

Human IT

Tidskrift för studier av IT
ur ett humanvetenskapligt perspektiv

Almqvist och Ibsen på webben två metoder för elektronisk publicering

av [Charlotte Wallberg](#)

This article is the result of a Master's thesis in library and information studies. The aim of the article is to study and compare to what extent the Hypertext Markup Language (HTML) and the Text Encoding Initiative (TEI) support bibliographic description, metadata and the markup of textual segments. The projects in the study comprise the Norwegian The Writings of Henrik Ibsen (Henrik Ibsens Skrifter), which applies the use of TEI, and The Royal Academy of Letters' The collected works of CJL Almqvist (Vitterhetssamfundets CJL Almqvist Samlade Verk), which applies the use of HTML.

The study shows that the Ibsen project provides its text with bibliographic description and metadata to a greater extent than does the Almqvist project. From the aspect of markup of textual segments, the Ibsen project presents the most detailed information. The amount of markup words used, were not only significantly higher in the texts of the Ibsen project, but the mark-up was also more specific. Much of the information provided by the use of markup in the Ibsen case was supplied by using hypertext links in the digitisation of Almqvist's works. However, these examples show that TEI offers more possibilities than HTML. In order to comprehend the difference in markup between the two projects it is necessary to keep in mind the different aims of HTML and TEI.

Innehåll

[1. Inledning](#)

[2. Genomförandet](#)

[3. Om projekten](#)

[4. Uppmärkning och märkspråk](#)

5. Resultat av jämförelsen

6. Vad kan vi dra för slutsatser av detta?

Om författaren

Redaktionell not: 2003-09-30

För att undvika eventuella missförstånd vill vi påpeka att artikeln utgår ifrån det stadium på vilket de beskrivna projekten befann sig vid artikelns tillkomst. Den elektroniska utgåvan av Henrik Ibsens skrifter är ännu inte *publicerad* på Internet. Projektet har nu kommit ungefär halvvägs, men en elektronisk publicering ligger ännu långt fram i tiden. De texter som det refereras till i artikeln var tidiga demonstrationsfiler som inte längre är tillgängliga. Sedan artikeln skrevs har projektet också gått över från SGML-TEI till XML-TEI, och följer nu TEIs senaste standard P4. Se <http://www.ibsen.uio.no/his/> för mer information.

1. Inledning

Den här texten utgör en bearbetad och förkortad version av en magisteruppsats i biblioteks- och informationsvetenskap som jag avslutade våren –00 vid Institutionen Bibliotekshögskolan, Högskolan i Borås. Artikeln är disponerad i tre delar, varav den första innehåller en bakgrund till studien med beskrivning av genomförandet. Del 2 presenterar viss litteratur och teori samt de båda märkspråken och del 3 de övergripande resultaten och diskussion av dessa. De många exempel som finns med i uppsatsen har med några få undantag utelämnats (en del finns med i form av fotnoter). Jag har här framför allt valt att betona de slutsatser arbetet ledde fram till.

Denna studie tog sin utgångspunkt i några tankar kring digitalisering av skönlitterära verk. Vilka olika möjligheter finns det idag till att göra om tryckta skönlitterära texter till ett format lämpligt för Internet? I framtiden kommer kanske all skönlitteratur att finnas tillgänglig digitalt, med eller utan tryckta motsvarigheter. Hur kan vi då på bästa sätt möjliggöra för den kommande tidens läsare att kunna ta del av vårt litterära arv? Två märkspråk, HTML, *Hypertext Markup Language*, och TEI, *Text Encoding Initiative*, används för att konvertera tryckt material till en elektronisk motsvarighet. Om likartade digitaliseringsprojekt använder sig av skilda märkspråk, vad ger det för konsekvenser? Har valet av märkspråk någon betydelse för projektens resultat? Vad är egentligen att föredra: ett och samma märkspråk på hela webben eller flera olika märkspråk anpassade för sina avsedda användningsområden? Är det inte så att HTML alltid utgör det bästa alternativet? Eller?

HTML, *Hypertext Markup Language*, är det märkspråk som används för World Wide Web, och brukar upplevas som ett enkelt sätt att publicera dokument över WWW. Det är idag mycket populärt att exempelvis skapa en egen hemsida. HTML är ett märkspråk som är relativt okomplicerat att förstå sig på och använda sig av. Dessutom kan HTML många gånger vara en bra introduktion till mer invecklade uppmärkningar. Men enligt många medför HTML stora begränsningar vid en mer detaljerad uppmärkning av en texts olika beståndsdelar (se bland andra Hockey, 1997).

TEI, *Text Encoding Initiative*, har rötter som går längre tillbaka än HTML. 1987 började framställningen av det som så småningom skulle bli TEI att diskuteras, medan HTML:s födsel brukar dateras till början av 1990-talet. TEI skapades främst för att på ett enhetligt sätt kunna digitalisera humanistiska och samhällsvetenskapliga texter, och erbjuder en mer detaljerad uppmärkning än HTML. Men medan det utan problem går att läsa HTML-märkta dokument över webben med en vanlig webbläsare krävs det i fallet med TEI en specifik mjukvara.

Syftet med min undersökning var att studera och försöka jämföra i vilken utsträckning två olika märkspråk, HTML och TEI, ger stöd för bibliografisk information och annan metadata samt uppmärkning av textuella segment. De studerade projekten utgörs av det norska projektet för digitaliseringen av *Henrik Ibsens skrifter* och Vitterhetssamfundets utgivning av *CJL Almqvist Samlade verk*¹. De använder sig av TEI, respektive HTML, vid uppmärkning.

([Åter](#) till början av artikeln)

2. Genomförandet

Det intressanta för den här studien är att beskriva och försöka förklara HTML och TEI utifrån den uppmärkning som gjorts hos två specifika projekt. De båda projekten uppvisar två praktiska exempel på HTML och TEI och uppmärkningen medför därför stora skillnader i jämförelse med hur märkspråken beskrivs i olika manualer. En studie av andra projekt skulle alltså med stor sannolikhet kunna uppvisa andra resultat.

Undersökningen görs utifrån studier av hur texterna gestaltas i uppmärkt form. Då texterna är uppmärkta enligt HTML kan detta göras i en texteditor, som till exempel Anteckningar / Notepad. De TEI-märkta texterna läses med hjälp av en särskild mjukvara. I studien har SoftQuad Panorama version 1.11 använts. Den valda versionen av Panorama hämtades hem gratis, men är en begränsad och inte komplett version då den inte erbjuder till exempel style sheets. För studiens syfte är den dock tillräcklig. Uppmärkningen kan studeras antingen direkt i Panorama, där det finns en funktion som medger att märkorden visas, eller i en texteditor. Eftersom funktionen i Panorama är begränsad har uppmärkningen, liksom var fallet med HTML-texterna, främst studerats i en texteditor, här Anteckningar / Notepad.

