

FAKSIMIL

Donald Broady, *Det nya handbiblioteket. Forskningsprogram februari 1993*.
Institutionen för numerisk analys och datalogi, KTH, Stockholm 1993

[Ansökan till Forskningsrådsnämnden och Skolverket 15 febr. 1993.]

I denna pdf-version avviker sidbrytningen från originalet
men i övrigt inga ändringar.

1	
Bilaga till ansökan till FRN och Skolverket 1993-02-15	
Det nya handbiblioteket	
Projektprogram februari 1993	
Donald Broady	
KTH Inst f. numerisk analys och datalogi 100 44 Stockholm	HLS Inst. f. pedagogik Box 34103, 100 26 Stockholm
Innehåll	
I. Projektplan	3
II. Handbiblioteket	10
III. System för arbete med lokala dokumentbaser, önskvärda egenskaper	14
IV. Om beskrivande märkning och SGML	20
V. Några fördelar med SGML-märkta dokument	26
Bilagor	
1. Kunskapsverkstaden — om lokala dokumentbaser som arbetsverktyg för lärare, <i>Datorn i Utbildningen</i> , årg. V, nr 2 1992, pp. 22-28. [Äv. utgiven under samma titel som IPLab-rapport nr 47, Institutionen för numerisk analys och datalogi, KTH, 1991.]	28
2. Lokala dokumentbaser för undervisningsbruk, Teknologi och kompetens/Technology and Competence. Proceedings from the 8 th Nordic Conference on Information and Documentation, 1992, pp. 59-63.	37
3. Darc — Document ARchive Controller. Institutionen för numerisk analys och datalogi, KTH, prel. version 15 febr 1993.	44
<small>D. Broady, Det nya handbiblioteket, projektprogram februari 1993 Bilaga till ansökan till FRN och Skolverket 1993-02-15</small>	

Det nya handbiblioteket

Projektprogram februari 1993

Donald Broady

KTH
Inst f. numerisk analys och datalogi
100 44 Stockholm

HLS
Inst. f. pedagogik
Box 34103, 100 26 Stockholm

Innehåll

I. Projektplan	3
II. Handbiblioteket	10
III. System för arbete med lokala dokumentbaser, önskvärda egenskaper	14
IV. Om beskrivande märkning och SGML	20
V. Några fördelar med SGML-märkta dokument	26

Bilagor

1. Kunskapsverkstaden — om lokala dokumentbaser som arbetsverktyg för lärare, <i>Datorn i Utbildningen</i> , årg. V, nr 2 1992, pp. 22-28. [Äv. utgiven under samma titel som IPLab-rapport nr 47, Institutionen för numerisk analys och datalogi, KTH, 1991.]	28
2. Lokala dokumentbaser för undervisningsbruk, <i>Teknologi och kompetens/Technology and Competence. Proceedings from the 8th Nordic Conference on Information and Documentation</i> , 1992, pp. 59-63.	37
3. Darc – Document ARchive Controller. Institutionen för numerisk analys och datalogi, KTH, prel. version 15 febr 1993.	44

Läsanvisning

Medel för projektet söks hos två anslagsgivare, Forskningsrådsnämnden (området Biblioteks- och informationsvetenskaplig forskning) och Skolverket (enheten för utveckling). Detta projektprogram lämnas till både FRN och Skolverket.

FRN och Skolverket har hittills under snart två år (bå 1991/92 och 1992/93) bidragit till ett projekt "Datorstöd för kunskapsarbete", som avslutas i juni 1993 och som det nu föreslagna projektet skall bygga vidare på. (FRNs Dnr är 910176:1, Skolverkets delprojekt är rubricerat "Kunskapsverkstaden".) Vi räknar inte med att FRN och Skolverket skall ta definitiv ställning till föreliggande nya projektförslag innan vi inlämnat en redogörelse för arbetet inom det tidigare projektet, och vi kommer snarast möjligt att sammanställa en sådan.

Det är inte nödvändigt att läsa alla de följande, alltför många sidorna.

Den egentliga projektplanen utgör avsnitt I.

Själva projektidén presenteras närmare i avsnitt II och III samt i två tidigare skrivna uppsatser som bifogas som bilaga 1 och 2.

Den som är obekant med principerna för beskrivande märkning i allmänhet och SGML-märkning i synnerhet hänvisas till avsnitt IV och V.

Bilaga 3 presenterar det mest påtagliga resultatet från projektet Datorstöd för kunskapsarbete, nämligen en uppsättning verktyg för dokumenthantering som vi räknar med att använda inom det nya projektet.

I. Projektplan

Sammanfattning

Projektets syfte är att utveckla och pröva redskap och metoder som tillåter användare att bygga upp, ordna och använda ett eget "handbibliotek", bestående av maskinläsbara texter som följer den internationella märkningsstandarden SGML (Standard Generalized Markup Language, ISO 8879).

Avsikten är att användarna skall kunna överblicka, organisera och genomsöka handbiblioteket, tillfoga egna kommentarer och korshänvisningar, välja ut material för framställning av t.ex. skraddarsydd läromedel, färdigställa texter för utskrift eller fotosättning, samarbeta kring en gemensam uppsättning dokument, utbyta dokument eller hela dokumentinsamlingar, nyttja behörighetskontroll som i möjligaste mån skyddar textens integritet.

Särskilt intresse ägnas den humanistiska och samhällsvetenskapliga undervisningens och forskningens behov, samt bibliotekens möjligheter att förmedla maskinläsbar litteratur.

Som en fallstudie prövas metoder för att framställa en SGML-märkt maskinläsbar version av Nationalupplagan av August Strindbergs samlade verk i enlighet med de riktlinjer som utarbetats av TEI (Text Encoding Initiative).

Projektet är förlagt till Institutionen för numerisk analys och datalogi vid KTH, och bedrivs i samarbete med Litteraturvetenskapliga institutionen vid Stockholms universitet och Institutionen för pedagogik vid Högskolan för lärarutbildning i Stockholm.

Syfte

Projektets syfte är att utveckla och pröva redskap och metoder som tillåter användare att bygga upp, ordna och använda ett eget "handbibliotek", bestående av maskinläsbara texter som följer den internationella märkningsstandarden SGML (Standard Generalized Markup Language, ISO 8879).

Utveckling av metodik för arbete med SGML-märkning

Dokumentet som ingår i handbiblioteket bör vara flyttbara. De skall med andra ord icke vara bundna till maskinvara, tillämpningsprogram eller nationella teckenuppsättningar. Dessutom bör märkningen vara beskrivande, för att dokumentens innehållsliga struktur (rubriknivåer, fotnoter, olika slag av citat eller källhänvisningar etc) skall kunna bevaras i alla led av hanteringen. Dokumentbeskrivningsspråket SGML tillfredsställer kraven på flyttbarhet och beskrivande märkning.

SGML är en internationell standard, antagen 1986, som blivit ett slags "esperanto" för märkning och spridning av maskinläsbar text. Bland pågående försök att anpassa SGML-standarden till olika användningsområden bör här framhållas projektet TEI (Text Encoding Initiative), vari humanister och samhällsvetare i många länder medverkar. Inom ramen för TEI är man i färd med att utveckla riktlinjer för beskrivande märkning av texter av intresse

för humanister och samhällsvetare, allt från medeltida handskrifter till diktsamlingar och vetenskapliga monografier. Den andra och (åtminstone tills vidare) definitiva versionen av TEIs riktlinjer beräknas föreligga i juli 1993, men redan nu levereras de olika kapitlen till denna andra version via de elektroniska näten i takt med att de färdigställs. De för märkningsarbetet mest centrala kapitlen är redan tillgängliga. TEI innebär förhoppningsvis genombrottet för en internationell standard för framställning och spridning av maskinläsbar humanistisk och samhällsvetenskaplig litteratur. Det är nu dags att pröva TEIs riktlinjer på svenska textmaterial.

Frågor kring beskrivande märkspråk har i Sverige hittills i hög grad varit en angelägenhet för dataloger och vissa specialister bland dokumentalister och inom förlags- och tryckeribranschen, samt för de stora exportföretagen (där man just nu på många håll är i färd med att utreda möjligheten att märka sin dokumentation i enlighet med SGML-standard; den utlösande faktorn är att utländska affärspartners kräver sådan dokumentation). Dessutom har svenska lingvister haft intresse av SGM-märkning som ett medel för att göra stora textcorpora tillgängliga för språklig analys. När inom kort TEIs riktlinjer fullbordas får frågorna aktualitet även för andra humanister och samhällsvetare.

Projektet skall utveckla och pröva metoder för framställning av en maskinläsbar version av Nationalupplagan av August Strindbergs samlade verk (eller delar därav) i enlighet med TEIs riktlinjer. Den pågående utgivningen av Strindbergs samlade verk är ett lämpligt experimentalfält för en sådan metodikutveckling, bl.a. eftersom skilda litteraturarter och genrer finns representerade (prosa, dramatik, lyrik och brev, kommentarer och ordförklaringar etc). Dessutom är Strindbergs författarskap av den digniteten att det tilldrar sig intresse utanför Sveriges gränser, vilket talar för att man snarast bör undersöka möjligheterna att ge ut en maskinläsbar version som ansluter till internationell märkningsstandard. Här kommer att krävas viss utveckling och utprovning av för detta märkningsarbete nödvändiga redskap, såsom SGML-editorer, program för validering av den färdiga TEI-märkningen samt behövliga konverteringsrutiner. Under förutsättning att TEIs riktlinjer visar sig användbara kommer projektet att åstadkomma en lathund som fortsättningsvis kan användas vid märkning av Strindbergsmaterialet och liknande material. För tydlighet skall bör framhållas, att det föreslagna projektet inte syftar till att framställa en färdig maskinläsbar version av Strindbergupplagan. Innan ett sådant arbete kan påbörjas måste metodiken utvecklas och prövas — härtill skall det föreslagna projektet bidra. (Här måste inflickas, att särskilda medel för detta arbete med Strindbergsmaterialet söktes hos HSFR, se avsnittet om finansiering nedan. Förutsatt att HSFR beviljar medel behöver således inte det här föreslagna projektets budget bära hela kostnaden för arbetet med att finna metoder för märkning av Strindbergsmaterialet — vilket inte hindrar att denna metodikutvecklingen kommer att vara av största betydelse för det projekt som nu föreslås FRN och Skolverket.)

Dessutom måste självfallet de upphovsrättsliga problemen lösas. De hinder som står i vägen för framställningen och spridningen av maskinläsbart material är i dag framför allt av affärsjuridisk och upphovsrättslig art. Lagstiftningen och standardavtalsjuridiken är föråldrad och duger inte när vi har att göra med maskinläsbara dokument. För utbildningsväsendets vidkommande (skolor, universitet och högskolor) är det en akut uppgift att få till stånd en motsvarighet till det kopieringsavtal som gäller för spridningen av pappersburen information. Utan ett fungerande regelverk för spridningen av maskinläsbara dokument kommer författare och förläggare även fortsättningsvis att dra sig för att göra sina alster allmänt tillgängliga. De upphovsrättsliga problemen faller utanför detta projektförslag, men i andra sammanhang har vi försökt komma någon vart med dem. Jag vill här särskilt nämna ett nyss påbörjat utredningsarbete (under ledning av Peter Seipel, Institutet för rättsinformatik, Stockholms universitet, och med stöd från NUTEK och BIBSAM/Kungl. biblioteket) rörande upphovs-

och spridningsrätt för maskinläsbar litteratur, i samband med vilket vi hoppas kunna använda Strindbergupplagan som ett pilotfall.

Det är en angelägen uppgift att på svensk botten skaffa erfarenheter av märkningsarbete enligt TEIs riktlinjer, och att utveckla verktyg som stödjer sådant arbete (framför allt verktyg som reducerar de rutinbetonade arbetsuppgifterna till ett minimum och som automatiskt validerar den färdiga TEI-märkningen). Projektet måste även behandla mer generella problem kring svenska tillämpningar av SGML-standarden: Vilka speciella krav ställer svenskt språkbruk och svenska typografiska konventioner? Vilka principer bör styra utvecklingen av DTD:er för t.ex. olika typer av undervisningsmaterial? I vilken grad är en rent beskrivande märkning tillräcklig, och hur kan information av typografisk art tillfogas?

Ytterligare en viktig uppgift är inventering av det internationella utbudet av andra, existerande och planerade, samlingar av maskinläsbara texter lämpade för humanistisk och samhällsvetenskaplig undervisning och forskning, samt undersökningar av hur tillgängliga dessa samlingar är. Frågan om tillgänglighet har flera aspekter: juridiska (upphovsrättsproblemen är en central fråga, många samlingar som i dag är fritt tillgängliga över universitetsnäten får endast användas för forskningsändamål), tekniska (anpassning till internationell märkningstandard, överföringsrutiner etc) och arbetsmässiga (att veta att materialet finns och varifrån det kan hämtas, att bedöma dess art och kvalitet, etc).

Utveckling och utprövning av byggstenarna i ett ändamålsenligt dokumenthanteringssystem

Arbetet inom projektet skall bidra till utvecklingen av system och arbetsmetoder som tillåter yrkesgrupper som lärare, läromedelsförfattare, forskare, redaktörer, bibliotekarie, dokumentalister och informatiker att utbyta och göra fruktbara bruk av maskinläsbara texter. En centralt syfte är att befrämja framställningen och spridningen av maskinläsbar litteratur för undervisnings- och forskningsändamål, företrädesvis inom samhällsvetenskap och humaniora, samt att utpröva märkningsrutiner enligt internationell standard som gör sådan litteratur lätt åtkomlig och mångsidigt användbar. De redskap och metoder som utvecklas inom projektet skall fungera som stöd för användare som

- märker dokument enligt SGML-standarden eller handskas med redan SGML-märkta dokument,
- införlivar dokumenten med sitt handbibliotek,
- ordnar, överblickar och genomsöker handbiblioteket och extraherar fram utsnitt därur,
- indexerar eller katalogiserar dokumenten,
- tillfogar kommentarer och korshänvisningar (s.k. hypertextlänkar),
- samarbetar kring en gemensam uppsättning dokument,
- utnyttjar behörighetskontroll som i möjligaste mån skyddar textens integritet,
- utbyter dokument och dokumentsamlingar med kolleger,
- överför dem för bearbetning i t.ex. text- eller bildbehandlingsprogram,
- färdigställer dem för utskrift eller fotosättning.

På alla dessa punkter kommer vi att kunna bygga vidare på erfarenheterna från projektet Datorstöd för kunskapsarbete, som vi under snart tre år bedrivit med medel från NUTEK och (under projektets två sista år) från Skolverket och FRN. De redskap som vi under denna period utvecklat kommer att användas inom det här föreslagna projektet. Se bilaga 3 för en presentation av det viktigaste resultatet, dokumenthanteringssystemet Darc (Document Archive Controller). Systemet visades för första gången upp i ett internationellt sammanhang

under konferensen ECHT'92 (The fourth ACM Conference on Hypertext), Milano, 30 nov—4 dec 1992, och vi har redan fått många förfrågningar från olika håll i världen. Systemet har försetts med avsevärt mycket mer funktionalitet än vad som ursprungligen planerats. Från början var systemet närmast tänkt att vara ett kitt mellan olika färdiga komponenter, men har nu utvecklas i riktning mot en generell tillämpning på ett just nu synnerligen expansivt område (hantering av SGML-kodade dokument) där efterfrågan på goda verktyg just nu är i växande. I Sverige har preliminära versioner av Darc demonstrerats bl.a. för medlemmar av Stimdi (Sveriges tvärvetenskapliga intresseförening för människa-dator-interaktion), för ett par hundra datalärare och, självfallet, för många besökare som kommit för att se projektets installationer vid KTH. En färdig version kommer att vara tillgänglig för betatestning (dvs. levereras till var och som vill pröva den under ordnade former) under mars månad 1993.

Det har varit tidskrävande och kostsamt att skapa dessa verktyg (finansieringen av programmeringsarbetet har framför allt kommit från NUTEK). Nu utgör de en grund för fortsatt arbete, vilket garanterar att det här föreslagna projektet får en flygande start.

Vad gäller tillämpningar i arbetslivet kommer särskilt intresse ägnas behoven hos två kategorier användare:

- lärare och läromedelsförfattare, som har behov av dokumentsamlingar vilka kan fungera som råmaterial för lokal framställning av skräddarsydda läromedel,
- bibliotekarier, dokumentalister och informatiker, som har behov av fulltextdatabaser bestående av strukturerade dokument, förberedda för någorlunda automatisk katalogisering och indexering samt flyttbara (att dokumenten är flyttbara innebär här att de kan levereras till mottagare med tillgång till olika slag av datorutrustning och programvara).

Finansiering, skälen till att vi vänder oss till Skolverket och FRN

Med tanke på de två nämnda användarkategorierna föreslås att projektet samfinansieras av Skolverket (enheten för utveckling) och FRN (området Biblioteks- och informationsvetenskaplig forskning). Detta projektprogram sänds till båda anslagsgivarna. Ansökningarnas kostnadsberäkning är utformad så att var och en av projekts kostnadsposter fördelas lika mellan Skolverket och FRN.

Att vi vänder oss till Skolverket motiveras av att lärare eller läromedelsförfattare skulle kunna ha nytta av ett innehållrikt, välordnat och flexibelt "handbibliotek". Lärare kan behöva ett sådant i sin egen dagliga fortbildning. En enskild lärare eller ett arbetslag kan i handbiblioteket samla och ordna undervisningsmaterial, provbanker, etc för utvecklingen av den lokala arbetsplanen, för kurs- och lektionsplaneringen — och kanske framför allt för den kontinuerliga lokala framställningen av skräddarsydda läromedel. Informationsteknologins utveckling kan komma att förändra den traditionella läromedelsmarknaden, så att förlag, läromedelsförfattare, forskargrupper och innovativa lärare distribuerar maskinläsbart "halvfabrikat" av hög kvalitet, varav lärare skapar skräddarsydda läromedel i enlighet med de lokala behoven. Detta halvfabrikat kan i dag enklast och billigast distribueras på kompaktskivor. (Här bör observeras de helt nya möjligheter till småskalig produktion som uppstått i och med att en inspelningsutrustning för individuella CD-ROM-skivor nu inte kostar mer än fotokopieringsmaskin, ca 8 000 dollar i USA. En tomskiva, som rymmer 600 MB, kostar mellan 25 och 40 dollar. Därmed är läromedels- och CD-ROM-marknadernas monopolistiska struktur inte längre tekniskt motiverad.) Även läroplanen och kursplanerna och andra styrdokument samt kommentar- och referensmaterial bör levereras i sådan maskinläsbar form att det i valda delar kan flyta in i lärarnas handbibliotek. (Ang. tillämpningar inom skolväsendet, se bilagorna 1 och 2 nedan).

Att vi vänder oss till FRNs biblioteks- och informationsvetenskapliga program motiveras av att biblioteken borde kunna vara knutpunkter i flödena av maskinläsbara dokument. En naturlig uppgift för bibliotekarier, dokumentalister och informatiker vore att inventera, samla och förmedla fulltextmaterial, men enligt en aktuell utvärdering förmedlar biblioteken världen över i dag blott 5% av alla offentligt tillgängliga maskinläsbara dokument lämpade för humanistisk undervisning och forskning.

Tillfogas bör, att jag med Lars Dahlbäck som medsökande dessutom den 15 januari 1993 till HSFR inlämnade en ansökan om medel för ett näraliggande projekt, "Hantering av maskinläsbar humanistisk och samhällsvetenskaplig litteratur". Det projekt som föreslagits HSFR är tvåårigt (bå 1993/94 och 1994/95) och gäller i högre grad de specifika problemen kring SGML-märkningen av Nationalupplagan av August Strindbergs samlade verk och annan humanistisk och samhällsvetenskaplig litteratur, medan det projekt som föreslås Skolverket och FRN är mer inriktat mot utveckling och utprovning av system för yrkesgrupper som lärare, läromedelsförfattare, bibliotekarier, dokumentalister, informatiker etc.

Personal, arbetsfördelning

Personal:	Donald Broady	Docent, tf prof vid HLS Institutionen för pedagogik och forskare vid KTH, Institutionen för numerisk analys och datalogi.
	Lars Dahlbäck	Docent, Litteraturvetenskapliga institutionen, SU. Huvudredaktör för Nationalupplagan av August Strindbergs samlade verk.
	Hasse Haitto	Civ.ing., doktorand vid KTH, Institutionen för numerisk analys och datalogi och forskningsassistent vid HLS, Institutionen för pedagogik.
	Eva Trotzig	Fil. lic., chefsbibliotekarie för Statens psykologisk-pedagogiska bibliotek
	Forskningsass.	En eller flera doktorander vid Litteraturvetenskapliga institutionen, SU. Arbetsuppgift: märkning av Strindbergmaterialet
	Programmerare, dokumentalist, försöksledare, timanställda	Som programmare önskar vi anlita Peter Lidbaum och Magnus Tobiasson, civilingenjörer med datalogisk inriktning. Som dokumentalist ev. Krister Lagerborg, informatiker vid SPPB. Som försöksledare sannolikt lärare vid HLS.

Projektet förläggs till IPLab (Interaktions- och presentationslaboratorium), Institutionen för numerisk analys och datalogi, KTH.

Sedan 1988 har undertecknad arbetat på deltid som forskare inom denna miljö, från och med juli 1990 som projektledare för "Datorstöd för kunskapsarbete" (även benämnt "Datorstöd för författande och samarbete i utbildning"). Finansiärer har varit NUTEK (ITYP projekt 90-02737P), Skolverket och Forskningsrådsnämnden. Inom detta projekt, som avslutas i juni 1993, har vi utvecklat redskap för arbete med lokala dokumentbaser, vilka är tänkta att fungera som ett slags handbibliotek av exempelvis ett lärarlag, en forskargrupp, en redaktionskommitté. I dessa dokumentbaser lagras och bearbetas både användarnas eget arbetsmaterial och material som hämtas hem från annat håll. Ett centralt användningsområde är framställning av skraddarsydda läromedel. Dokumenten hanteras företrädesvis i SGML-format, och vi har ägnat stort intresse åt att följa SGML-standardens utveckling och dess implementeringar runt om i världen. Det projekt som härmed föreslås Skolverket och FRN kan med andra ord utnyttja detta tidigare arbete.

