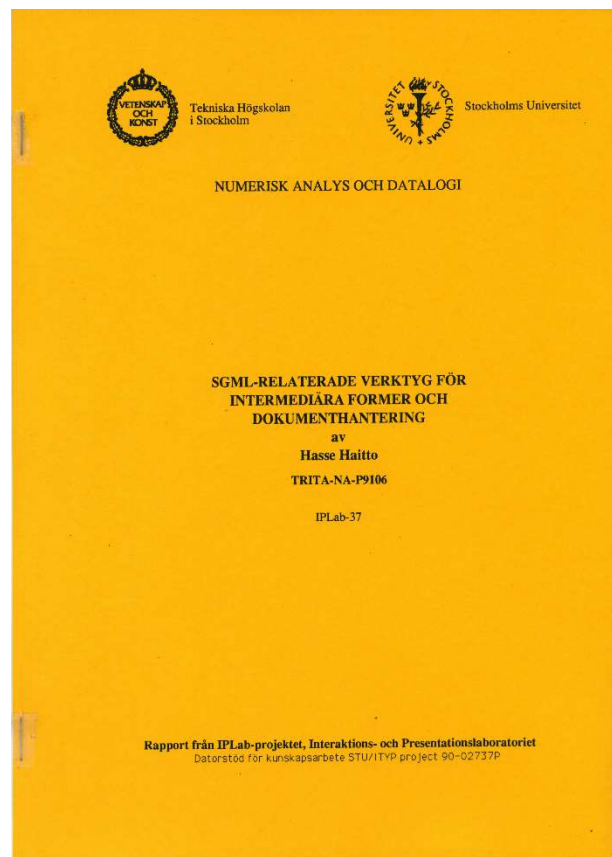


## Faksimil

D. Broady, "Förord", pp. 2f i Hasse Haitto, *SGML-relaterade verktyg för intermediära former och dokumenthantering*. Rapport från Datorstöd för kunskapsarbete, STU/ITYP projekt 90-02737P, IPLab-37. Institutionen för numerisk analys och datalogi, KTH, Stockholm, 1991.



Datorstöd för kunskapsarbete  
STU/ITYP projekt 90-02737P

# **SGML-relaterade verktyg för intermediära former och dokumenthantering**

## Förord

Denna rapport är av teknisk natur, men berör problem av intresse för många som arbetar med komplicerade dokument. Därför kan några förklarande ord vara på sin plats.

Vissa yrkesgrupper — låt säga en grupp handläggare, ett lärarlag, en tidskriftsredaktion, en forskargrupp, en studerandegrupp — har behov av att bygga upp och arbeta med en egen "decentraliserad" informationsbank, en dokumentbas. Inom det STU-finansierade projektet "Datorstöd för kunskapsarbete" studerar vi metoder för att organisera och hantera sådana dokumentbaser.

I en praktisk tillämpning kan en tänkbar arbetsgång vara den följande.

- Dokumenten kan ha skapats av någon av gruppens medlemmar, med hjälp av de vardagliga arbetsverktygen (text- och bildbehandlingsprogram, struktureditor, hypertextsystem etc). De kan även ha hämtats utifrån, skannats in från ett pappersdokument, eller hämtas från en extern databas, från något nät, från CD-ROM, från ett bibliotekssystem etc.
- Vid behov måste dokumenten först överföras till ett lämpligt hanterbart format.
- I nästa steg överföres dokumenten till ett intermediärt format. Poängen med ett intermediärt format är att alla dokument, oavsett hur de skapats och hur de senare kommer att användas, föreligger i ett och samma format. Därmed kan de sammanföras i en enda dokumentbas, och de är omedelbart tillgängliga för export när de skall användas i andra sammanhang. Vi har valt att arbeta med SGML-format (Standard Generalized Markup Language), som har den stora fördelen att dokumentens innehållsliga struktur i stor utsträckning bevaras (för att välja ett enkelt exempel: i en vanlig sekventiell text bevaras rubriknivåer, fotnoter etc).
- En viss förarbetning kan vara erforderlig: versionskontroll, "remissbehandling", revidering.
- Därefter lagras dokumenten i dokumentbasen.
- Olika redskap behövs för att medarbetarna skall kunna skaffa sig en överblick över dokumentbasens innehåll.
- Dokumentbasen skall vara omedelbart tillgänglig från medarbetarnas arbetsstationer. Här behövs således automatiska översättningsrutiner som låter medarbetarna importera dokumenten till sina ordbehandlingsprogram, hypertextsystem etc.
- Vidare skall dokumenten kunna matas ut till exempelvis textbehandlings-, bildbehandlings- eller layout-program för vidare befordran till utskrift eller publicering. Ett typiskt användningsområde är när ett lärarlag skapar skraddarsydda läromedel för sina elever eller en elevgrupp skriver ett specialarbete.

Det är givetvis inte nödvändigt att varje arbetsplats är utrustad med alla de nämnda faciliteterna, men på många håll kan medarbetarna ha behov av att hantera åtminstone några av stegen i den ovan skisserade arbetsgången.

För att ett dylikt system skall fungera, och vara mer till hjälp än till hinder, krävs att så många rutiner som möjligt är automatiserade. Det är också väsentligt att systemet är integrerat med medarbetarnas vardagliga arbetsmiljö. Man skall t ex kunna använda sitt vanliga textbehandlingsprogram för att författa dokument som lagras i basen eller för att bearbeta dokument som hämtas upp ur basen. Över huvud taget skall medarbetarna ägna möda åt sina egentliga arbetsuppgifter, inte åt teknikaliteterna. I denna rapport, som härrör från det inledande skedet av arbetet med projektet "Datorstöd för kunskapsarbete", har Hasse Haitto sammanfört information om i dag tillgänglig programvara som kan fungera som beståndsdelar i ett ramverk för arbetet med dokumentbaser. Tyngdpunkten ligger vid överföring till och från SGML-format, men rapporten behandlar även verktyg för bearbetning av SGML-kodade dokument, versionshantering, överföring från hypertext till sekventiell text och vice versa, m.m.

