av Sveriges kapacitet av ingenjörsutbildning och teknisk forskning på högskolenivå. KTHs verksamhet svarar mot de höga krav, som samhälle och näringsliv ställer och är slagkraftig även i ett internationellt perspektiv.

(Figur 14. Sökresultat för "kth.se" från november 1996)

---

Bilden visar hur webbsidan kth.se såg ut november 1996, den tidigaste kopian som finns sparad. Det går även att klicka på de olika länkarna och de flesta fungerar vilket visar att på noggrannheten hos the Internet Archive. Det går inte att söka direkt på länkarna.

## 4.5 Slutsats

*Yahoo!:*

Sökverket *Yahoo!* började som en lista över nya intressanta webbsidor. I takt med att nya tips kom in blev listan svår att överskåda och i ett försök att kontrollera informationen delades webbsidorna in i olika kategorier. *Yahoo!'s* webbkatalog innehåller sammanlagt cirka 100 kategorier som i sin tur innehåller subkategorier. Systematiseringstekniken att dela upp information i ämnehierarkier bygger helt på Linnés kortindexsystem. Istället för att enligt Principle of Provenance bevara den ordning i vilken informationen uppkommit så är informationen i *Yahoo!'s* webbkatalog uppdelad efter ämne och oberoende av tiden. Gränssnittet för avancerad sökning visar att *Yahoo!'s* fokus ligger på att analysera webbsidor efter deras textinnehåll. Precis som kortindexsystemet fokuserar på objekten i sig så beskrivs en webbsida av dess innehåll och inte av faktorer runtomkring som t.ex. tiden för dess uppkomst.

Finns ingen möjlighet att söka information från ett visst datum utan den enda möjlighet att specificera sökning i tiden är att söka uppdaterad information. Det går inte att jämföra olika kopior av en webbsida eftersom när en webbsida uppdateras så är det bara den senast uppdaterade webbsidan som visas. Funktionen att söka på uppdaterad information är inte att söka äldre information utan att begränsa urvalet.

*Yahoo!* har därför ingen information om när ett dokument uppkommit och det går inte att försöka återskapa de förhållanden som skapat informationen eller jämföra med yngre versioner för att undersöka hur en sida ändrats över tiden.

*Google:*

Istället för att endast ranka resultat beroende på hur många gånger ett visst sökord återfinns på en sida tolkar tekniken PageRank relevansen hos webbsidor genom antalet länkar de har. Precis som med en bok som flest refererar till i en citeringsdatabas visar att den är mest relevant, så visar antalet länkar till en viss webbsida att den är mest relevant där en länk ses som en röst från Internets användare. Webbsidor länkas beroende på textinnehåll och PageRank liknar kortindexsystemet eftersom att informationen vävs samman beroende på ämne. En webbsida beskrivs inte av tiden för dess uppkomst utan det är webbsidans innehåll som ger den dess relevans, men skillnaden mot en webbkatalog är att länkarna bestäms av Internets användare. Utöver PageRank hittar *Google* relevant information åt den som söker genom att analysera texten på en webbsida, där både webbsidans och de länkade webbsidornas analyseras.

Svarsresultaten i sökning på tid skiljer sig beroende på om de är webbsidor som uppdaterats efter det datum som söks. De sidor som uppdaterats visar den uppdaterade informationen och inte den informationen den innehöll det datum som söks. De webbsidor som visar den information som webbsidan innehöll det sökta datumet är de som inte uppdateras längre. De webbsidor som inte uppdateras längre betraktas som i tiden diskreta enheter som har en fast plats i historien precis som ett arkiv innehåller information som inte längre cirkulerar i samhället. Vare sig en sida fortfarande uppdateras eller inte så går det bara att söka den senast uppdaterade versionen vilket gör att inte går att söka äldre versioner av webbsidor för att jämföra hur information ändrats över tiden. *Google's* fokus ligger på den senaste information, där sökning i tiden alltid utgår från nuet. Sökning på tid har funktionen att den begränsar urvalet genom att inte visa gammal information. Även sökning på ett diskret tidsintervall som i sökning på "anpassad period" visar den information som var ny då, inte hur information som är ny idag

---

såg ut då.