Arbetet sker i samverkan med Litteraturvetenskapliga institutionen, Stockholms universitet. Docent Lars Dahlbäck ansvarar för den vetenskapliga ledningen av det arbete inom det föreslagna projektet som är av litteraturhistorisk och filologisk karaktär. Lars Dahlbäck är huvudredaktör för Nationalupplagan av August Strindbergs samlade verk, vars redigeringsprinciper bygger på den utredning Dahlbäck genomförde i slutet av 1970-talet. Via Dahlbäck har projektet kontakt med redaktörerna för de olika volymerna i Strindbergutgåvan. Vidare kommer en doktorand med anknytning till Strindbergutgåvan att anlitas för att ombesörja åtskilligt av det kontinuerliga märkningsarbetet.

I fråga om de pedagogiska implikationerna (såsom möjligheterna att ur en dokumentbas med SGML-märkt material framställa skraddarsydda läromedel, referensmaterial etc) samverkan med Högskolan för lärarutbildning i Stockholm, Institutionen för pedagogik.

I projektet medverkar även fil. lic. Eva Trotzig, chefsbibliotekarie för Statens pedagogisk-psykologiska bibliotek, som i synnerhet bevakar bibliotekstillämpningar och anknytningar till bibliotekssystem.

Bland övriga kontakter som kommer att vara värdefulla för projektarbetet bör särskilt nämnas professor Gunnel Engwall, romanska institutionen, Stockholms universitet, som dels är ordförande för ledningsgruppen för Nationalupplagan av August Strindbergs samlade verk, dels från svensk sida medverkat i utarbetandet av TEIs riktlinjer.

Tidsplan

Projektet beräknas starta 1993-07-01 och vara slutfört 1996-06-30.

De ovan nämnda deluppgifterna är av den karaktären att de i stor utsträckning måste genomföras parallellt. Därför vore det vilseledande att här presentera en tidsplan där projektets arbete styckas upp i olika perioder. Dock kan några arbetsuppgifter nämnas som måste prioriteras redan i projektets inledningsskede:

- Vissa systemutvecklingsinsatser som krävs för att "handbiblioteket" skall kunna utnyttjas fullt ut i realistiska arbetssituationer: utveckling av redskap som tillåter att utsnitt ur dokumentbaser omedelbart omvandlas till "kompendier" (i dag måste användaren ta omvägen över t.ex. textbehandlings- och layout-program); utveckling av fler konverteringsrutiner mellan SGML och vissa textbehandlingsformat; utveckling av

möjligheterna att inkludera icke-SGML-märkt material i dokumentbasen; utveckling av stöd för bildhantering. Allt detta förutsätter utveckling av mer eller mindre automatiserade rutiner som innebär att användaren så litet som möjligt skall behöva bekymra sig om SGML-märkningens krav.

- För att utveckla samordningen mellan lärares eller forskares "handbibliotek" och befintliga bibliotekssystem krävs studier av hur de bibliografiska basernas poster kan hanteras med SGML-stöd. Å ena sidan bör lärare eller forskare kunna hämta poster från bibliotekssystemet och läsa in valda delar i sitt handbibliotek, å andra bör de maskinläsbara dokument som lärare och forskare levererar till biblioteken vara utformade så att upprättandet av posterna underlättas, och kanske till och med kan automatiseras. På detta område finns en del förarbeten från TEI-projektet att bygga vidare på (TEI's riktlinjer för bibliografisk information är anpassad till olika standarder som bygger på ISBD, International Standard Book Description), men mycket återstår; anpassningen till LIBRIS-formatet är en särskilt angelägen uppgift. Ännu mer näraliggande för det här föreslagna projektet är att bidra till utvecklingen av bibliotekens möjligheter att ta emot och förmedla fulltext-dokument, och vi avser att först och främst bygga bryggor till UNIX-baserade bibliotekssystem såsom TinLib.
- Utprovning av TEI's riktlinjer för dokumentmärkning, utveckling av de konverteringsrutiner som behövs för återanvändningen av de databand som som kontinuerlig framställts i samband med tryckningen av Nationalupplagan av August Strindbergs samlade verk, samt undersökningar av de speciella krav som uppkommer vid märkningen av detta material.

II. Handbiblioteket

En tumregel är att datortekniken kommer till korta, eller till ingen användning alls, om den förhindrar människor att göra det de brukar göra med exempelvis pappersburen information. En humanist eller samhällsvetare utvecklar med åren sofistikerade sätt att läsa och skriva och att hålla ordning på sina anteckningar, pappershögar och samlingar av böcker och tidskrifter. Ett i sammanhanget brukbart dokumenthanteringssystem måste bereda utrymme för dessa värdefulla och mödosamt förvärvade förmågor — självfallet utan illusioner om att de maskinläsbara dokumenten skulle ersätta böcker eller tidskrifter. Vid mer intensivt umgänge med en text brukar man önska tillgång till den såväl på papper, för överblickens och de estetiska kvaliteternas skull, som i maskinläsbar form, för kunna bearbeta den, kommentera, söka, extrahera, indexera, kompilera etc.

Ett bra handbibliotek är en oskattbar tillgång för den som arbetar som exempelvis lärare, forskare, redaktör, översättare eller dokumentalist/informatiker. Med handbibliotek avser jag helt enkelt den samling av material — böcker och tidskrifter, papper i travar och i holkar och pärmar, kortlådor och disketter — som man önskar ha nära till hands i arbetsrummet därhemma eller på arbetsplatsen.

Handbiblioteket utgör ofta hjärtat i, låt säga, en lärares eller forskares arbetsmiljö. Jag har visserligen mött filosofer som påstår sig behöva föga mer än sitt huvud och papper och penna för att kunna arbeta, men vanligare är att man är beroende av ett rum eller åtminstone ett par bokhyllor med det material man önskar ha på armlängds eller några stegs avstånd: speciallitteratur, allmän referenslitteratur såsom handböcker, lexika och ordböcker, mapper med egna och kollegers skrifter, elevarbeten, egna halvfärdiga texter ...

Flertalet humanister och samhällsvetare torde utnyttja sitt handbibliotek ungefär på det sätt som illustreras av Figur 1 nedan. In flyter böcker, tidskrifter, fotokopior, rapporter och manuskript som man ordnar på hyllor och bord, samt datafiler som man skaffat sig och som vanligen lagras på något magnetiskt eller mer sällan optiskt medium (diskett, skivminne i den egna datorn eller i en server i ett nätverk, kompaktskiva). Utöver det egentliga handbiblioteket har man på sina bord och hyllor material under arbete, dvs. ofärdiga manuskript och utkast, anteckningar och excerpter, material med anknytning till pågående forsknings- och undervisningsuppgifter, elevarbeten etc, allt detta på papper eller på magnetiskt medium. Man har även tillgång till externa källor såsom bibliotek, arkiv, databasvärdar, men sådant material finns inte till hands i varje ögonblick utan kräver att man beger sig någonstans eller anlitar postverket eller telefonnätet; det viktigaste materialet kan man sedan välja att införliva med handbiblioteket genom hemlån eller fotokopiering eller nedladdning av filer. Ut flyter papper eller datafiler med manuskript till artiklar och böcker, rapporter, undervisningsmaterial etc, varav något exemplar brukar införlivas med handbiblioteket.

Figur 1 visar således flödena i ett arbetsrum vars innehavare inte gör särskilt mycket bruk av datortekniken annat än möjligen för textbehandling, databearbetning, lagring, utskrift, uppringning av externa databaser etc. Figur 2 visar byggstenar som skulle kunna ingå i ett dokumenthanteringssystem med väl utbyggt datorstöd. Även här är "handbiblioteket" en hjärtpunkt. Det rör sig då inte om pappersburen information utan om maskinläsbart material samlat i en lokal dokumentbas. Denna dokumentbas är lokal såtillvida att en användare (eller en grupp användare) i alla avseenden rör över materialet, ungefär som över innehållet i de egna bokhyllorna.

Genom att använda handbiblioteket som metafor för den lokala dokumentbasen vill jag framhäva tillgängligheten. Dokumenten skall vara tillgängliga i flera avseenden. De skall vara lätt åtkomliga (kanske lokalt lagrade på skivminne i den egna datorn eller i ett lokalt nätverk) och de skall vid behov ofördröjligen kunna bearbetas med användarens egna

favoritprogram (textbehandlingsprogram, bildbehandlingsprogram, kalkylprogram etc). De skall kunna bearbetas, kompletteras, förses med randmärkningar och korshänvisningar (s.k. hypertextlänkar). Samlingen skall vara överblickbar, man skall utan stort besvär erhålla automatiskt genererade innehållsförteckningar eller kartor över över dokumentbasens aktuella bestånd, och man skall på enkelt sätt kunna revidera inte blott de enskilda dokumenten utan även själva dokumentbasens struktur (ungefär som när man flyttar om böckerna på hyllorna eller sorterar pappren i pärmarna).

Materialet i den lokala dokumentbasen kan hemföras från många håll och via många medier: via näten, från böcker och tidskrifter via skanner kombinerad med program för optisk teckenigenkänning och bildbehandling, samt förstås material som man själv och kollegerna framställer.

Dessutom kan material hämtas från externa dokumentbaser, varmed jag här avser material som inte är omedelbart integrerat med den lokala dokumentbasen. Sådant material skulle kunna tillhandahållas av exempelvis forskargrupper, förlag eller bibliotek. Redan i dag blir det, särskilt bland dataloger, allt vanligare att forskargrupper gör sina uppsatser och rapporter allmänt tillgängliga via de internationella universitetsnäten. Förutsatt att de f.n. besvärliga upphovsrättsproblemen löses kommer utan tvivel allt mer maskinläsbar litteratur att saluföras, i dag billigast och enklast på kompaktskivor, och därmed fungera som externa dokumentbaser, från vilka användaren kan välja material att införliva med sin lokala dokumentbas. Så skulle en maskinläsbar version av Strindbergs samlade verk kunna användas.

Var och en som försöker använda sådant material i större skala vet hur svårt det är eftersom formaten är så mångskiftande. Att få ut dokumenten på papper brukar inte vålla större besvär, de anländer vanligen som ASCII- eller PostScript-filer eller i något annat bekant format, men om det rör sig om många eller komplexa dokument krävs mycket tid och datorkunnande för att stöpa om dem så att de passar min maskin och mina favoritprogram — och trots alla ansträngningar händer ofta att dokumentens innehållsliga struktur (dvs relationerna mellan enheter som kapitel, rubriknivåer, fotnoter etc) eller de nationsspecifika tecknen och vissa specialtecken deformerar. Först i och med SGML, den internationella standarden för beskrivande märkning, blir det möjligt att jämförelsevis problemfritt och felfritt låta dokument med bibehållen struktur cirkulera tvärs över nationsgränserna, oberoende av olikheter i hårdvara, operativsystem, tillämpningsprogram, tillåtna filnamn etc, och (så länge vi håller oss till västerländska språk) oberoende av skillnader mellan nationella teckenuppsättningar.

Det är väsentligt att den lokala dokumentbasen är nära integrerad med material under arbete. Således skall dokumentbasens material snabbt och smärtfritt kunna överföras till exempelvis användarens ordinära ord- eller bildbehandlingsprogram, där det kan redigeras, kommenteras, sammanföras med annat material under arbete etc. Omvänt skall det finnas möjligheter att införliva material av egen tillverkning med den lokala dokumentbasen, en överföring som dock med nödvändighet blir mer komplicerad och knappast kan automatiseras fullt ut.

Från denna arbetsmiljö levereras materialet på skilda sätt: för direkt skärmpresentation (med hjälp av program som DynaText eller de presentationsverktyg vi själva utvecklat vid KTH), till enkel utskrift på skrivare eller till t.ex. LaTeX för mer sofistikerad laserutskrift eller fotosättning, för distribution på magnetiskt medium (diskett, band) eller över näten meddelst elektronisk post eller konferenssystem.

Den lokala dokumentbasen kan även göra tjänst som hjärtpunkten i en skrifvarverkstad, där ett lärarkollegium, en redaktionsgrupp, en grupp läromedelsförfattare eller en forskargrupp framställer mer genomarbetade och omfattande dokumentsamlingar vilka förtjänar att spridas

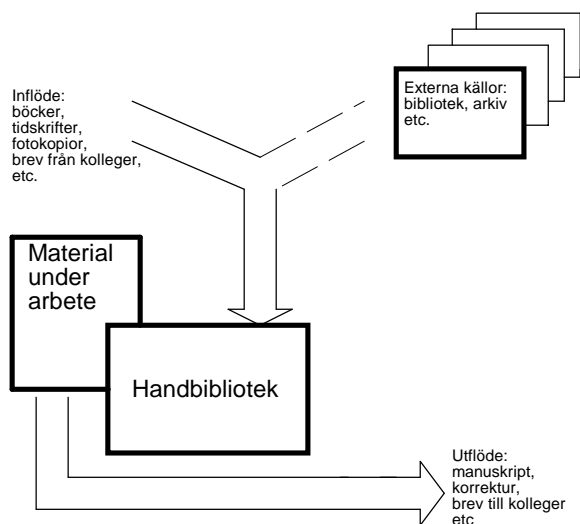
till andra. Dessa samlingar kan därmed, för att använda terminologin i figur 2, ingå i utbudet av offentligt tillgängliga "externa dokumentbaser".

Ett naturligt användningsområde för en lokal dokumentbas vore att ur den hämta råmaterial eller halvfabrikat med vars hjälp lärare förfärdigar skraddarsydda läromedel, dvs. kompendier där utsnitt ur dokumentbasen (med automatiskt genererade "källhänvisningar") kompletteras med lärarnas egna tillskott. Förlagen eller biblioteken, författare, redaktörer och lärarkolleger skulle kunna tillhandahålla dokumentbaser som är rikhaltiga och av god kvalitet (låt säga Strindbergs samlade verk tillsammans med kommentarer och sekundärlitteratur), varur läraren framställer undervisningsmaterial.

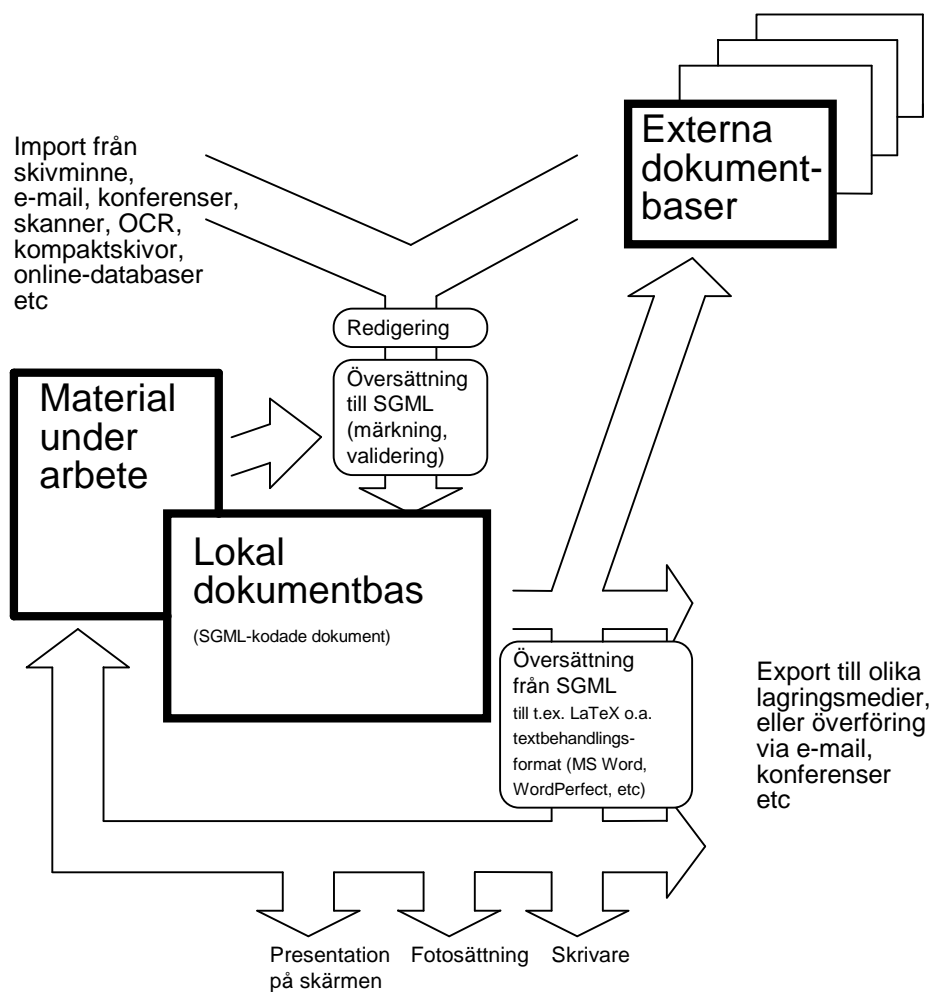
Ett annat användningsområde är att bibliotekarier, dokumentalister eller informatiker samlar material i en lokal dokumentbas, för att därifrån leverera maskinläsbara dokument till användarna. Detta borde vara en naturlig arbetsuppgift för bibliotek, men så är knappast fallet. Bland offentligt tillgängliga maskinläsbara dokument lämpade för humanistisk undervisning och forskning som finns offentligt tillgängliga världen över uppskattar man i en aktuell inventering att blott 5% tillhandahålles av bibliotek (enligt MRTH, Rutgers Inventory of Machine-Readable Texts in the Humanities). Man kunde tänka sig att ett skolbibliotek bygger upp en lokal samling att användas i arbetet med lokala arbetsplaner, terminsplaneringar, läromedelsframställning etc, eller att biblioteket vid en universitetsinstitution samlar elektroniska versioner av uppsatser, rapporter, undervisningsmaterial som medarbetarna där framställer, eller att forskningsbiblioteken beställer och tillhandahåller mer omfattande dokumentbaser.

Närmare samarbete mellan lärare och bibliotekarier vore önskvärt. En rimlig framtidsvision att lärarna och forskarna från sin ordinarie arbetsstation har nätförbindelse med sin institutions eller skolas bibliotek, hämtar maskinläsbart material därifrån och omvänt förser biblioteket med de uppsatser och rapporter som förtjänar att inlemmas med beståndet av offentligt tillgängliga dokument. De maskinläsbara dokument som biblioteken tar emot bör självfallet vara märkta på så sätt att katalogiseringen i största möjliga utsträckning kan automatiseras.

De dokument som flyter till och från biblioteken bör levereras i flyttbart format, så att mottagarna utan alltför mycket besvär kan använda dokumenten på sin egen maskin och med hjälp av sina favoritprogram. Det är dessutom väsentligt att dokumentens struktur bevaras, dvs att de olika innehållsliga element varav dokumentet består (sådant brödtext, fotnotstext, rubriknivåer, källhänvisningar, "front matter" såsom titel, författarnamn o.a. bibliografiska upplysningar) när mottagaren i det skick som författaren avsett — det gäller att undvika situationen att överföringen resulterar i en lång teckensträng, som mottagaren får ägna mycken möda åt att redigera. Allt detta talar för en märkning enligt SGML-standarden.



Figur 1. Flöden i ett ordinärt arbetsrum, utan särskilt mycket datorstöd



Figur 2. Flöden vid arbete med maskinläsbara, företrädesvis SGML-märkta dokument

III. System för arbete med lokala dokumentbaser, önskvärda egenskaper

- Så få antaganden som möjligt om hur dokumenten kommer att användas.
- Så få antaganden som möjligt om vilka maskiner och program som skall ta hand om dokumenten.
- Stöd för arbete med — inte enbart för presentation av — dokumenten.
- Stöd för att revidera, ordna och överblicka dokumentbasen.
- Stöd för samarbete. Behörighetskontroll avseende både enskilda dokument och dokumentbasens organisering.
- Stöd för hantering av stora och komplexa dokument.
- En och samma dokumentsamling skall kunna användas av både noviser och specialister.
- Modifierbart och utbyggbart system: fler medier och fler program skall kunna tillföras.
- Åtskilligt arbete med dokumenten skall kunna ske med enkel utrustning.
- Systemet skall stödja hantering av strukturerade dokument. Beskrivande märkning, ej procedurmärkning.

Här följer en lista över krav som kan ställas på byggstenarna i ett dokumenthanteringssystem som det ovan skisserade. (I systemet Darc som utvecklats vid KTH har vi efter förmåga försökt att tillfredsställa de flesta av dessa krav, se bilaga 3.)

1. Systemutvecklarna bör göra så få antaganden som möjligt om hur dokumenten kommer att användas.

Användaren kanske vill läsa dokumenten på skärmen eller på papper, redigera, komplettera, extrahera fram material till ett kompendium, eller formge materialet typografiskt för utskrift på skrivare eller för vidare befordran till ett fotosätter. Dokumentsamlingen skall också kunna användas för sökning, kompilering eller innehållslig, semantisk, syntaktisk eller lexikalisk analys etc. Systemkonstruktören skall i så ringa utsträckning som möjligt försöka förutsäga och styra användarnas behov.

2. Dokumenten skall var flyttbara.

Systemutvecklarna bör göra så få antaganden som möjligt om vilka maskiner och program användaren har eller föredrar. Användaren skall till vardags kunna arbeta med t.ex. sin favoritordbehandlare eller sin favoritgrafikeditor under sitt favoritoperativsystem (i dag vanligen MS DOS, MS Windows, UNIX eller Macintosh OS), samtidigt som dokumenten skall kunna flyttas till andra miljöer. Systemet bör uppmuntra spridning av materialet, dvs källtexterna själva bör vara åtkomliga och inte levereras i kompilerade versioner som man behöver särskilda program för att komma åt.