Donald Broady, projektledare

## Innehållsförteckning

Förord .....	2
Innehållsförteckning .....	4
Introduktion.....	6
Översättningsprogram.....	7
Kommersiell programvara .....	7
FastTAG .....	7
TextTagger Workstation Edition.....	8
Mark-It.....	9
SGML Translator DCF Edition .....	10
XGML Translator .....	11
Fritt tillgänglig programvara .....	12
Glammar.....	12
ICA Toolset .....	13
NIST NCSL OSE SGML parser.....	14
Editorer och presentationsverktyg .....	15
DynaText Indexer & Browser .....	15
Grif .....	17
SoftQuad Publishing Software .....	18
SoftQuad Author/Editor .....	19
SoftQuad RulesBuilder .....	20
SGML TextWrite OS/2 Edition.....	21
SGML TextWrite Tools OS/2 Edition.....	22
Xerox DocuBuild för PC .....	23
XGML CheckMark .....	24
Write-It .....	25
Verktyg för versionshantering.....	26
DocuComp .....	26
RCS .....	27
Ett prototypamverk för dokumenthantering.....	28
Import av dokument .....	28
Remissbearbetning .....	28
Initial analys och redigering .....	29
Komplex analys och presentation .....	29
Export av dokument .....	29
Referenser .....	30
Appendix.....	31
Addendum 920206.....	31

## Introduktion

I [Broady90] utmålas en trolig utveckling av datorstött författande och samarbete i utbildning. Ett scenario för framtida datorstött samarbete bör utgå från att såväl datorstöd som program kommer att växla mycket beroende på arbetsplats och -uppgifter. Det är likväl uppenbart att ett välfungerande samarbete kräver att personliga preferenser i val av verktyg beaktas. Den därvid uppkomna variationen i fråga om dokumentformat, hårdvarubindningar m m skapar en del problem: Datorstött material — såsom strukturerade texter eller hypertext — måste kunna gå en remissrunda eller motsvarande, flyttas mellan olika system, och enkelt uppdateras eller redigeras.

Ett sätt att hantera de aktuella problemen är att skapa en gemensam intermediär form som bevarar strukturen hos materialet. En möjlig intermediär form är dokument som kodas i enlighet med SGML (Standard Generalized Markup Language), en ISO-standard [SGML86] för dokumentbeskrivning. En central organisation för arbete med denna inriktning är *Chameleon*-konsortiet, som utvecklar verktyg (s k ICA Toolset) för att understödja flyttbarhet i dokumenthantering [Mamrak90].

I en strävan att skapa en aktuell (och någorlunda fullständig) översikt av SGML-relaterade tillämpningar, både kommersiella och fritt tillgängliga, sammanställs här material från ett antal källor. En del av programmen har vi provat själva. Ett av syftena med översikten är att identifiera direkt tillgängliga "off-the-shelf" programvaror, där de olika komponenterna sätts ihop som byggstenar. Sammanställningen indelar tillämpningarna i olika klasser beroende på användningsområde:

*Översättare:* För att transformera dokument från ett format till ett annat krävs olika slags verktyg. De transformationer som är aktuella är t ex konverteringar till och från SGML, eller mellan skilda representationer.

*Editorer:* SGML-markeringen av dokument bör understödjas av textbehandlingsprogrammet. Ideellt tillåter editorn interaktiv och inkrementell validering under författandet, gärna med någon form av *wysiwyg*-återkoppling som svarar mot dokumenttypdefinitionen.

*Presentationsverktyg:* Eftersom SGML-dokument är väl strukturerade går de att bearbeta på sätt som skiljer sig från traditionell dokumenthantering. Markeringen gör det t ex möjligt att navigera i informationsmassan, liksom att skapa dynamiska elektroniska böcker. En klass av presentationsverktyg är formaterare, vilka även kan hänföras till översättningsverktygen.

Genom att SGML-standarden i första hand efterfrågats av tung industri och militären har de flesta av programvarorna en koppling mot CALS (Computer-aided Acquisition and Logistics Support).<sup>1</sup> Denna översikt tar snarare upp andra aspekter på programmen.

Den som vill läsa mer om SGML refereras till [Goldfarb90], som är en gedigen och komplett referens till standarden. Charles Goldfarb har lett utvecklingen av SGML och fortsätter nu arbetet med standardiseringen av andra medier än text med hjälp av SGML. ANSI-kommittén X3V1.8M utvecklar SMDL (Standard Music Description Language,

---

<sup>1</sup>CALS är en uppsättning av standarder som täcker olika aspekter av dokumentbehandling (t ex grafikformat, överföring etc) och där SGML används för markering av text.

ANSI projekt X3.542-D) som en utvidgning av SGML, liksom HyTime, (Hypermedia/ Time-based Structuring Language, ANSI projekt X3.749-D).

## Översättningsprogram

För att överbrygga klyftan mellan olika dokumentrepresentationer behövs programvara som översätter dokumenten från ett format till ett annat. Förutom att fungera som filter bör programmen även hjälpa till med att skapa en SGML-kodad representation när källdokumentet är ren text, t ex när den skapats med OCR-läsare. Nedan följer en genomgång av sådana program och likartade verktyg.

