

PROJEKT RNEBERG, PIONJÄRPROJEKT OCH OFRIVILLIG DIGITALISERING AV NORDISK LITTERATUR

bis intervjuar projektets grundare Lars Aronsson, som är kritisk till hur digitaliseringsfrågorna hanterats, av de statliga institutionerna inom kulturarvssektorn. Hos det id
Runeberg – som träget scannat, bearbetat och publicerat nordisk litteratur på webben s
digitaliseringsflitens lampa.

Hej Lars! Projekt Runeberg startade redan 1992. Hur har förutsättningarna för verksamheten förändrats under de här åren?

Det tragiska är väl att de inte har förändrats så mycket som de hade kunnat. Om statliga institutioner hade tagit digitalisering på allvar, hade Projekt Runeberg idag varit en överspelad pionjärinsats. Eller om det hade funnits privata finansiärer, hade det inte längre behövt vara ett hobbyprojekt. Men i viss mån har Projekt Runeberg kört fast och kommer inte längre, men behåller ändå ledartröjan.

Vår lilla webbsajt har mellan 1 och 2 miljoner sidvisningar per månad, rankas bland de 100.000 mest besökta i världen och bland de 1000 mest besökta i Sverige. Det är siffror som borde få de flesta svenska statliga institutioner att bli avundsjuka, och utbildningsministern att bli angelägen. Men de flesta som läser det här vet inte om det är någon skillnad på tio, hundra eller tusen när det kommer till webbstatistik. Man har så lite att jämföra med.

Wikipedia, där jag också har varit mycket engagerad, har samma problem i ännu högre grad. De har 132 miljoner sidvisningar per månad i Sverige, omkring hundra gånger mer än Projekt Runeberg, och räknas som den 6:e mest besökta webbplatsen, både i Sverige och globalt. Men vi hör inte Jan Björklund kommentera att alla statliga webbsatsningar på kunskap är så totalt akterseglade av detta amatörprojekt. Och vi hör inte heller oppositionen kritisera honom för detta. Det är som elefanten mitt i

rummet, som ingen varken hör eller ser. Betyg hit eller dit är hur viktigt som helst, men tillgången till kunskap på nätet finns inte på dagordningen. Är det inte märkligt?

Vad beror det på, tror du? Har vi fortfarande inte förstått Internet och internetkulturer?

Det kanske finns ett gap mellan ungdomar, som lär sig (Google!) och knyter nya kontakter (Facebook!), och de vuxna beslutsfattare som förvaltar våra befintliga institutioner. Att institutioner som museer och bibliotek en gång skapades för att underlätta lärandet, har med tiden kommit i skymundan av att bibehålla personalstruktur och investerat kapital.

Det är ju inte institutionsbyggare som Artur Hazelius (Skansens grundare) och Gustaf Klemming (Kungliga bibliotekets chef och pionjär inom Fornskriftsällskapet) som idag är myndighetschefer. Dagens chefer har kanske aldrig mött en institutionsgrundare. Den senaste institutionsbyggaren var kanske var Sven Allerstrand, som 1979 grundade Ljud- och bildarkivet. Men det finns ju inte kvar som egen institution längre.

*

Projekt Runeberg startade som ett sätt att visa hur Internet fungerar. Tidigare hade nätet använts för att överföra filer och skicka e-post. Men att publicera information utan att ha en i förväg känd mottagare, det var något nytt på nätet. Det vi idag känner som webben, skapat av Tim Berners-Lee, var ännu ett litet obetydligt system. Ett lite större hette Gopher och det var där Projekt Runeberg började. Under 1993 och 1994 gick vi gradvis över till

*Projekt Runeberg startade som ett
sätt att visa hur Internet fungerar. Tidigare hade
nätet använts för att överföra filer och skicka e-post.
Men att publicera information utan att ha en i
förväg känd mottagare, det var något nytt
på nätet.*