([Åter](#) till början av artikeln)

3. Om projekten

Texterna inom ramen för *Almqvist Samlade Verk* ges ut i både tryckt och elektronisk form. Vintern 2000 fanns åtta verk uppmärkta i HTML, i version 3.2, och tillgängliga på webben². Projektet *Henrik Ibsens Skrifter* startades 1998 med syftet att ge ut en tryckt pappersutgåva samt en elektronisk utgåva av Ibsens samlade verk. Vintern 2000 fanns fem verk uppmärkta med hjälp av TEI, varav fyra var skådespel och en bestod av ett antal dikter³. Dessa båda projekt liknar varandra på många olika sätt, såsom exempelvis litteraturformerna de avser kon-vertera till elektronisk form. Vad har det då för betydelse att de har använt sig av två olika märkspråk? Innan vi kommer fram till ett

svar på den frågan bör vi göra klart för oss vad ett märkspråk egentligen är; hur det är uppbyggt och hur det används. Vi ska även titta på några teorier om uppmärkning och digitalisering av texter.

([Åter](#) till början av artikeln)

4. Uppmärkning och märkspråk

Sperberg-McQueen⁴ skrev 1991 en artikel där han på ett teoretiskt plan diskuterar uppmärkning och märkspråk utifrån av honom nio formulerade teser, så kallade *axiom*. Dessa axiom grundas till viss del i arbetet med framställningen av TEI:s riktlinjer, och bör enligt författaren tas i beaktande vid konstruktionen av ett nytt märkspråk. Centralt för Sperberg-McQueens artikel är idén om att den elektroniska representationen och uppmärkningen av en text redogör för en åsikt om vad som är det essentiella i just den texten. I axiomet *Markup reflects a theory of the text* menar författaren att ett omsorgsfullt utarbetat märkspråk återspeglar själva teorin bakom texten, och visar med andra ord förhållandet mellan tolkningen av en text och utformningen av den textens uppmärkning. Vidare säger Sperberg-McQueen att *No finite markup can be complete*, det vill säga ett märkspråk kan aldrig ses som avslutat och komplett eftersom det inte finns en enda fastställd och begränsad form av texter som kan märkas upp. För att undvika dessa problem bör det vara möjligt att vidareutveckla ett märkspråk. Vid den händelse att någon arbetar på en ny typ av text, måste denne ha en möjlighet att tillfoga nya märkord till märkspråket som är relevanta för den aktuella texten. Innebörden i axiomet *Texts are linguistic objects* är att texters lexikaliska och fonetiska information ska lyftas fram. Märkorden bör enligt Sperberg-McQueen kunna beskriva en texts språkliga struktur och möjligheter till att betona en texts frasstruktur, rim i dikter och ortografiska omskrivningar bör finnas tillgängliga.

Allen Renear (1997) presenterar i en artikel olika strategier för uppmärkning av texter⁵. Utifrån en historisk genomgång kommer han fram till de idag framträdande strategierna, *de innehållsbaserade strategierna*. Med utgångspunkt i dessa har tre riktningar av teorier om textualitet utvecklats. Renear kallar dem *platonism*, *pluralism* och *antirealism*. Karaktäristiskt för platonismen är att se texter som uttryck för den hos upphovsmannen underliggande intentionen, vilken uppmärkningen sedan beskriver så troget som möjligt. En elektronisk text formas med andra ord utifrån den person som genomfört uppmärkningen och ger därmed uttryck för den personens avsikter. SGML, som utgör ett exempel på platonism, syftar till att användas på många typer av texter. Bakom platonismen ligger just denna generalitetsprincip. Pluralismen, till vilken TEI hör, menade att det istället för att använda ett enda märkspråk på alla texter bör utvecklas en specifik DTD för varje litteraturform. Enligt både platonismen och pluralismen styr texterna utformningen av uppmärkningen, medan företrädarna för antirealismen däremot hävdar att texter utgör produkter av märkspråkens strukturer, och att alltså uppmärkningen styr utformningen av de elektroniska texterna.

En stor skillnad mellan de ovanstående tre teorierna är alltså synen på förhållandet mellan text och märkspråk. Enligt platonismen och pluralismen styr texterna utformningen av uppmärkningen, det är texterna som ger form åt märkspråket. Antirealismens förhållningssätt är att teorier om uppmärkning och märkspråk styr

texterna och definierar hur de ser ut i uppmärkt form.

Vad är ett märkspråk?

Det vanligaste märkspråket är idag förmodligen HTML. Ett annat allt mer omtalat märkspråk är XML, men det får kommenteras vid ett senare tillfälle. Många har idag hört talas om HTML, och vet också att det står för Hypertext Markup Language. Men vad innebär ordet markup, och vad är egentligen ett märkspråk? Enligt Broady och Juliusson (1997) karaktäriseras ett märkspråk av en "uppsättning regler som anger hur man tydliggör strukturen och innehållet i informationen genom att förse den med ytterligare information". Den ytterligare informationen utgörs då av märkord⁶. Anders Möller (1994, s 34) talar om två sorter av märkspråk. Han menar att å ena sidan handlar det om vilket syfte informationen har i en text, och å andra sidan om hur den här informationen presenteras. De här båda kallar han för logisk respektive typografisk struktur. Möller tar exemplet med en bok för att beskriva de här båda strukturerna. Bokens titel, författare, förlagsnamn, innehållsförteckning, kapitel, rubriker, tabeller och så vidare står för den logiska strukturen, medan den typografiska strukturen bestämmer teckenstorlek, typsnitt, radavstånd och andra liknande aspekter. De två begrepp Möller använder sig av, logisk respektive typografisk struktur, benämns ofta (och även i denna artikel) beskrivande märkning samt procedurmärkning. Beskrivande märkning visar alltså vad själva informationen är och vad den består av. Den koncentrerar sig på strukturen och innehållet hos en text och låter själva utseendet vara odefinierat. Procedurmärkning används däremot för att återge textsegment med olika utseenden på skärmen.

Element, entiteter, attribut och dokumenttypsdefinitioner

Element, entiteter och attribut är tre komponenter som utgör grunden i ett märkspråks struktur. Dokumenttypsdefinitionen (DTD) sammanför och visar på förhållandena mellan dessa komponenter. En *entitet* utgörs enkelt beskrivet av ett stycke symboler. De används främst för att kunna använda sig av särskilda symboler som till exempel svenskans å, ä, och ö. Även fast de här tre bokstäverna utan problem kan infogas i ett dokument från ett tangentbord i Sverige kan de framstå som helt oläsliga när dokumentet ska läsas på andra sidan jordklotet. I stället för att skriva bokstaven å och riskera att den ser oförståelig ut används en teckensträng som fungerar som en kod för den gällande bokstaven. Medan en entitet representerar enkla symbolsekvenser innehåller elementen objekt som till exempel paragrafer, titlar och rubriker. Ett *element* utgörs av ett märkord samt dess innehåll (Broady & Juliusson, 1997). Märkordet omsluts av tecken som visar var elementet börjar < > och slutar </ >. Ett element kan se ut på följande sätt: <h1>Rubrik</h1>. Ett *attribut* har som uppgift att tilldela elementen ytterligare egenskaper. De här egenskaperna kan bestå av detaljerade aspekter, såsom storlek eller färg. Följande är ett exempel på ett element med ett attribut: <h1 align=center>Rubrik</h1>. Innebörden av det här elementet är att rubriken ska vara justerad. Attributets *värde*, här center, förklarar att justeringen ska vara centrerad.

