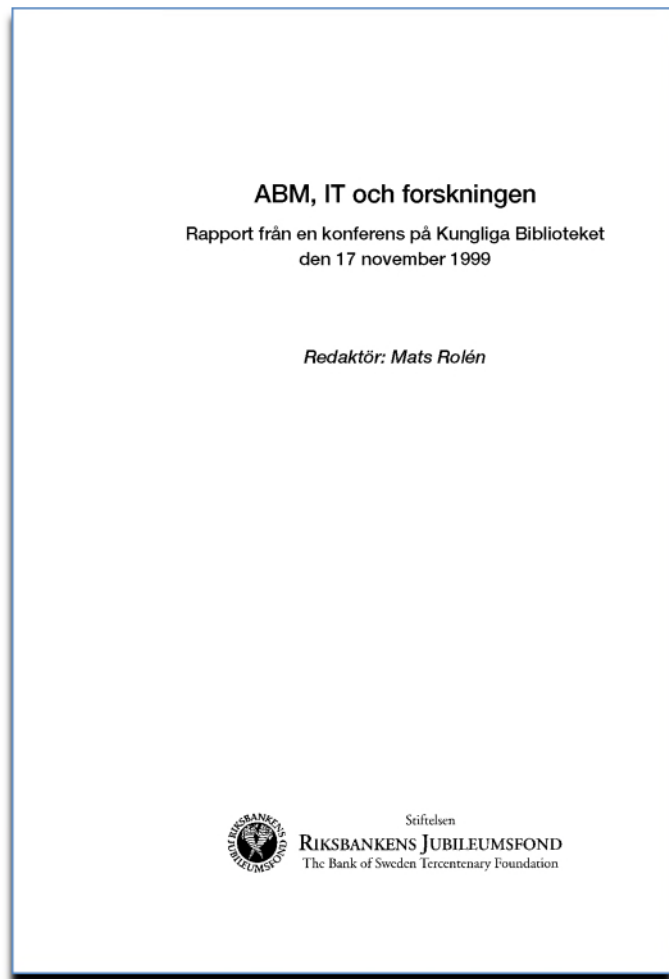


FAKSIMIL

Donald Broady, "Hur kodar man Röda Rummet?", pp. 48-61 i *ABM, IT och forskningen. Rapport från en konferens på Kungliga Biblioteket den 17 november 1999*. Red. Mats Rolén. Stockholm: Riksbankens Jubileumsfond, 2000.



Hur kodar man Röda Rummet?

En fungerande infrastruktur förutsätter anpassning till relevanta standarder såsom SGML (Standard Generalized Markup Language) och XML (Extensible Markup Language). Inom olika områden växer överenskommelser fram om hur standarderna skall utnyttjas. Ett exempel är TEI (Text Encoding Initiative), som utarbetar riktlinjer för representation av litterärt och historiskt källmaterial.

Fyra slags digital litteratur på nätet

Frågan om tillgänglighet till arkivens, bibliotekens och museernas samlingar är på väg att ändra karaktär. Det handlar inte längre enbart om att slå upp dörrarna, låna ut böckerna, exponera föremålen, inbjuda till besök på webbplatser. Att tillgängliggöra materialet innebär i dag, och än mer i framtiden, även att erbjuda människor möjlighet att hemföra digitalt material – transkriberad litteratur, arkivförteckningar, avbildningar och annat – och att göra något *eget* därav. Det kan gälla lärare som framställer skraddarsydda läromedel eller föreningar som bygger upp ett eget referensbibliotek inom sitt intresseområde.

Informationstekniken har givit somliga brukare möjlighet att själva bygga upp sina samlingar av digital information. Materialet måste därför tillhandahållas i format som tillåter många slags användningar, däribland användningar som vi inte kan förutse.

Därmed får standardiseringsarbetet nya syften. Det handlar inte längre bara om överenskommelser som tillåter arkiv, bibliotek och museer att utbyta information sinsemellan och underlätta för bru-

karna att hitta det de behöver. Därutöver syftar standardiseringsarbetet till att utforma det digitala materialet så att det låter sig flyttas till brukarna.