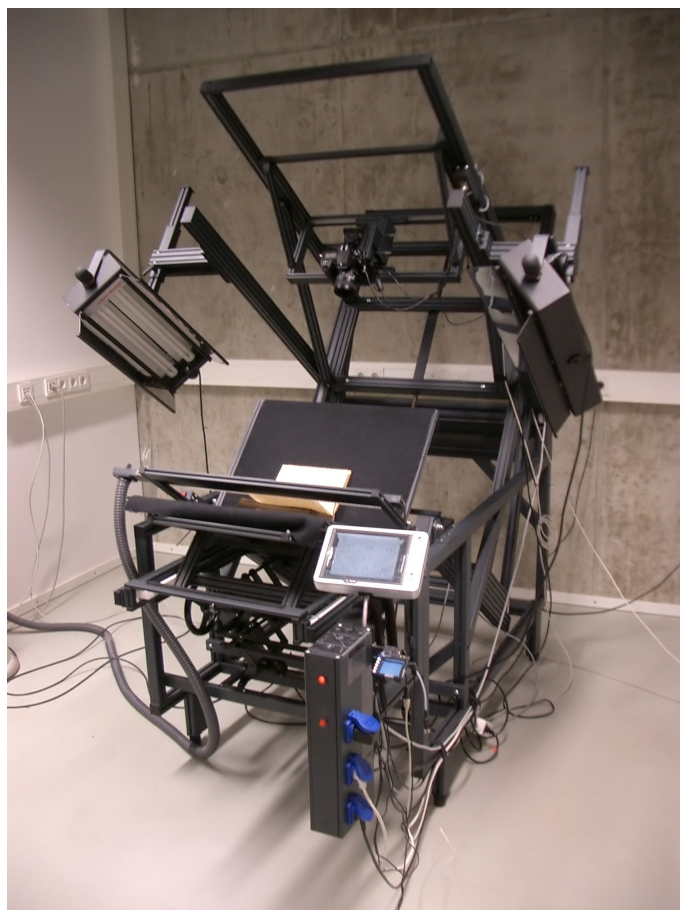
LEDARE INOM

*Frågorna hanterats, eller
Självt drivna Projekt
Sedan 1992 – lyser dock*

webben. Sett i det tidsperspektivet har förstås allting ändrat sig. Nätet finns idag i vars och ens kontor, hem och ficka. Webben är dess huvudsakliga användning. Böcker och tidningar på papper är i många fall redan utdöda, om man tänker på tryckta encyklopedier, kartböcker, telefonkataloger, stavningsordlistor, teknisk dokumentation och vetenskapliga publikationer. Allt sådant var självklart tryckta publikationer 1992 och är självklart inte tryckta idag. Samtidigt ges romaner och dagstidningar fortfarande ut i tryck; där har det elektroniska genomslaget inte varit lika stort. När det gäller äldre böcker, det vi samlar i våra bibliotek, förlitar vi oss fortfarande på trycket, trots alla fördelar som en digitalisering skulle kunna innebära. Där finns fortfarande mycket kvar att göra. Men det har visat sig svårt för dagstidningarna att tjäna pengar på webben och det är också svårt för bibliotek att komma ur startgroparna med digitaliseringsprojekt. De få som har digitaliserat något, har inte blivit belönade efter förtjänst. Det offentliga finansierar arkiv, bibliotek och museer som håller öppet i sina befintliga lokaler, men de som tar emot digitala besökare får inte samma uppskattning. Eller rättare sagt: De institutioner som struntar i att digitalisera, blir inte straffade för det.

Det verkar behövas ett samlat, nationellt grepp på digitaliseringsfrågorna, både vad gäller ekonomi och uppdrag. Men Digisam, samordningssekreteriat för digitaliseringsfrågor där Riksarkivet, Kungliga biblioteket och Riksantikvarieämbetet är representerade, är väl ett steg i rätt riktning?

Nej, jag ser ingen nytta komma ut av Digisam. Om det finns någon nytta med att digitalisera, så borde man sätta igång och digitalisera, i stället för att utreda i flera år. Att utreda är bara detsamma som att förhala.



Digitaliseringsmaskin från Digitalisierungszentrum der SLUB Dresden (Pondus, 2009, CC-BY-SA)

Hur fungerar Projekt Runeberg? Jag har förstått att användardeltagelse spelar en central roll, både vad gäller scanning, bearbetning och korrekturläsning? Hur finansierar ni utrustning och annat som kostar?

Ända sedan starten har Projekt Runeberg varit ett hobbyprojekt med hemvist inom en datorförening vid Linköpings universitet som heter Lysator. Några gånger har vi fått pengar ur fonder till små projekt, men vi har inte haft någon stabil finansiering till hela verksamheten. Att universitetet stöttar den här föreningen med en datorhall, elektricitet och nätanslutning har varit en grundförutsättning. Men i övrigt har projektet måst utformas för att klara sig på en nollbudget. Vår server kör enbart fri programvara så vi har inga kostnader för programlicenser. Vi har blivit väldigt konservativa i utformningen av tekniken, så att den ska klara sig själv utan någon daglig bemanning. Det mesta innehållet

på sajten är helt statiskt. Det ändrar sig sällan. Men även när fel behöver rättas till eller nytt innehåll läggs till, överläter vi så mycket som möjligt av det på utomstående att göra själva. Vi tillämpar versionshantering för alla filer, precis som programutvecklare gör, så att alla ändringar sparas och kan återställas. När vi digitaliserar en bok, delar vi upp arbetet i tre moment: Scanna bilder, OCR-tolka bilderna för att skapa en text och korrekturläsa texten. Varje steg kan göras av frivilliga utomstående och de kan göras av olika personer, så att man kan specialisera sig på det man gör bäst.