HTML och TEI

HTML och TEI är två märkspråk som båda utvecklats från SGML. SGML, eller *Standard Generalized Markup Language*, är ett exempel på ett beskrivande märkspråk. Även HTML, *Hypertext Markup Language*, är ursprungligen ett beskrivande märkspråk, trots att det idag används som om det vore en form av procedurmärkning. Varningar om en användning av HTML som ett redskap för grafisk formgivning är inte ovanliga. Argumenten som förs fram är att ett dokument utseende på webben inte kan garanteras. Elementen beskriver den uppmärkta textens innehåll, inte dess grafiska utseende, och det är varje webbläsares uppgift att avgöra hur uppmärkningens utseende kommer att ta form på skärmen. *Hypertext Markup Languages* ursprung dateras till 1991, efter det att en av CERNs⁷ anställda, Tim Berners-Lee⁸ (1989), skrivit en artikel som beskrev ett förslag om hur man skulle kunna förbättra hanteringen av information inom organisationen. Det här förslaget innebar teorin om ett system baserat på hypertext.

Metadata, data om data, sägs vara en lösning till de problem som i dag råder med att återfinna dokument över webben. Metadata kan bestå av en katalogpost i ett bibliotek, eller information om titel, upphov och nyckelord till ett elektroniskt dokument. Möjligheter till att tilldela en text uppmärkt i HTML metadata finns inom ramen för elementet `<head></head>`. Med hjälp av `<title>` kan titeln beskrivas, och elementet `<meta>` kan i kombination med olika namnattribut formas så att många olika aspekter kan beskrivas. Inom ramen för HTML finns inga på förväg utarbetade och fastställda namnattribut, utan det är upp till var och en att använda sig av attributet, helt i enlighet med egna behov⁹.

TEI, *Text Encoding Initiative* arbetades fram i slutet av 1980-talet med syfte att märka upp humanistiskt och samhällsvetenskapligt material. En text uppmärkt enligt TEI:s riktlinjer består av en Header, som befinner sig allra först i dokumentet, samt efter denna den uppmärkta texten. TEI Header kan sägas motsvara det som utgör den bibliografiska informationen om en text. Den bidrar med metadata som kan vara till nytta för användare som behöver information om texten, såsom till exempel forskare eller bibliotekarier. Headern består av fyra delar som kallas descriptions, det vill säga beskrivningar. Dessa är i tur och ordning filbeskrivningen (file description), kodbeskrivningen (encoding description), profilbeskrivningen (profile description) och revisionsbeskrivningen (revision description). Det existerar sedan för själva texten ett antal olika märkordsgrupper, som kallas kärnmärken, basmärken, tilläggsmärken och hjälpmärken. Gruppen av basmärken väljs utifrån vilken litteraturform eller genre som ska märkas upp.

Manualen som beskriver TEI:s alla märkord kan verka ganska svåröverskådlig och antalet märkord kan tyckas vara lite väl många. Få av de 400 olika märkorden är dock obligatoriska (Hockey, 1994). TEI:s schema erbjuder stor valfrihet, det är upp till var och en att avgöra i hur pass stor utsträckning en text ska märkas upp. TEI betraktas som en DTD av SGML, men erbjuder inte en enda fastställd DTD utan en möjlighet för användaren att själv kombinera ihop en DTD anpassad till de dokument som ska märkas upp. TEI:s riktlinjer erbjuder även möjligheter för användarna att modifiera märkorden för ett efter egna behov anpassat bruk. Nya element kan läggas till och de fast definierade märkorden samt deras attribut kan ändras så att de stämmer överens med användarens syften. (Guidelines for Electronic Text Encoding and Interchange, 1994, s 35ff).

5. Resultat av jämförelsen

Bibliografisk information och annan metadata

Begreppen bibliografisk information och metadata kan ha skiftande innebörd. Här lades koncentrationen på beskrivning av nyckelord, information om källmaterial, samt titel, upphovsmän och ursprung för den elektroniska utgåvan av det uppmärkta verket.

Två tydliga skillnader kan urskiljas. Den första skillnaden är att det ena projektet (Ibsen) följer ett för ändamålet utarbetat schema, medan det andra projektet (Almqvist) inte har använt sig av ett speciellt och konkret system¹⁰. Den andra skillnaden ligger i att Ibsenprojektets uppmärkning i det här hänseendet är konsekvent, samma detaljerade nivå genomsyrar samtliga texter, och TEI Headerns schema följs så att samtliga obligatoriska delar är inkluderade. Almqvistprojektets uppmärkning, däremot, är inte lika homogen.

Undersökningen visade att Ibsenprojektet i större utsträckning än Almqvistprojektet tilldelar sina texter bibliografisk information och metadata. Enda undantaget är nyckelord, som endast anvisats Almqvists texter¹¹. Möjligheten till att med hjälp av TEI Header delge nyckelord finns, men har i fallet med Ibsenprojektet inte tagits i bruk. En förklaring till detta kan ligga i att den del av TEI Header inom vilken nyckelord placeras, *profilbeskrivningen*, inte är obligatorisk. För att en text ska kunna sägas vara uppmärkt enligt TEI:s riktlinjer krävs nämligen endast en minimal Header innehållandes ett antal obligatoriska delar i filbeskrivningen (Giordano, 1995, s 76). Kan det i det här fallet vara så att de ansvariga bakom Ibsenprojektet framför allt strävat efter att genomföra en korrekt uppmärkning utifrån TEI:s riktlinjer, och därför främst inkluderat de föreskrivna delarna av Headern? Istället för att överväga vilka komponenter som kan vara till användning för ett digitaliseringsprojekt?

Texterna inom Almqvistprojektet tilldelas viss information om titel, författare, information om källmaterial och den elektroniska utgåvan, men uppmärkningen har inte visat sig vara konsekvent, en del texter innehar mer information än andra, och speciella märkord används inte. Elementet <title> inkluderar här både författare och titel, även fast det med HTML är möjligt att med ett specifikt märkord beskriva författarens namn¹². Elementet <meta> kan tilldelas vilket attribut som helst, vilket innebär att det går att lämna information om många olika aspekter, så i Almqvists fall hade författaren tydligt och uttryckligen kunnat specificeras, liksom många andra komponenter som här inte getts en anpassad uppmärkning. HTML erbjuder i det här avseendet en oerhörd flexibilitet och gränslöshet, men frågan är om inte grundläggande kunskaper om katalogisering krävs av den som genomför uppmärkningen för att möjligheterna ska kunna användas? Den stora valfriheten som ges kan skapa förvirring istället för en lättnad av att kunna delge det personen i fråga önskar. Ett förslag är att ett fåtal exempel skulle kunna utformas så att det blir lättare att tilldela ett HTML-dokument metadata.