Dit har vi ännu inte nått. I dag levereras material oftast färdigförpackat och fungerar dåligt när det flyttas ut ur förpackningen. En utställning av museiföremål på webben eller en utgåva av något författarskap på CD-ROM kan fungera väl så länge brukaren uppehåller sig på webbplatsen eller låter CD-ROM-skivan snurra i datorn. Men så snart någon försöker införliva materialet med sitt eget handbibliotek, går mycket av den viktiga informationen förlorad: strukturmärkning och innehållsmärkning, index, länkinformation, metadata, upplysningar om ursprung, senaste revisionsdatum, innehållsdeklarationer, kvalitetsstämplar etc.

Ett inspirerande exempel på flyttbar information är aldinerna, böckerna som skapades av den venetianske boktryckaren Aldus Manutius kring århundradeskiftet 1500. De var tryckta med lättlästa och vackra typsnitt. De var billiga, studenter hade råd med dem. En orsak till att de är så sällsynta i dag är att de var bruksföremål som slets ut. Upplagorna var ovanligt stora, kanske tusen exemplar.

Aldus anlidade lärda korrekturläsare. Hans böcker betraktades som pålitliga standardutgåvor. Över hela den lärda världen kunde alla identifiera ett citat hos någon latinsk författare med ledning av en hänvisning till en viss sida i en viss aldin. Och böckerna var flyttbara. De trycktes i oktavformat som gick att stoppa i fickan i stället för de gängse kvarto- eller folioformaten.

Så fungerar digital litteratur sällan i dag. Pålitliga vetenskapliga utgåvor är föga spridda och ibland mycket dyra. Flertalet gratisutgåvor är av usel kvalitet. Merparten är inte särskilt flyttbara.

Det finns fyra slag av utgåvor på nätet.

För det första finns lågbudgetutgåvor utan anspråk på att vara etablerade på anständigt vis. De levererades tidigare som regel i ASCII-format, i dag oftast i HTML-format.

För det andra finns utgåvor för forskningsbruk, vilka oftast endast är tillgängliga i färdigpacketerat icke flyttbart skick på en CD-ROM eller webbplats eller på en server hos något lärosäte eller någon forskningsinstitution.

För det tredje finns leveransformat som presenterar utseendet

hos en boksida (oftast återgiven med hjälp av PDF) eller avfotograferade tryckta sidor eller handskrifter.

Och för det fjärde finns utgåvor som bygger på SGML, den viktigaste standarden för att i dag åstadkomma flyttbar information.

SGML (Standard Generalized Markup Language, ISO 8879:1986) är ett metaspråk för beskrivning av informationens innehåll och struktur. SGML är föregångaren till dagens välbekanta märkspråk HTML och XML som används för publicering på webben.

HTML är ett enklare språk som främst används för att styra typografin på skärmen eller pappersutskriften. SGML och XML tillåter märkning som passar bättre för exempelvis vetenskapliga utgåvor. De sist nämnda standarderna används ofta för att märka text men man kan också märka annat, dvs beskriva sådant som bilder, ljud, länkar eller programkod.

SGML och XML erbjuder fördelen att märkningen kan beledsaga dokumenten när de sprids. Tidigare har man tvingats arbeta med externa nyckelord, registerfiler eller index, vilka varit knutna till någon databas eller webbplats men skalats bort så snart dokumentet flyttas för att användas i något annat sammanhang.

Märkning

Märkning tillgår så att *märkord*, även kallade *taggar* (av engelska *tags*) fogas in i dokumentet. Märkorden brukar avgränsas av hakparenteser och innesluter delar av texten som kallas *element*. Exempelvis:

```
<p>Det var en afton i <tid>början av maj</tid>. Den lilla trädgården på <plats>Mosebacke</plats> hade ännu inte blivit öppnad för allmänheten.
```

Vi har här avgränsat tidsangivelsen "början av maj" som ett element genom att stoppa in märkordet `tid` i början och märkordet `/tid` efteråt. På samma sätt anges att Mosebacke är ett element som anger en plats. Element är alltså byggstenar. Även hela stycket är ett element, som inleds med märkordet `p` vilket utläses paragraph. När stycket är slut stoppar man in `</p>`.