Har ni något hum om hur många av era besökare som också bidrar med arbete? Ibland är det en hög tröskel från att bara läsa till att själv bidra. Jag, till exempel, har bara redigerat någon enstaka artikel på Wikipedia, trots att jag är en flitig läsare.

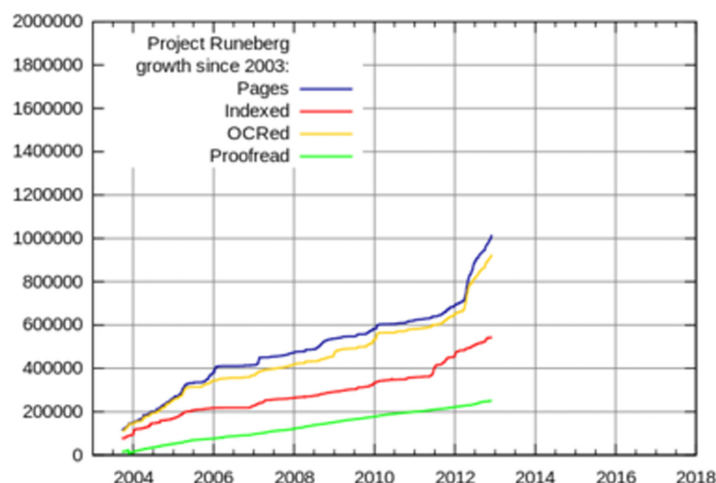
Både i Projekt Runeberg och Wikipedia är det en mycket liten andel av besökarna som blir aktiva medhjälpare.

De flesta frivilliga medhjälparna i Projekt Runeberg är korrekturläsare och deras insats är att korrekturläsa cirka 25.000 boksidor per år. Den siffran är väldigt konstant över tiden, vilket gör att så fort vi scannar mer än 25.000 sidor per år, så kommer korrekturläsningen att halka efter. Om vi hade pengar, skulle scanningen kunna skötas med avlönad personal, medan korrekturläsningen är alldeles för tidskrävande och dyrbar för att avlöna.

Under 2012 har vi varit fyra personer som scannat och kanske 40 som har korrekturläst. Det kan jämföras med att vår webserver under en vanlig dag tar emot anrop från 10.000 olika IP-adresser. (Många av dem är engångsbesökare som fått en märklig träff vid Google-sökning och som backar ut och aldrig återvänder. Men ändå.)

Digitalisering av tryckt material är ju en mycket tidskrävande process och därmed också resurskrävande. Hur ser du på enskilda aktörers roll i denna process i förhållande till offentliga institutioner (till exempel biblioteken)? Vid sidan om ideella initiativ som Projekt Runeberg finns det ju även kommersiella aktörer, t.ex. Google Books – spelar det någon roll vem som digitaliserar?

Google är ett kommersiellt företag, där annonser drar in pengar, och övriga projekt kostar. Jag vet inte hur de räknar med att någonsin tjäna pengar på digitalisering av äldre böcker, men jag är glad så



Projekt Runebergs utveckling. Graf hämtad från: URL www.runeberg.org

länge deras ägare har lust att bära den kostnaden. En annan stor aktör i USA är Internet Archive, som är ett ideellt organ, grundat av Brewster Kahle, med pengar som han tjänade på att sälja sitt datorföretag. Skillnaden är alltså inte så stor. Både Google och Internet Archive kan ses som privat filantropi med rötterna i företagsvinster. I båda fallen är det visionära män från datorbranschen som ligger bakom, och inte statliga institutioner inom kultur eller utbildning.

Om Google hade varit ett svenskt företag, hade de kanske betalat mer i skatt och vi hade förväntat oss att bokscanningen skulle skötas av statliga institutioner som Kun-
gliga biblioteket och Nationalmuseum.

Det finns några få länder, där staten tar digitalisering av äldre böcker på allvar: Frankrike och Norge är två exempel. Men trots höga skatter, ligger Sveriges offentliga institutioner långt efter inom digitalisering. Ingen har koll på vem som digitaliserar vad, och ingen verkar bry sig.