*Google* gör det möjligt att söka på information som inte uppdaterats sen 1 januari 1970 och har då funktionen av ett arkiv och i försök att som i Principle of Provenance återskapa de tidshistoriska förhållandena så har jag försökt kombinera länk- och tidsökning. Som datummärkning visar så går det inte att söka länkar i tiden och *Google* ger ingen möjlighet att se hur länkarna till en viss webbadress förändras över tiden.

*The Wayback Machine:*

The Internet Archive har som uttalat mål att bevara ett kulturellt arv av digital information för att skapa minne av det förflutna, utan ett sådant minne har ett samhälle inga mekanismer att lära sig. Sökverktyget sparar äldre webbsidor för att göra det möjligt att söka information från webben som den såg ut tidigare. De har spindlat webben sen 1996 och databasen innehåller idag 10 miljarder webbsidor.

Precis som enligt Principle of Provenance så bevarar *The Wayback Machine* den ordning i vilken informationen uppkommit där informationen är fixerad vid en linjär tidsskala. Webbsidorna beskrivs av tiden för deras uppkomst dvs. datumet då de spindlades och webbsidorna som ses som resultatet av en viss tid där det är förhållandena runt omkring som särskiljer dem.

Fokus ligger på sökning i tiden och meningen är att göra det möjligt att återskapa de förhållanden som skapat informationen. Eftersom sökverktygets index innehåller tidigare versioner av webbsidor går det att jämföra olika versioner av en sida för att se hur den ändrats över tiden. Det går bara att jämföra hur en webbadress förändrats, och även om flera av länkarna också finns sparade går det inte att söka på dem specifikt inte anpassat för att kunna skapa ett objektivt förhållande till den ursprungliga källan.

## 5. slusats

Texten börjar med att ställa frågan om vad som utgör kontexten i en samling information och om det finns ett optimalt sätt att systematisera information. Robert M. Colomb skriver i boken "Information Spaces" att samlingar av information i allmänhet organiseras beroende på två faktorer: "maintaining context" och "respecting time". Med det första menar han att bevara länkarna mellan de enskilda dokumenten för att stödja en förståelse av kontexten, och med det andra att göra det enkelt för forskaren att veta när ett dokument uppkom för att underlätta att sätta in det i en viss tidssekvens (Colomb, 2002).

Om man ser till den ursprungliga betydelsen av ordet kontext som kommer från det latinska ordet *Contextus* och betyder att väva samman, "con- 'together' + *texere* 'to weave'" (Oxford Dictionaries) blir uppdelningen mellan "maintaining context" och "respecting time" felaktig. Systematisering av en samling information handlar just om att väva samman information men uppfattningen av vad det är som väver samman informationen har skiftat genom historien. I vissa tidsperioder har information vävts samman beroende på ämnesinnehåll och i andra perioder beroende på när informationen uppkommit. Uppfattningen av vad som väver samman information är föränderlig och Colomb's uppdelning mellan "maintaining context" och "respecting time" kommer från den betydelse det engelska ordet *context* kom att få under sen medeltid, "denoting the construction of a text"(Ibid.), då ordet kontext relaterades till en texts innehåll.

---

Det finns inget optimalt sätt att systematisera information. Genom historien har det skapats tekniker beroende på uppfattningen om vad som väver samman en samling information.

De två olika systematiseringstekniker som dominerat modern tid är *Kortindex* och *Principle of Provenance*. *Principle of Provenance* introducerades 1881 som en reaktion mot kortindextekniken och den tidlöshet tekniken fick till följd. Kortindex skapade ett nytt sätt att se på helheten hos en samling information där informationen inte relateras till tiden för dess uppkomst. *Principle of Provenance* skapade ett nytt sätt att se på objektet i en samling där objektet beskrivs av tiden för dess uppkomst. Genom historien har systematiseringsteknikerna växlat mellan dessa två traditioner vilket också går att se i sättet att systematisera information på webben.