Rent praktiskt går märkningen till så att märkord och attribut skrivs in i textfilen med Strindbergs *Röda rummet*, kanske i vanlig ordbehandlare eller enklare med speciell SGML-editor som bland annat tillhandahåller menyer för de märkord som är disponibla på det ställe där markören befinner sig. Vissa otillåtna märkord blir därmed otillgängliga. Man får inte placera en fotnot inuti en fotnot eller rubriknivå 2 omedelbart före rubriknivå 1.

Elementen kan även förses med attribut. Att en plats befinner sig i Stockholm kan anges med `<plats lokaliserad="Sthlm">`.

SGML-standarderna i sig är ett metaspråk, en syntax för hur konkreta märkspråk kan utformas. Standarderna innebär alltså inte i sig något påbud om användningen av just märkorden `tid`, `rum` eller `p` eller ens utseendet på parenteserna. De här märkorden kanske väljs av en litteraturvetare som vill överblicka rörelserna i Stockholm som geografiskt rum eller tidsplanen. *Röda rummet* utspelas under ett år, från maj till maj, med några kapitel i slutet om vad som hände sedan.

Inom olika vetenskapliga discipliner, liksom inom olika industribranscher och andra verksamhetsområden, har man kommit överens om egna uppsättningar märkord och attribut och deras relationer till varandra. Dessa överenskommelser finns formaliserade i regelsamlingar som kallas DTDer, vilket utläses dokumenttypdefinitioner. Skilda slag av dokument behöver skilda DTDer, exempelvis

- TEI DTD (Text Encoding Initiative) för humanistisk litteratur
- EAD DTD (Encoded Archival Description) för arkivdata
- ISO 12083 för publicering på papper.

DTDer talar om vilka uppsättningar märkord och attribut som får användas och hur de förhåller sig till varandra. Reglerna kan således innebära sådant som att rubriknivå 2 inte får förekomma omedelbart före rubriknivå 1, att det skall finnas ett och högst ett element *Abstract* i en akademisk avhandling och att detta får förekomma endast inuti elementet *Front matter*.

Text Encoding Initiative

Text Encoding Initiative (TEI) är ett internationellt projekt som sedan 1987 engagerar humanister och dataloger i många länder i syfte att utarbeta riktlinjer för organisering, märkning och utbyte av all slags humanistisk litteratur: editioner av medeltida handskrifter, diktsamlingar, romaner, pjäser, språkkorpora, lexika etc.

Enligt min mening är TEIs riktlinjer en välsignelse. För den som önskar utnyttja sig därav existerar nu en internationell överenskommelse om hur humanister skall koda, arkivera, katalogisera, distribuera och hänvisa till digital litteratur tvärs över gränserna mellan nationalspråk, datorplattformar och tillämpningar. Övriga, som av någon anledning väljer andra lösningar, kan utnyttja TEIs riktlinjer som en gemensam referenspunkt.

TEIs förslag är avsedda att tjäna som inlägg i den bredare diskussionen om hur texters egenskaper skall representeras. Även den som för egen del föredrar andra sätt att representera redaktionella ingrepp, dikters rytm och meter eller lakuner i källor, kan hämta inspiration från TEIs detaljerade behandling av otaliga sådana märkningsproblem. Och till sist, även om man ändå beslutar sig för en mycket egen lösning, kan det tänkas att TEI erbjuder ett utbytesformat som gör materialet flyttbart så att det blir tillgängligt för omvärlden.