Här kolliderar lärarnas och forskarnas behov med de kommersiella intressena. Av naturliga skäl önskar programvaruleverantörer att man använder deras produkter. Därför brukar den som vill publicera material i maskinläsbar form, exempelvis på CD-ROM, mötas av förslag i stil med: —Utmärkt, vårt företag tar hand om era dokument och stoppar in dem i vårt system som ger suveräna möjligheter till indexering, informationsåtervinning, sökning etc och kostar någon tusenlapp per CD-ROM. (Därmed förfuskas idén med flyttbara dokument. Det är som om en författare som går upp på ett förlag med sitt manuskript skulle få beskedet: — Visst, vi ger gärna ut din bok, som vi för att den ska bli riktigt bra trycker

med osynligt bläck av vår egen förnämliga tillverkning, vilket förutsätter att läsekretsen införskaffar våra specialglasögon för tusen kronor styck.)

Ett annat sätt att uttrycka samma sak är att hävda att dokumentet bör sättas i centrum. Hittills har man inom personatorvärlden på sätt och vis satt programmet i centrum. Man har definierat dokumentet i termer av det tillämpningsprogram i vilket det skapats eller senast lagrats. Man talar således om ett WordPerfect-dokument eller ett Excel-dokument. Vidare definierar man sina dokumentsamlingar med hänvisning till maskinvara och operativsystem: man säger att man arbetar under DOS eller Windows eller Macintosh OS. I många sammanhang vore det bättre att kunna sätta dokumentet i centrum. Dagens teknik erbjuder viss möjligheter att bearbeta ett och samma dokument eller en och samma dokumentsamling på mångahanda sätt och med många slag av program på många slag av maskiner. För en humanist eller samhällsvetare är detta ofta ett naturligt sätt att arbeta: jag har en intervju eller en källtext och det är den jag önskar bearbeta med skilda redskap. Jag vill inte först starta en tillämpning och därefter öppna en fil som har mening uteslutande för mitt operativsystem eller rentav blott för ett enskilt program.

3. Systemet bör stödja samarbete. Flera medarbetare skall kunna använda samma eller varandra överlappande lokala dokumentbaser.

Här är behörighetskontroll väsentlig, och denna bör gälla såväl de enskilda dokumenten som hela dokumentbasens organisering. I en redaktionsgrupp eller forskargrupp kan huvudredaktören eller den vetenskaplige ledaren tilldelas vittgående befogenheter, kanske t.o.m. rätt att organisera hela dokumentbasen och att ändra allt innehåll, medan andra medarbetare med mer begränsad befogenhet endast har rätt att redigera vissa dokument, och åter andra endast rätt att läsa vissa dokument, att för eget bruk infoga korshänvisningar och att skapa egna personliga kartor över basens innehåll.

4. Systemet bör stödja arbete med — inte enbart presentation av — dokumenten.

Någon läsare kanske undrar varför ordet multimedia inte förekommit i denna framställning. Det kan synas en smula otidsenligt att lägga så stor vikt vid texthantering som här sker.

Jag tror att det är olyckligt och onödigt att utvecklingen av de basala arbetsverktygen för hantering av text och stillbilder har hamnat en smula i skuggan av den mer uppseendeväckande utvecklingen av presentationsverktygen. När termen "multimedia" introducerades innebar den helt enkelt att en personator presenterade ljud, stillbilder, animerade bilder och videosekvenser, vilket lät sig åstadkommas genom att videospelare och högtalare kopplades till datorn. I dag har ambitionerna vuxit. Man strävar efter att bygga digitala system som mest av allt liknar en professionell videoproduktionsstudio. Man skall med andra ord samtidigt hantera ett flertal kanaler med video- och ljudinformation som kan spelas in, bearbetas, spelas in på nytt, spelas upp och sammanfogas till ett enda program. Ett sådan utrustning ger förnämliga möjligheter när det gäller att presentera ett färdigt material. Det är dock få förunnat att kunna hantera sitt eget arbetsmaterial på så sätt. Att skapa en fullfjädrad multimediassession kräver knappast mindre kompetens (inte blott tekniskt kunnande, utan även i fråga om formgivning, bild- och ljudhantering etc) än att producera en musikvideo. Bl.a. på grund av att ljudmaterial och videosekvenser är tidsberoende information och inte "maskinläsbar" i samma mening som text och vanlig grafik, är det dessutom svårt att automatisera sökning, informationsåtervinning, konvertering och liknande rutinuppgifter. Utan tvivel kommer det med tiden att bli allt lättare även för amatörer att någorlunda snabbt redigera och kombinera digitalt lagrad ljudinformation och digitaliserade rörliga bilder, men i dag utgör redan kraven på lagringsminne ett svåröverkomligt hinder. Digitalt lagrat ljud i full bandvidd kräver ca 5 MB per minut för monoåtergivning och 10 MB

per minut i stereo, och en minut digitaliserad videosekvens i färg och fullskärmsformat kräver inte mindre än 1½ Gigabyte i lagringsutrymme och därtill en synnerligen snabb datorutrustning. Trots att komprimeringsalgoritmerna förfinas och lagringsutrymmet blir billigare torde det dröja ännu någon tid innan lärare och andra ordinära datoranvändare får plats med särskilt mycket digitaliserad ljud- och videoinformation på det lagringsminne där de har sitt material under arbete.

Utveckling av system som företrädesvis arbetar med tidsberoende information, dvs rörliga bilder och ljud, är självfallet viktig men om ambitionen är att utveckla system där användarna utvecklar kvalificerat material av eget fabrikat, så är det det fråga om ett långsiktigt arbete. Däremot kan förbättrade verktyg för hantering av dokument innehållande strukturerad text ge stora omedelbara vinster redan i dag. Även grafik och stillbilder kan hanteras någorlunda bekymmersfritt; med de nya effektiva komprimeringsprogrammen är det möjligt att med acceptabel visuell kvalitet lagra stillbilder med komprimeringsförhållandet 1:50 eller mer, vilket drastiskt minskat minneskraven (ett exempel är JPEG från Joint Photographic Experts Group, ett komprimeringsprogram som behåller den slags information som betyder mest för det mänskliga ögat och skär bort den information som ögat knappast uppfattar, såsom smärre skiftningar i tämligen enfärgade ytor etc). Därför förefaller det tills vidare angeläget att — vilket är uppgiften för det här föreslagna projektet — ägna kraft åt problemen rörande texthantering och enklare bild- och grafikhantering. Om inte dessa basala problem finner tillfredställande lösningar, och om inte rikhaltiga text- och bildbaser av god kvalitet blir tillgängliga i det dagliga arbetet, kommer maskinläsbara dokument aldrig att bli naturliga inslag i exempelvis lärares och elevers vardag, trots aldrig så välgjorda animeringar och hur vackert ljudet än klingar. Här krävs avsevärda satsningar med tanke på utbildningsväsendets behov — de stora investeringarna i dokumenthanteringssystem gäller i dag sådant som underhållsmanualer hos industriföretagen eller handlingar hos myndigheter.

5. Systemet skall hjälpa användaren att revidera, ordna och överblicka materialet i den lokala dokumentbasen.

Här vill jag i synnerhet lyfta fram överblicksproblemet, som var och en som samlar sitt arbetsmaterial på ett skivminne är förtrogen med. Om det vill sig illa fungerar hårddisken som en *geniza*, det slutna rum med en liten öppning i taket där förbrukade papper och pergament med hebreisk skrift kastades ned eftersom det som skrivits på detta heliga språk enligt judisk tradition inte fick utplånas.

Även i fråga om de enskilda dokumentens innehåll gör sig överblicksproblemet snabbt påmint. Så länge jag handskas med pappersburen information utnyttjar jag mer eller mindre medvetet en rad knep för att skaffa mig överblick. Jag hittar rätt bok tack vare att jag minns dess plats i hyllan eller pärmens färg. Redan genom att ta den i handen erhåller jag en uppfattning om hur mycket den innehåller. Jag vet av erfarenhet ungefär var det lönar sig att leta efter tryckåret, innehållsförteckningen eller litteraturreferenserna. Om jag tidigare tittat i boken kanske jag bevarat ett synminne av att något stod högst upp på en höggersida. Under läsningen vet jag hela tiden om jag befinner mig i början, mitten eller slutet av boken. Jag stoppar vid behov tummen i ett uppslag dit jag önskar återvända, jag lägger komihåglappar i boken, stryker under, skriver anteckningar i marginalen, viker hundöron eller spiller kaffe — ett helt spektrum av visuella och taktila orienteringsmöjligheter som saknas när informationen uppträder på skärmen. Nya slag av hjälpmedel behövs.

För att elektroniskt lagrad information verkligen skall bli *tillgänglig*, krävs nya redskap som hjälper både "författare" och "läsare" att orientera sig och att överblicka och organisera materialet. De i dag invanda sätten att presentera och använda pappersburen information är minst av allt naturgivna. Det dröjde sekler innan dagens typografiska konventioner för presentation av tryckt text fann sin form. De första generationerna boktryckare satte en ära i

att imitera handskriften, och medeltida handskrifter ter sig för oss sena tiders barn nära nog ogenomträngliga redan på grund av textens massivitet. Konventionen att indela texten i stycken är en innovation (den medeltide skrivaren kunde föredra att fylla ut en halvfull rad med betydelselösa tecken för att erhålla en jämn högermarginal), liksom indelningen i kapitel och den systematiska användningen av rubriknivåer, olika typsnitt, sidnumrering, innehållsförteckningar, sakordsregister eller fotnoter. Allt det som vi idag förväntar oss av en läslig och läsbar tryckt text är frukten av en lång utveckling av såväl typografin som människors varseblivningsförmåga. I fråga om elektronisk information har en motsvarande utveckling av formgivningskonventioner och formkänsla nätt och jämt börjat.

Systemet bör förse användaren med orienteringshjälpmedel, sådant som kartor över hur innehållet i dokumentbasen är organiserat, eller hypertextlänkar, dvs korshänvisningar eller vägvisare mellan olika ställen i informationsmassan. Hypertextlänkar används ofta som vägvisare för att leta upp isolerade faktauppgifter o.likn. Jag tror att mer möda borde ägnas åt att utveckla länkar som tjänar som orienteringshjälpmedel med vilkas hjälp man skaffar sig överblick hur informationen är ordnad.

Översiktskartorna bör vara ett hjälpmedel inte bara för att överblicka hur dokumentbasen är ordnad för tillfället, utan även för att stuva om i den, ungefär som när jag flyttar böckerna i mina bokhyllor. Skilda användare eller skilda grupper av användare bör kunna ordna en och samma dokumentbas på olika sätt i enlighet med sina särskilda behov. Vidare bör ett och samma dokument vid behov kunna placeras på flera ställen i dokumentbasen (ungefär som när jag skulle behöva kopior av ett och samma papper i flera pärmar avsedda för skilda arbetsuppgifter).

Flertalet systemkonstruktörer tycks i dag favorisera en motsatt princip — man talar om "information at your fingertips", med Bill Gates formulering, eller om "seamless information environment" — som innebär att användaren skall finna det han behöver utan att fundera över hur informationen är ordnad. Denna princip är rimligt i då man är på jakt efter isolerade faktauppgifter eller enstaka dokument, men i mitt eget handbibliotek vill jag ha kontroll över hur samlingarna är ordnade.

6. Systemet bör stödja hanteringen av stora och komplexa dokument, samt ordnade samlingar av dokument.

Den populära kort-metaforen och annat som fragmenterar materialet och hotar att slå sönder dokumentens och dokumentksamlingarnas struktur bör bannlysas. Systemet måste utan långa väntetider som hejdar tankens flykt tillåta arbete med dokument omfattande många hundra sidor.

7. En och samma dokumentsamling skall kunna användas på olika nivåer.

Såväl noviser som specialister, skolelever likaväl som universitetsstudenter, skollärare likaväl som forskare, skall med hjälp av olika slags filter och urvalshjälpmedel kunna använda en och samma dokumentbas och överföra vad de har bruk för till sitt handbibliotek. Hellre än att producera separata editioner för olika målgrupper bör leverantörerna om möjligt förse alla med samma rikhaltiga material. Skoleleven som arbetar med första kapitlet ur Röda rummet och doktoranden i litteraturvetenskap bör båda ha tillgång till hela den maskinläsbara Strindbergsutgåvan med kommentarer och konkordanser. Självfallet behöver noviser och specialister olika texter men den mest rimliga lösningen, som garanterar stora upplagor och bred spridning och som ger användaren största möjliga frihet, är att dessa texter samlas på t.ex. samma kompaktskiva. Det finns knappast några skäl (annat än möjligen kommersiella) att producera maskinläsbara motsvarigheter till de handböcker eller läromedelpaket som enbart är avsedda för studenter eller skolelever.

(Jfr argumenteringen nedan, bilaga 1, för att lärarna och inte bara eleverna måste kunna använda den maskinläsbara litteratur som produceras för skolbruk. Det är min fasta övertygelse att alla, även eleverna, i det långa loppet vinner på att lärarna använder systemen och programmen i sitt eget arbete, i sin dagliga "fortbildning", i kursplaneringen, i lektionsplaneringen. Först då finns det chans att gedigna och rikhaltiga material matas in i systemen. Här som annars är jämförelser med biblioteksväsendet upplysande: tänk hur trist det vore om skolbiblioteken blott innehöll elementära läroböcker som lärarna knappast hade något att lära av. Dessutom får eleverna en bättre uppfattning om hur datorn kan användas som arbetsredskap när de möter lärare med egna erfarenheter därav.)

Just nu planerar många förlag att återanvända sina maskinläsbara dokumentbaser, t.ex underlaget för de tryckta encyklopedierna, genom därur plocka urval att saluföra till bestämda målgrupper: allt om hundar i en liten bok och allt om Afrika i en annan. Vid spridning på billigt magnetiskt eller optiskt medium finns knappast några skäl (återigen: fränsett de kommersiella) för ett sådant förfarande. Bättre vore att leverera hela encyklopedin, förutsatt att mottagaren har möjlighet att sortera fram artiklarna om hundar.

Principen bör således vara: tillhandahåll ett så rikhaltigt material som möjligt, utan tvång att allt skall användas. Med en översättning kunde gärna följa arbetet på originalspråk (i fråga om bokutgivningen kan man i Europa skönja en intressant tendens till att parallellutgåvor med texten på både målspråket och originalspråket blir allt vanligare). En kortare introducerande text om ett ämnesområde kunde beledsagas av utförligare framställningar. Här finns stora samordningsvinster om specialister och amatörer kunde använda samma dokumentsamlingar, på olika nivåer. Så sker sällan i dag. Ett exempel är det skåp som Apoteksbolaget tagit fram för att placera i apotekens kundutrymme. Skåpet döljer en skrivare kopplad till en persondator. Meningen är att kunden skall trycka på knappar märkta "huvudvärk", "hemorrojder" etc, varefter skrivaren skriver ut ett informationsblad med goda råd. Ett annat exempel är det ADB-system som många privatpraktiserande tandläkare använder och som bl.a. håller reda på patientjournalerna. I det ena fallet betjänas lekmännen, i det andra specialisterna. Det förefaller rimligare att ge fler grupper tillgång till olika utsnitt ur en och samma materialbank. Apotekens personal och kunder kunde använda samma dokumentdatabas, och hos tandläkaren skulle även patienten kunna erhålla en utskrift av valda delar av sin journal, kompletterad med översättningar av svåra ord.

8. Systemet bör vara modifierbart och utbyggbart.

Fler medier och fler program skall kunna tillföras. Systemet bör innebära så få kommersiella bindningar som möjligt, dvs. man skall inte vara alltför beroende av enskilda hårdvaru- och programleverantörer.

9. Enkel utrustning.

En användare bör utan tillgång till sofistikerade arbetsstationer eller teknisk personal kunna läsa, skriva och revidera dokument och helst också arbeta med hela dokumentbaser eller utsnitt ur sådana. En stor del av arbetet skall kunna ske med enkel utrustning som man vid behov kan ha i sommarstugan eller låna ut till en medarbetare på annan ort. Så är sällan fallet i dag, eftersom de mer avancerade dokumenthanteringssystemen är avsedda för stora företag eller förvaltningar. Över huvud taget begränsas humanister och samhällsvetare av att många datorlösningar på universitet och högskolor närmast befrämjar vad amerikanerna kallar "administrative computing". Motsatsen, "academic computing", förutsätter ett rörligare arbetssätt.

10. Systemet måste kunna hantera strukturerade dokument.

Att dokumentet är strukturerat innebär att inte blott innehållet i snävare mening, texten och bilderna, är betydelsebärande. Även det sätt på vilket innehållet är ordnat, i kapitel och avsnitt, i rubriknivåer, med olika slag av textelemt för olika slag av innehåll (brödtext, fotnotstext, infogat citat, litteraturreferens, författarnamn, tryckort och tryckår etc) har betydelse för hur ett strukturerat dokument tolkas.

En konsekvens av denna kravlista är att det är ändamålsenligt att dokumenten är märkta enligt SGML-standarden. Därömr handlar nästa avsnitt.

IV. Om beskrivande märkning och SGML

<p>SGML (Standard Generalized Markup Language)</p> <p>ISO 8879, antagen 1986 Standardverk: Charles F. Goldfarb, <i>The SGML Handbook</i>. Oxford: Clarendon Press, 1990 "Dotterstandarder" till SGML: DSSSL för typografering HyTime för hypertext samt för tidsbaserade dokument (multimedia) TEI (Text Encoding Initiative), riktlinjer för märkning av humanistisk och samhällsvetenskaplig litteratur</p>
--

SGML (Standard Generalized Markup Language) är ett språk för beskrivande märkning av maskinläsbara dokument. Jag skall inledningsvis med ett par exempel illustrera skillnaden mellan procedurmärkning och beskrivande märkning.

Procedurmärkning talar om hur den märkta informationen skall hanteras, exempelvis vad datorn och skrivaren eller fotosättningsmaskinen skall göra med en textfil. Var och en som infogat sättningsanvisningar i ett manuskript vet vad det innebär. En viss anvisning kan ange att vissa ord skall tryckas i 12 p kursiv stil. I detta fall innebär procedurmärkningen en instruktion till sätteriet.

En beskrivande märkning ger i stället besked om av vilka element texten består. Några ord med ett och samma typografiska utseende, t.ex. ord satta med 12 p kursiv, kan utgöra vitt skilda slag av element, exempelvis 1. en rubrik, 2. något författaren vill framhäva, 3. ett insprängt ord på främmande språk, 4. en boktitel. Om vi tillämpar procedurmärkning skulle i följande fiktiva text samtliga de fyra nämnda slagen av element anges på samma sätt (12 p Courier kursiv):

II.3.1. Om Descartes metod

Descartes har skapat ett mönster för vad filosofisk *metod* vill säga. Men i själva verket är Descartes berömda text ett *préface* till tre naturvetenskapliga och matematiska studier (tillgängliga i René Descartes: *Discours de la méthode*. Texte et commentaire par Étienne Gilson. Paris: Vrin, 6 uppl. 1987).

Med beskrivande märkning, som skiljer på dessa fyra slag av innehållsliga enheter, undviker man att en och samma märkning betyder olika saker. Om vi tillämpar SGML-märkning enligt TEIs riktlinjer, första versionen (mer därom strax) skulle märkningen se ut på följande sätt. Rubriknivå nr 3 markeras med ett inledande märkord `<h3>` och avslutas med märkordet `</h3>`. (Observera att ingen numrering av avsnittsrubrikerna behöver sättas ut; avsnittets plats i dokumentet gör att vi ändå vet att det rör sig om andra kapitlet, tredje avsnittet, första underavsnittet). Framhävningen avgränsas av märkorden `` och ``. Bokstäverna `em` skall utläsas "emfas". (Observera att framhävning ofta representeras typografiskt av kursiv

stil, men den kan även representeras på annat sätt, förr i världen med spärrad stil, i dag ibland med rak stil, exempelvis i ett förord där brödtexten är satt med kursiv stil.) Ett insprängt ord på främmande språk, här franska, markeras med de inledande märkorden `<gloss><foreign lang=Fr>` och avslutas med märkorden `</foreign></gloss>`. En hänvisning till en boktitel kan ske med korsreferens till ett unikt ställe; om den fullständiga referensen till Descartesutgåvan återfinns i en litteraturlista i slutet av det aktuella dokumentet och där försetts med märkordet DES87 räcker det med att i slutet av texten ovan infoga `<xref RID=DES87>`. (`xref` uttydes korsreferens, `RID` uttydes referensidentifikation). Så här ser ser texten ut då den på detta sätt försetts med beskrivande märkord:

```
<h3>Om Descartes metod</h3>
Descartes har skapat ett mönster för vad filosofisk
<em>metod</em> vill säga. Men i själva verket är Descartes
berömda text ett <gloss><foreign
lang=Fr>préface</foreign></gloss> till tre naturvetenskapliga
och matematiska studier (tillgängliga i <xref RID=DES87>).
```

Detta var ett exempel på att man med hjälp av beskrivande märkning undviker att ett och samma typografiska utseende (kursiv stil) betecknar olika innehållsliga element. Omvänt undviker man också att ett och samma slag av innehållslig enhet märks på olika sätt. Tag citat som exempel. Citat kan utmärkas typografiskt på en rad sätt. Redan citationstecknens utseende och placering varierar:

Tyska	»XXXXX« >XXXXX< „XXXXX“
Engelska	“XXXXX” ‘XXXXX’
Franska	« XXXXX »
Svenska	”XXXXX” »XXXXX»

Även svenska typografiska konventioner kan variera. Ett citat som utgör ett eget stycke markeras ibland med med indragen vänstermarginal, ibland minskad grad, ibland citationstecken och ibland inte. Tag följande citat ur Viktor Rydbergs *Bibelns lära om Kristus*:

Det andra af de båda föregifna intygen lemnas af Cypriani skrift “Om kyrkans enhet“ och har följande lydelse:
 “Herren säger: jag och fadren äro ett. Och återigen är det skrifvet om fadren och sonen och den helige andre: och tre äro ett.”
 Onekligen har detta intyg ett viss sken för sig. Tvenne olika bibelställen äro här åsyftade. Det ena är, liksom hos Tertullianus, Joh. 10, 30; det andra är utan tvifvel 1 Joh. 5, 8.

som med dagens typografiska konventioner skulle kunna se ut på flera sätt, exempelvis:

Det andra af de båda föregivna intygen lämnas af Cypriani skrift ”Om kyrkans enhet” och har följande lydelse:

”Herren säger: jag och fadren är ett. Och återigen är det skrivet om fadern och sonen och den helige andre: och tre är ett.”