### Kommersiell programvara

#### FastTAG

*Avalanche Development Company*  
947 Walnut Street, Boulder, Colorado 80302, U.S.A.  
Tel: + (303) 449-50323 Fax: + (303) 449-3246

FastTAG är en produkt för automatiserad översättning mellan olika representationer. Den klarar av att läsa följande dokumentformat: ASCII, WordPerfect, Microsoft Word, KDOC, Interleaf ASCII markeringar, DCA/RFT, OCR/ICR-filer, utskriftsfiler (Epson, IBM, Diablo, ASCII List 1403), samt Calera PDA och Shaffstall filer. Dessa format är alltså fördefinierade och direkt tillgängliga.

Det inlästa materialet markeras sedan i enlighet med instruktioner som definierats av användaren, som alltså får "lära upp" programvaran i ett inledningsskede. Dessa instruktioner specificerar bl a hur olika textobjekt såsom stycken, rubriker, listor, tabeller, tabellceller, sidhuvuden, och operiodiska textelement som korsreferenser, understruken text, kursiverad text o s v skall markeras vid utmatning. Utmatningen från FastTAG är användardefinierbar, så produkten kan i princip understödja godtyckligt markeringsformat, t ex en bestämd SGML-dokumentmall eller något annat godtyckligt format.

Enligt tillverkaren är översättningsprocessen intelligent och baserad på textanalys, i princip analog med mönstermatchning. Detta innebär att skannade dokument markeras i enlighet med deras visuella intryck, produkten känner s a s igen den logiska strukturen.

*Understödda hårdvaruplattformar/systemkrav:*

IBM PC/AT, 640K, MS-DOS 3.0  
Sun-3, Sun-4, Sun SparcStation under Sun OS 4.0  
Apollo DN-3000 under SR 10.0  
DEC VAX, VMS 5.1, VAX Ultrix 4.0, RISC Ultrix 4.0

*Prisbild:*

FastTAG kostar \$3185 för PC-versionen, \$4030 för de övriga plattformarna. Avalanche Development Company ger f n inga högskolerabatter.



## TextTagger Workstation Edition

*IBM Svenska AB*

*S-163 92 Stockholm*

*Tel: (08) 793 10 00 Fax: (08) 793 10 20*

IBM Publishing Systems TextTagger Workstation Edition — hädanefter benämnd TextTagger — är avsedd till att översätta icke-SGML kodade dokument till CALS SGML. TextTagger tolkar format och struktur i textfiler och skapar en SGML-märkerad fil. Källdokumenterna kan t ex vara dokument från ordbehandlare, ren text (ASCII), inskannade dokument, eller filer avsedda för utskrift (bl a för IBM 1403 skrivare).

TextTagger fungerar som ett första steg i konverteringsprocessen och den genererade utmatningen kan sedan valideras och ev modifieras.

*Understödda hårdvaruplattformar/systemkrav:*

IBM PS/2, OS/2 eller MS-DOS

*Prisbild:*

TextTagger kostar ca 43 000:- SEK.

*Anmärkning:*

Det finns en motsvarande tillämpning under VM och MVS kallad *ProcessMaster TextTagger Feature*. Förutom CALS SGML kan den generera utfiler med märkord enligt BookMaster eller SGML Starter Set.<sup>2</sup>

---

<sup>2</sup>Se SGML Translator DCF Edition.

## Mark-It

*Yard Software Systems  
Avonbridge House, Bath Road,  
Chippenham, Wiltshire SN15 2BB  
Tel: + (0249) 656194 Fax: + (0249) 655723*

Mark-It är en huvudsakligen batchorienterad SGML-parser som utgående från SGML-märkerade dokument kan generera användardefinierad utmatning före och efter varje källkodselement. Mark-It har ett eget språk för att styra denna process: Instruktioner insprängda som SGML-kommentarer i LINK-deklarationer tolkas av Mark-It för att styra utmatningen. Instruktionerna skiljs från vanliga kommentarer genom att de inleds med nyckelordet APP. Mark-It kan även använda reguljära uttryck i parsing-processen. På så sätt kan man översätta källdokument till måldokument med hjälp av kontextberoende, SGML-baserad markering.

Mark-It understöder alla tillval i SGML och har även en del utvidgningar enligt tillägget (s k *amendment 1*) som föreslagits till ISO-standarden.

*Understödda hårdvaruplattformar/systemkrav:*

IBM PC/AT, 640K, MS-DOS 2.11 eller högre  
DEC VAX-VMS, Intergraph

*Prisbild:*

MS-DOS versionen är fritt tillgänglig för akademiskt bruk.

*Anmärkning:*

Förutom ovannämnda plattformar finns Mark-It tillgänglig till andra (ospecificerade) miljöer, bl a Unix, och portas på begäran, mot betalning. Mark-It är erkänt en av marknadens mest fullständiga parsers.

## SGML Translator DCF Edition

*IBM Svenska AB*

*S-163 92 Stockholm*

*Tel: (08) 793 10 00 Fax: (08) 793 10 20*

Tillämpningen SGML Translator DCF Edition fungerar under VM och MVS. Den består av en validerare, som kontrollerar att ett dokument följer en valfri dokumenttypdefinition. Nästa steg är en översättare som konverterar det SGML-kodade dokumentet till IBMs dokumenthanteringssystem DCF/GML enligt en översättningstabell. Slutligen ingår en tillämpning för matematiska formler att användas med SGML Starter Application, som levereras med systemet.

*Understödda hårdvaruplattformar/systemkrav:*

VM/SP release 5 eller 6, VM/SP HPO release 5, VM/XA SP release 2 eller VM/IS release 6, samt MVS (MVS/XA alt MVS/ESA).

*Prisbild:*

Grundlicens kostar ca 150 000 sek exkl. moms, universitetsrabatten är ca 40%.

*Anmärkning:*

SGML Translator DCF Edition levererar utmatning till IBMs Document Composition Facility (DCF), vilken formaterar dokumentet på papper. Trots namnet är dock SGML Translator inte bundet till DCF utan den är en generell översättare där instruktionerna kan skrivas i REXX eller C.