*Yahoo!'s* webbkatalog bygger helt på Linnés kortindexteknik från mitten av 1700-talet där information delas in i olika ämneskategorier. Insamling av ny information utgår från den redan indelade informationen och ser vad i omgivningen som kan delas in i systemet. Webbkatalogen saknar sökning i tiden eftersom målet är att skapa mening oberoende av tid.

*Google* liknar kortindextekniken i användandet av textjämförelsetekniker där information främst söks genom matchning av en sökterm mot texten i en webbsida och i de webbsidor som länkar till webbsidan.

*Google* ger en begränsad möjlighet att söka i tiden eftersom äldre information försvinner när den uppdateras och det går därför inte att få en tydlig bild av vilka webbsidor som gick att söka vid ett visst tillfälle i historien. *Google* har som mål att "organize the world's information and make it universally accessible and useful" (Google.com, 2010-08-25). Idén att organisera all världens information är densamma som i kortindextekniken men för att göra all världens information tillgänglig i tiden måste den sparas enligt *Principle of Provenance*. *Google* fokuserar på nuet och organiserar information i tiden genom att spara den information som inte uppdateras längre.

Man kan se *The Wayback Machine* som en reaktion mot de andra sökverktygen på samma sätt som introducerandet av *Principle of Provenance* år 1881 var en reaktion mot kortindextekniken. *The Wayback Machine* bygger på idén att innehållet på en viss webbadress visar olika information vid olika tidpunkter och att de olika webbsidorna måste sparas i tiden för att informationen inte ska försvinna. Enligt *Principle of Provenance* ska information vävas samman beroende på tiden den skapades för att kunna återskapa de tidshistoriska förhållandena och det är av samma anledning som *The Wayback Machine* bevarar webbsidor i deras ursprungliga ordning. Sökverktyget sparar olika versioner av webbsidorna och beskrivs av datumet för deras uppkomst eftersom de ses som resultatet av en viss tid. Enligt *Principle of Provenance* kan allt förklaras gentemot historien, vilket gör att allt kan bli användbart i framtiden. All data blir intressant och sparas på hög medan kortindextekniken utgår från den insamlade information och ser vad som passar in i den redan systematiserade informationen. Enligt *Principle of Provenance* dras gränsen mot omgivningen bredare än i kortindexsystemet då allt i historien är relevant och utövar ett inflytande på systemet vilket också är anledningen till att *The Wayback Machine's* index innehåller 10 miljarder webbsidor jämfört med *Google's* som innehåller drygt en miljard webbsidor.

Det går att använda länkarna hos de sparade webbsidorna i *The Wayback Machine* vilket visar att även en webbsidas länkar sparas i tiden, men det går inte att specifikt söka på dem med länksökning. *Google* gör det möjligt att söka på vilka länkar en webbsida har men bara i nuet.

Enligt hermeneutiken skapar människor mening genom tolkning dvs. informationens innebörd går inte att definiera i någon absolut mening. Möjligheten att skapa mening av informationen på webben är beroende av både dess ursprung och nutiden eftersom den får olika betydelser genom historien. För att kunna tolka ett historiskt sammanhang mot historien själv måste det gå att etablera en länk med den tid då informationen skapats. *The Wayback Machine* har som mål att göra det möjligt att få tillgång till informationens källa tillskillnad från de andra sökverktygen som absorberas i nuet.