Onekligen har detta intyg ett viss sken för sig. Tvenne olika bibelställen är här åsyftade. Det ena är, liksom hos Tertullianus, Joh. 10, 30; det andra är utan tvivel 1 Joh. 5, 8.

eller:

Det andra af de båda föregivna intygen lämnas af Cypriani skrift ”Om kyrkans enhet” och har följande lydelse:

Herren säger: jag och fadren är ett. Och återigen är det skrivet om fadern och sonen och den helige andre: och tre är ett.

Onekligen har detta intyg ett viss sken för sig. Tvenne olika bibelställen är här åsyftade. Det ena är, liksom hos Tertullianus, Joh. 10, 30; det andra är utan tvivel 1 Joh. 5, 8.

I franskt tryck är sägesatserna inte sällan infogade innanför citationstecknen, som här på ett ställe i Lévi Strauss' *La pensée Sauvage*:

« Ce procédé, dit Boas, y est plus fréquent que dans tout autre langage connu de moi. »

och i engelskt tryck förekommer ofta att ett komma omedelbart efter citatet placeras före citationstecknet, som här i Paul Feyerabend's *Against method*:

‘If any metaphysics goes,’ writes Dr Hesse in her review of an earlier essay of mine, ‘then the question arises [...]’

Vid användning av beskrivande märkning bryr man sig i princip inte om sådana närmast estetiska spörsmål, man nöjer sig med att sätta märkordet <q> före citatet och </q> efter citatet. Sedan ankommer det på hur man styr det typografiska formateringsprogrammet och tillgängliga utskriftsmöjligheter hur den tryckta texten gestaltas.

De exempel på märkning jag här givit ansluter till den internationella standarden för rent beskrivande märkning, SGML (Standard Generalized Markup Language), antagen som ISO-standard (ISO-8879) år 1986. Standardverket är Charles F. Goldfarb, *The SGML Handbook*, Clarendon Press, Oxford 1990, som bl.a. innehåller den fullständiga ISO-texten. Arbetet går längre tillbaka i tiden. Ett förstadium till SGML var märkspråket GML, som Charles Goldfarb m.fl. från och med slutet av 60-talet utvecklade vid IBM. Tanken var då att dokument (det rörde sig om juridiska dokument) skulle kunna märkas på ett enhetligt sätt, så att ett och samma dokument kunde matas in i skilda system för textbehandling, formatering och informationsåtervinning.

Dagens teknik öppnar vissa möjligheter att bearbeta ett och samma dokument eller en och samma dokumentsamling på mångahanda sätt och med många slag av program på många slag av maskiner. En sådan utveckling förutsätter ett generaliserat dokumentbeskrivningsspråk, ett slags esperanto om man så vill, som gör dokumenten oberoende av maskinvara och programvara och nationell teckenuppsättning. Ett och samma dokument kan med andra ord stoppas in i en SUN-maskin, en Macintosh eller en DOS/Windows-maskin och bearbetas med olika program. På detta område pågår ett livaktigt

internationellt arbete med utvecklingen av SGML och med en hel svit av "dotterstandarder" såsom DSSSL och HyTime.

DSSSL (Document Style Semantics and Specification Language, DIS 10179; akronymen uttalas vanligen "dissel") styr den typografiska formgivningen, dvs. sørjer för att layouten blir någorlunda densamma vid utskrift på olika medier och med olika utrustning. DSSSL förutsätter att dokumenten är SGML-märkta. DSSSL antogs som Draft International Standard i augusti 1991 och omröstning om en andra draft-version är planerad till april 1993.

HyTime (ISO 10744), som antogs i april 1992 efter en rekordkort förberedelsetid, eftersom behovet ansågs akut, är den första internationella standarden för överföring av hypertextdokument och multimediadokument (eller för att vara mer exakt: tidsberoende dokument, innehållande sådant som ljud och rörliga bilder). Även HyTime bygger på SGML-standarderna.

Detta standardiseringsarbete öppnar nya möjligheter att överskrida de gränser — nationsgränser, skillnader mellan olika slag av datorer och program, gränser mellan utbildningsväsendets nivåer och ämnen och mellan forskning och undervisning, avstånden mellan datorentusiasterna och de övriga — som i dag förhindrar samlade ansträngningar att skapa och sprida rikhaltiga dokumentbaser av god kvalitet.

Redan möjligheten att komma förbi de till synes triviala men enerverande problemen med olika teckenuppsättningar är ett stort framsteg. I SGML-standarderna (dvs ISO 8879) ingår en *reference concrete syntax*, som är utgångspunkten för översättningar mellan olika datormiljöer, nationalspråk, teckenuppsättningar m.m. I denna syntax ingår en teckenuppsättning (kallad *the base character set*) som överensstämmer med standarden ISO 646 (dvs en 7-bit-standard känd som IRV = International Reference Version). Detta är det som i egentlig mening är ASCII-teckenuppsättningen, dvs tecknen 0-127, som ligger till grund för en mängd andra internationella teckenstandarder. Med de 128 tecken som däri ingår och som nästan alla maskiner förstår kan snart sagt varje tecken i de europeiska skriftspråken representeras. Den mest flyttbara flyttbara representationen av å,ä,ö är enligt regelboken (Charles F. Goldfarb, *The SGML Handbook*, 1990, p. 506f):

Å	översätts med	Å
å	översätts med	å
Ä	översätts med	Ä
ä	översätts med	ä
Ö	översätts med	Ö
ö	översätts med	ö

SGML föreskriver egentligen inte regler för hur man skall koda dokument. I stället är SGML en internationell överenskommelse om en uppsättning regler för hur den som märkt ett dokument skall berätta för andra hur denna märkning gått till, en berättelse som är omedelbart läsbar för mottagaren (antingen det är en människa eller en maskin).

Att SGML är ett beskrivande språk innebär att SGML-märkningen inte säger något om den typografiska formgivningen. Vid utskrift eller fotosättning eller visning på skärmen vidarebefordras det SGML-märkta dokumentet till ett formateringsprogram som sørjer för att utseendet blir det önskade.

Det är inte meningen att den ordinäre användaren skall behöva se eller bekymra sig om SGML-märkningen. Han eller hon ser den version som passerat genom ett formateringsprogram (som ersatt Å med bokstaven Å etc). Ett alternativ är att SGML-märkta dokument konverteras till användarens favorittextbehandlingsprogram och bearbetas där. Flera av de mest spridda textbehandlingsprogrammen (Microsoft Word, WordPerfect) kommer enligt tillverkarnas löften inom kort att förses med stöd för SGML.

Ett SGML-dokument består av tre delar. För det första en SGML-deklaration som talar om på vilket sätt man byggt ut eller modifierat ISO 8879. (Exempelvis anger TEIs riktlinjer att märkorden får vara upp till 128 tecken långa, till skillnad från ISO 8879 som föreskriver högst 8 tecken). För det andra en DTD (Document Type Definition) som anger hur föreliggande typ av dokument märkes. En DTD kan avse affärsbrev, ytterligare en annan romaner, ytterligare en annan vetenskapliga monografier etc. För det tredje själva innehållet i dokumentet. Det är inte nödvändigt att SGML-deklaration och DTD medföljer varje enskilt dokument, men de måste (som regel som egna filer dit pekare i dokumentet hänvisar) vara tillgängliga någonstans i det system där man handskas med dokumentet, och de måste medfölja vid export.

Några ord bör sägas om det internationella projekt, TEI, Text Encoding Initiative, som tagit sig an uppgiften att utarbeta rutiner för kodning av humanisk och samhällsvetenskaplig litteratur, och som för det syftet har samlat humanister och samhällsvetare från många länder i skilda arbetsgrupper. TEIs syfte är att "formulate and disseminate guidelines for the encoding and interchange of machine-readable texts intended for literary, linguistic, historical, or other textual research."¹ Det första planeringsmötet skedde i november 1987. Projektet stöds av Association for Computers and the Humanities (ACH), Association for Computational Linguistics (ACL), Association for Literary and Linguistic Computing (ALLC), U.S. National Endowment for the Humanities (NEH); Andrew W. Mellon Foundation, samt Europakommissionen (XIIIe direktoratet). En första preliminär version (kallad "P1") av TEIs riktlinjer publicerades i juli 1990.² En andra och tills vidare slutgiltig versionen (kallad "P2") av riktlinjerna beräknas vara helt klar i juli 1993. När detta skrives har dock de för märkningsarbetet viktigaste kapitlen redan gjorts tillgängliga via universitetsnäten, och det finns ingen anledning att vänta med att pröva riktlinjerna på svenska material.³

På nästa sida återfinns som en avslutande illustration ett hemsnickrat exempel på SGML-märkning som följer den första versionen, daterad oktober 1990, av TEIs riktlinjer. Det är början av dokumentet vi ser. SGML-deklaration och DTD behöver som sagt inte alltid finnas med och de är här utelämnade (DTD'n representeras av en pekare `TEI . 1`, som är en av TEI föreslagen normal-DTD). Däremot är varje TEI-dokument, även detta, försett med inledande information som fungerar som "elektronisk titelsida och elektroniskt förord", bestående av tre delar [1] file description, [2] encoding declarations, [3] revision history — ett värdefullt underlag för bibliografiskt arbete eller för automatisk eller halvautomatisk katalogisering i ett bibliotekssystem. Sedan följer själva texten, märkt `<text>`, som inleds med en rubrik på den översta nivån, märkt `<h1>`, samt brödtextens stycken, märkta `<p>`. Avslutande märkord behövs inte eftersom element som rubriker och stycken förutsättes avslutade när ett nytt element introduceras. En rubrik är slut där vissa andra element, t.ex. en underrubrik eller första stycket i brödtexten, inleds. SGML-märkningen på följande sida är inte fullt genomförd. För läsbarhetens skull har i detta exempel (och de tidigare) de svenska tecknen behållits.

¹ Citerat efter Susan Hockeys och Donald A. Walkers presentation under konferensen SGML'92 i Danvers, Mass., USA, 26-29 oktober 1992.

² C.M. Sperberg-McQueen and Lou Burnard (Eds): *Guidelines For the Encoding and Interchange Of Machine-Readable Texts*. Draft, Version 1.1, October 1990.

³ Kapitel 22, som blev tillgängligt redan i augusti 1992, innehöll de inte minst för bibliografiska syften väsentliga riktlinjerna för skapande av "TEI headers", dvs det avsnitt som skall inleda varje TEI-dokument och som innehåller uppgifter om titel, författare, källa, version, redaktionella principer och allt annat som kan behövas, exempelvis för automatisk katalogisering, för upprättande av index över dokumenten i ett bestånd etc. Kapitel 7 blev tillgängligt i oktober 1992 och innehöll "the bas tag set" för prosa. Kapitel 6, tillgängligt i december 1992, beskrev element som kan finnas i alla slags TEI-dokument.

```

<TEI.1>
<TEI.header>
  <file.description>
    <title.statement>
      <title>Alva Myrdal & Gunnar Myrdal: Uppfostran i
      samhälllets mitt. En maskinläsbar transkription</title>
      <statement.of. responsibility>
        <role>etablerad av</role>
        <name>Donald Broady</name>
      </statement.of. responsibility>
    </title.statement>
    <extent.statement>116.300 Kb</extent.statement>
    <publication.statement>
      <creation.date>nov 1991</creation.date>
      <publication>Forum för pedagogisk historia</publication>
      <distribution>Ej för spridning</distribution>
      <release>ver. 1.0</release>
    </publication.statement>
    <source.description>
      <author>Alva Myrdal & Gunnar Myrdal</author>
      <title>Kontakt med Amerika</title>
      <imprint>
        <publ.city>Stockholm</publ.city>
        <publisher>Bonniers</publisher>
        <publ.date>1941</publ.date>
        <citn.detail>kap IV, pp. 88-132</citn.detail>
        <comment>Inflytelserik text inför reformeringen av
        den svenska skolan</comment>
      </imprint>
    </source.description>
  </file.description>
  <encoding.declarations>Standardförfarande</encoding.declarations>
  <revision.history>Denna version från nov 1991 är den som är rejält
  korrekturläst</revision.history>
</TEI.header>
<text>
  <h1>Uppfostran i samhälllets mitt
<p>
Direkt kräver den amerikanska trosbekännelsen frihet, jämlikhet och
broderskap, indirekt kräver den uppfostran. Det ena ger formen och målet,
det andra ger innehållet och medlen. Det ena är det evigt konservativa i de
amerikanska idealen, det andra det evigt radikala.
<p>
Statsgrunden, samhällsidealen har stått som oföränderliga genom en längre
period än i något annat land. Men medlen till att fördjupa och befästa
idealerna är ständigt växlande. Det är i sökandet efter dem amerikanens
experimentalistiska livsinställning röjer sig. Denna paradox mellan vad som
är gammalt och nytt, stabilt och föränderligt, odiskutabelt och
diskutabelt, vad som är konservativt och radikalt i den världsdelen, som
kallar sig den Nya Världen men som samtidigt är den äldsta demokratin, har
sällan blottlagts. Det ivrigt nya gäller aldrig bottenvärdena. Men heller
ingenting annat än bottenvärdena får i Amerika fortleva utan ständiga
frågetecken.

```

V. Några fördelar med SGML-märkta dokument

- Regelstyrd beskrivande märkning ger entydig tolkning.
- Få antaganden om hur dokumenten kommer att användas.
- Dokumenten oberoende av plattformar, tillämpningsprogram och teckenuppsättningar.
- Dokumenten kan i princip redigeras med den enklaste ordbehandlare och skrivas ut på vilken skrivare eller skärm som helst.
- Märkningen är någorlunda begriplig för människor.
- Dokumenten är flata textfiler som kan användas som databaser.
- Dokumenten kan återanvändas.
- Det finns hopp om en ny Aldus Manutius.

1. Regelstyrd beskrivande märkning ger entydig tolkning.

Program — eller människor — kan på ett entydigt sätt tolka ett dokument som är korrekt SGML-märkt. Man undviker med andra ord att en och samma märkning har olika betydelse, eller omvänt att en och samma betydelse representerar av olika märkningar (jfr exempen kursivstil resp. citat ovan).

2. Få antaganden om hur dokumenten kommer att användas.

Några exempel på tänkbara användningar är följande. Dokumenten kanske skall överföras till textbehandlingsprogram och där redigeras; formges typografiskt för utskrift på skrivare eller för vidare befordran till fotosätter; införlivas med en fulltextdatabas eller en databas över abstracts; användas vid strängsökning (för att letar upp enstaka ord eller fraser) och/eller struktursökning (varvid sökspråket utnyttjar relationerna mellan de element varav dokumenten eller dokumentbasen är uppbyggda); användas för lingvistisk analys (lexikalisk, syntaktisk eller semantisk); utgöra underlag för indexering, för upprättande av konkordanser etc; tappas på viss specifik information med hjälp av sök- och informationsåtervinningsprogram; utgöra underlag för automatisk eller halvautomatisk katalogisering eller framställning av bibliografier; förses med hypertextlänkar. Listan kunde göras hur lång som helst. Nu senast har jag stött på flera stora projekt i olika länder som utvecklar metodik för att ur SGML-märkningen generera utskrift på punktskriftsplatta för synskadade.

3. Dokumenten är flyttbara.

Kompleta SGML-märkta dokument är i princip oberoende av plattformar och tillämpningsprogram och kan (förutsatt att systemen innehåller programvara som ger SGML-stöd) flyttas mellan dessa. Spridningen är ej heller begränsade av nationalspråk, av teckenuppsättningar (så länge de är västerländska) eller av vad som låter sig skrivas in från tangentbordet.

4. Måttliga hårdvarukrav.

Dokumentet kan i princip redigeras med den enklaste ordbehandlare och skrivas ut på vilken skrivare eller skärm som helst. (Från början hade man tanken att SGML-märkning skulle kunna åstadkommas med skrivmaskin.)

5. SGML-märkningen är begriplig för människor.

Märkorden är mnemotekniska och återfinns omedelbart intill det textparti de avser. Detta är en viktig skillnad i jämförelse med exempelvis flertalet ordinära textbehandlingsfiler, vilka är bemängda med koder och pekare som maskinen förstår men som är obegripliga för nästan alla människor.

Det är självfallet inte meningen att man skall behöva sitta och genomföra SGML-märkningen manuellt. Ett naturligt arbetssätt är att använda sitt textbehandlingsprogram, varefter en konverteringsrutin ombesörjer omvandlingen till SGML-märkning och ett annat program validerar märkningen, dvs undersöker om den uppfyller SGML-kraven. Men den märkning som finns är sådan att människor med en smula träning lär sig begripa det viktigaste.

6. SGML-märkta dokument är flata textfiler som kan användas som databaser.

Man kan exempelvis i ett dokument eller en dokumentsamling söka upp alla element av ett visst slag (exempelvis alla litteraturreferenser, eller alla huvudrubriker, alla abstracts, alla citat på grekiska) som innehåller vissa ord eller uppfyller andra kriterier. Man kan extrahera fram alla element som innehåller författarnamn och sedan göra vad man vill med dem. Och så vidare.

En för bibliotekarier och andra välkommen möjlighet är att TEIs riktlinjer föreskriver att all information som behövs för katalogisering och bibliografering enligt internationell standard (såsom ISBD, International Standard Book Description, etc) skall finnas omedelbart tillgänglig som väl åtskilda och väl definierade element i början av SGML-dokumentet (jfr den TEI-märkta Myrdaltexten ovan). Sådana element är författarnamn, redaktör, titel, utgivningsort, publiceringsdatum, förlag, etc, och motsvarande uppgifter avseende maskinläsbara dokument. Därmed beledsagas varje text av information som utgör underlag för en i bästa fall alltigenom automatisk katalogisering, och som även är till stor hjälp i de ordinära bibliografiska sysslor som tillhör forskares eller lärares arbete.

7. Dokumenten kan återanvändas.

Att samma dokument kan användas för de mest skilda syften innebär bl.a. att vi kan räkna med att dokumenten överlever de operativsystem och tillämpningsprogram i vilka de skapats.

8. Efterlyses: en Aldus Manutius i vår tid.

Aldus Manutius var den boktryckare som i Venedig i slutet av 1400-talet och början av 1500-talet utgav en mängd antika författare i den tidens "flyttbara format": behändiga kvartovolymer som kunde stoppas i fickan, standardiserade och lättlästa typsnitt, texten kvalitetskontrollerad av lärda korrekturläsare, en med tidens mått extraordinärt stor upplaga (tusen exemplar). Under sekler betraktades dessa vackra volymer i hela den lärda världen som de pålitligaste utgåvorna av de latinska auktorerna. Jag har en vision av en framtida maskinläsbar litteratur som på samma sätt kunde förena flyttbarhet, tillgänglighet, bred spridning och hög kvalitet. För detta krävs det internationellt accepterade dokumentbeskrivningsspråk som SGML är på väg att bli.

Bilaga 1

D. Broady: "Kunskapsverkstaden — om lokala dokumentbaser som arbetsverktyg för lärare", *Datorn i Utbildningen*, årg. V, nr 2 1992, pp. 22-28. [Tidigare utgiven under samma titel som IPLab-rapport nr 47, Institutionen för numerisk analys och datalogi, KTH, 1991.]

Kunskapsverkstaden

Om lokala dokumentbaser som arbetsverktyg för lärare

Donald Broady

Inst. f. ped., HLS

100 26 Stockholm

IPLab, NADA, KTH

100 44 Stockholm

DEPAC,

Box 34103,

Lärarna behöver nya slag av arbetsverktyg. Skolväsendets decentralisering, övergången från regelstyrning till mål- och resultatstyrning, nya läroplaner och kursplaner som inte detaljreglerar undervisningen, de lokala arbetsplanernas ökande betydelse — samverkar till att ge lärarna mer frihet och mer ansvar. Det kommer i framtiden inte att vara lika lätt att luta sig mot detaljerade kursplaner eller prefabricerade läromedelspaket. Lärarna skall efter eget huvud och i samverkan med eleverna välja och organisera undervisningsstoffet.

Den uppgiften kräver ändamålsenliga arbetsverktyg. Sådana verktyg är läroplanen, kursplanerna, och de arbetsplaner som utformas vid den egna skolan. De framtida kursplanerna kommer sannolikt att kompletteras med referens- och kommentarmaterial varur läraren hämtar information och inspiration, håller sig à jour med utvecklingen inom vetenskapliga och andra kunskapsområden och stiftar bekantskap med goda pedagogiska exempel. Åter andra verktyg är vetenskaplig och populärvetenskaplig litteratur som läraren samlar på sig, tidningar och tidskrifter, olika slag av undervisningsmaterial etc. Även lärarutbildningen och fortbildningen förser lärarna med arbetsverktyg.

För den enskilde läraren är det ingen lätt uppgift att samla in, organisera, överblicka och använda detta material, vartill kommer allt det material som han eller hon själv ackumulerar i sin dagliga "fortbildning" eller framställer i samband med kurs- och undervisningsplaneringen, jämte resultaten av elevernas arbete.

En tänkbar väg, som prövas inom projektet Datorstöd för kunskapsarbete, är att utveckla system för arbete med lokala dokumentbaser vari enskilda lärare eller arbetsenheter lagrar

både arbetsmaterial de själva (och eleverna) förfärdigar och sådant material som de hämtar in från annat håll. Bland annat kan sådana lokala dokumentbaser användas för framställning av skraddarsydda läromedel.

Projektarbetet utgår från de överväganden som nedan sammanfattas i tio punkter. Jag skriver med tanke på läsare som är måttligt intresserade av de datortekniska aspekterna. Uppgiften att utveckla system för arbete med lokala dokumentbaser är inte enbart, eller ens i första hand, en teknisk fråga. Vi är tacksamma för synpunkter, i synnerhet från erfarna lärare, forskare, redaktörer, författare.