Medan Almqvistprojektet har uppvisat några problem med att beskriva upphov med

hjälp av HTML, kan man hos Ibsenprojektet se hur en lösning på detta problem kan ta form. Författaren <author> och övriga upphovsmän separeras här tydligt från titlarna med hjälp av specifika märkord. Utöver <author> har projektet även använt sig av ytterligare märkord, såsom <funder> för uppgifter om finansiering, samt <resp> och <name> för att själv kunna beskriva olika former av upphov¹³. Med andra ord medger även TEI en stor flexibilitet och valfrihet vid tilldelandet av information om upphovsmän. Klart är att riktlinjerna för TEI på ett tydligt sätt visar hur denna valfrihet kan brukas. Detta är kanske något som inte beskrivs lika ingående i olika HTML-manualer?

Sammanfattningsvis kan konstateras att Ibsenprojektets användning av metadata är mer omfattande och konsekvent genomförd än den som Almqvistprojektet uppvisar. Specifika märkord används utifrån deras avsedda syften och de täcker dessutom till stor del intressant information. Att Almqvistprojektets verk genomgående uppvisar ojämnheter tror jag kan bero på att flera olika personer genomfört arbetet och att ingen konsensus rått emellan dem angående hur uppmärkningen ska genomföras. Har en diskussion överhuvudtaget förts om vilken form av metadata som ska tilldelas?

Uppmärkning av textuella segment

Textuella segment är ett dunkelt uttryck. Det har här använts för att titta på förekomsten av element som märker upp stycken, titlar, rubriker, olika ord och fraser samt komponenter karaktäristiska för skådespel och dikter.

Almqvistprojektet använder sig av avsedda märkord för att märka upp stycken, <h> för rubriker och titlar och för den resterande delen av de undersökta textuella segmenten av <i>, , samt ¹⁴. Av de tre sista utgör <i> den i störst utsträckning nyttjade. Ibsenprojektet brukar särskilda märkord för uppmärkning av stycken, enkla verslinjer, titlar och rubriker av olika form, <emph> för att tilldela emfas till ord eller fraser, <castList> för att presentera roller i skådespel, samt specifika märkord vid talare inom skådespel och sådana som bidrar till att beskriva en dikts struktur. Undersökningen visade att Ibsenprojektet tillämpar flera olika märkord avsedda för särskilda syften, och innehar därför en mer detaljerad uppmärkning än Almqvistprojektet.

En avsevärd skillnad ligger här i möjligheten att märka upp vissa delar. Medan Ibsenprojektet uppvisar ett nyttjande av ett stort antal olika märkord, är användningen hos Almqvistprojektet koncentrerad till ett fåtal. Framträdande i uppräknningen av Almqvists texter är bruket av elementen <i> och , som fått representera många och olikartade ord och fraser. Den utbredda användningen av de båda märkorden visar att någon form av markering av vissa begrepp och ord varit önskvärd. Det verkar vid en första titt som om <i> och inte har använts på ett kompletterande vis. Skillnaden mellan de här båda elementen är att <i> enbart används för att återge ett segment i kursiv stil, medan 's funktion består av att ge ett ord eller en fras en betoning. I själva verket är en form av beskrivande märkning, medan <i> placeras hos de märkord som hör till procedurmärkning. Elementet <i>'s avsikt är endast att återge ord eller fraser i kursiv stil på skärmen. Hur ord uppmärkta med återges på skärmen är inte som för italic det primära syftet. Jag tror inte att denna skillnad är allmänt känd. Inga andra märkord för att specifikt markera speciella grupper av ord eller fraser har använts inom Almqvistprojektet. Det går ej heller att

urskilja något särskilt tillvägagångssätt vid uppmärkning av olika komponenter i skådespel såsom rollistor, talares namn och repliker (vilket är naturligt eftersom HTML inte har den funktionen). Dessa har liksom övriga ord och fraser markerats med hjälp av <i>. Varför, kan man då fråga sig, ser uppmärkningen i det här fallet ut som den gör?

En förklaring kan vara att uppmärkningen utgör en spegling av hur den tryckta utgåvan ser ut. Kanske redogör den även för hur författaren själv, Carl Jonas Love Almqvist, önskade se sin text? Almqvistprojektets texter har dessutom paginerats, vilket kan tolkas som att en strävan efter att i så stor utsträckning som möjligt försöka låta den elektroniska versionen likna en tryckt bok varit rådande. En ytterligare förklaring kan vara att det i Almqvistprojektets fall förmodligen inte varit av stor vikt att göra en detaljerad uppmärkning. Mycket av det som uppmärkning av textuella segment kan bidra till, exempelvis hos Ibsen, har nämligen lösts med hjälp av HTMLs triumfkort: länkning. Almqvists verk återges i fyra olika versioner. En version består av en ren text medan de andra tre innehåller *accidentalier*, *substantieller* och *ordförklaringar*. Dessa varianter uppmärkning avviker från den okommenterade texten i och med att de innehåller länkar. Åtskilliga ord eller fraser som märkts upp med <i> är länkade i versionen *ordförklaringar*. Det speciella med det uppmärkta textuella segmentet ges där sin förklaring. I volym 9 av *Törnrosens bok, Band XII, Poesi och politik*, åskådliggörs detta då ordet *patois* i den okommenterade texten har märkts upp med <i>, och sedan i versionen som benämns *ordförklaringar* tilldelats en länk som pekar på följande förtydligande: (*fra.*) *dialekt, landsmål*. Här förtydligas ordets språkliga ursprung, *fra*, att det handlar om ett ord som kan härledas till det franska språket. Dessutom ges översättningen av ordet, *dialekt, landsmål*. Det är i det här fallet alltså inte nödvändigt att med hjälp av uppmärkning visa att ordet har ett utländskt ursprung, utan frånvaron av specifika märkord kompenseras genom en användning av länkar. Medan Ibsenprojektet infogar sådan här information i samma fil, så hänvisar Almqvistprojektet den alltså till separata webbsidor. En sådan metod kan ha betydelse för analyserna av verken. Textkritiska och litteraturvetenskapliga tolkningar är alltid tidsbundna och föränderliga. En vetenskaplig tolkning som gjordes för hundra år sedan är bunden till sin historiska epok, liksom de tolkningar som görs idag innehar sina historiska villkor. Almqvistprojektets tillvägagångssätt att hålla själva verket fritt från kommentarer och placera dessa på separata sidor medför här en klar fördel. De separata webbsidorna kan betraktas som ett föränderligt lager som hela tiden kan bytas ut allt eftersom forskningen förändras, medan själva verket i sig ligger kvar i ursprunglig form. Som vi sett kombinerar Ibsenprojektet verket med kommentarer, vilket innebär att i det fall en ny tolkning önskas göras, skulle kanske hela det uppmärkta verket behövas tas i beaktande. Ett stort plus till Almqvistprojektet, alltså.