Heidegger menar att tiden är grunden för varat, att tiden är horisonten för vår varaförståelse. Möjligheten att skapa mening av vår vardag genom tolkning utgörs av tiden och att skapa mening genom att smälta samman horisonterna, delen och helheten, går ut på att placera sig i en tradition där tillvaron uppkommit. Heidegger menar att meningsskapande inte är en algoritm

---

som inom vetenskapen utan en ständigt pågående process. Systematiseringsteknikern *kortindex* och Principle of Provenance är två synsätt som växlat i historien och Heideggers filosofi befinner sig i en tradition som står närmare Principle of Provenance. Kortindextekniken skapades i en tid då information var något som byggde på mätbarhet och effektivitet, ett synsätt som präglar det s.k. vetenskapliga tänkandet. Fokus på exakthet och sanningar oberoende av tid står i kontrast till läran om tolkning och vetenskapens syn att mening uppstår när information samlats in och organiserats, kommenterade Heidegger med: "Science does not think" (Heidegger, 1968).

Heidegger skiljer på tid såsom den uppfattas inom vetenskapen respektive filosofin och istället för tid som en linje vars intervall kan mätas med en klocka och som prioriterar tiden i nuet menar Heidegger tid som i kultur och vår kulturella bakgrund. Med tid menas inte numerisk tidsbestämning som i kalendariskt datum eller "klocktid" utan att tid är något vi delar med de andra med vilka vi delar världen. Tillvaron har inte uppkommit genom en subjektiv process utan är allmän och bunden till vissa händelser, som Heidegger skriver så existerar tillvaron i form av *medvaro* med andra. Information står inte i förhållande till världsliga "sanningar" enligt kortindexteknikens synsätt utan är snarare händelser vilka *Principle of Provenance* försöker återskapa.

*The Wayback Machine* gör det möjligt att söka i tiden, men med tid som kalendariskt datum vilket i sig inte räcker för att placera informationen i en tradition. *Google's* teknik PageRank bygger på att skapa mening genom att systematisera informationen efter hur webbans användare länkar till varandra. En enskild webbsidas relevans skapas av Internets användare och sökresultatets ordning beror på vad användarna finner meningsfullt. Länkar ändras i tiden precis som information ändrar mening i historien, men eftersom information om länkar inte sparas går det inte att se hur användarnas uppfattningar ändrats i tiden. PageRank bygger på analys av webbsidors innehåll där användarnas "röster" baseras på textinnehåll, där webbsidorna (objekten) precis som i kortindextekniken beskrivs av deras innehåll. Webbsidornas mening ändras i tiden vilken systematiseringstekniken Principle of Provenance bygger på, men *Google* sparar inte information i samma grad som *The Wayback Machine* och ger ingen möjlighet att se hur webbsidor eller länkarna till en viss webbadress förändras över tiden.

*The Wayback Machine* visar hur webbsidor förändras över tiden men det finns ingen länksökning utan en webbadress söks för sig och resultaten beror inte av Internets användare. *Google's* specifika länksökning i kombination med *The Wayback Machine's* fokus på att spara information i tiden skulle ge en möjlighet att återskapa informationens källa och underlätta att skapa mening genom tolkning. Att genom den hermeneutiska processen sammansmälta horisonterna skulle ge användning åt informationen på webben genom att skapa effektiv historia istället för att historien blir en samling information som bara ackumulerar i arkiven eller helt försvinner.

De tekniker som sökverktygen använder för att skapa mening av informationen på webben bygger på det tillvägagångssätt som kännetecknar vetenskapen där information organiseras i olika kategorier med målet att dela in världen i ett system innehållande sanningar oberoende av tid. Kortindextekniken är en del av den traditionen tillskillnad från Principle of Provenance som är en del av den hermeneutiska traditionen som bygger på att vi skapar mening av information och förstår varat genom tolkning. De tidigaste sökmotorerna som introducerades i början av 1990-talet fokuserade på webbsidors textinnehåll och skapade en standard för hur sökmotorer skulle komma att fungera. Undantaget till de sökverktyg som skapats i den datalogiska traditionen är *The Wayback Machine* från 1996 som fokuserar på tiden en viss information uppkom och går att se som en reaktion mot de andra sökverktygen på samma sätt som Principle of Provenance kom som en reaktion mot kortindextekniken.