1. Datorstödet bör underlätta lärarnas och elevernas **arbete** med kunskapstoffet, och inte minst deras möjligheter att framställa eget material.

Hittills har den undervisningsteknologiska utvecklingen i alltför hög grad varit inriktad på system som presenterar ett färdigt stoff. Stick i stäv med all pedagogisk erfarenhet har magnetiskt och optiskt burna läromedel utformats som presentationsredskap snarare än som undersökningsverktyg. Åtskilligt av det undervisningsmaterial som i dag cirkulerar bland datorintresserade lärare liknar mest skioptikonvisningarna i forna tiders skola, eller kanske sextiotalets självinstruerande läroböcker (om du vet svaret, gå vidare till sidan 67). Ofta kan eleverna inte göra annat än att ge ett kommando som ersätter en skärmbild med annan.

Klyftan är stor mellan å ena sidan de klatschiga skärmbilderna, de allt läckrare videosekvenserna, det allt mer välklingande ljudet och å andra sidan det påvra och sönderstyckade vetande som brukar meddelas. Ofta serveras eleverna en enkel bild eller några få rader faktauppgifter. Det paradoxala är att lärarna för egen del aldrig skulle nöja sig med så torftiga arbetsverktyg. De letar förhoppningsvis fram den bästa facklitteraturen i ämnet.

Den ensidiga tonvikten vid presentationsmöjligheterna innebär dessutom att de interaktiva möjligheterna försummas. Läraren och eleverna erbjuds föga stöd för sin egen bearbetning av stoffet, skärmbilderna blir ett måhända aptitligare sätt att servera sådant som ungefär lika gärna kunnat inhämtas i en bok — ja, boken har åtminstone fördelen att den tillåter understrykningar och marginalanteckningar.

2. Datorstödet bör tjäna som verktyg i **lärarnas** dagliga arbete.

Lärarna bör ha glädje av datorstödet i sitt eget arbete: i samband med den egna kontinuerliga "fortbildningen", för framställning av läromedel, i kurs- och lektionsplaneringen, i genomförandet av undervisningen, i utvärderingen. Hittills har diskussionen om datorn i skolan alltför ensidigt fokuserat elevernas behov. Det finns skäl att ifrågasätta uppfattningen att datorn primärt skulle vara elevernas arbetsredskap.

En mer framkomlig väg, förmodligen i det långa loppet mer fruktbar även för eleverna, vore att ägna mer uppmärksamhet åt hur datortekniken kan fungera som stöd i lärarnas arbete. Lärare som funnit sätt att utnyttja datorer för att göra sitt eget arbete mer meningsfullt — och inte bara eleverna klokare — får lust och möjlighet att vägleda eleverna till vettiga datoranvändningar.

3. Lärare och elever har glädje av ett datorstöd som är **generellt**.

Lärare har behov av ett datorstöd som är mångsidigt och kan användas för att samla, ordna och bearbeta skilda slag av material. Även eleverna är betjänta av generella och öppna system, som tillåter att man skriver och ritar, bearbetar text och bilder, hämtar hem och skickar iväg material, genomför simuleringar, bygger upp egna dokumentbaser, etc. Det är en återvändsgränd att försöka specialskriva program för specifika undervisningsmoment, av typen program som lär ut Hallands floder eller något särskilt avsnitt av fysik- eller biologikursen. Med rimliga krav på kvalitet är sådana program synnerligen kostsamma att framställa. Erfarenheten visar dessutom att de sällan används av andra än upphovsmännen själva, och i de fall när de kommer till användning bidrager de till att befästa en steril indelning av kursplanen i på förhand fixerade moment. Härav kan vi dra slutsatsen att läraren och eleverna i stället bör förses med redskap som tillåter dem att skraddarsy sina egna tillämpningar.

Det är helt enkelt dålig moral när elever får arbeta med program som är så till den grad begränsade och begränsande, att läraren aldrig skulle drömma om att använda dem för att sköta sitt eget arbete.

4. Datorstödet bör främja **samarbete**.

För hela arbetslivet gäller att systemutvecklingen hittills alltför ensidigt prioriterat det individuella arbetet. Med "fleranvändarsystem" har oftast avsetts att medarbetarna delar på gemensamma centralenheter, program och datafiler, skrivare och lagringsmedia. Med konferenssystem har oftast avsetts tekniker för överförande av korta budskap mellan människor som i övrigt var och en sköter sitt. Datorstödet har med andra ord tjänat till att i olika avseenden effektivisera resursutnyttjandet eller kommunikationen, medan det mer intensiva samarbetet skett utan datorstöd, i sammanträdesrummet eller vid kaffebordet. Även forskningen inom det område som kallas "människa/dator-interaktion" har hittills överbetonat det individuella perspektivet. Regeln har varit, att man koncentrerat intresset till mötet mellan å ena sidan systemet, å andra sidan en individ som får representera människosläktet i största allmänhet. Man har bortsett från de sociala sammanhang i vilka arbetet ingår. Först på senare år har "datorstött samarbete" hamnat på systemutvecklarnas och forskarnas dagordning.

I skolans värld har datorstödet framför allt tagit sikte på att underlätta den enskilde elevens inläring. Mer intresse bör ägnas åt system som främjar samarbetet inom t.ex. en arbetsenhet, en elevgrupp eller ett kontaktnät mellan likasinnade lärare, läromedelsförfattare och forskare.

Vad gäller kontaktnäten har skollärarna olyckligtvis hamnat i bakvattnet. Vid universitetet och högskolorna är informationsutbytet med kolleger i landet och runtom i världen en självklarhet för många lärare och forskare, vilkas arbete har blir roligare och mer fruktbarande i och med möjligheten att snabbt och billigt utbyta meddelanden och material via de nationella och internationella universitetsnäten. I dag har en skollärare som saknar kontakt med en universitetsinstitution svårt att ens skaffa sig en datoradress. I samband med universitetsnätens fortsatta utbyggnad är det hög tid att väcka frågan om skolväsendets behov.

5. Datorstödet bör tillåta arbete med och överföring av **strukturerade** dokument.

En generell svaghet hos dagens kommunikations- och samarbetsystem är att det är svårt att utbyta strukturerad information. Att utbyta korta meddelanden eller att leta upp isolerade faktauppgifter i någon databas är jämförelsevis problemfritt, och många av oss är fortfarande så imponerade av möjligheten att utnyttja elektronisk post, konferenssystem eller externa databaser, att vi inte tänker på vad vi går miste om.

Svårigheterna tornar upp sig så snart vi har att göra med en informationsmassa som är strukturerad. Tag något så trivialt som ett bokkapitel eller en tidskriftsuppsats med olika rubriknivåer, fotnoter, referenslista etc. Att överföra den typografiska formateringen brukar gå någorlunda bra; det finns en uppsjö av rutiner och hjälpprogram som sørjer för att en rubrik i 14 punkter halvfet Times ser ungefär likadan ut när en text forslas mellan olika program eller plattformar. När man intresserar sig för "strukturerade dokument" avser man dock inte hur texten ser ut, utan av vilka "innehållsliga" element den är uppbyggd. Man bryr sig kanske inte om rubrikens stilsort och grad, men man vill att systemet skall hålla reda på att det är en rubrik. Detsamma gäller för andra "innehållsliga" element. Man vill veta att detta är ett författarnamn, detta är en boktitel, detta är ett publiceringsår, detta är en kapitelrubrik, detta är rubriknivå 2, detta är rubriknivå 3, detta är ett citat, detta är en litteraturreferens, detta är en korshänvisning till ett annat ställe i texten, detta är redaktörens infogade kommentar... I dag är det närmast regel att det slaget av information går förlorad när dokumenten överförs mellan program eller plattformar, skickas via elektronisk post eller formateras för vidare befordran till fotosättning.

Den gyllne regeln bör vara: dokumenten skall så långt som möjligt bearbetas och cirkulera med bevarad struktur. Den innehållsliga uppbyggnaden skall inte gå förlorad i kretsloppet. Om jag per diskett eller elektronisk post vill sända ett dokument på remiss till kolleger som skriver in sina ändringsförslag, eller om jag vill överföra dokumentet till ett nytt format eller till en ny plattform, eller färdigställa det för utskrift på en skrivare eller för fotosättning, skall jag ständigt med bevarad struktur kunna återföra dokumentet till, låt säga, mitt textbehandlingsprogram. Jag skall slippa att råka ut för att den innehållsliga märkningen ramlar bort när dokumentet cirkulerar. För att välja ett numer vardagligt exempel: vid telefaxöverföring, som innebär att dokumentet läses in som en bild, måste systemet tillåta att jag vid behov också sänder en beledsagande strukturerad textfil. Om jag omvänt till min egen dokumentbas vill hämta in information från t.ex. en extern databas, från en kompaktskiva, eller genom att skanna en tidskriftsartikel, skall det finnas verktyg som ger mig kontroll över strukturen. För att uttrycka saken i tekniska termer: systemen bör understödja format, förslagsvis SGML (Standard Generalized Markup Language), som tillåter att dokumentens struktur återskapas i alla led av hanteringen.

Åtskillig systemutveckling krävs innan det blir naturligt och lätt för människor som sitter på olika håll och kanske använder olika slag av utrustning och program att utbyta och arbeta med strukturerade dokument. Detta är dock ett nödvändigt arbete, om inte de dokumentbaser som byggs upp av lärare och andra skall bli isolerade öar.

6. Lärare och elever bör kunna använda datorstödet för att bygga upp och använda **egna** dokumentbaser.

Läraryrket är krävande. Den pågående reformeringen av skolväsendet innebär att läraren förutsättes medverka i skapandet av arbetsplaner, att de aktivt väljer undervisningsstoff etc. De centrala kursplanerna och de färdiga läromedelspaketen kommer i framtiden inte att tillåtas styra i samma omfattning som tidigare. Det är inte lätt att leva upp till dessa fordringar. Ett av lärarens ständiga bekymmer är individualiseringsproblemet: eleverna är olika och kan inte utsättas för samma undervisning. Ett annat problem är svårigheterna med att utvecklas i arbetet. Det är mödosamt att förnya sin undervisning och sitt undervisningsmaterial. Erfarenheten visar, att det framför allt är de yngre och mest energiska lärarna som orkar leta upp och stencilera ut nytt material, medan andra tenderar att upprepa samma lektioner år efter år. Ett därmed besläktat problem är beroendet av de tillgängliga läromedelspaketen. Ett fjärde problem är att lärararbetet ofta är isolerat. Det är trögt att få igång arbetsenheter som verkligen samarbetar, eller samverkan över ämnesgränserna eller mellan stadier och skolor.

Ett förnuftigt utformat datorstöd skulle kunna hjälpa lärarna att hantera bl.a. de fyra nämnda problemen. Så skulle exempelvis en arbetsenhet kunna skapa dokumentbaser som innehåller undervisningsmaterial (författat av lärarna själva och hämtat från annat håll), referensmaterial och kanske elevarbeten. Sådana dokumentbaser kan underlätta individualiseringen: skilda elevgrupper kan förse med skräddarsytt undervisningsmaterial och kanske själva tillföra material från sina egna undersökningar. För det andra kan dokumentbaserna bidra till att göra lärarnas arbete mer kumulativt. Det arbete som under ett läsår ägnas ett visst kursmoment finns lätt tillgängligt och kan revideras eller kompletteras till nästa år. För det tredje skulle en gemensam dokumentbas kunna fungera som en rik och flexibel "elektronisk lärobok", ur vilken undervisningsmaterial kan hämtas, i form av kompendier eller på annat sätt. För det fjärde skulle välfungerande gemensamma dokumentbaser underlätta och utveckla samarbetet mellan lärare.

De tekniska möjligheterna finns. Redan i dag erbjuder ett av de stora amerikanska bokförlagen skräddarsydda läromedel, hämtade ur förlagets egen omfattande dokumentbas. Kunderna är lärare, som i en katalog över vad dokumentbasen har att bjuda kryssar för de dokument som kan fungera som "kapitel" i en lärobok för en bestämd kurs och en bestämd elevgrupp. Läraren skickat in sin beställning och erhåller efter någon vecka per post en bunt läroböcker, kompletta med omslag (där det kan stå "i urval av adjunkt X"), löpande paginering, innehållsförteckning etc. Tack vare den automatiserade produktionstekniken blir det inte dyrare att beställa 23 skräddarsydda exemplar till sin klass än att köpa färdiga läroböcker i bokhandeln.

Ur lärarens synvinkel är nackdelen med ett sådant förfarande — som vi utan tvivel kommer att se mer av i framtiden — att förlaget, inte läraren själv, kontrollerar dokumentbasen. Lärarna borde ha möjlighet att bygga upp sina egna dokumentbaser. Materialet kan självfallet köpas från bokförlag (i framtiden förhoppningsvis i större skala i takt med att de upphovsrättsliga problemen löses), men det kan också hemföras från annat håll. Det kan skapas av lärarna eller eleverna, hämtas in från tryckta dokument med hjälp av skanner, från kolleger på andra skolor via elektronisk post och konferenssystem, från Skolverket och andra verk och myndigheter, från externa databaser etc.

Det kommer självfallet att dröja innan sådana "kunskapsverkstäder" finns på varje skola. Vi får här tänka oss olika ambitionsnivåer. En fullt utbyggd verkstad bör tillåta medarbetarna att arbeta med sina favoritprogram under alla de mest spridda operativsystemen (f.n. DOS, Windows, UNIX, Macintosh OS), och här skall finnas en rik uppsättning redskap för att

hämta hem, bearbeta och offentliggöra material. Kanske kan exempelvis vissa lärarhögskolor utrusta sig med en sådan arbetsmiljö. På skolorna behövs enklare och mer robusta system — men även om det är osannolikt att man där kommer att behöva alla de byggstenar som ingår i en utbyggd verkstad, är det sannolikt att man med tiden kommer att behöva några av dem. Man kan också tänka sig en för en kommun eller region gemensam kunskapsverkstad, dit intresserade lärarna kommer för att under sakkunnig ledning sammanställa och duplicera sina läromedel för nästkommande termin.

7. Datorstödet bör främja ett **undersökande** arbetssätt, och samtidigt öppna porten till kunskapstraditioner.

Många pedagoger hyllar ett undersökande arbetssätt. Kunskapssökandet får inte onödigtvis begränsas av ett traditionellt stoffurval, eller av skrankorna mellan vetenskapliga discipliner eller mellan skolämnen. Tidigare har lärare ofta känt sig bundna av hur stoffet presenteras i den tillgängliga läroboken. I dag skulle rika och välstrukturerade dokumentbaser kunna underlätta ett arbetssätt som är mer "problemorienterat", för att använda ett frekvent honnörsord.

Samtidigt får problemorienteringsprincipen inte drivas för långt. Att studera är att ärva, att göra sig förtrogen med skilda slag av kunskapstraditioner — inte bara de traditioner som representeras av olika vetenskapliga discipliner, utan även litterära och konstnärliga traditioner, yrkeskunnande etc.

En väl organiserad dokumentbas skulle kunna vara till hjälp i denna besvärliga balansgång mellan problemorientering och traditionsmedvetenhet. Systemet kunde exempelvis tillåta att man väljer mellan (eller själv skapar) olika alternativa uppsättningar s.k. hypertextlänkar, dvs. vägvisare till olika ställen i dokumentbasen. Man kunde då välja mellan å ena sidan hierarkiska länkar, som från en översiktlig grundkurs visar vägen till fördjupningar, till monografiska framställningar, till klassiska originaltexter, å andra sidan "laterala" länkar som överskrider disciplinränsar etc och med utgångspunkt i bestämda problem visar vägen till relevant material på skilda håll i dokumentbasen. I många sammanhang behövs en kombination. En dokumentbas utan hierarkisk organisering blir lätt till en labyrint där man varken hittar ut eller in, medan en renodlat hierarkisk organisering av ett kunskapsområde om det vill sig illa fungerar som en korrektionsanstalt (ett välbekant bekymmer är att grundskolans kursplaner ibland varit utspädda versioner av gymnasiet, vilka i sin tur varit utspädda versioner av universitetskurserna).

8. Datorstöd för skolbruk bör underlätta lärarnas (och elevernas) **tillgång till arbetsmaterial**.

Till arbetsmaterialet i vidaste mening hör vetenskaplig litteratur, referenslitteratur, undervisningsmaterial, men även de gällande kursplanerna, det därtill knutna referens- och kommentarmaterialet, de lokala arbetsplanerna, relevant material som producerats på andra skolor, samt givetvis det material som läraren själv (eller eleverna) sammanställer.

All dylik information bör så långt som möjligt göras tillgänglig i sådan maskinläsbar form att den smärtfritt kan flyttas in i lärarens eller arbetsenhetens egen dokumentbas. Så är icke fallet i dag. Regeln är i stället att informationen, exempelvis större databaser som nås via telenätet eller de encyklopedier som säljs på kompaktskivor, är utformad med tanke på den

användare som önskar leta fram isolerade sakupplysningar. Det är förståeligt att sökningsproblemet hamnat i fokus: inför de snabbt växande informationsmängderna behövs system som hjälper en att hitta det man behöver (och som låter en slippa hitta det man inte behöver). Men konsekvensen har blivit en djungel av oförenliga söksystem och databasarkitekturer som står i vägen för det arbete vi här talar om, nämligen när användaren önskar hemföra valda delar av materialet för att därav bygga sin egen dokumentbas. Ur lärarens perspektiv skulle arbetet bli mer givande om han eller hon på så sätt ur ett rikhaltigt "referensbibliotek" kunde välja material som på ett enkelt sätt kunde införlivas med det egna "handbiblioteket".

Ur statens perspektiv skulle en mer tidsenlig styrning av skolan kunna främjas. Tidigare har undervisningen i hög grad styrts av dels timplanerna, dels detaljerade kursplaner vilkas momentindelning speglats i läromedelspaketens uppbyggnad. I framtiden kommer andra styrinstrument att behövas. Idealet är att skolmyndigheterna litar på lärarnas professionella kunnande, vilket innebär att det goda exemplet skulle få styra. Skollärarna skulle därmed åtnjuta något av den *Lehrfreiheit*, den frihet att välja undervisningens form och innehåll, som nyhumanisterna såg som universitetets adelsmärke.

Man kan tänka sig att Skolverket en gång per termin distribuerar — med dagens teknik lämpligen på kompaktskivor, snart kanske på annat sätt — de offentliga styrdokumentet tillsammans med material som förslagsvis kan kallas referensdokument. Dessa referensdokument kan förslagsvis vara av tre slag: 1. kommentarer till styrdokumentet, 2. didaktiskt material, dvs. olika slag av undervisningshjälpmedel, såsom instuderings- och övningsmaterial, diagnostiska prov, inspirerande exempel på hur man vid andra skolor arbetat med arbetsplaner och undervisningsplanering, etc, och 3. saklitteratur, såsom vetenskaplig och populärvetenskaplig litteratur, parallellöversättningar, konkordanser, forskningsöversikter, bibliografier, handböcker, lexika etc. Härur klipper lärarna vad de har bruk för och infogar det i sina egna dokumentbaser.

Det kan visa sig rimligt att införa en arbetsdelning, där Skolverket ansvarar för spridningen av de två förstnämnda slagen av referensdokument, dvs. kommentarer och didaktiskt material, medan skolorna hämtar saklitteratur direkt från förlag, universitetsinstitutioner etc.

Ungefär som i fråga om biblioteksväsendet är det ett offentligt intresse — som organ som forskningsråden, Forskningsrådsnämnden, UHÄ och Skolverket borde försvara — att central litteratur för forsknings- och undervisningsändamål blir lätt och billigt tillgänglig i maskinläsbar form. En möjlighet är att litteratur av nationellt intresse (Nationalencyklopedin, Nationalatlasen, Nationalutgåvan av Strindbergs verk...) "friköpes" en gång för alla, att användas fritt inom hela utbildningsväsendet. Ett annat alternativ är att materialet vid behov inköpes av kommuner, rektorsområden, enskilda skolor, enskilda lärare; i så fall behövs något slags standardiserat kopieringsavtal som reglerar den fortsatta spridningen. Ett tredje alternativ är olika former av licensering, ungefär så som företag eller institutioner i dag köper licenser för datorprogram. Tack vare enskilda entusiasters insatser cirkulerar dessutom redan i dag på universitetsnäten en hel del material (exempelvis Shakespeares samlade skrifter), kostnadsfritt tillgängligt för envar med en dator, ett modem och tillgång till näten. Det rör sig dock än så länge i huvudsak om engelskspråkiga texter, dessutom en försvinnande liten del av all den litteratur som vore av intresse i undervisningen och forskningen.

Dessvärre finns en rad olösta juridiska och ekonomiska problem som förhindrar att värdefull litteratur i maskinläsbar form distribueras i större skala. Så länge oklarheterna kvarstår rörande den juridiska upphovsrätten och rörande den ekonomiska ersättningen till förlag, författare, redaktörer, konstnärer och fotografer, kommer förlag eller bildbyråer även fortsättningsvis att vara obenägna att offentliggöra sitt material i maskinläsbar form — naturligt nog, så snart det börjar cirkulera på näten kan det mångfaldigas i obegränsad

omfattning. Även frågan om den ideella upphovsrätten måste redas ut: det bör finnas vissa garantier (juridiska regler eller, som inom den vetenskapliga världen, en hederskodex) som förhindrar att t.ex. författarens namn försvinner i hanteringen, eller att den ursprungliga texten börjar cirkulera i deformerat skick.

Problemen är svårhanterliga men de måste lösas. Ur lärarnas och forskarnas synvinkel innebär dagens situation ett otillständigt intellektuellt resursslöseri, eftersom så gott som allt tryckt material numer existerar i maskinläsbar form innan det skickas till sättning.