## XGML Translator

*Software Exoterica Corporation  
383 Parkdale Avenue, Suite 406  
Ottawa, Ontario, CANADA K1Y4R4  
Tel: + (613) 722-1700 Fax: + (613) 722-5706*

XGML Engine är kärnan i Software Exotericas SGML-produkter. Bland dessa produkter är XGML Translator en översättare som huvudsakligen utgår från SGML-märkade dokument men som även med fördel kan användas på andra sorters dokument, t ex oformaterad text (ren ASCII) eller dokument formaterade med sedvanlig ordbehandlare. Programmet består av två delar, en kompilator och en interpretator.

XGML Translator är i princip likvärdig med FastTAG beträffande funktionalitet, med den viktiga skillnaden att XGML dessutom samtidigt utför SGML-validering. När källdokumentet inte är ett SGML-kodat dokument översätter XGML Translator det först till en intermediär, SGML-baserad form för vidare bearbetning. Detta SGML-kodade mellansteg är även tillgängligt för användaren.

Översättningsprocessen bygger på mönstermatchning och styrs av instruktioner skrivna av användaren, i ett språk som kallas XTRAN (Se [XGML90]). Ett XTRAN källprogram kompileras till SGML som sedan interpreteras tillsammans med en dokumenttypdefinition (DTD) av interpretatorn. XGML Translator har använts till att översätta SGML-kodade dokument till bl a XICS, TROFF, Intergraph, T<sub>E</sub>X, RTF; översättning åt motsatt håll, liksom mellan olika format, går givetvis lika bra.

En del andra företag, såsom Intergraph<sup>3</sup>, Xerox, ArborText, Frame m fl har skaffat licens för att bygga in XGML Engine i sina egna produkter.

### *Understödda hårdvaruplattformar/systemkrav:*

IBM PC/AT/386, MS-DOS, OS/2, MS Windows 3.0, SCO Unix  
Macintosh (4MB)  
Sun  
DEC VAX  
Intergraph

### *Prisbild:*

Software Exoterica ger f n inga högskolerabatter. Enanvändare, d v s versionen för PC (MS-DOS, OS/2, MS Windows 3.0) och Macintosh, betalar \$3995. Fleranvändarsystem under Unix, Sun, VAX m fl behöver speciella licenser och prissätts mellan \$14995 och \$19995 (samt uppåt).

### *Anmärkning:*

Förutom ovannämnda plattformar finns XGML Translator enligt uppgift tillgänglig till andra (ospecificerade) miljöer.

---

<sup>3</sup>Intergraph har bl a CALS-programvaran DP/SGML Publisher.

## **Fritt tillgänglig programvara**

Förutom de kommersiella produkterna finns ett antal fritt tillgängliga och/eller olicensierade program (i stil med GNU-projektets eller X-Windows konsortiets program).

### **Glammar**

*phoibos.cs.kun.nl:/pub/GLASS/glammar.tar.Z [131.174.81.1]  
ericv@phoibos.cs.kun.nl (Eric Vos)*

Universitet i Nijmegen, Holland, har utvecklat Glammar, som är ett verktyg för att generera översättare (text to text translator generator) under Unix. Det kan hämtas via anonym ftp på adress "phoibos.cs.kun.nl" [131.174.81.1]. Aktuell version är en beta-version. Dokumentationen är ca 100 sidor lång och levereras som en dvi-fil.

Glammar kan bäst beskrivas som ett syntaxstyrt översättningsverktyg, eller som ett språk för att transformera text på basis av mönsterigenkänning. Glammar-beskrivningar bygger på en utbyggd contextfri grammatik kallad EAG och resulterar i C-program, vilka utför själva översättningen. Enligt upphovsmannen är Glammar tillämpbar på ett brett spektrum av områden, inklusive naturliga språk.

## ICA Toolset

*S.A. Mamrak & C.S. O'Connell,  
Department of Computer and Information Science,  
The Ohio State University,  
2036 Neil Avenue Mall,  
Columbus, Ohio 43210, U.S.A.  
Tel: + (614) 292-5813*

Chameleon-konsortiet utvecklar verktyg för översättning mellan olika dokumentslag. Den huvudsakliga målsättningen har varit att automatiskt generera översättare till och från intermediära former, vilket har resulterat i tre grupper av verktyg. Dessa grupper svarar mot översättningsprocessens olika faser: 1) specifikation av intermediära former, 2) konstruktion av översättare, och 3) bruk av översättare. Verktygen betecknas under samlingsnamnet ICA Toolset, där ICA står för *Integrated Chameleon Architecture*. ICAs intermediära former bygger på kontextfria grammatiker, vilka automatiskt översätts till SGML-dokumenttypdefinitioner.

Den första fasen består av tre verktyg för att skapa intermediära former. Det första av dessa, *DeveGram*, understödjer skapandet av formerna, det andra verktyget är en editor för redigering av instanser svarande mot formen, och det tredje verktyget validerar SGML-dokumenttypdefinitioner som specificerats externt, d v s utom systemet. *DeveGram* kan generera syntax svarande mot Yacc eller SGML-dokumenttypdefinitioner. Dokumenttypdefinitioner genererade inom systemet är garanterat korrekta.

Den andra fasen består i att specificera avbildningen mellan en dokumentkodning (t ex *TEX*) och den intermediära formen. Avbildningen måste specificeras åt bägge riktningarna, till och från den intermediära formen, samt för varje dokumentslag i aktuellt kodningsförfarande. Översättarna består av två komponenter: en *igenkännare* för t ex nyckelord och viktiga strängar, och en *parser* som kontrollerar att funna symboler är legala i sammanhanget.