---

# litteraturlista

Google.com: Retrieved 2010-08-25 from (<http://www.google.sh/intl/en/corporate/>)

Robert M. Colomb: *Information spaces: the architecture of cyberspace* (Springer-Verlag London limited 2002), p. 136.

Staffan Mueller-Wille: *Carl Linnaeus Invented The Index Card* (*ScienceDaily* 2009, June 16)  
Retrieved May 12, 2010, from  
<http://www.sciencedaily.com/releases/2009/06/090616080137.htm>

T.R Schellenberg: *The Management of Archives* (New York: Columbia University Press, 1965), p. 93.

Sven Spieker: *The Big Archive. Art from Bureaucracy* (Cambridge/Mass.: MIT Press, 2008), p. 18.

P. Schoderbek, C. Schoderbek, G. Kafalas: *Management Systems* (McGraw-Hill Companies, Ic. 1998), p. 13.

P. Schoderbek, C. Schoderbek, G. Kafalas: *Management Systems* (McGraw-Hill Companies, Ic. 1998), p. 24.

Martin Heidegger: *What is Called Thinking?* (New York, Harper & Row, 1968), p. 8.

T. Winograd, F. Flores: *Understanding Computers and Cognition, A New Foundation for Design* (Ablex Publishing Corporation 1986,1987), p. 27-28.

Martin Heidegger: *Varat och tiden* (Daidalos AB, 1992), s. 195.

Martin Heidegger: *Varat och tiden* (Daidalos AB, 1992), s. 195-196.

Hans-Georg Gadamer: *Thruth and Method* (Sheed & Ward Ltd. 2004), p. 258.

T. Winograd, F. Flores: *Understanding Computers and Cognition, A New Foundation for Design* (Ablex Publishing Corporation, 1986, 1987), p. 30.

Martin Heidegger: *Varat och tiden* (Daidalos AB, 1992), s. 38.

---

Martin Heidegger: *Varat och tiden* (Daidalos AB, 1992), s. 211-212.

Martin Heidegger: *Varat och tiden* (Daidalos AB, 1992), s. 219.

Martin Heidegger: *Varat och tiden* (Daidalos AB, 1992), s. 216-7.

### ***Yahoo!***

<http://yhoo.client.shareholder.com/press/history.cfm> 2010-08-25

<http://yhoo.client.shareholder.com/press/overview.cfm> 2010-08-25

[http://help.yahoo.com/l/us/yahoo/search/basics/basics08.html;\\_ylt=AiByZ52ZdDvS7\\_Sjx9mlfzguqCN4](http://help.yahoo.com/l/us/yahoo/search/basics/basics08.html;_ylt=AiByZ52ZdDvS7_Sjx9mlfzguqCN4) 2010-08-25

A. Glossbrenner, E. Glossbrenner: *Search engines for the World Wide Web* (Peachpit Press, 1998), p.157

### ***Google***

<http://www.google.sh/intl/en/corporate/> 2010-08-25

[http://www.google.com/intl/sv/why\\_use.html](http://www.google.com/intl/sv/why_use.html) 2010-06-26

[http://www.google.com/intl/sv/why\\_use.html](http://www.google.com/intl/sv/why_use.html) 2010-06-26

[http://www.google.com/intl/sv/why\\_use.html](http://www.google.com/intl/sv/why_use.html) 2010-06-26

[http://www.google.com/intl/sv/why\\_use.html](http://www.google.com/intl/sv/why_use.html) 2010-06-26

### ***The Wayback Machine***

<http://www.archive.org/about/faqs.php#Archive-It> 2010-08-25

<http://web.archive.org/collections/web.html> 2010-08-25

<http://www.archive.org/about/faqs.php#103> 2010-08-25

Robert M. Colomb: *Information spaces: the architecture of cyberspace* (Springer-Verlag London limited 2002), p. 193.

<http://www.google.sh/intl/en/corporate/> 2010-08-25

Martin Heidegger: *What is Called Thinking?* (New York, Harper & Row, 1968), p. 8.