9. Datorstödet bör bidra till att **knyta samman** kunskapsområden och skilda nivåer inom samma kunskapsområde.

Den informationsteknologiska utvecklingen har inneburit att det nu existerar verktyg som medger en synnerligen flexibel användning av lokala dokumentbaser. Man kan tänka sig att sådana baser innehåller en förhållandevis fast och mångsidig "kärna" — låt säga en elektronisk version av den svenska Nationalencyklopedin, kompletterad med ordböcker, lexika och annan referenslitteratur — omgiven av ett "skal" av material anpassat till specifika behov (ett visst skolämne, ett visst skolstadium etc), samt ytterst ett flexibelt och ständigt föränderligt "skal" innehållande lärarens, arbetsenhetens eller elevgruppernas aktuella arbetsmaterial.

En annan modell är en dokumentbas som är "skiktad" på så sätt att den blir tillgänglig för användare med varierande förkunskaper och ambitionsnivå. På den översta nivån finns det aktuella arbetsmaterialet. Närmast därunder finns "innehållsförteckningar" eller andra slag av ingångar till dokumentbasen (exempelvis i form av den lokala arbetsplanens uppdelningar i ämnesområden, moment eller teman). I nästa skikt finns material av "grundkurskaraktär", det vill säga korta och didaktiskt väl tillrättalagda framställningar som vid behov även kan fungera som repetition eller som översikt. Härifrån kan läraren eller mer försigkomna elever leta sig ned i djupare skikt med allt mer omfattande och komplex information.

Åtskilliga läromedelsförfattare som idag lider under tvånget att förenkla skulle förmodligen attraheras av uppgiften att bidra till att bygga upp "skiktade" dokumentbaser. När de i dag sammanställer ett tryckt läromedel erbjuder de i bästa fall tillfälle att servera läsaren förslag till vidare läsning. Det vore bättre om fördjupningslitteraturen helt enkelt ingick i läromedlet, eller åtminstone vore lätt åtkomlig för den intresserade. Ytliga framställningar behövs, men vore ur både författarens och läsarens synvinkel mer acceptabla om mer kvalificerat material funnes tillgängligt under ytan.

Med olika slag av "skal" eller "ytterskikt" borde samma eller överlappande dokumentbaser kunna användas på olika nivåer inom utbildningsväsendet, från förskolan till universitetet och den fria bildningsverksamheten. Det är lätt att inse de arbetsmässiga och ekonomiska fördelarna med ett sådant sambruk av information. Förlagen, författarna och översättarna skulle kunna räkna med en månghövdad marknad, vilket torde vara en förutsättning för att omfattande och genomarbetade dokumentbaser över huvud taget skall se dagens ljus. Revideringar av innehållet skulle kunna ske snabbt och till rimlig kostnad. Eftersom all erfarenhet visar att elektroniska informationsbaser måste var synnerligen innehållsrika för att utgöra ett alternativ till pappersburen information, är det nödvändigt att informationens "styckepris" hålls nere.

10. Datorstödet är **ingen ersättning** för goda lärare eller goda böcker.

Det borde inte behöva sägas, att datortekniken varken skall eller kan träda i stället för läraren. Datorstödet är just ett stöd, som om det fungerar väl låter lärarens yrkeskunnande komma till sin rätt. Lärarens arbete blir kanske inte lättare, men åtminstone mer givande och fruktbarande.

Digitalt lagrad information skall heller inte tränga ut den pappersburna. Praktiska och estetiska skäl gör att böcker, tidskrifter och tidningar är oersättliga kunskapskällor. Det finns dessutom ofta anledning att skriva ut digitalt lagrad information på papper. Det gäller exempelvis när lärare väljer material ur sin lokala dokumentbas och därav skräddarsyr läromedel i form av kompendier som delas ut till eleverna. Därmed kan eleverna slippa sådana läromedelspaket och studiehäften som saknar bestående värde. I stället kan läromedelskontot användas för att införskaffa riktiga böcker.

Bilaga 2

D. Broady: "Lokala dokumentbaser för undervisningsbruk", publicerad i *Teknologi och kompetens/Technology and Competence. Proceedings from the 8th Nordic Conference on Information and Documentation*, 1992, pp. 59-63.

Lokala dokumentbaser för undervisningsbruk

Title: Local Document Bases in Education

Keywords: Structured documents, hypertext, SGML, collaborative authoring, educational applications.

Abstract: Collections of structured documents are powerful resources in education, e.g., as raw material for production of tailor-built course material. DARC (Document ARchive Controller) is a system supporting import, organization, editing, revision, and export of such documents. The ISO-standard SGML is used as a machine-independent intermediary document format.

Historikern Roger Chartiers, en framstående kännare av läsandets och bokutgivningens historia i Frankrike, inleder sin senaste bok med en fråga: "hur försökte människorna i Västerlandet mellan senmedeltiden och 1600-talet att bemästra den mångdubblade mängd texter som spreds i först handskrivna och sedan tryckta böcker? Att hålla räkning på titlarna, klassificera böckerna och tillskriva dem en författare var operationer som gjorde det möjligt att bringa ordning i det skrivna ordets värld. Vår egen samtid är direkt arvtagare till detta oerhörda arbete som utlöstes av att människorna fått sina cirklar rubbade. Under dessa avgörande sekel, då den för hand kopierade boken gradvis ersattes av boken som trycktes med lösa typer, befästes de gester och tänkesätt som alltjämt är våra. Uppfinnandet av författaren som grundläggande princip för hur man hänvisar till texter, drömmen om ett universellt bibliotek, existerande eller immateriellt, innefattande alla verk som någonsin skrivits, uppkomsten av en ny definition av boken såsom en ouplöslig förening av ett föremål, en text och en författare, utgör några av de innovationer som, före och efter Gutenberg, omvandlade människors förhållande till texterna." (Chartiers, 1992, p. 7)

Framtidens historiker kommer förmodligen att fälla liknande omdömen om det sena tjugonde seklet. Den magnetiskt eller optiskt lagrade informationen och de digitala överföringsteknikerna är inte blott tekniska nyheter. De ackompanjeras av en begynnande omvälvning av vår hållning till skrivandet och läsandet.

Det kommer att dröja länge innan de nya sätt att skriva och läsa som just nu växer fram har funnit sin form. Vi får inte glömma att det tog hundratals år att utveckla de konventioner vi nu finner självklara beträffande tryckta alster: paginering, textens indelning i kapitel och

stycken, titelsidans, innehållsförteckningens, sakordsregistrets och litteraturreferensernas utformning och placering, att boken låter sig identifieras med hjälp av sådant som författarnamn, titel, förlagsort, tryckår, ISBN-nummer, signum etc. Alla dessa konventioner för innehållslig organisering, typografisk formgivning och bibliografisk klassificering måste omprövas när vi har att göra med elektronisk i stället för pappersburen information.

I vissa avseenden innebär paradoxalt nog den nutida informationshanteringen en återgång till tiden före den gutenbergska eran. Elektroniska texter uppträder ofta i många individuella versioner, ungefär som handskrifterna, till skillnad från den massproducerade boken där varje exemplar ur en och samma upplaga är det andra likt. Ett annat exempel är bläddringsförfarandet, där läsaren i fönstret på skärmen ofta överst ser de rader som fanns längst ned på förra skärmbilden, en princip som var vanlig i kodexar och även i en del av de första tryckta böckerna. (Detta illustrerar i miniformat ett av de svåra problemen i samband med övergången till elektronisk information: att de flesta av de möjligheter till orientering och överblick som den tryckta information erbjuder går förlorade och måste ersättas med andra slag av stöd.) Ett tredje exempel är att man ånyo drömmer om det totala biblioteket som ger tillgång till all världens litteratur, i dag dock inte längre på hyllor utan via online-förbindelser. Ted Nelson (se t.ex. 1987) är den tänkare som under de senaste decennierna mest radikalt aktualiserat denna dröm, som är mycket gammal (jfr särsk. femte kapitlet i Luciano Canforas bok från 1986 om biblioteken i Alexandria och Pergamon under den hellinistiska epoken).

Även om dagens informationsteknologi erbjuder både nygamla och tidigare helt oprövade möjligheter, måste systemutvecklarna samtidigt ta hänsyn till att det existerar sofistikerade metoder för arbete med skriftliga dokument. Den författare eller läsare som är van att handskas med omfattande och komplexa textmassor, som sätter värde på att kunna följa en författares argumentation tvärs genom en hel bok, som kanske t.o.m vet att utnyttja textkritiska utgåvor, konkordanser och parallellöversättningar, är föga trakterad av datorprogram av den sort som inbjuder till att informationen hackas upp i små tuggor. Jag skall här, med lärarnas arbete som åskådningsexempel, redogöra för ett pågående försök att utveckla stöd för arbete med omfattande samlingar strukturerade dokument.

De följande övervägandena är utgångspunkter för projektet "Datorstöd för kunskapsarbete" (även benämnt "Datorstöd för författande och samarbete i utbildning"), som stöds av NUTEK (insatsprogrammet ITYP), Forskningsrådsnämnden och Skolverket. Projektet är ett samarbete mellan IPLab (Informations- och presentationslaboratorium) vid Institutionen för numerisk analys och datalogi, KTH, och Institutionen för pedagogik, Högskolan för lärarutbildning i Stockholm. Bland medarbetarna bör nämnas Hasse Haitto, samt Peter Lidbaum och Magnus Tobiasson som tagit sig an programmeringsproblemen. Rapporten *Kunskapsverkstaden*, 1991, ger en kort presentation av projektets syfte.

Läraren som författare och redaktör

Den svenska skolan genomgår just nu genomgripande förändringar som kommer att påverka lärarnas arbetssituation. Hittills har innehållet i skolans undervisning framför allt styrts av dels de centrala timplanerna och kursplanerna, dels läromedlen. De senare har i sin tur ofta byggt på kursplanernas förslag till undervisningsinnehåll. Dessa arbetsvillkor kommer förmodligen att omvandlas som en följd av pågående organisatoriska förändringar. Decentraliseringen och kommunaliseringen, övergången från regelstyrning till mål- och resultatstyrning, det nya statsbidragssystemet som innebär att kommunerna får en enda

klumpsumma i stället för ett antal separata öronmärkta bidrag, nya mindre detaljstyrande läroplaner, nya profilerade skolor med eget undervisningsutbud, kravet att lärarna på varje skola skall utveckla lokala arbetsplaner för den egna undervisningen, avskaffandet av den statliga läromedelgranskningen liksom av principen om gåvoläromedel — allt samverkar till att ansvaret förpassas längre ned i systemet. En tänkbar konsekvens är att lärarna blir friare att välja undervisningsstoff efter eget huvud. De kan inte längre förlita sig på detaljerade kursplaner eller prefabricerade läromedelspaket.

Samtidigt med dessa organisatoriska förändringar har den teknologiska utvecklingen skapat nya möjligheter att hämta undervisningsmaterial från olika håll och att sammanställa skraddarsydd läromedel. De nya redskapen låter läraren fungera som redaktör och författare.

Det är i och för sig ingen nyhet att somliga lärare (och elever) sökt alternativ till de färdiga läromedlen, men detta arbete har varit mödosamt. Mycken tid har spillts vid fotokopieringsapparaten, många timmar har ägnats åt renskrift vid skrivmaskinen eller datorn och åt klippande och klistrande. Ett tungrovt arbetssätt, som sällan blivit riktigt kumulativt: det har varit svårt att utvidga och förbättra materialsamlingen från det ena året till det andra, och lärarna har haft svårt att finna formerna för samarbete kring en gemensam materialbank.

Kunskapsverkstaden

Låt oss för ett ögonblick tänka bort datorerna, och föreställa oss all den information på papper som en ambitiös lärare samlar i sitt arbetsrum. På hyllorna finns ett handbibliotek: böcker och tidskrifter med understrykningar och marginalanteckningar, handböcker, encyklopedier och lexika. Här finns samlingar av fotokopior, klipp och särtryck. Här finns excerpater ur lästa böcker och material från genomgångna kurser. Här finns läroplaner och annat offentligt tryck, diagnostiska prov och olika slag av läromedel, både inköpta och av egen tillverkning. Och här finns, i A4-pärmar eller mappar eller kortlådor, andra papper som framställts av läraren själv, kollegerna eller eleverna: de lokala arbetsplanerna, kurs- och lektionsförberedelser, skrivningsfrågor, anteckningar om eleverna, elevernas grupparbeten.

Låt oss sedan tänka oss en tekniskt välutrustad "kunskapsverkstad" för en enskild lärare eller en arbetsenhet. I en sådan måste man kunna hantera material som det ovan nämnda. Till begränsningarna hos det mesta av den hittillsvarande undervisningsteknologin hör att den så ensidigt syftat till att betjäna eleverna. Jag tror att de system som introduceras i skolorna i första hand bör fungera som stöd i lärarnas dagliga arbete. Då och först då kommer lärarna att ha lust och möjlighet att visa eleverna vägen till vettiga användningar av datortekniken.

Kärnan i en sådan verkstad kan vara en lokal dokumentbas, dvs. en lättåtkomlig och flexibel informationsbank som lärarna ständigt har till hands och kan anpassa till sina egna behov. (Med attributet "lokal" avses här inte att dokumentsamlingen i fysiskt avseende behöver lagras i användarens närhet, utan blott att den, ungefär som litteraturen i handbiblioteket i arbetsrummet, skall väljas och ordnas med omsorg och vara lätt tillgänglig, till skillnad från externa informationskällor som begagnas mer sporadiskt.) Så kan exempelvis lärarna samla undervisningsmaterial av god kvalitet i dokumentbasen för att därur göra utsnitt som kan stöpas om till läromedel, kanske i form av kompendier som delas ut till eleverna. Men dokumentbasen bör duga till mer än så. Den bör vara ett ställe där läraren samlar och ordnar referenslitteratur och saklitteratur, excerpater från sina egna självstudier, material från kurs- och lektionsplaneringen etc.

En lokal dokumentbas skulle kunna innehålla två slag av dokumentsamlingar. För det första finns här "material under arbete", som lagras i varierande format, oftast anpassade till

de program för textbehandling, registerhantering, bildbehandling, layout etc som användaren föredrar. Ett särskilt slag av "material under arbete", som kräver specifika samarbetsverktyg (versionshantering, stöd för remissförfaranden etc), är det som bearbetas av flera medarbetare, exempelvis en framväxande kollektivt författad lärobok eller material som elever använder i ett pågående grupparbete.

För det andra innehåller dokumentbasen ett antal "arkiv", bestående av dokument som (tills vidare) är färdiga och "statiska". Somliga arkiv av mer privat natur existerar blott i ett fåtal kopior hos den enskilde lärare eller arbetsenhet som framställt dem. Andra arkiv av mer allmänt intresse är offentliga. En grupp lärare kan låta guldkornen ur det egna undervisningsmaterialet cirkulera bland kolleger vid andra skolor. Redan i dag finns forskargrupper som genom ftp-förbindelse eller på annat sätt gör sina uppsatser och rapporter allmänt tillgängliga via universitetsnäten. Författare eller förlag kan (på kompaktskivor, via näten etc) publicera material som lärare kan inlemma i sina dokumentbaser och kanske använda som halvfabrikat varav valda delar infogas i exempelvis skraddarsydda läromedel.

Beskrivningen ovan skiljer sig inte så mycket från hur somliga datorintresserade lärare eller forskare arbetar redan i dag. Man bygger upp rikhaltiga samlingar av maskinläsbart material, både eget och sådant man erhållit från kolleger eller införskaffat på annat sätt. En växande litteratur och många databaser finns tillgängliga via näten, och förlagen distribuerar encyklopedier och annan saklitteratur på kompaktskivor. Men var och en som använder sin dator på det sättet vet hur mycket tid som krävs för att stöpa om materialet så att det blir omedelbart tillgängligt. Eftersom producenterna använder de mest olikartade dokumentformat, databashanterare och söksystem, tvingas den som önskar nyttja materialet till att bygga upp ett eget handbibliotek ägna mycken möda åt formatkonverteringar och andra teknikaliteter. Uppenbarligen vänder sig producenterna främst till kunder som nöjer sig med att leta upp isolerade sakuppgifter. Ett honnörsord i sammanhanget brukar vara "information at your fingertips" — användaren skall få sina frågor besvarade utan att behöva bekymra sig om hur informationsmassan är organiserad. Den väg vi här föreslår är den motsatta. Användaren skall ha full kontroll över hur det "handbibliotek" han eller hon önskar daglig tillgång till är organiserat.

Systemkrav

Inom projektet Datorstöd för kunskapsarbete utvecklar vi prototyper för "kunskapsverkstäder" under de f.n. vanligaste operativsystemen, dvs. DOS, Windows, UNIX och Macintosh OS.

Den som arbetar i persondatormiljö är ofta utlämnad åt tillämpningsprogrammen. Regeln är att ett dokument är knutet till vissa tillämpningsprogram, vanligtvis till det program i vilket dokumentet skapats. Vi eftersträvar i stället system där arbetet organiseras med dokumentet i centrum. Ett och samma dokument skall kunna betraktas ur en mängd perspektiv och bearbetas med en mängd verktyg.

En fullt utbyggd kunskapsverkstad bör erbjuda en rik uppsättning verktyg som tillåter att material hämtas hem från skanner, CD-ROM eller WORM, tape, disketter, eller via teleförbindelse från externa databaser, från biblioteks-, konferens- eller postsystem etc. Valda delar av detta material sparas i dokumentbasens arkiv. Vidare behövs konverteringsprogram och andra verktyg för utförsel av material från arkiven: till fortsatt bearbetning i ett textbehandlings-, layout- eller hypertextprogram, till laserskrivare eller till fotosätter, till näten etc. Dessutom behövs verktyg för indexering, dokumentjämförelser och mycket annat.

Ett akut bekymmer så snart man arbetar med större mängder information rör svårigheten att skaffa sig överblick. De orienterings- och överblickshjälpmedel vi vant oss vid i samband med pappersburen information måste ersättas. Inom projektet Datorstöd för kunskapsarbete har vi ägnat särskilt intresse åt de s.k. hypertextsystemen, som tillåter att man lägger in ett slags vägvisare som förbinder olika platser i informationsmassan. Sådana hypertextlänkar används ofta som korshänvisningar mellan isolerade stycken information, men vi är här mer intresserade av hur länkningsmekanismerna kan utnyttjas för att skänka överblick över innehållet i dokumentbasen. Inom projektet har vi installationer av några av de kraftfullare hypertextsystemen (Intermedia, Xerox NoteCards m.fl.), vilka dock kräver ganska mycket i fråga om både utrustning och användarnas datorkunnande. Ett än större hinder är att så många av de mer utvecklade hypertextsystemen var för sig utgör en i hög grad slutna värld. Därför har vi fastnat för ett par av de mest "öppna" hypertextsystemen: Guide, Microcosm, samt framför allt DynaText. Det sistnämnda systemet erbjuder en lång rad fördelar vid arbete med dokumentbaser av det här skisserade slaget. Visserligen är DynaText främst avsett för författande och presentation av statiska dokument, men den UNIX-version 1.5.1 som blev tillgänglig i början av 1992 ger läsaren vissa interaktiva möjligheter att anpassa materialet till sina behov: att infoga egna (privata eller offentliga) annoteringar, att genomföra struktursökningar, att skapa egna "stigar" som genomkorsar materialet, etc. DynaText finns även för Windows, när detta skrivs fortfarande i betaversion.

Dessutom bygger vi inom projektet ett eget system, DARC (Document ARchive Controller), med vars hjälp en lokal dokumentbas kan byggas upp, inspekteras och ordnas. DARC stödjer exempelvis framställningen av ett slags elektroniskt "kompendium", dvs. en sammanställning av utsnitt ur dokumentbasen kompletterade med automatiskt genererade "källhänvisningar". Framställning av skräddarsydda läromedel är ett tänkbart användningsområde. DARC skrivs i C++ och existerar i en UNIX-version för SUN arbetsstationer samt i en persondatorversion under Windows.

Det vore orealistiskt att förvänta sig att varje lärare har tillgång till alla de nämnda faciliteterna. Vi får tänka oss olika ambitionsnivåer. Rikt utrustade UNIX-baserade verkstäder kan finns vid vissa universitetsinstitutioner eller lärarhögskolor, hos vissa läromedelproducenter, eller kanske på något centralt ställe i kommunen dit lärare beger sig för att under sakkunnig ledning framställa läromedel för nästkommande termin. Persondatorbaserade installationer kan finnas på enskilda skolor eller hemma hos särskilt intresserade lärare.

Dokumentformat

Ett centralt systemkrav är att samma dokumentarkiv skall kunna användas i många skilda sammanhang. Det innebär att man, utan att den innehållsliga strukturen går förlorad, skall kunna forsla dokumenten mellan olika plattformar och olika program. Att strukturen bevaras innebär att systemet håller reda på vad som är brödtext, huvudrubrik och underrubrik på skilda nivåer, namnet på dokumentets författare, grafiskt element av en viss sort, fotnot, litteraturreferens, och så vidare. (Vi talar här inte om den typografiska utformningen. Det väsentliga är inte att en rubriktext skrivs ut med 14 p Times och en viss kägel, utan att textelementet i fråga markeras som t.ex. rubriknivå 3). För det syftet behövs ett generellt, av maskinvara och tillämpningsprogram oberoende dokumentbeskrivningsspråk. Därmed kan alla dokument, oavsett hur de skapats och oavsett hur de är avsedda att senare användas,

föreligga i ett och samma intermediära format, omedelbart tillgängliga för export till, låt säga, ett textbehandlingsprogram. Närhelst ett nytt textbehandlingsformat el.likn. blir aktuellt, kan systemet kompletteras genom att man skriver översättningsprogram till och från det intermediära formatet (till skillnad från den anarki som uppstår vid direktkonvertering mellan ett otal tillämpningsprogram). Det bästa alternativet är av allt att döma dokumentbeskrivningsspråket SGML (Standard Generalized Markup Language), en etablerad ISO-standard som är väl ägnad för syftet att definiera de strukturella elementen hos komplicerade dokument. De system vi bygger inom projektet Datorstöd för kunskapsarbete är avpassade för dokumentbaser där flertalet dokument lagras i SGML-format, och vi använder ett stort antal konverterings- och presentationsprogram och andra byggstenar (däribland det nämnda hypertextsystemet DynaText) avsedda för hantering av just SGML-märkta dokument. (För en översikt över den på sistone snabbt växande floran av verktyg för hantering av sådana dokument, se Haitto, 1991.)