Med riktlinjerna vill TEI bidra till gemensamma internationella överenskommelser om hur all slags litteratur av intresse för humanister kan lagras och spridas i digital form. Av ett så generellt syfte följer att riktlinjerna inte får vara alltför styrande. En viktig princip är att riktlinjerna skall vara "konkurrensneutrala" i förhållande till olika skolbildningar och forskningsinriktningar. De får med andra ord inte favorisera exempelvis de forskare som vill lyfta fram författarens intentioner eller de forskare som vill studera historiska och sociala omständigheter eller dem som prioriterar tematisk eller stilistisk analys.

En annan princip är att det finns ytterst få tvingande påbud i TEIs riktlinjer, utöver det tvång som SGML-standarden i sig innebär. För att riktlinjernas krav skall vara tillfredsställda är det framför allt "meta-informationen" (upplysningar om dokumentets rubrik, upphovsman, redaktionellt ansvar, publiceringsdatum, re-

visionshistoria etc, dvs sådant som ingår i den *TEI header* som skall medfölja varje dokument) som måste ingå och märkas på föreskrivet sätt för att möjliggöra identifiering, automatisk katalogisering, kvalitetsbedömning etc. I övrigt har man stor frihet att i enlighet med egna behov utforma sitt märkningsschema. Man är dock tvungen att i dokumentet precisera principerna för det märkningsschema man har valt, och TEIs riktlinjer innebär ytterst just ett elaborerat förslag om hur märkningsprinciper kan fastställas så att andra (mänskliga läsare eller datorprogram) begriper dem.

TEIs riktlinjer består av en rikhaltig uppsättning (cirka 400) märkord och attribut samt principer för hur dessa kan utnyttjas för att beskriva olika egenskaper hos skilda slag av texter.

Egenskaperna kan vara *strukturella*, dvs beskriva textens uppbyggad i termer av titel, författare, rubriknivåer, blockcitat, fotnoter, redaktionella tillägg, scenanvisningar och dialog i en pjäs etc. Flertalet av de egenskaper som fångas upp av TEIs märkord är dock mer *innehållsliga*.

Med deras hjälp kan man ange om Strindberg skrev med bläck, blyerts eller rödpenna, man kan märka ut personnamn, geografiska lokaliteter eller datum, tematiska och stilistiska egenskaper hos texten, metriska egenskaper hos en dikt etc.

Om de märkord och attribut TEI föreslår inte duger, är det jämförelsevis enkelt att modifiera dem eller lägga till egna. Materialet är fortfarande *TEI conformant* förutsatt att man i eller i anslutning till själva dokumentet på föreskrivet sätt har angivit hur ändringarna genomförts. Och i synnerhet åtskilliga av attributen är redan i TEIs riktlinjer definierade så att de kan ges vilka värden som helst.

Det torde vara ändamålsenligt att arkiv, bibliotek och förlag levererar standardutgåvor med sparsam märkning. Sedan kan forskargrupper och andra tillfoga märkord för egna syften. Det är så riktlinjerna är avsedda att användas, inte som tvångströja utan som flexibelt ramverk att anpassas till de egna kraven.

Ett exempel på SGML-märkning

Tag följande text ur *Röda Rummet*, början av kapitel 7:

SJUNDE KAPITLET Jesu Efterföljelse.

Följande morgon väcktes han av städerskan som framlämnade ett brev vilket befanns vara av följande lydelse.

Timot. Kap. X, v. 27, 28, 29. Första Korint.
Kap VI, v. 3, 4, 5.

Dyre Br!

Vår H:s J. Kr. Nåd och Frid, Fadrens kärlek och D. H. A:s delaktighet etc. Amen!

Jag såg av Gråkappan i går afton att Du ämnar utgiva Försoningsfacklan. Sök mig i min verksamhet i morgon bittida före 9.

din återlöste
Nathanael Skåre

Nu förstod han Lundells gåtor, till en del! Han kände visserligen icke den store gudsmannen Skåre personligen och visste intet om Försoningsfacklan, men han var nyfiken och beslöt att hörsamma den närgångna kallelsen.