Den sista fasen omfattar egentligen översättningen från en given kodning till den intermediära formen, eftersom motsatt riktning alltid kan göras automatiskt. Vid översättning från en kodning uppstår vanligtvis tvetydiga situationer som kräver någon form av åtgärd eller beslut. Interaktionen hanteras smidigt i en fönster/mus-omgivning.

*Understödda hårdvaruplattformor/systemkrav:*

ICA-verktygen är skrivna i C och har implementerats i sin helhet på Sun Sparc arbetsstationer under Unix och X-Windows. De är fortfarande i prototypstadiet, vilket innebär att dokumentationen är bristfällig och att mjukvaran troligen är behäftad med fel. Inom konsortiet används ICA Toolset till att bygga översättare för dokument kodade med Scribe, L<sup>A</sup>T<sub>E</sub>X och troff.

*Prisbild:*

ICA Toolset kommer att vara fritt tillgängligt (enligt samma villkor som X-Windows konsortiet tillhandahåller X-Windows).

## NIST NCSL OSE SGML parser

*swe.ncsl.nist.gov:/sgml/sgml.cpio*

*Jim Heath*

*NIST Bldg. 225, Rm. B-266*

*Gaithersburg, MD 20899, U.S.A.*

*Tel: + (301) 975-3350*

Vid National Institute of Standards and Technology (NIST), National Computer System Laboratory (NCSL), Office Systems Engineering (OSE) Group, har Jim Heath skrivit en SGML-parser. C-källkoden är fritt tillgänglig via anonym ftp från *swe.ncsl.nist.gov*. Under filkatalogen */sgml* finns två filer:

```
-rw-r--r--  1 ftp           1731 Aug 25  1989 sgml.cheater
-rw-r--r--  1 ftp        652800 Aug 25  1989 sgml.cpio
```

Dokumentation saknas tyvärr och programmen har uppenbarligen inte uppdaterats sedan 1989. Filen *sgml.cheater* ger en handledning i hur man skall packa upp arkivet *sgml.cpio* och räknar upp vad detta innehåller; förutom en README-fil finns sju kataloger:

```
'incs'      --- include files
'parse'     --- parse  module source amd sample DTD
'parse1'    --- parse1 module source and Makefile
'parse1a'   --- parse1a module source
'parse2a'   --- parse2a module source
'parse2b'   --- parse2b module source and Makefile
'parse3'    --- parse3  module source and Makefile
```

För den hugade programmeraren finns alltså möjligheten att implementera en egen SGML-baserad översättare utgående från några existerande parserprogram. Tipset beträffande mjukvaran lämnades i Usenet-konferensen *comp.text.sgml*.

## Editorer och presentationsverktyg

Genom att SGML är en såpass ung standard har det inte ännu uppstått en stor mängd editorer som direkt bygger på SGML. [Herwijnen90] tar upp ett antal punkter som är önskvärda hos SGML-baserade editorer; bl a tar han upp aspekten om navigering med hjälp av SGML-strukturen och behovet av att kunna slå SGML-valideringen av och på efter behov: Valideringen bör inte inkräkta på författandet.

I sammanhanget nämns även möjligheten att utföra SGML-kodning av dokument med vanliga editorer, för att sedan validera dem med en parser. I vissa textbehandlare kan man skapa mallar (ev i kombination med gömd text) som implementerar delar av en SGML-dokumenttypdefinition, vilket väsentligt underlättar konverteringsprocessen. En viktig fördel med denna lösning är att författaren tillåts använda sin egen textbehandlare. Konverteringsprocessen kan givetvis alltid utföras med hjälp av av översättningsverktyg.

### DynaText Indexer & Browser

*Electronic Book Technologies, Inc.,  
One Richmond Square,  
Providence, Rhode Island 02906, U.S.A,  
Tel: + (401) 421 - 9550 Fax: + (401) 421 9551*

DynaText är ett SGML-baserat presentations- och navigeringsverktyg som skapar elektroniska böcker ur SGML-dokument. DynaText körs på Sun 4 arbetsstationer under X Windows och har två beståndsdelar: Ett indexprogram (*Indexer*) som läser SGML-källdokument och grafik i ett antal olika format, samt en "bläddrare" (*Browser*) som tillåter användare att utforska och läsa de resulterande elektroniska böckerna. Man skulle kunna säga att DynaText är en andra generationens SGML-programvara genom att den visar på nya sätt att arbeta med SGML-kodade dokument. Den kombinerar koncept från hypertextsystem, fulltext sökning, idébehandling och *wysiwyg* skärmpresentation.

Programmet *DynaText Indexer* bygger upp ett fulltext index ur indata, extraherar SGML-strukturen för innehållsförteckningen samt fastslår länkarna till tabeller, figurer, korsreferenser och fotnoter. Grafik kan importeras i ett flertal olika format, bl a CCITT, TIFF, PICT eller Sun raster format. Systemet använder internt pbm-formatet men levereras med en delmängd av Jeff Poskanzers pbmplus grafikpaket, som är en fritt tillgänglig och kraftfull samling program för att översätta mellan olika grafikformat och bearbeta bilder<sup>4</sup>. Fotnoter, bilder o dyl presenteras i egna fönster när läsaren klickar på motsvarande referens.

Med bläddraren kan man sedan utforska den resulterande boken. Användare kan navigera genom att klicka i den dynamiska innehållsförteckningen som expanderar eller krymper ihop motsvarande avsnitt såsom i en struktureditor eller idébehandlare. Flera fönster mot boken kan samtidigt vara öppna, i vilka man använder olika mallar för skärmpresentationen av boken. Formateringen sker dynamiskt och automatiskt utgående från fördefinierade mallar. Slutligen tillåter fulltextindexet snabba sökningar av böckerna, där antalet förekomster av ett sökt ord även visas på motsvarande ställe i innehållsförteckningen.