Vid jämförelse mellan de båda projekten blir det intressant att uppmärksamma tillvägagångssättet vid uppmärkning av talare, repliker samt olika roller inom skådespel. Ibsenprojektet använder sig i det här hänseendet av specifikt avsedda märkord, medan Almqvistprojektet, i den mån uppmärkning av sådana delar gjorts, löst det med hjälp av <i>. Text Encoding Initiatives schema skiljer vid uppmärkning på olika genrer eller typer av litteratur i den grupp märkord som kallas basmärken. Inom Ibsenprojektet har texter från två skönlitterära genrer märkts upp: skådespel och dikter. Dessa verks uppmärkning grundas därför i stor utsträckning i basmärkena för drama respektive vers. Inom *basmärkena* ryms alltså en uppsättning element som är framtagna efter ett

dramas olika komponenter¹⁵. Trots förekomsten av stycken som har karaktären av skådespel inom Almqvists texter används inga särskilda märkord för att upplysa om vem som är talare och vad denne säger, vilket har sin förklaring i HTML:s uppsättning element, där sådana komponenter inte tagits i beaktande. TEI:s struktur speglas på samma sätt som vid skådespelen i Ibsens *Dikt*, som med hjälp av basmärkena för texter på vers kunnat märkas upp utifrån för dikter typiska aspekter.

Almqvistprojektet fokuserar alltså på den fysiska, tryckta utgåvan, och bruket av `<i>` samt `` är i det fall de speglar författarens eller utgivarens intentioner berättigat. En annorlunda inställning kan vara att försöka tolka och lista ut vad författaren till de verk som ska digitaliseras avsett med typografiska markeringar såsom kursivering av ord eller fraser. Ibsenprojektet, utifrån dess tillämpning av TEI, kräver en gissning av vad författaren avsåg med sin text. Märkorden speglar här alltså till en viss utsträckning hur de ansvariga bakom Ibsenprojektet tolkat författarens intentioner. Tolkningen kan tillskrivas en skala, inom vilken det går att skilja mellan märkord som kräver större tolkning än andra, vars innebörd kan vara mer uppenbar. Det är givetvis problematiskt att uttala sig om i vilka fall en ansenlig tolkning gjorts samt i vilka fall uppmärkningen kan sägas vara objektiv. Indelningen i stycken kan tyckas vara harmlös, medan uppmärkningen av en dikt utifrån metrisk struktur och rytmiska mönster däremot förutsätter en fylligare analys. En sådan här analys kan beroende på grad av tolkning vara vanskelig och ibland, beroende på vilka intentioner ett projekt har, kan en objektiv kopiering av originalet vara det tryggaste tillvägagångssättet.

Konklusionen av jämförelsen utifrån den här aspekten är att Ibsenprojektet tillämpar en mer detaljerad uppmärkning än Almqvistprojektet, som däremot speglar de begränsningar HTML besitter angående uppmärkning av textuella segment typiska för skönlitteratur. Användningen av `<i>` samt `` är svår att behandla, eftersom det inte handlar om en specifik uppmärkning av olika ord eller fraser, utan förmodligen speglar layouten i den tryckta utgåvan. Ibsenprojektet samt Almqvistprojektet reflekterar båda strukturen i de märkspråk de brukar, vilket i Ibsens texters fall leder till en detaljrik uppmärkning och motsvarande i Almqvists texters fall till en enkel och inte så komplicerad uppmärkning. Almqvistprojektet har dock kompletterat med hjälp av länkning, och förlägger på ett praktiskt sätt informationen och tolkningen till separata filer.

Vilka konsekvenser får användningen av märkspråken?

Många, har den här studien visat. Vilka konsekvenserna är och vad de beror på beskrivs nedan.

Skillnaderna mellan HTML och TEI ...

Visst skiljer sig de båda märkspråken från varandra, och visst har det stor betydelse. För det första: riktlinjerna för TEI beskriver ett för syftet utarbetat schema för tilldelning av metadata, TEI Header. HTML föreslår däremot inte ett specifikt tillvägagångssätt för att tilldela ett dokument exempelvis nyckelord, utan dessa bör antingen själv utformas inom ramen för märkordet `<meta>`, eller så bör de sökas i ett utarbetat och fristående metadataschema. För det andra: TEI erbjuder en mer detaljerad uppmärkning och ett större antal märkord anpassade för särskilda litteraturformer än

HTML.

Men samtidigt som TEI erbjuder fler möjligheter än HTML är det ett komplicerat märkspråk som kräver stora kunskaper för att kunna brukas på ett korrekt sätt. Trots att användningen av TEI är anpassningsbar efter enskilda behov och kan förenklas till en relativt enkel uppmärkning så bör dess komplexitet inte föringas. En inblick i de olika alternativen anpassade enligt genre och de detaljerade märkorden kräver både tid och ett stort engagemang. Som vi har sett är HTML begränsat i den bemärkelsen att uppmärkning av specifika aspekter inte är genomförbar i samma utsträckning som TEI. Ur en beskrivande uppmärknings synvinkel medför HTML i förhållande till TEI få möjligheter. HTML är dock ett vida spritt märkspråk vars struktur är betydligt enklare än TEI, vilket leder till att en hel del tid på inläsning och förbearbetning kan sparas.

En ytterligare skillnad är tillgängligheten. För tillfället finns visserligen texter uppmärkta i TEI tillgängliga över webben, men dessa kräver en särskild programvara. Dokument uppmärkta i HTML läses idag över webben med de vanligaste webbläsarna. Texter som är tillgängliga över webben med hjälp av HTML kan dessutom lätt samlas på en (vanlig) länklista.

...beror på deras syften och historik...

För att förstå skillnaderna mellan de båda märkspråken är det viktigt att lyfta fram deras olikartade syften. HTML och TEI har skapats med grund i två skilda behov, och deras användning och spridning har vidare lett till att syftena med de båda fortsatt vara separerade. HTML arbetades fram med utgångspunkt i SGML i början av 1990-talet på CERN i Schweiz. Målet var att förbättra hanteringen av information och lösningen hette hypertext (Berners-Lee, 1989). TEI som också togs fram utifrån SGML svarade mot behovet av att kunna märka upp humanistiska och samhällsvetenskapliga texter. Under en längre tid hade en enda metod för att märka upp sådana texter efterfrågats, och förhoppningen hos företrädarna var att TEI skulle fylla det här behovet (Ide & Sperberg-McQueen, 1995, s 5). Att HTML inte förblev en intern affär inom CERN har de senaste tio årens utveckling visat oss. HTML:s användning idag ligger i att göra alla möjliga dokument tillgängliga via webben, och syftar inte till att märka upp en specifik form av texter, vilket TEI däremot gör.