Behovet av ett generellt dokumentbeskrivningsspråk kan inte nog understrykas. Om forskare och lärare, förlag och myndigheter publicerade sitt material i förslagsvis SGML-format, skulle detta med bevarad innehållslig struktur jämförelsevis problemfritt kunna införlivas med användarens egen materialsamling, exempelvis i en lokal dokumentbas som den här skisserade. (Hindren är här inte i första hand tekniska, utan kommersiella och juridiska. Flaskhalsen som förhindrar att maskinläsbart material göres allmänt och enkelt tillgängligt är den rådande oklarheten rörande upphovsrätten och kopieringsrätten i fråga om elektroniskt material.)

Inom projektet strävar vi efter att skapa integrerade miljöer, där systemet i så hög grad som möjligt ombesörjer översättningen till och från SGML-format och kontrollerna i samband därmed. Det innebär att SGML-märkta dokument, lagrade i den lokala dokumentbasens "arkiv" eller hemförda från annat håll, mer eller mindre automatiskt skall kunna överföras till användarens vardagliga arbetsmiljö (SUN, PC eller Mascintosh) och favoritprogram, och omvänt.

Slutord

Jag har resonerat om skollärares och läromedelsförfattares behov, men system av det slag som här beskrivits kan användas även av elever som genomför ett grupparbete, eller av vem som helst som i sitt arbete dels samlar och ordnar större dokumentbaser, dels därur framställer egna skrifter: forskare och utredare, högskole- och universitetslärare, dokumentalister, redaktörer, skriftställare och många andra.

Referenser

D. Broady (1991): *Kunskapsverkstaden — om lokala dokumentbaser som arbetsverktyg för lärare*. IPLab-rapport 47, TRITA-NA-P9132. Stockholm: KTH, Institutionen för numerisk analys och datalogi.

L. Canfora (1986): *La biblioteca scomparsa*. Palermo: Sellerio editore. (Fransk. övers. *La véritable histoire de la bibliothèque d'Alexandrie*. Paris: Éditions Desjonquères, 1988.)

R. Chartiers (1992): *L'ordre des livres. Lecteurs, auteurs, bibliothèques en Europe entre XIV^e et XVIII^e siècle*. Aix-en-Provence: Éditions ALINEA.

H. Haitto (1991): *SGML-relaterade verktyg för intermediära former och dokumenthantering*. IPLab-rapport 37, TRITA-NA-P9106. Stockholm: KTH, Institutionen för numerisk analys och datalogi.

Th. H. Nelson (1987): *Literary Machines*, edition 87.1 (eget förlag).

Darc – Document ARchive Controller. Institutionen för numerisk analys och datalogi, KTH,
 prel. version 15 febr 1993

Darc

Document Archive Controller

Donald Broady
 broady@nada.kth.se

Hasse Haitto
 haitto@nada.kth.se

Peter Lidbaum
 pli@nada.kth.se

Magnus Tobiasson
 tobima@nada.kth.se

IPLab/NADA*
 Royal Institute of Technology
 S-100 44 STOCKHOLM
 SWEDEN

IPLab Report 63, TRITA-NA-P9306

March 19, 1993

Keywords: SGML, hypertext, information retrieval, CSCW

Abstract


Darc is a multi-user, cross-platform (Sun SPARC/X11 & PC/Windows 3.1) database and information retrieval application designed for storing, reusing, and querying large quantities of SGML¹-encoded documents. It consists of three major parts:

- An SGML document on-line delivery tool (i.e. a viewer) with annotation and interactive hypertext facilities.
- A set of document databases with access control of users, and network-based support for Computer Supported Cooperative Work (CSCW) through hierarchical groups.
- A virtual filing tool to specify hierarchical relationships between documents.

*Phone: Int +46 8 790 6000 (Voice) or Int +46 8 790 0930 (Fax)

¹STANDARD GENERALIZED MARKUP LANGUAGE, an ISO-standard for document markup.

Background and Funding

 (Document Archive Controller) is the principal outcome of the *Computer Supported Knowledge Work* research project, a three-year joint effort between the Interaction and Presentation Laboratory (IPLab) at the department of Numerical Analysis and Computing Science of the Royal Institute of Technology, and the Department of Educational Research, Stockholm Institute of Education (Sweden's largest Teachers' College). The project has been funded by grants from the Swedish National Board for Industrial and Technical Development (NUTEK)², the Swedish National Board of Education (Statens Skolverk), and the Swedish Council for Planning and Coordination of Research (FRN).

This report is an overview which emphasizes features of the *Darc* system. Design rationale and technical aspects will be covered in subsequent papers.

Introduction

There are currently three commonplace techniques for organizing document collections:

- Information Retrieval (IR) methods
- Hierarchical file systems
- Hypertext solutions

The *Darc* system applies all of these techniques in one coherent framework for document management.

About The Overview

This overview will explain the *Darc* system from the viewpoint of its users, who belong to one of five categories. These categories form a hierarchy whose levels are numbered 1 to 5 (in ascending order). These levels are described in the next five sections. Some knowledge of SGML is assumed of the reader, but is not absolutely necessary.

1 The Guest

Darc has password-controlled access of database contents, with users belonging to one of five categories. At the lowest level in the *Darc* system, a *guest* has the least privileges. This category is for the occasional user, as a guest cannot modify anything in the database and is allowed access only to documents available to everyone.

²Formerly known by the acronym STU. The project's Swedish title is "Datorstöd för kunskapsarbete," project ITVP 90-02737P.

1.1 Document Databases

Being a database system, **DatE** is designed to handle vast quantities of documents (literally tens of thousands). Preferably, the documents are marked-up with the ISO standard SGML [7], since it is an SGML-based system and has support for full-text presentation and navigation of such documents. Naturally, the system can also be used as a repository for documents coded in other formats but without the benefits of the SGML support.

A guest can perform traditional, index-based bibliographical searches and view documents on-line. The guest will probably encounter and use *views*. These are covered in the next section.

2 The Reader

Anyone who uses **DatE** regularly should be granted at least *reader* privileges, as they include the right to create *views*—a mechanism which allows documents to be ordered and accessed in hierarchical structures.

2.1 A Virtual Filing Tool

Views are based on the metaphor of how one organizes the *hierarchical* file structure of a hard disk, but are more flexible. In effect, they allow users to create a personalized interface to database contents, and reduce the need to hunt for documents using traditional index-based searches.

A view (as shown in figure 1) is a set of labeled boxes, or *nodes*. Each node may contain documents as well as other nodes, commonly called subnodes or children. Of course, this representation has nothing to do with the actual physical storage of the documents. Views can be stored and accessed like any other database object.

Furthermore, views are not *static*. Nodes can be cut, copied and pasted. In fact, each user may tailor a view to his or her specific needs, add new nodes, delete obsolete ones, etc. A node can be closed (collapsed) so that it hides its children. The visual aspects of a node are covered in the next two sections.

2.1.1 Labels

A label is the name assigned to a node, and is supplied by the user when the node is created. The label can be any string, but normally reflects the nature of the documents stored within the node, or the methodology used in the hierarchical ordering. Views can be string searched by label.

The user is informed of all accessible views that contain a particular document³. This is done when the document is selected, e.g. as the result of a bibliographical search.

2.1.2 Child Indicator

Nodes with subnodes are prefixed with a child indicator. This is a plus sign to the left of the node label when the node is closed, and a minus sign when it is open.

³As illustrated in figure 3.

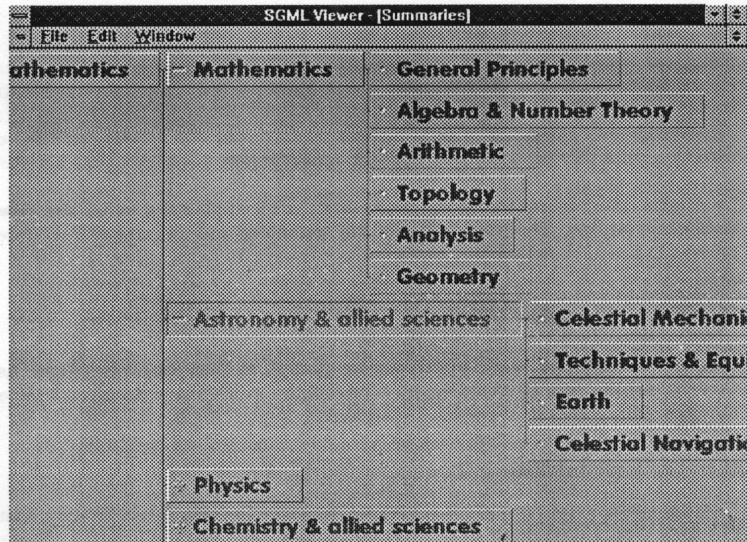


Figure 1: Views allow users to organize documents in hierarchical relations. Unlike the other figures in this paper, this screen snapshot is from the PC/Windows version of *DATE*. Both versions are functionally equivalent and the document databases are binary compatible across platforms.

Nodes without + or - are leaves, i.e. nodes without children⁴. The same node may exist in several places in a view.

2.1.3 Document Interchange

Views are somewhat similar to symbolic links under UNIX or aliases under Apple System 7, but are more flexible. Finally, the view mechanism is used to *export* a subset of the document database. A selected node, its subnodes and all documents stored therein can be compiled into a new document database. This packaging is a convenient method of database-level document interchange.

2.2 On-line Browsing, Annotating, and Linking

Any user may also read documents on-line. Beginning at the reader level and upwards, the presentation tool supports *browsing*, *annotating*, and *linking* of the documents in the database. These facilities will be covered in detail below.

⁴Leaves are prefixed with a cool 3D bullet. :-)

from the xterm window and 'paste' them into the new specification next to their newly named equivalents. The X Window System supports cutting and pasting among windows by clicking and dragging over the text with the left mouse button to cut and clicking on the middle button to paste. This facility can be used to facilitate data entry and update in any of

Figure 2: The highlighted section has been annotated. Clicking the icon will open up a separate window to display the contents of the annotation. Like all document objects, annotations can be shared publically with group members or kept private.

3 The Author

Documents may be added to the database (and in some circumstances removed) by users who have *author* privileges or higher.

3.1 Adaptable SGML-based Filing

When dealing with material marked-up with SGML, a system should make use of the existing mark-up to simplify document management. This is done in *DATE* by adapting database import to various SGML document type definitions. E.g., it is a reasonable assumption that bibliographical data elements will vary for each document class: Therefore, there is an end-user modifiable interface which specifies what elements in a document type definition are to be treated as bibliographical data so that any document will be able to *file itself*⁵.

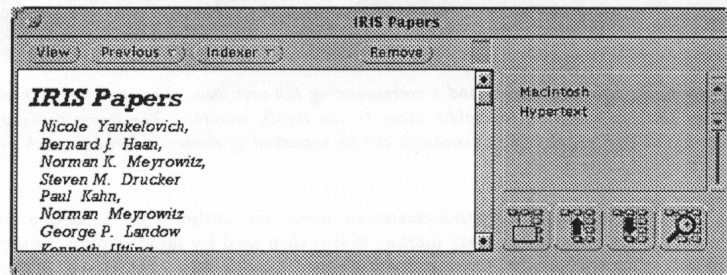


Figure 3: Bibliographical elements are shown in a window of their own. To see the full-text view of a document one selects the View button at the top left. The list at right displays the labels of nodes in which the document is represented in, if any.

Importing a document marked-up in SGML to the *DATE* system is extremely simple: First, the document is parsed in a matter of seconds, then the bibliographical data is presented in a separate window (see figure 3). At this point, the user may file the document (i.e. add it to the database), and/or browse through it using the

⁵See section 3.2.3 for how the SGML markup is used for indexing.

on-line viewer. It is thus not necessary to pre-compile or to file a document before reading it on-line.

3.2 An SGML On-line Delivery Tool

The pioneer efforts of Douglas Engelbart's *Augment* at the Stanford Research Institute demonstrated the benefits of organizing files into hierarchical structures, with outline-style access, that can be arbitrarily referenced and linked on-line [4]. The advent of descriptive markup, especially SGML, has since paved the way for on-line viewers to build on this rich heritage of ideas, see e.g. [3, 9]. In the same vein, *Date* includes an interactive browser to view SGML-encoded documents and network-based support for CSCW through its notion of groups (see section 4).

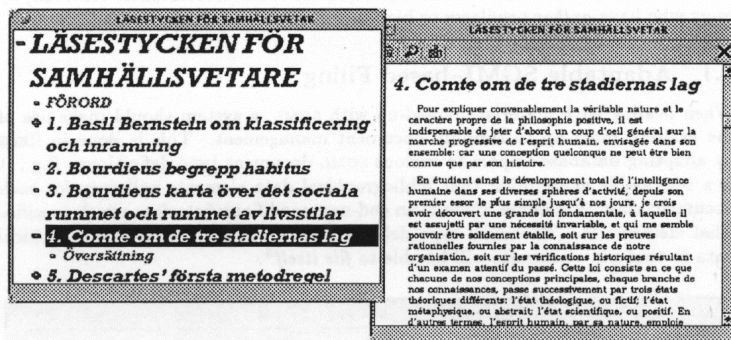


Figure 4: A table of contents and a corresponding full-text view. The user has clicked in the table of contents to rapidly move to the fourth chapter. The items prefixed with a plus-sign in the table of contents can be expanded to show subelements such as sections, subsections, etc.

The *Date* on-line delivery tool creates an interactive, outline-style, table of contents from the hierarchical SGML markup that is then used for rapid scrolling of the full-text window contents. Any document element may be defined to appear in the table of contents (or derivatives, e.g., lists of figures). Table of contents entries are made expandable/collapsible (just like nodes in a view) for easy navigation through a document. This is illustrated in figure 4.

3.2.1 Style Sheet Formatting

As SGML separates the format from the content of a document, the screen presentation is governed by (separately stored) style sheets. The style sheets define context-sensitive formatting and are easily redesigned interactively. Differently formatted versions of a document may be open simultaneously.

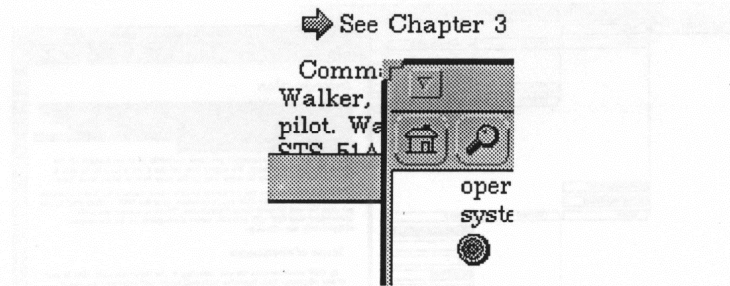


Figure 5: The arrow at the top left is a markup-based cross-reference displayed as a hypertext link. The referred element will have a 'target' icon, such as the one shown at the lower right. Links can be navigated bidirectionally.

3.2.2 Markup-based Linking

The *Dare* presentation tool resolves, from the SGML mark-up, intra-document cross-references⁶ into live hypertext links. The targets of mark-up based links are indicated, so that a referred element is also linked to the elements that refer to it (see figure 5).

The SGML structure is displayed on demand (in a separate window, as shown in figure 6), in what is usually called a *tree*. This provides another way of maneuvering through a text, as well as a way of relating the contents of the full-text window to its hierarchical SGML context.

3.2.3 Markup-based Indexing

Section 3.1 describes how bibliographical data is extracted from the markup. In the same spirit, one can specify what element contents are to be indexed, what the corresponding indexing field should be referred to as, and whether the indexing should be case sensitive. One example would be indexing the contents of, e.g., <AU> tags to an index referred to as 'Author', but only when that tag is found in the front matter. Using the full-text index capability of *Dare* one can look up all documents containing a particular word.

Because the viewer and the database cooperate, one can use the database indices from within the viewer: Any selected word in the text can be looked up *as if it were* an author name, a title word, a keyword, etc. Search parameters are stored so that one can easily redo a search.

3.2.4 String Searching

Naturally, one can also do string searches of document contents. The result of such searches are displayed in a pick list, as illustrated in figure 7. When one clicks in

⁶Encoded with the ID, IDREF and IDREFS attribute mechanism; typically references to elements such as footnotes, tables etc.

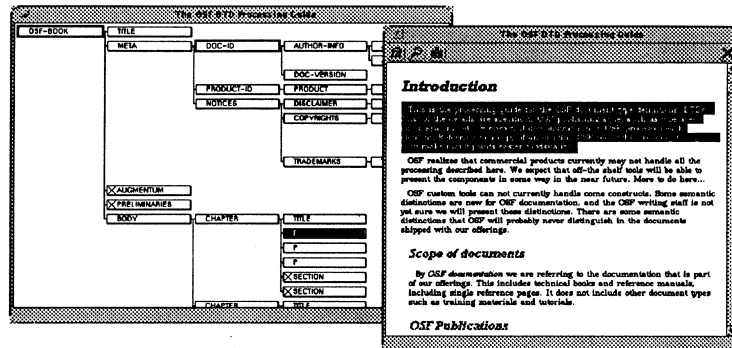


Figure 6: The SGML structure of a document can also be shown as a tree, or hierarchy of nodes, to be used both for navigation and for style sheet access. The marked section in the full text view to the right corresponds to the marked element in the SGML tree. Node elements of the tree can be expanded and collapsed just as the nodes in views (see figure 1).

the pick list, the corresponding line will be scrolled to the top of the full text view. All matching occurrences are shown highlighted. Just as for index-based searches (in the previous section) string search parameters are kept in a list so that one can swiftly redo a search.

3.2.5 Annotation and Layered Hypertext Support

In addition to the links automatically derived from the mark-up, documents can be annotated and linked by hypertext links which are external to the document markup. These user-created links can span documents and are decoupled from restrictions pertaining to SGML element boundaries.

A special form of annotation is *highlighting*, where blocks are highlighted as if by a marker pen.

Webs

Annotations and links are connected to blocks of contiguous selections or *anchors*. This additional data⁷ is stored separately in what is usually called *webs*.

Branching Bidirectional Links

A link can fork to a multi-way branch, leading to a choice of several destination anchors. Just as the markup-based links, the dynamically created links can be followed from either direction.

⁷I.e. anchor, link, and annotation information.

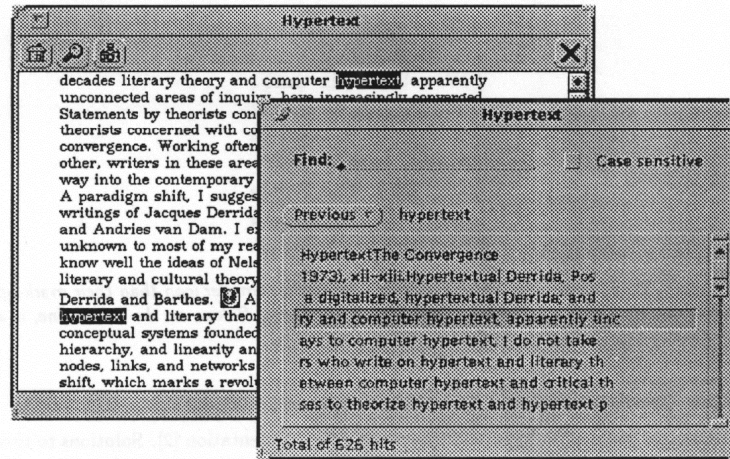


Figure 7: String searches result in a pick list, where the matched string is displayed concordance-style with its surrounding words (here the word 'hypertext' has been sought). The button marked Previous is a list of search strings which have been entered during the session. Notice also the highlighted icon in the full text view; this is a footnote icon, and it has been highlighted because the string being sought occurs in the footnote contents (as can be deduced from the pick list).

The linking functionality outlined above was pioneered by Brown University's Institute for Research in Information and Scholarship (IRIS) in the *Intermedia* system [6]. We have expanded this functionality in three ways:

- **Unmodal webs:** The externally stored webs can be opened and closed at any time while reading a document. The on-line display will adapt itself accordingly.
- **Concurrent webs:** Several webs, i.e. collections of annotations and/or links, can be open at the same time. A useful metaphor is to think of each open web as a transparency layer upon which the links and annotations are attached; the document is displayed as if it were seen through these layers of transparencies. The anchors to which annotations and links are connected can reside in different webs.
- **Enhanced user feedback:** The user is informed of which documents are contained in a web, as well as the reverse ("in which webs are there links to this particular document?").

ie File menu ↗ in e:
 s a New option, in ac
 put in the tools that c
 ated by the ↗ fact th
 les that have been ge
 nvoled, before their
 gure ↗). The Repla
 apter ↗ both need e
 l and the Manning S

Figure 8: Web-based links are shown with a somewhat smaller icon than their mark-up based counterparts, see the link at the top of the figure. Below, on the fourth line, is a multi-way link (indicated by double arrows).

User Disorientation and The Web Manager

A classical problem in hypertext systems is user disorientation [2]. Solutions to this problem have usually taken the form of maps [5], sometimes in conjunction with a history list or *path* [6]. Brown [1] explores the problem of providing *hierarchical* and *cross-referencing* links (or equivalently, *structural* and *unstructured* links) with respect to the *Guide* system.

In *Guide*, hierarchical links are used to encapsulate document sections that are expanded at will. In contrast, *DEAR* uses the natural document structure inherent in the markup, allowing the user to manipulate this structure by interacting with the table of contents or the graphical SGML tree. The cross-referencing links have three variants:

- **Markup-based cross-references.** The presentation tool resolves intra-document links from the SGML markup as explained in section 3.2.2.
- **Targets of string searches.** Navigation as a result of string searches is also a form of unstructured linking. *DEAR* uses a pick list (see section 3.2.4 and figure 7) as the starting point for such navigation. All links remain available when one follows such a link, and previous search strings are kept during the session.
- **Web-based cross-references.** This is the most powerful (and potentially most disorienting) link mechanism, as it allows unrestricted linking among all available documents. The *Web Manager* is the user interface to these links: All links and annotations⁸ are accessible in a list, and any link endpoint anchor contents can be previewed without effectuating the jump.