---

<sup>4</sup>Den senaste versionen är alltid tillgänglig via anonym ftp ([expo.lcs.mit.edu:contrib/pbmplus.tar.Z](ftp://expo.lcs.mit.edu/contrib/pbmplus.tar.Z) eller <ftp://ee.lbl.gov/pbmplus.tar.Z>).



DynaText har ett programmeringsgränssnitt benämnt ERTL (EBT runtime library). ERTL kommunicerar med DynaText genom RPC-nätverksprotokoll, men som ur programmerarens synvinkel fungerar helt transparent, som om användaren själv interagerat med programmet via det vanliga gränssnittet. Diverse små hjälpmedel ingår i programpaketet, som filter för att modifiera SGML-filer, räkna märkord o s v. En editor för malleditering finns i en betaversion.

En planerad utbyggnad av Browsern kommer att låta läsaren att namnge, editera och spara stigar i dokument; sätta in bokmärken för att markera passager, och även annotera böckerna. Browsern kommer att spara en logg av hur dokumentet har navigerats.

*Understödda hårdvaruplattformar/systemkrav:*

Sun 4 under SunOS 4.03 eller högre. X-Windows med valfri ICCCM fönsterhanterare.

*Prisbild:*

DynaText Indexer \$10000, DynaText Browser för fem simultana användare \$2500. Akademisk rabatt kan erhållas.

*Anmärkning:*

En version av Browsern för PC (under MS Windows) resp. Macintosh är under utveckling. PC-versionen sägs bli klar kring första kvartalet 1991, Macintosh versionen i en oviss framtid.

## Grif

*Gipsi S.A.,  
Immeuble "Le Florestan",  
2, Boulevard Vauban, BP 268,  
78053 St Quentin en Yvelines Cedex, FRANCE  
Tel: + (1) 30 60 75 00 Fax: + (1) 30 60 75 90*

Grif är en interaktiv *wysiwyg*-editor för strukturerade dokument. Den lagrar dokumentens struktur internt med en SGML-lik markering och representerar dokumenten genom mallar (presentationsregler). Man kan ha flera mallar till en och samma dokumenttyp. Översättning från det interna formatet till andra format utförs genom översättningsmallar. Instruktionerna till dessa skrivs i ett eget språk kallat T; analogt med presentationsmallarna är det möjligt att ha ett flertal översättningsmallar för ett givet dokumentslag. Översättning till SGML finns inbyggd i systemet.

Grif körs på Unix arbetsstationer under X Windows. Dokument skrivs ut med hjälp av PostScript: De tillgängliga teckensnittsfamiljerna är Times, Helvetica, Courier och Symbol (som innehåller matematiska symboler och det grekiska alfabetet). Teckensnittens grad går från 6 till 72 punkter Grif i stilarna rak, kursiv eller fet. Grif har f n ingen inbyggd grafisk editor men kan importera grafik (CGM, CCITT, gr4, TIFF) och filer kodade med de vanligaste förekommande sidbeskrivningsspråken (bl a PostScript och HPGL).

Grif ger bra stöd för tekniska dokument med t ex tabeller och matematiska formler. Tabeller editeras valfritt på cellnivå, per kolumn eller rad. Rader och kolumner kan vara enkla eller sammansatta, nästlade; celler får innehålla text, bilder, korsreferenser och formler. Formler skapas och editeras *wysiwyg* som övrigt material. Storleken hos, och placeringen av, symboler såsom integraler, kvadratroter, bråkstrecker m m anpassas automatiskt till innehållet.

*Understödda hårdvaruplattformar/systemkrav:*

Sun, IBM RS/6000 m fl.

*Prisbild:*

Det akademiska priset för Grif är 12000 FF.

*Anmärkning:*

Grif härstammar från en prototyp med samma namn utvecklad vid INRIA.

## SoftQuad Publishing Software

*SoftQuad Inc.,  
720 Spadina Avenue,  
Toronto, CANADA, M5S 2T9  
Tel: + (416) 963-8337 Fax: + (416) 963-9575*

SoftQuad Publishing Software är en uppsättning dokumentbehandlingsverktyg som bygger vidare på UNIX Documenter's Workbench (DWB) från AT&T. Den är avsedd för batchformatering av dokument i produktionsmiljö. Den viktigaste beståndsdel i DWB version 2.0 är *ditroff* (device independent *troff*), som kompletteras med diverse program för hantering av bl a tabeller (*tbl*), matematiska formler (*eqn*), grafik (*pic*). Softquads miljö omfattar förbättrade versioner av alla dessa program från DWB och nya diagnostik- (*trace*) och avlusningshjälpmedel (*dbg*) för dokumenthantering.

Utmatningen från SoftQuad Publishing Software understöder laserskrivare och fototypsättare som tolkar PostScript, imPRESS och ytterligare ett antal format. För kunder av systemet tillhandahåller SoftQuad skräddarsydda rutiner baserade på SGML-dokumenttypdefinitioner för översättning till kundspecificerat format.

### *Understödda hårdvaruplattformar/systemkrav:*

Programpaketet finns till ett brett spektrum av såväl Unix-versioner (t ex AIX, A/UX, BSD 4.2, System V, Xenix) som -plattformar (Altos, AT&T, DEC, IBM PC (AT/RT), Macintosh II, Motorola, Pyramid, Sun m fl). Mjukvaran tar upp ca 3 MB skivminne och kräver minst 1 MB RAM.

### *Prisbild:*

SoftQuad ger 25% högskolerabatt. Priset varierar från \$995 till \$7500 beroende på plattform.