... samt olika teoretiska perspektiv på uppmärkning och märkspråk

Skillnaderna mellan HTML och TEI samt Almqvist- och Ibsenprojekten kan även förklaras utifrån de teoretiska synsätt som ligger bakom märkspråken. Dessa teorier har även samband med syftena som beskrivs ovan. Sperberg-McQueen och Renear företräder olika teoretiska perspektiv på uppmärkning av texter, vilka kan ställas mot HTML och TEI samt projektens tillämpning av de båda.

En framträdande skillnad mellan HTML och TEI ligger i den förbarhet och valfrihet deras respektive schema erbjuder en användare. TEI har en möjlighet för användaren att själv utforma en för syftet relevant användning och använda sig av de märkord som önskas, och kunna tilldela egna valda värden eller till och med utforma egna märkord. Hos en del av Ibsenprojektets texter används attributet *rend* i samband med märkordet <emph>. Det finns inga konkreta förslag på vad *rend* kan tänkas ha för olika värden, utan det är upp till den som gör uppmärkningen att anvisa ett värde. I det

här fallet har värdet "*spr*" tilldelats, alltså `<emph rend="spr">`. Det uppmärkta ordet har här presenterats i kursiv stil. Sperberg-McQueen (1991, s 36) anser med axiomet *No finite markup can be complete* att det bör vara möjligt att vidareutveckla ett märkspråk, genom att till exempel efter egna behov tillfoga nya märkord eller attribut, vilket tilldelandet av värde till *rend* inom Ibsenprojektet ovan visar exempel på. Sperberg-McQueens axiom kan här alltså delvis förklara TEI:s struktur.

TEI:s flexibilitet i det här fallet kan även sägas vara en följd av dess förankring inom pluralismen och önskan att svara mot SGML:s generalitetsprincip och förhoppning att kunna användas för alla former av dokument. Istället för att tillämpa ett enda märkspråk på flera olika texter ansågs det inom pluralismen vara mer fruktbart att utarbeta en DTD för varje genre eller litteraturform. TEI uppfyller de här önskemålen, vilket möjligheterna till en egen tillämpning visar. Den som genomför uppmärkningen innehar här också en central roll. Renear (1997, s 122) hävdar att uppmärkning är en mänsklig aktivitet och att en elektronisk texts struktur till stor del beror på vad den som gjort uppmärkningen önskat betona. Uppmärkningen är således en personlig process inom vilken tolkningar och intressen är av stor betydelse. Sperberg-McQueen (1991, s 35) tilldelar liksom Renear den person som genomför uppmärkningen en framträdande roll i sitt första axiom, *Markup reflects a theory of a text*, och antyder därmed ett liknande betraktelsesätt.

Renear nämner inte inom vilken teoretisk riktning HTML kan placeras. Den här studien har dock visat att HTML är mer generell än TEI, det går exempelvis inte att anpassa uppmärkningen efter genre. En tolkning kan vara att HTML i enlighet med Renears uttalanden hör till *platonismen* och syftar till att kunna tillämpas på ett stort antal olikartade texter. HTML erbjuder ingen möjlighet till utformning utan snarare en fast uppsättning märkord som avses brukas i samtliga med hjälp av HTML uppmärkta texter. Det är framför allt texterna som anpassas till märkspråket och dess möjligheter och inte tvärtom. Almqvistprojektet får hålla till godo med de märkord HTML erbjuder och kan, bortsett från tilldelandet av attribut på metadata inom `<head>`, inte begagna sig av egna värden såsom Ibsenprojektet. Att TEI:s schema består av ett så stort antal märkord av vilka de för ändamålet mest relevanta kan väljas leder till att en ytterligare anpassning kan noteras, som däremot inte är möjlig i HTML. Den teoretiska riktning som Renear (1997, s 122) kallar *antirealism* betraktar elektroniska texter som produkter av de märkspråk de valt begagna sig av. En tolkning av Almqvistprojektets uppmärkning är att texternas utformning är beroende av HTML och dess möjligheter, vilket skulle innebära att HTML även kan härledas till det antirealistiska synsättet.

Skillnaden mellan uppmärkning av textuella segment betonades tidigare, och det visade sig att TEI:s schema i större utsträckning än HTML medger en detaljerad uppmärkning av olika ord och fraser. Den här skillnaden framgår tydligt projekten emellan vid digitalisering av skådespel. Ibsenprojektets beskrivning av skådespel och även av versers rytmiska mönster i dikter speglar det axiom Sperberg-McQueen (1991, s 37) benämner *Texts are linguistic objects*, och visar att uppmärkning av frasstrukturer samt lexikalisk och fonetisk information är hanterbart i TEI men inte i HTML. Som nämnts tidigare innehar emellertid HTML en klar fördel genom möjligheten till att kunna länka samman olika dokument över webben.

Vilket märkspråk ska ett projekt då helst använda?

Det är en fråga som det inte går att ge ett entydigt svar. Valet av märkspråk borde vara beroende av respektive digitaliseringsprojekts syfte. En komplex uppmärkning bör fylla sitt syfte och ha en funktion. Att lägga ner mycket tid på att genomföra en detaljerad uppmärkning utan att veta varför är inte en god lösning.

TEI bör inte användas oreflekterat för alla skönlitterära texter, utan för att de projektansvariga anser att den uppmärkning som är möjlig med hjälp av TEI:s schema är givande för det aktuella projektet. En konsekvens för de i den här undersökningen studerade projekten är som jag tidigare påpekat texternas tillgänglighet. *CJL Almqvist Samlade Verk* läses, utan att några åtgärder måste vidtas, direkt över webben. För att läsa de digitaliserade texterna i *Henrik Ibsens Skrifter* bör däremot en specifik mjukvara såsom exempelvis Panorama inskaffas och installeras. Förslagsvis bör HTML användas för digitalisering då man önskar sprida de skönlitterära texterna till en stor mängd läsare och då möjligheten att i stor utsträckning kunna märka upp olika textuella segment inte ses som ett primärt önskemål. TEI däremot kommer till sin rätt då en komplex uppmärkning av skådespel, dikter eller romaner för olika motiv anses vara fruktbar. TEI kan här bidra till en vidare bearbetning av texterna som inte anses vara lika viktig för användarna av HTML. Dessutom är en användning av TEI givande då ett önskemål om att arbeta fram en egen utformning av märkspråket finns. Slutligen bör TEI användas då målgruppen är begränsad och har kännedom om TEI och dess möjligheter.