⁸Optionally filtered, e.g. so that only annotations or links of one's own are displayed.

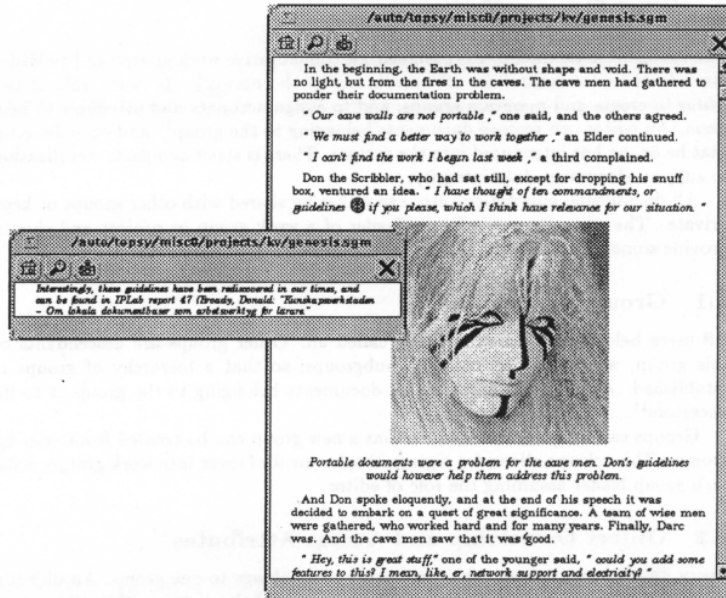


Figure 9: Any document element may be displayed either in-line or in a separate window. The latter display is usually preferred for floating elements such as tables, figures and footnotes. In this figure, a color picture is displayed in-line while a footnote, indicated by an icon (to the left and above in the picture), has its contents displayed in the external top window.

3.3 Inline Display of Elements

Any element content can be shown in-line or displayed—on demand—in separate windows⁹. In the latter case, the elements are indicated with a clickable icon (see figure 9). This behavior is of course tailorable through the mechanism of style sheets.

3.3.1 Graphics Support

The presentation tool has built-in support for common graphic formats, currently Sun raster, bmp, GIF, JPEG/JFIF, pbm/pgm/ppm, pcx, and xbm.

⁹Elements typically presented this way are footnotes and tables.

4 The Group Editor

The **DARE** database manager is designed for collaborative work groups and provides multi-level access control for all objects across the network. It is the role of the *editor* to create and maintain groups, and to assign accounts and privileges to new users. An editor can remove documents belonging to the group¹⁰ and exclude users that he or she has introduced into the system. There is strict compartmentalization to support document data integrity.

All documents, annotations etc. can either be shared with other groups or kept private. The editor is normally the leader of a work group or project and should provide some *quality control* for documents belonging to the group.

4.1 Group Hierarchies

All users belong to a common group called *all*. Other groups are descendants of this group, which in turn can have subgroups, so that a hierarchy of groups is established: A group has access to all documents belonging to the group or to its ancestors¹¹.

Groups can also be combined, so that a new group can be created from existing groups. This scheme allows for a gradual refinement of users into work groups, with each group leader assuming the role of editor.

4.2 Object Ownership and Access Attributes

Every **DARE** object has precisely one owner and belongs to one group. An object's access is controlled by setting two *attributes* to one of three states. Only the owner can modify the attributes, which are *Read Access* and *Modify Access*. The attributes can be set to:

Public A public object is available to all users. A *public read access* allows everyone access to the object. If *modify access* is set to public, anyone can edit the object, e.g. a view or a web¹².

Protected A protected object is restricted to members of the same (or a descendant) group.

Private A private object is restricted to its owner.

An owner can also give away ownership.

5 The System Administrator

The *system administrator* is allowed unrestricted access to all database objects, regardless of their access status, and is the initial user who assigns accounts to editors (and assists them if they forget their passwords).

¹⁰A privilege shared with the owner of the document.

¹¹One could say that the groups inherit previous access rights but deny access from their parent groups.

¹²*Modifiable* with respect to SGML documents implies only that these can be removed from the database. There is not yet an SGML-validating editor built into the system.

Summing Up Key Concepts

The **Dart** system combines methodologies from the fields of information retrieval, hypertext, and CSCW, to create a coherent framework for SGML-based document management.

- SGML is used to maximum advantage to simplify document processing in e.g. formatting, filing and indexing. Documents in SGML can be examined directly without preprocessing.
- The information structure conveyed by the somewhat static hierarchical ordering is complemented by the powerful *layered web* mechanism of the presentation tool. Advanced hypertext facilities allow any continuous selection to be made into an anchor, to be highlighted, annotated, or linked to.
- Network-based *group support* is built into the system, distributing the responsibility of user maintenance without sacrificing security considerations or data integrity. Any database object such as a document, view, style sheet, or web, can be shared.
- *Customization* is encouraged in every aspect such as style sheet design, creation of personal views, live hypertext linking, annotation, and highlighting.
- The system emphasizes the use of *hierarchical organization*, as it is an ordering concept which is powerful yet familiar. It is employed in creating the hierarchical categories of users (the five levels), and again, in the user group construct. It is also used in views to organize documents. Looking at the document element level (corresponding to the SGML markup), hierarchical structure is inherent in the table of contents, and (of course), in the graphical tree representation of SGML-encoded documents.

Acknowledgements

The authors thank Rand Waltzman for perceptive criticism of this report.

Copyright Information

© 1993 by the authors. Reproduction and copying without fee, in any form or by any means, is permitted provided that the copies are not made or distributed for direct commercial advantage and credit to the source is given. Abstracting with credit is permitted.

References

- [1] Brown, Peter: *Linking and Searching Within Hypertext*, Electronic Publishing, Vol. 1(1), April 1988.
- [2] Conklin, Jeff: *Hypertext: An Introduction and Survey*, IEEE Computer, Vol. 20, No. 9, September 1987.
- [3] Electronic Book Technologies: *DynaText System, Reader Guide*, Providence, RI, 1991.
- [4] Engelbart, Douglas C. & English, William K.: *A research center for augmenting human intellect*, Proceedings of the 1968 Fall Joint Computer Conference, Montvale, N.J., 1968.
- [5] Envos Corporation, *NoteCards User's Guide*, Release 1.1M, Mountain View, CA, 1989.
- [6] Institute for Research in Information and Scholarship, Brown University, *User's Guide: IRIS Intermedia*, Release 3.0, Providence, RI, 1990.
- [7] International Organization for Standardization: *Information Processing—Text and Office Systems—Standard Generalized Markup Language (SGML)*, ISO 8879-1986 (E), 1986.
- [8] Pirinen, Joakim: *Socker-Conny*, Tago Förlag, 1985.
- [9] Raymond, Darrell R.: *Flexible Text Display with Lector*, IEEE Computer, August 1992.

<i>CONTENTS</i>	15
-----------------	----

Contents

1 The Guest	2
1.1 Document Databases	3
2 The Reader	3
2.1 A Virtual Filing Tool	3
2.1.1 Labels	3
2.1.2 Child Indicator	3
2.1.3 Document Interchange	4
2.2 On-line Browsing, Annotating, and Linking	4
3 The Author	5
3.1 Adaptable SGML-based Filing	5
3.2 An SGML On-line Delivery Tool	6
3.2.1 Style Sheet Formatting	6
3.2.2 Markup-based Linking	7
3.2.3 Markup-based Indexing	7
3.2.4 String Searching	7
3.2.5 Annotation and Layered Hypertext Support	8
3.3 Inline Display of Elements	11
3.3.1 Graphics Support	11
4 The Group Editor	12
4.1 Group Hierarchies	12
4.2 Object Ownership and Access Attributes	12
5 The System Administrator	12

PS till bilaga 3: Något om hur principerna realiserats

D. Broady, H. Haitto 1993-02-15

De krav som nämns i projektprogrammets tredje avsnitt (punkterna 1-10, s. 13-17 ovan) har fungerat som riktlinjer vid konstruktionen av det dokumenthanteringsystem Darc som beskrives i bilaga 3. Här följer för var och en av punkterna några exempel på de lösningar vi stannat för.

1. Systemutvecklarna bör göra så få antaganden som möjligt om hur dokumenten kommer att användas.

Genom att understödja den internationella standarden SGML kan Darc ur samma dokument skapa en version (vars utseende helt styrs av användaren) som är lämplig för läsning på skärm, vars material kan sökas, skrivas ut, samt inlemmas i en databas som sedan exporteras.

2. Dokument skall vara flyttbara.

Genom användningen av SGML erhålles dokument som är maskinoberoende. Själva tillämpningens databasfiler är binärkompatibla (dvs helt flyttbara) mellan de plattformar som understöds (f.n. PC och Sun, ev. kan Macintosh tillkomma).

3. Systemet bör stödja samarbete. Flera medarbetare skall kunna använda samma eller varandra överlappande lokala dokumentbaser.

Fleranvändarstöd har byggts in i Darc på ett sätt som uppmuntrar till samarbete, genom att var och en kan skraddarsy egna utsnitt (vyer) för eget behov, men även dela dessa med andra. Den hierarkiska indelningen av användare i kategorierna gäst, läsare, författare, redaktör samt databasadministratör, uppfyller behörighetskontrollkrav utan att göra avkall på funktionaliteten. Den hierarkiska gruppindelningen är även ett smidigt sätt att skapa samsarbetsgrupper kring gemensamma dokumentbaser. Att förlägga ansvaret för gruppadministration och grupper på redaktörerna och inte på någon systemansvarig tekniker är ett led i strävan att förlägga kontrollen där den hör hemma — hos lärare, läromedelförfattare, redaktörer, dokumentalister etc.

4. Systemet bör stödja arbete med — inte enbart presentation av — dokumenten.

Systemet tillåter användning av vävar av hypertextlänkar för att infoga individuella anteckningar och korshänvisningar inom och mellan dokument.

5. Systemet skall hjälpa användaren att revidera, ordna och överblicka materialet i den lokala dokumentbasen.

Databastjänsterna ger förutom sedvanliga söktjänster på valfria indextermer möjligheten till fulltextsökning och tillhandahåller virtuella vyer, som gör att utsnitt i databasen kan utformas ungefär på samma sätt som när man arbetar med material på en hårdskiva (men utan att man egentligen flyttar filerna). Varje användare kan skapa sina individuella vyer över innehållet i en och samma dokumentbas.

6. Systemet bör stödja hanteringen av stora och komplexa dokument, samt ordnade samlingar av dokument.

I och med SGML-märkningen klargöres dokumentens struktur . Systemet hanterar med lätthet dokument som motsvarar flera hundra sidor. Beträffande det totala antalet dokument i dokumentbaserna är de fysiska gränserna satta mycket högt, flera tiotusental dokument erbjuder inga problem.

7. En och samma dokumentsamling skall kunna användas på olika nivåer.

Hypertextlänkarna och möjligheten att annotera utan att påverka originalets integritet ger vida ramar för sådan bearbetning.

8. Systemet bör vara modifierbart och utbyggbart.

Darc är modulariserad, så att man enkelt skall kunna ersätta valda delar med annan mjukvara. Mycket är inställningsbart av användaren.

9. Enkel utrustning.

Darc finns till såväl arbetsstationer (Sun) som till enkla PC-maskiner utan att göra avkall på funktionaliteten. Konverteringsrutiner har utvecklats till viss del för att prova integrationen med sedvanliga ordbehandlare. Programtillverkarna har dock redan inlett en anpassning av sina produkter till SGML; Word Perfect finns redan i en version med SGML-stöd.

10. Systemet måste kunna hantera strukturerade dokument.

SGML-standarden är gjord för sådana dokument. Darc utnyttjar SGML-kodningen i hela dokumenthanteringen: för att plocka ut bibliografiska data, för att skapa index, samt för navigation vid läsning på skärm, etc

1993-02-15

Till Skolverket

Ansökan om utvecklingsmedel för projektet "Det nya handbiblioteket"

Härmed ansökes om 580 000 kr för vart och ett av de tre budgetåren 1993/94, 1994/95 och 1995/96. Denna summa utgör hälften av de nedan angivna budgeterade kostnaderna för projektet.

Den andra hälften av kostnaderna sökes från Forskningsrådsnämnden (området området Biblioteks- och informationsvetenskaplig forskning, handläggare Margareta Larsson, tel 08-610 06 30).

Huvudansvarig sökande:

Donald Broady, docent, verksam på två arbetsplatser: som tf professor vid HLS, Institutionen för pedagogik, samt som forskare vid KTH, Institutionen för numerisk analys och datalogi.

Medsökande:

Lars Dahlbäck, docent, verksam vid Litteraturvetenskapliga institutionen, Stockholms universitet och huvudredaktör för Nationalupplagan av August Strindbergs samlade verk.

Eva Trotzig, fil. lic, chefsbibliotekarie för Statens psykologisk-pedagogiska bibliotek.

Hasse Haitto., civ.ing., doktorand vid KTH, Institutionen för numerisk analys och datalogi och forskningsassistent vid HLS, Institutionen för pedagogik.

Adress till projektet: KTH, NADA, 100 24 Stockholm

Tel: 08-737 56 02 (HLS), 08-790 62 99 (KTH), 08-767 71 27 (D.Broady, hem)

E--mail broady@nada.kth.se

Fax 08-737 56 63 (HLS), 08-790 09 30 (KTH).

Tidsram: Projektet beräknas starta 1993-07-01 och vara slutfört 1996-06-30

Projektitel: Det nya handbiblioteket

Sammanfattande projektbeskrivning:

Projektets syfte är att utveckla och pröva redskap och metoder som tillåter användare att bygga upp, ordna och använda ett eget "handbibliotek", bestående av maskinläsbara texter som följer den internationella märkningsstandarden SGML (Standard Generalized Markup Language, ISO 8879).

Avsikten är att användarna skall kunna överblicka, organisera och genomsöka handbiblioteket, tillfoga egna kommentarer och korshänvisningar, välja ut material för framställning av t.ex. skraddarsydda läromedel, färdigställa texter för utskrift eller fotosättning, samarbeta kring en gemensam uppsättning dokument, utbyta dokument eller hela dokumentsamlingar, nyttja behörighetskontroll som i möjligaste mån skyddar textens integritet.

Särskilt intresse ägnas den humanistiska och samhällsvetenskapliga undervisningens och forskningens behov, samt bibliotekens möjligheter att förmedla maskinläsbar litteratur.

Som en fallstudie prövas metoder för att framställa en SGML-märkt maskinläsbar version av Nationalupplagan av August Strindbergs samlade verk i enlighet med de riktlinjer som utarbetats av TEI (Text Encoding Initiative).

Projektet är förlagt till Institutionen för numerisk analys och datalogi vid KTH, och bedrivs i samarbete med Litteraturvetenskapliga institutionen vid Stockholms universitet och Institutionen för pedagogik vid Högskolan för lärarutbildning i Stockholm.

(För en utförlig beskrivning, se bifogat projektprogram)

Tidigare beviljade bidrag: Det här föreslagna projektet är ett nytt projekt, som dock kommer att bygga vidare på vunna erfarenheter och utnyttja den system- och programutveckling som åstadkommit under snart tre års arbete inom projektet "Datorstöd för kunskapsarbete", som påbörjades 1 juli 1990 och skall avslutas 30 juni 1993.

Skolverket har under bå 1991/92 och 1992/93 (400 kkr för vardera året, sammantaget 800 kkr) stött ett projekt "Kunskapsverkstaden", som varit ett delprojekt inom projektet "Datorstöd för kunskapsarbete", vilket även erhållit stöd från NUTEK (ITYP 90-02737P) och Forskningsrådsnämnden (Dnr 910176:1). Merparten av stödet har kommit från NUTEK. Tack vare Skolverkets medverkan har arbetet kunnat inriktas mot skolväsendets behov. Vi räknar inte med att Skolverket skall ta definitiv ställning till föreliggande nya projektförslag innan vi inlämnat en redogörelse för arbetet inom det tidigare projektet, och vi kommer att sammanställa en sådan snarast möjligt.

Personal, lönekostnader:

Namn, titel	lön febr 1993	procent av heltid	lönekostnad 1993/94 kkkr	lönekostnad 1994/95 kkkr	lönekostnad 1995/96 kkkr
Donald Broady, doc., tf prof vid HLS Institutionen för pedagogik och forskare vid KTH, Institutionen för numerisk analys och datalogi	23.000	50	138	138	138
Lars Dahlbäck, doc, Litteraturvetenskapliga institutionen, SU. Huvudredaktör för Nationalupplagan av August Strindbergs samlade verk.	(arbete i tjänsten)		—	—	—
Hasse Haitto, civ.ing., doktorand vid KTH, Institutionen för numerisk analys och datalogi och forskningsassistent vid HLS, Institutionen för pedagogik.	18.600	50	112	112	112
Eva Trotzig, fil. lic., chefsbibliotekarie för Statens psykologisk-pedagogiska bibliotek	(arbete i tjänsten)		—	—	—
Forskningsass (som arbetar med märkning av Strindbergmaterialet, trol. doktorand vid Littvet. inst, SU)	ca 16.000	50	96	96	96
Programmerare, dokumentalist, försöksledare, timanställda (Som programmare önskar vi anlita Peter Lidbaum och Magnus Tobiasson, civilingenjörer med datalogisk inriktning. Som dokumentalist ev. Krister Lagerborg, informatiker vid SPPB. Som försöksledare sannolikt lärare vid HLS.)	ca 18.000	motsv 100%	216	216	216
Lönekostnadspålägg 42,5%			239	239	239
Summa lönekostnader			801	801	801

Övriga kostnader

	1993/94 kkkr	1994/95 kkkr	1995/96 kkkr
Kostnader för dataöverföring, inköp av textmaterial, litteratur, förbrukningsmaterial	20	20	20
Tryckkostn.			30
Resor Bl.a. resor till de viktigare relevanta internationella konferenser i Europa och USA	30	30	30
Apparatur Komplettering av datorutrustning och programbibliotek vid NADA, KTH och Inst. f. ped. HLS, samt inköp av dator och programvara till littvet. inst, SU	70	70	40
Förvaltningsavgift 12% av totalkostnaden	131	131	131
Del i institutionens gemensamma drift 10 % (sekreterare, intendent, telefon, kontorsmaterial, visst datorstöd)	105	105	105
Totalsumma	1157	1157	1157

Beräknad finansiering av projektet

För projektet "Det nya handbiblioteket" föreslås en samfinansiering. Hälften av medlen (för varje budgetår 580 kkr) sökes från Skolverket (utvecklingsenheten) och lika mycket från Forskningsrådsnämnden.

Dessutom bör nämnas, att jag (med Lars Dahlbäck som medsökande) till HSFR den 15 januari 1992 inlämnat en ansökan om ett tvåårigt projekt, "Hantering av maskinläsbar humanistisk och samhällsvetenskaplig litteratur", som anknyter till det här föreslagna. Dock är det sist nämnda projektet "Hantering..." i högre grad inriktat mot de specifika problemen kring SGML-märkningen av Nationalupplagan av August Strindbergs samlade verk, medan projektet "Det nya handbiblioteket" är mer inriktat mot utveckling och utprovning av system för yrkesgrupper som lärare, läromedelsförfattare, bibliotekarier, dokumentalister etc.

Härmed intygas att Institutionen för numerisk analys och datalogi, KTH, har möjlighet att ställa lokaler och andra nödvändiga resurser till förfogande för projektet "Det nya handbiblioteket"

.....INGRID MELINDER)

Prefekt, Institutionen för numerisk analys och datalogi, KTH

Stockholm den 15 februari 1993

Huvudansvarig sökande.....DONALD BROADY

Bilaga: Projektprogram

Stockholm den 15 februari 1993

Till HSFR

Tillägg till ansökan om medel för projektet
"Hantering av maskinläsbar humanistisk och samhällsvetenskaplig litteratur"

Som nämndes på den ansökningsblankett för ovan nämnda projekt som inlämnades till HSFR 1993-01-15, avsåg vi att föreslå FRN ett näraliggande projekt.

Vi lämnar således i dag en ansökan till FRN (området Biblioteks- och informationsvetenskaplig forskning, handläggare Margareta Larsson, tel 08-610 06 30) om ett projekt rubricerat "Det nya handbiblioteket". Där föreslås en samfinansiering mellan FRN och Skolverket. Samma projektplan lämnas samtidigt till FRN och Skolverket (där ansökan sannolikt tas om hand av utvecklingsenheten).

Dessa två projekt — "Hantering av maskinläsbar humanistisk och samhällsvetenskaplig litteratur", som föreslås HSFR, och "Det nya handbiblioteket", som föreslås FRN och Skolverket — är nära besläktade i fråga om själva projektidén. Skillnaden gäller det tyngdpunkten i det arbete som skall genomföras. Inom det projekt som föreslås HSFR kommer arbetet i högre grad att gälla de specifika problemen kring SGML-märkningen av Nationalupplagan av August Strindbergs samlade verk och annan humanistisk och samhällsvetenskaplig litteratur, medan det projekt som föreslås Skolverket och FRN är mer inriktat mot systemutveckling, närmare bestämt utveckling av arbetsverktyg för yrkesgrupper som lärare, läromedelsförfattare, bibliotekarier, dokumentalister.

Dock överlappar de två projekten varandra, vilket innebär att det projekt som föreslås HSFR skulle kunna bedrivas med framgång även om HSFRs forskningsbidrag blir av mindre omfattning än vad vi föreslog i ansökan av den 15 januari.

Donald Broady

Institutionen för numerisk analys och datalogi, KTH, 100 44 Stockholm

Tel: 08-737 56 02 (HLS), 08-790 62 99 (KTH), 08-767 71 27 (hem)

E-mail broady@nada.kth.se

Fax 08-737 56 63 (HLS), 08-790 09 30 (KTH).