Detta textparti kan SGML-märkas enligt TEIs riktlinjer på följande sätt.

```
<div1 type="chapter" n='7'>
<pb n="81">
<head type="n">SJUNDE KAPITLET</head>
<head type="main">Jesu Efterföljelse.</head>
<p rend="noindent">Följande morgon väcktes han av städerskan
som framlämnade ett brev vilket befanns vara av följande lydelse.</p>
<div2 type="letter">
<opener rend="center">Timot. Kap. X, v. 27, 28, 29. Första Korint.
Kap VI, v. 3, 4, 5.</opener>
```



```

<salute>Dyre Br!</salute>
<p>Vår H:s J. Kr. Nåd och Frid, Fadrens kärlek och D. H. A:s
delaktighet etc. Amen!</p>
<p>Jag såg av Gråkappan i går afton att Du ämnar utgiva
Försoningsfacklan. Sök mig i min verksamhet i morgon bittida före
9.</p>
<close rend="right">
<salute>din återlöste</salute>
<signed>Nathanael Skåre</signed>
</close>
</div2>

<div2 type="section">
<p>Nu förstod han Lundells gåtor, till en del! Han kände
visserli&shy;gen icke den store gudsmannen Skåre personligen och
visste intet om Försoningsfacklan, men han var nyfiken och beslöt att
hörsamma den närgångna kallelsen.
</p>

```

Märkningen inte alls uttömmande eller ens helt korrekt. Bokstäverna å, ä och ö borde bytas ut mot standardiserade tecken så att texten kan flyttas mellan alla plattformar.

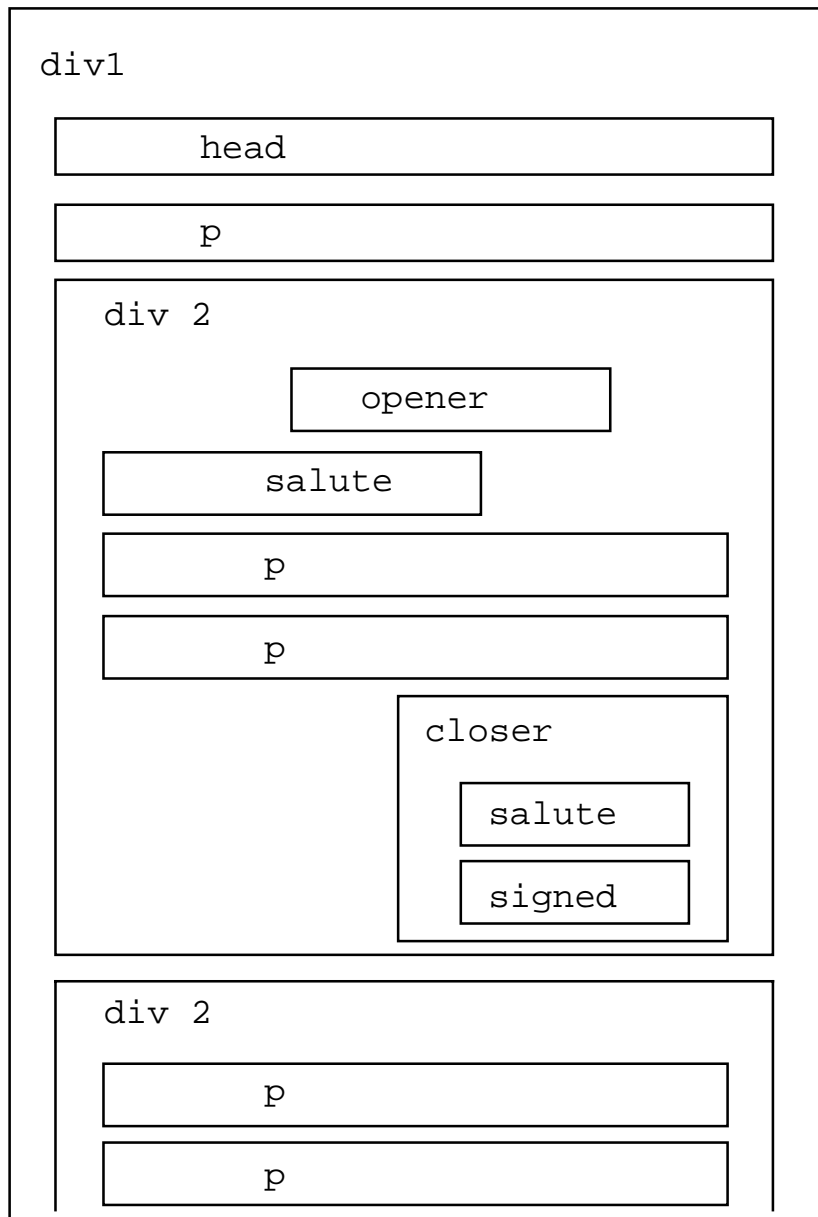
Låt oss först granska märkorden som beskriver textens strukturella eller formella egenskaper.