Ovanstående tankegångar speglar innebörden i Sperberg-McQueens första axiom, *Markup reflects a theory of a text*. HTML:s och TEI:s respektive syften återger i Almqvist- respektive Ibsenprojekten i själva verket teorin bakom de texter de avser märka upp, och projektens tillämpning av märkspråken kan därför sägas ge information om vad som enligt de projektansvariga utgör det essentiella i de uppmärkta texterna. Sperberg-McQueen (1991, s 34) menar att en elektronisk text består av en representation av originaltexten, och att en sådan representation avslöjar författarens (eller den som gjort uppmärkningen) medvetna eller omedvetna tolkningar av texten i fråga.

([Åter](#) till början av artikeln)

6. Vad kan vi dra för slutsatser av detta?

Undersökningen har visat att TEI möjliggör en mer detaljerad uppmärkning än HTML. Uppmärkning av olika textuella segment, såsom komponenter typiska för skådespel och dikter, är genomförbar med hjälp av TEI men däremot inte i HTML. En tolkning skulle här kunna vara att TEI är ett för syftet, det vill säga att märka upp skönlitterära texter, bättre och mer passande märkspråk än HTML. Ett sådant ställningstagande är befogat med tanke på att TEI har ett snävare användningsområde och syfte än HTML, nämligen att märka upp litterära texter för humanistiska studier. Det bör här emellertid påpekas att TEI inte behöver vara en utmärkt lösning utan kan medföra en hel del problem för en eventuell producent. TEI är inte ett enkelt märkspråk; det tar tid och engagemang att sätta sig in i dess komplicerade schema och i vissa situationer är arbetet helt enkelt inte värt mödan. HTML har ett bredare användningsområde och når via webben ut till en större grupp människor, samtidigt som det är lättare att sätta sig in i och använda. Medan texter uppmärkta i HTML läses med väl spridda programvaror,

kräver TEI däremot en särskild sådan.

En fördel med TEI är att det bara finns en enda manual att studera då man ska lära sig märkspråkets struktur. Det här är tvärtom en nackdel vad gäller HTML: där finns det ett hav med manualer och webbkonstruktionsguider. Hur ska man veta vilken guide som är bäst att använda sig av? Naturligtvis väljer en nybörjare hellre en pedagogiskt upplagd och förenklad bok över HTML framför en manual som beskriver märkspråkets alla möjligheter. Och vem kan klandra någon för det valet?

TEI och HTML erbjuder olika fördelar och nackdelar, och valet av märkspråk bör därför grundas i ett projekts syfte. Det bör hållas i åtanke att de båda märkspråkens roller kan komma att förändras. Eftersom webben ständigt utvecklas kan dess struktur vara betydligt förändrad inom ett antal år, och att idag förutsäga HTML:s respektive TEI:s framtida roller är inte möjligt. XML, Extensible Markup Language, blir för var dag allt mer omdiskuterat och spritt. Liksom TEI tillåter det här märkspråket att användaren själv formar en egen användning, så framtiden kanske inte ligger i vare sig HTML eller TEI, utan just XML?

[\(Åter till början av artikeln\)](#)

Om författaren

Charlotte Wallberg har nyligen avslutat sina studier i Biblioteks- och Informationsvetenskap vid Institutionen Bibliotekshögskolan, Högskolan i Borås. Den här artikeln utgör en omarbetad version av en magisteruppsats (20 poäng) med titeln *Uppmärkning av skönlitterära verk: en jämförelse mellan användningen av Text Encoding Initiative och Hypertext Markup Language*.

[\(Åter till början av artikeln\)](#)

Noter

1. Det bör påpekas att det ingalunda är själva webbplatserna som helhet som studeras, utan enbart de skönlitterära verken. Övriga webbsidor med exempelvis projektinformation granskas ej utifrån hur dess uppmärkning ser ut. För att begränsa omfattningen på undersökningen sattes en tidsmässig precisering. De båda projekten studerades såsom de såg ut under perioden vintern 99/00. ([Åter till texten](#))

2. Den tryckta upplagan av *Almqvist Samlade verk* planeras bestå av totalt 51 volymer, varav den första volymen kom ut 1993. Parallellt med bokutgivningen läggs de elektroniska motsvarigheterna av verken ut på Internet. När digitaliseringen skulle genomföras valde de projektansvariga att arbeta med "det mest använda och mest flexibla formatet, HTML, vilket gör att 'Samlade verk' enkelt blir tillgängliga i World Wide Webs hypertextmiljö – och därmed kan länkas till andra hemsidor och hypertexter". (Svedjedal, 1997, s 61ff). Under vintern –00 fanns följande åtta verk utlagda på webben: 8. Törnrosens bok. Duodesupplagan. Band VIII-XI; 9. Törnrosens bok. Duodesupplagan. Band XII-XIV; 21. Det går an. – Hvarför reser du? (1838 års versioner); 23. Amalia Hillner; 25. Tre Fruar i Småland; 26. C.J.L. Almqvist. Monografi; 30. Herrane på Ekolsund. Del 1 och 2; 31. Herrane på Ekolsund. Del 3. ([Åter till texten](#))

3. Mellan tolv och tretton volymer beräknas ge ut i tryckt form, medan det i elektronisk form kommer att publiceras betydligt mycket mer än så. Texterna märks upp med hjälp av TEI, och görs sökbara på komponenter såsom genre, namn, och rim (Jørgensen, 1999). Följande fem verk fanns under vinterhalvåret 99/00 utlagda på Internet: Catilina, 1850; Norma, 1851; Catilina, 1875;

Vildanden, 1886; Dikt. ([Åter till texten](#))

4. C.M. Sperberg-McQueen har sedan 1988 varit *editor-in-chief* för Text Encoding Initiative. ([Åter till texten](#))

5. Termerna *uppmärkning* och *märkning* används ibland synonymt med *kodning* (jämför med engelskans *markup* respektive *encoding*). I den här texten föredras de båda förstnämnda benämningarna. Likadant gäller för termen *märkspråk* som föredras framför *kodningsspråk* eller *uppmärkningsteknik*. Renear använder t.ex termerna *encoding* respektive *text encoding*, men för ett konsekvent språkbruk översätts dessa här med *uppmärkning*. ([Åter till texten](#))

6. Istället för märkord ser man ibland bara märke, eller tagg, från engelskan tag. Här används främst begreppen märkord samt märke, men även element. ([Åter till texten](#))

7. European Organization for Nuclear Research, beläget i Geneve. URL: <http://www.cern.ch> ([Åter till texten](#))

8. Tim Berners-Lee brukar benämnas mannen som låg bakom skapandet av The World Wide Web. Han arbetar idag för the World Wide Web Consortium. ([Åter till texten](#))

9. Om någon efter att ha skapat ett HTML-dokument önskar låta det bli indexerat av en sökmotor bör denne alltså formulera något i stil med nedanstående inom elementet <head>. Orden efter *content* utgörs av indexeringsorden.

```
<meta name="keywords" content = "markup, uppmärkning, HTML">
```