Det blir skandal om böcker stjäls från KB eller om de skulle börja mögla för att lokalerna vore dåliga. Men frånvaron av digitaliseringsprojekt eller bristerna i de projekt som finns, är inget som leder till klagomål. När Jan Björklund föreslår reformer av betygssystemet då blir det ett ramaskri. Men när svenska skolbarn söker på nätet och bara hittar utländska källor, som om alla våra bibliotek och arkiv hade brunnit upp, då är det ingen som reagerar.

Min upplevelse är att svenska folkbibliotek inte arbetar med förmedling av fria e-böcker (från t.ex. Projekt Runeberg och Project Gutenberg) i någon större utsträckning, utan ofta nöjer

sig med att erbjuda Elibs katalog (även om man nu kan notera ett ökat missnöje med nuvarande avtal). Hur skulle biblioteken kunna arbeta för att ytterligare synliggöra fritt tillgänglig litteratur i digitalt format?

Det som kallas e-böcker är en ansträngning att försvåra kopieringen, genom att använda särskilda format med inlagda kopieringsskydd och tidsbegränsningar. Eftersom Projekt Runeberg sysslar med fritt innehåll, har vi inte prioriterat att använda oss av de här formaten. Vi har ju inte heller någon säljare som åker runt till biblioteken för att övertyga dem om att förmedla våra böcker. Wikipedia har samma problem. De flesta bibliotek informerar sina besökare bara om de kommersiella alternativen. Men vanligt folk använder Google och hittar fram till både Wikipedia och Projekt Runeberg. Jag tror att det här är ett större problem för biblioteken än för de fria projekten. För femtio år sedan hade biblioteken en gemensam inköpscentral som kunde samordna information, men av detta blev företaget BTJ som idag är sålt och är en leverantör i mängden. De har inget kommersiellt intresse av att informera om Wikipedia och andra fria projekt på nätet. I avsaknaden av samordning, har biblioteken satt sin förhoppning till att Kungliga biblioteket ska axla sin "nya roll" som samordnare, men tiden tycks gå långsammare i Humlegården än i resten av världen.

Projekt Runeberg betonar tydligt att materialet är fritt från DRM (Digital Rights Management), varför är det viktigt? I den debatt som funnits om folkbibliotek och e-böcker på senare tid har det knappt talats alls om konsekvenserna av en låst infrastruktur för förmedling av digital litteratur, men vilken roll spelar egentligen öppna filformat och standarder för digital litteratur? Och är det någon skillnad om vi talar om nya texter gentemot äldre texter, där upphovsrätten löpt ut?

Jag är övertygad om att "information vill vara fri", i den betydelsen att den mest framgångsrika långsiktiga strategin är att utgå ifrån att information är gratis och kan kopieras fritt. Vi ser folk köpa musik och filmer, som de sedan inte kan spela upp på mer än en av sina apparater. Varför krångla till det? Datorer och läsplattor kan vara nog krångliga som de är.

Man blir aldrig glad när man upptäcker att ett bibliotekslån har gått ut och man måste betala en straffavgift, eller om en e-bok försvinner från läsplattan när man har ett kapitel kvar. Är det någon på biblioteken som har tagit reda på hur ofta det händer? Kommer den personen in och lånar fler e-böcker sedan? Eller går den lille mannen sin väg och kommer aldrig mer igen?

Slutligen, hur ser du på läsningens framtid med tanke på den ökade användningen av digitala medier?

Jag tror nog, kanske blåögt, att läsningen blir viktigare och att det gäller fler personer och inte färre. Men jag tror att svenska språket snabbt håller på att tappa mark till engelska. Alla som har ett någorlunda ovanligt intresse, har genom nätet fått möjlighet att utbyta erfarenheter med en person i Chile och en annan i Bulgarien. Deras gemensamma språk kommer att vara engelska. Kanske är en god engelskundervisning den bästa investeringen, och då är det kanske rätt och riktigt att inte lägga en massa energi på att digitalisera äldre böcker på svenska språket, som våra barnbarn kanske ändå aldrig kommer att använda. Eller om vi gör det, så borde vi samtidigt bifoga en översättning till engelska. För vad är ett läroverk, ett prebende, ett sockenstämmoprotokoll egentligen? How do you explain those terms in English? Redan inom Projekt Runeberg, vars ambition är att omfatta de nordiska språken och länderna, använder vi oftast engelska för att beskriva hur man scannar en bok.

Martin Persson