<div1 type="chapter" n='7'> anger att här börjar ett avsnitt. TEIs riktlinjer föreslår att märkorden `div0`, `div1`, `div2` etc används för att representera avsnitt på olika nivåer, exempelvis motsvarande en romans delar (`div0`), kapitel (`div1`) och sektioner inne i kapitel (`div2`) eller kanske diktsamlingar (`div0`) och däri ingående diktsviter (`div1`) osv. Avsnitten kan således vara inkapslade i varandra, och det är i linje med TEIs filosofi att inte på förhand föreskriva att de skall kallas kapitel, avdelning, diktsvit, essä, novell eller något annat. Här rör det sig om ett kapitel, varför attributet `type` till elementet `div1` tilldelas värdet `chapter`.

Märkordet <head> anger rubrik. Attributet `n` anger att kapitlet är numrerat. Attributet `main` anger att det är ett fråga om kapitlets egentliga rubrik. Inget av dessa attribut föreskrives av TEIs riktlinjer utan väljes på det sätt som bäst passar texten i fråga, men principerna måste redovisas i den TEI header som inleder dokumentet.

<p> och </p> omsluter ordinära stycken (engelska *paragraph*).

Med hjälp av dessa strukturella märkord har vi beskrivit text-avsnittets hierarkiska uppbyggnad som vi kan föreställa oss som lådor instoppade i varandra (jfr figuren nedan). Det största lådan (div1-nivån) representerar det sjunde kapitlet som i sin tur inne-



Figur. Röda Rummet, början av kapitel 7, som lådor i lådor.

håller dels rubrik, dels olika avsnitt på närmast lägre nivå (`div2`). Det första elementet på `div2`-nivån är ett brev innehållande ett antal märkord som identifierar hälsningsfraser; vissa av dessa element är inkapslade i varandra. Nästföljande element på nivån `div2` är ett mer ordinärt avsnitt som innehåller ett antal stycken.

Jag har även tagit med en entitet, `­`. Sådana känner man igen på att de alltid inleds med tecknet `&` och avslutas med tecknet `;` (semikolon). De står där för att bli utbytta mot något annat, t ex mot ett tecken i en systemberoende teckenuppsättning, innehållet i en teckensträng, eller en fil (t ex en liten bildfil).

Entiteten `­` uttydes *soft hyphen* och används här för att representera bindestreck som uppstått i samband med radbrytningen i den tryckta utgåvan av Samlade Verk. Det gäller att hålla reda på dessa, bland annat eftersom Strindbergs egna bindestreck på ett eller annat sätt måste särskiljas från andra bindestreck (vilket inte sker i Samlade Verk, men detta är lättare att åstadkomma i en digital version).

I en fullständig SGML-märkning måste även bokstäverna å, ä, ö bytas mot entiteter för att inte fördärvas vid överföring till system med andra teckenuppsättningar.

Attributet `rend` uttydes *rendition* och används bland annat för att vid behov styra typografin på skärmen eller på papper. Här förekommer attributet `rend` med värdet `center` som anger att bibelcitaten i brevets motto skall centreras.