([Åter till texten](#))

10. Ett exempel på ett utarbetat system för att tilldela dokument metadata är *Dublin Core Data Element Set*. Dublin Core kan användas för många olika dokument, bland annat elektroniska sådana som finns tillgängliga över webben. ([Åter till texten](#))

11. Nedanstående element visar ett exempel (från volym 9, *Törnrosens bok*) på hur tilldelandet av nyckelord genomförts i Almqvistprojektet. Författarens samliga förnamn, initialer, efternamn samt utgivare och titel är de komponenter som getts möjlighet att bli sökbara.

```
<meta http-equiv="keywords" content ="Carl, Jonas, Love, C.J.L., Almqvist, Vitterhetsamfundet, Törnrosens bok, Duodesupplaga">
```

([Åter till texten](#))

12. I elementet <title> ingår förutom verkets titel ibland namnet *C.J.L. Almqvist*. En anledning till att man inom projektet valt att inkludera även författarnamnet i dokumentets titel kan vara att det som står inom title-elementet är det som visas om man t ex sparar dokumentets URL som ett bokmärke i webbläsaren. <title> placeras liksom <meta> inom <head></head>. Förutom att visa ett dokumentets titel återges dess innehåll även överst i webbläsarens fönster.

Exemplet nedan är från volym 8, *Törnrosens bok*.

```
<title>C.J.L. Almqvist. Samlade verk 8</title>
```

För att göra en författares namn sökbar utifrån dess egenskap som upphovsman, kan man formulera något liknande nedanstående:

```
<meta name="author" content="Almqvist">
```

([Åter till texten](#))

13. Nedanstående exempel visar hur Ibsenprojektet på egen hand har formulerat olika upphov (från *Catilina, 1850*):

```
<respStmt>
```

```
<resp>redigert i Ibsenkonkordansprojektet</resp><name>Harald Noreng</name>
```

```
<resp>Dataansvarig i Konkordansprojektet</resp><name>Knut Hofland</name>
```

```
<resp>Konvertering til TEI, mai 1998</resp><name>Espen S. Ore</name>
```

```
</respStmt>
```

([Åter till texten](#))

14. Vid uppmärkning av stycken används inom HTML <p> </p>. För att märka upp rubriker kan

<h1> modifieras så att det hos de flesta webbläsare är möjligt att bestämma storleken på rubrikerna. Inom Almqvistprojektet har rubrikerna märkts upp som <h1></h1> till och med <h5></h5>. <h1> utgör den största rubriken. ([Åter till texten](#))

15. För att märka upp roller i ett skådespel används elementen <castList> och <castItem>. <castList> innehåller en lista med roller, och <castItem> en enda roll. Exemplet nedan, från *Catilina*, 1850, beskriver rollen Lucius Catilina, en romersk adelsman.

```
<castList>
```

```
<castItemID="Catilina"><emph rend="spr">Lucius Catilina</emph>, romersk Adelsmand.
```

```
</castItem>
```

([Åter till texten](#))

Referenser

Berners-Lee, Tim, (1989): *Information Management : a proposal*. Geneve: CERN
URL: <http://www.w3.org/History/1989/proposal.html>, senast uppdaterad 1998, senast granskad 000727

Broady, Donald & Juliusson, John, (1997): *Introduktion till SGML* URL: <http://www.nada.kth.se/~broady/digitlit/sgml-intro/sgmlintr.html>, senast uppdaterad 970801, senast granskad 000727

Burnard, Lou, (1995): "What is SGML and how does it help?" // *Text encoding initiative : background and context* / edited by Nancy Ide and Jean Véronis. Dordrecht : Kluwer. S. 41- 50

CERN : European organization for nuclear research URL: <http://cern.web.cern.ch/>, senast uppdaterad 000329, senast granskad 000727

Giordano, Richard, (1995): "The TEI Header and the documentation of electronic texts" // *Text Encoding Initiative: background and context* / edited by Nancy Ide and Jean Véronis. Dordrecht: Kluwer. S. 75- 84

Guidelines for Electronic Text Encoding and Interchange (1994) / edited by C.M. Sperberg-McQueen and Lou Burnard. Chicago, Oxford: TEI P3 Text Encoding Initiative URL: <http://ftp.sunet.se/pub/etext/ota/TEI/>

Henrik Ibsens skrifter / kommentarer Espen S. Ore. Bergen: HIT-senteret, Universitetet i Bergen URL: <http://ibsentexts.hit.uib.no/nor-1.htm>, senast uppdaterad 000427, senast granskad 000727

Hockey, Susan, (1994): *Describing Electronic Texts: the Text Encoding Initiative and SGML* URL: <http://www.loc.gov/catdir/semidigdocs/hockey.html>, senast uppdaterad 941213, senast granskad 000727

Hockey, Susan, (1997): "Electronic Texts: the Promise and the Reality" // *American Council of Learned Societies Newsletter*. - Vol. 4: 4 URL: <http://www.acls.org/n44hock.htm>, senast uppdaterad 990713, senast granskad 000727

Ide, Nancy & Sperberg-McQueen, C.M., (1995): "The Text Encoding Initiative : its history, goals, and future development" // *Text Encoding Initiative : background and context* / edited by Nancy Ide and Jean Véronis. Dordrecht : Kluwer. S. 5-16

Jørgensen, Jon Gunnar, (1999): "Henrik Ibsen's writings" // *The complete Ibsen: existing databases, new digital archives, and new digital editions: linking to the future* [Konferensabstract från Digital Resources for the Humanities Conference 1999] URL: <http://www.kcl.ac.uk/humanities/cch/drhahc/drh/abst30.htm>, senast granskad 000727

Möller, Anders, (1994): *SGML: en introduktion till Standard Generalized Markup Language*. Lund: Studentlitteratur

Renear, Allen, (1997): "Out of Praxis : three (meta) theories of textuality" // *Electronic Text: investigations in method and theory* / Sutherland, Kathryn, (ed.). Oxford. S. 107-126

Sperberg-McQueen, C.M., (1991): "Text in the electronic age : textual study and text encoding, with examples from medieval texts" // *Literary and Linguistic Computing*. Vol.6, No.1. S. 34-46

Svedjedal, Johan, (1997): "Almqvist på Internet : om publicering av en textkritisk edition som digital hypertext " // *Tidskrift för litteraturvetenskap*. 26(1997):2. S. 60-74 URL: <http://g3.spraakdata.gu.se/vittsam/ALMINTER.html>

Svenska vitterhetssamfundet: *C.J.L. Almqvist samlade verk i hypertextversion* URL: <http://g3.spraakdata.gu.se/vittsam/almqvist.html>, senast uppdaterad 991001, senast granskad 000727

([Åter](#) till början av artikeln)

© Charlotte Wallberg 2000

[Åter till Human IT 2-3/2000](#)