### *Anmärkning:*

En previewer under X11 finns till X-terminaler med följande värdplattformar: Sun 3, Macintosh II, Sequent Symmetry och Interactive Systems. Kostnaden är \$300.

## SoftQuad Author/Editor

*SoftQuad Inc.,  
720 Spadina Avenue,  
Toronto, CANADA, M5S 2T9  
Tel: + (416) 963-8337 Fax: + (416) 963-9575*

SoftQuad Author/Editor är en textbehandlare och struktureditor avsedd för SGML-formaterade dokument. Den levereras antingen med regler som motsvarar den dokumenttypdefinition som AAP (American Association of Publishers) har tagit fram, eller med regler som uppfyller CALS.

För att skapa nya regler (förkompilerade dokumenttypdefinitioner) måste man skaffa SoftQuad RulesBuilder till Macintosh eller SoftQuad BuildRules till DOS/Unix, eller (tillverkarens rekommendation) Mark-It<sup>5</sup>. SoftQuad och Yard Software Systems, som gjort Mark-It, samarbetar och representerar varandra i Nordamerika resp. Europa.

SoftQuad Author/Editor på Macintosh har kvasi-wysiwyg egenskaper på så sätt att användaren väljer märkorden ur menyer och motsvarande element formateras automatiskt enligt mallar. Man kan välja att ha två fönster, ett som visar dokumentets SGML-struktur, och ett som visar dokumentinnehållet. Om man väljer ett element i strukturfönstret kan man med ett par musclick expandera motsvarande del i det andra fönstret. Det går alltså att förflytta sig i dokumentet med hjälp av strukturen. Beträffande valideringen så kan denna ske interaktivt eller i ett pass efter författandet.

De filer som produceras blir rena ASCII-filer, men det går även att importera icke-SGML material såsom bilder (MacDraw, MacPaint, PostScript, inkapslad PostScript och PICT) för vidarebefordran åt formateraren. Det är möjligt att skriva gömd text och definiera egna makron; avancerade sökfunktioner med bl a mönstermatchning ingår. Förutom dessa egenskaper är Microlytics' program för stavningskontroll inbyggd. En besynnerlig begränsning är att tabeller får vara högst elva kolumner breda.

### *Understödda hårdvaruplattformar/systemkrav:*

Version 1.1: Apple Macintosh Plus, Macintosh SE, Macintosh II inkl. A/UX. (1MB)  
Version 2.0: MS-DOS, Sun 3, Sun 4, Sun Sparcstation, DECstation 3100.

### *Prisbild:*

SoftQuad ger 25% högskolerabatt. Priset varierar beroende på plattform: \$495 för Macintosh, \$695 för MS-DOS, och från \$995 till \$1695 på Unix-plattform. CALS-versionen kostar ungefär mellan \$500 och \$1000 mer.

---

<sup>5</sup>Detta torde innebära att Mark-It och SoftQuad använder samma format på de kompilerade typdefinitionerna.

## SoftQuad RulesBuilder

*SoftQuad Inc.,  
720 Spadina Avenue,  
Toronto, CANADA, M5S 2T9  
Tel: + (416) 963-8337 Fax: + (416) 963-9575*

SoftQuad RulesBuilder är en interaktiv editor och parser för redigering och författande av SGML-typdefinitioner på Macintosh. Dessa typdefinitioner kontrolleras av RulesBuilder som kompilerar om dem till regler vilka används av SoftQuad Author/Editor. Liksom Author/Editor på Macintosh har RulesBuilder kvasi-wysiwyg egenskaper. Motsvarande tillämpning till DOS och Unix kallas för BuildRules.

*Understödda hårdvaruplattformar/systemkrav:*

Macintosh, MS-DOS, Sun 3, Sun 4, Sun Sparcstation, DECstation 3100.

*Anmärkning:*

Det framgår inte ur dokumentationen huruvida man kan flytta regler kompillerade i DOS/Unix miljö till Macintosh-versionen av Author/Editor (eller vice-versa).

*Prisbild:*

SoftQuad ger 25% högskolerabatt. Priset varierar beroende på plattform: \$995 för Macintosh, \$1195 för MS-DOS, och \$1995 på Unix-plattform.

## SGML TextWrite OS/2 Edition

*IBM Svenska AB*

*S-163 92 Stockholm*

*Tel: (08) 793 10 00 Fax: (08) 793 10 20*

IBM Publishing Systems SGML TextWrite OS/2 Edition — hädanefter benämnd TextWrite — är en editor för att redigera och validera SGML-märkerad text. TextWrite körs under Presentation Manager och tillåter att man har flera filer öppna samtidigt eller att man har fler fönster mot samma dokument. Validering mot den använda dokumenttypdefinitionen sker kontinuerligt. TextWrite har mallar som används för att avbilda SGML-strukturen på skärm och vid utskrift på papper. På skärmen kan man välja mellan att ha texten visad med märkord eller formaterad utan märkord; vid formatering används bl a typsnitt för att markera strukturen, märkorden expanderas och eventuella indrag och numreringar utförs. Om viss annan programvara finns installerad kan även illustrationer visas i separata fönster genom att man klickar på illustrationens namn i dokumentet.

Förutom dokumenttypdefinitioner för CALS ingår motsvarande för DCF GML Starter Set och ytterligare en DCF-tillämpning och en för matematiska formler (enligt ISO, teknisk rapport 9573). Import av ASCII-filer kan göras, och utskrifter kan även göras på PostScript-skrivare.

*Understödda hårdvaruplattformar/systemkrav:*

IBM PS/2, OS/2 Standard Edition 1.2 alt. Extended Edition 1.2 med PM 1.2, minst 5MB RAM och 5MB skivminne.

*Anmärkning:*

Programmet kan anpassas så att t ex vanliga kommandon tilldelas funktionstangenter.