Elementet `<pb n="81">` i början är ett s k tomt element, dvs det tjänar som platshållare och inrymmer ingen text. Förkortningen `pb` skall utläsas *page break*. Elementen beskriver sidbrytningen i den tryckta upplagan Samlade Verk. Vi tänker oss att läsare kommer att vilja arbeta med både pappersversionen och digital version.

Man kan gå vidare och märka språkliga aspekter eller mer innehållsliga aspekter, exempelvis framställningen av romanfigurerna. Där personer förekommer i romantexten skulle märkorden kunna förses med ID-attribut som låter läsaren spåra en och samma person trots att denne omtalas på olika sätt, med för- och efternamn, smek- eller öknamn, omskrivningar, eller enbart ett personligt pronomen: *Arvid Falk, den unge mannen, författaren, han*.

Vidare kan man märka ut platser, årtal och datumangivelser och så vidare i all oändlighet. Möjligheterna är obegränsade och kan

förstås inte uttömmas, när materialet redigeras för allmän spridning. I stället måste läsarna ges möjlighet att komplettera med egna märkord: somliga Strindbergforskare kan ha intresse av att tillföra märkning som möjliggör lingvistisk analys, andra önskar skapa sig ett underlag för tematisk eller stilistisk analys, åter andra vill identifiera anspelningar på personer eller händelser i Strindbergs samtid. Det är tänkbart att skilda forskningsprojekt med olika inriktning som ett led i sitt arbete åstadkommer en märkning för eget bruk som senare överlämnas som gåva till forskarsamhället.

HTML och SGML

SGML-märkningen ovan kan jämföras med en ordinär HTML-märkning:

```
<h2 align="center">SJUNDE KAPITLET</h2>
<h1 align="center">Jesu Efterföljelse.</h1>
<p>Följande morgon väcktes han av städerskan som framlämnade
ett brev vilket befanns vara av följande lydelse.</p>
<p align="center">Timot. Kap. X, v. 27, 28, 29. Första Korint.
Kap VI, v. 3, 4, 5.</p>
<p>Dyre Br!</p>
<p>Vår H:s J. Kr. Nåd och Frid, Fadrens kärlek och D. H. A:s
delaktighet etc. Amen!</p>
<p>Jag såg av Gråkappan i går afton att Du ämnar utgiva
Försoningsfacklan. Sök mig i min verksamhet i morgon bittida
före 9.</p>
<p align="right">din återlöste</p>
<p align="right">Nathael Skåre</p>
<p>Nu förstod han Lundells gåtor, till en del! Han kände visserligen
icke den store gudsmannen Skåre personligen och visste intet om
Försoningsfacklan, men han var nyfiken och beslöt att hörsamma den
närgångna kallelsen.
</p>
```

HTML används framför allt för att styra utseendet på skärmen. Därför är det inte ovanligt att man som här placerar rubriknivå 2 före rubriknivå 1, eftersom det ger ett önskvärt utseende i webb-läsaren. Därmed blir det omöjligt att hantera textens struktur.

Man kan exempelvis inte, som när man har tillgång till SGML-märkt text, söka genom Strindbergs författarskap på jakt efter efter brev inuti kapitel inuti romaner.

En annan skillnad är att HTML enbart erbjuder några tiotal förbestämda märkord att välja på. Det innebär att själva brödtexten som regel enbart märks som stycken, med attribut som styr typografin på skärmen.

Eftersom HTML framför allt brukar användas används för att styra utseendet har man ofta struntat i att följa HTMLs DTD. Det gör inte så mycket om HTML-sidorna bara skall visas upp i någon webb-läsare. Dessa läsare är mycket förlåtande. (När någon för ett par år sedan gjorde experimentet att validera en stor mängd HTML-filer på Internet, visade det sig att några få procent följde HTML-standarden.) Men det blir problem om man vill göra något annat med materialet än att visa upp det på skärmen eller skriva ut det på papper.

Sedan något år existerar en ny standard, XML (Extensible Markup Language), som liksom HTML utvecklas av World Wide Web Consortium. XML är ett slags förenklad dialekt av SGML avsedd främst för märkning av information på WWW (Word Wide Web). XML är således ett mellanting mellan SGML och HTML. Det är mindre komplext än SGML och mer kraftfullt än HTML.

Till skillnad från HTML tillåter XML användaren att definiera egna märkord och att hantera strukturerad information (dvs elementen kan inneslutas i varandra, som lådor i lådor). Förmodligen kommer HTML, XML och SGML att under ytterligare någon tid existera parallellt:

- HTML för färskvara, som man vill ha ut snabbt och enkelt på nätet, och för korta och enkla dokument
- XML för mer komplext material med längre livslängd, där det finns behov att definiera egna märkordsuppsättningar, mer precisa länkningsmekanismer, metadata.
- SGML för beständigt material och krävande tillämpningar, t ex vetenskapliga utgåvor, samt för internt bruk inom organisationer, företag och olika forskningsspecialiteter.

Den som arbetar med beständigt material bör hålla i minnet att det torde vara ganska problemfritt att konvertera från en mer

komplex märkning till en fattigare, dvs från full SGML till XML till HTML, medan motsatt väg kan bli mycket arbetskrävande. Det kan således vara lämpligt att använda SGML som arkivformat, XML som utbytesformat och HTML som presentationsformat.

I och med informationsteknikens utveckling har arkiven, biblioteken och museerna ett nytt slags uppgift. Hittills har frågorna rört sådant som digitaliseringsmetoder, val av egna databasprogram och sökmotorer och design av den egna webbplatsen. I framtiden tillkommer uppgiften att utforma materialet så att nyttjarna kan hantera det med hjälp av sina verktyg.

Några akronymer och adresser

SGML (Standard Generalized Markup Language, ISO 8879:1986) är ett metaspråk för beskrivning av informationens innehåll och struktur. ISO-standard sedan 1986.

Standardverk: Charles F. Goldfarb *The SGML Handbook*. Clarendon Press, Oxford 1990.

För mer information om SGML, se t ex

<http://www.oasis-open.org/cover/>

SGML användarföreningen i Sverige, se <http://www.sgml.a.se/>

XML (Extensible Markup Language) är ett metaspråk för märkning av information på WWW (Word Wide Web). XML är en förenklad dialekt av SGML. Utvecklas av W3C (World Wide Web Consortium). Version 1.0 av XML antogs i februari 1998

<http://www.w3.org/XML/>

HTML (Hypertext Markup Language) det i dag vanliga märkspråket för information på WWW. Utvecklas av W3C. HTML kan betraktas som en tillämpning av SGML.

<http://www.w3.org/>

DTD (Dokumenttypdefinition) är en regelsamling för märkning enligt SGML-standarderna (samt för XML och HTML). Det finns olika DTDer för olika användningsområden, t ex EDAs DTD och TEIs DTD (se nedan).

EAD (Encoded Archival Description) är en SGML-tillämpning för märkning av arkivdata.

<http://lcweb.loc.gov/ead/> är EADs hemsida hos Library of Congress.

TEI (Text Encoding Initiative) är en SGML-tillämpning inom humaniora. Sedan 1987 utvecklad av projektet TEI med deltagande av humanister och dataloger i många länder. Den dominerande de facto-standarden för digitala textkritiska utgåvor av litterärt och historiskt källmaterial.

TEIs riktlinjer finns utgivna som Lou Burnard/C.M. Sperberg-McQueen (Eds) *Guidelines for Electronic Text Encoding and Interchange (TEI P3)*, Volume I-II. Chicago University/Oxford University, 1994

Den senast reviderade versionen av riktlinjerna finns på adressen <http://etext.lib.virginia.edu/TEI.html>

Se även TEI-konsortiets webbplats <http://www.tei-c.org/>