*Prisbild:*

TextWrite kostar ca 13 000:- SEK. Högskolerabatt kan erhållas.

## SGML TextWrite Tools OS/2 Edition

*IBM Svenska AB*  
*S-163 92 Stockholm*  
*Tel: (08) 793 10 00 Fax: (08) 793 10 20*

IBM Publishing Systems SGML TextWrite Tools OS/2 Edition är ett stödprogram för att skapa och ändra dokumenttypdefinitioner. Det används vidare för att skapa mallarna som TextWrite använder vid formatering: Man kan t ex ange teckensnitt, teckengrad, färg m m. TextWrite Tools körs under Presentation Manager.

*Understödda hårdvaruplattformar/systemkrav:*

IBM PS/2, OS/2 Standard Edition 1.2 alt. Extended Edition 1.2 med PM 1.2, minst 5MB RAM och 5MB skivminne.

*Prisbild:*

TextWrite Tools kostar ca 43 000:- SEK. Högskolerabatt kan erhållas.

## Xerox DocuBuild för PC

*Rank Xerox AB,  
Hemvärnsgatan 9,  
163 87 Stockholm  
Tel: (08) 795 1000 Fax: (08) 7951440*

Xerox DocuBuild är ett integrerat system för dokumenthantering och layout inom GlobalView-omgivningen. Systemet är avsett för teknisk dokumentation och är väl lämpat i nätverksmiljö (ethernet). GlobalView finns tillgängligt dels på PC, dels på Sun Sparc. I DEC VAX-miljö finns DocuBuild SGML som tillåter validering av dokument i enlighet med SGML-dokumenttypdefinitioner.

### *Understödda hårdvaruplattformar/systemkrav:*

GlobalView kan köras genom ett instickskort på IBM PC AT, eller PS/2 försedd med AT-buss. Vidare krävs 1 MB RAM, 80 MB skivminne, mus, 19" skärm samt ethernetkort. Ett annat alternativ är Xerox' egen arbetsstation 6085 eller GlobalView på Xerox 6520 (Sun Sparc). En ny version av kortet (för MicroChannel arkitekturen) är under utveckling.

### *Prisbild:*

GlobalView på PC kostar ca 46000 exkl. moms. Här tillkommer kostnaden för programvaran, som sägs bli tillgänglig i slutet av 1991. Högskolerabatt kan erhållas.

### *Anmärkning:*

En översättningsprogramvara, Xerox DocuTran, är under utveckling.



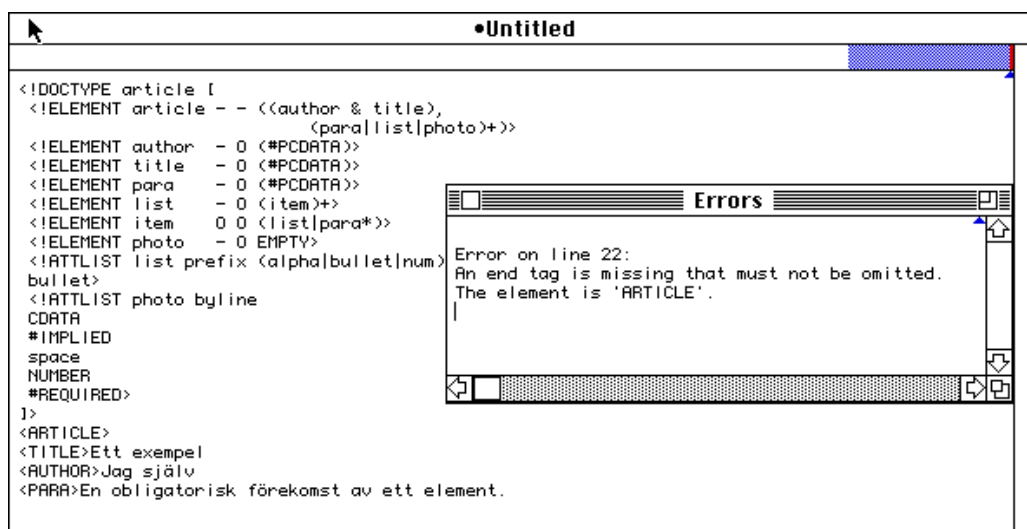
## XGML CheckMark

*Software Exoterica Corporation*  
383 Parkdale Avenue, Suite 406  
Ottawa, Ontario, CANADA K1Y4R4  
Tel: + (613) 722-1700 Fax: + (613) 722-5706

XGML CheckMark är en Macintosh-baserad SGML-dokumenteditor. Den är enligt tillverkaren utvecklad för att ge särskilt stöd för att konvertera existerande dokument till SGML-markering, i synnerhet stora dokument.

Användare kan sätta in tillåtna märkord antingen från menyraden (menyerna visar då endast de i sammanhanget tillåtna märkorden) eller genom att skriva in dessa själv med tangentbordet. XGML CheckMark har interaktiv validering av dokument; valideringen kan utföras i bakgrunden (under MultiFinder) eller medan man författar — upp till fyrtio fönster kan vara öppna. Valideringsprocessen bygger kontinuerligt en fellista som användaren sedan kan bearbeta i valfri ordning, eller så kan valideringen avbrytas för varje fel som hittas. En viktig finess är att XGML CheckMark kan validera partiella dokument: Användaren identifierar då till vilken del av dokumenttypdefinitionen ett givet avsnitt tillhör och avsnittet kontrolleras sedan oberoende av de övriga dokumentdelarna.

XGML CheckMark understödjer samtliga minimeringstillval i SGML samt SUBDOC och FORMAL. Textredigeringsfaciliteterna tillåter reguljära uttryck för sökning i en eller flera filer (Jfr Unix grep). Slutligen kan XGML CheckMark "normalisera" dokument vilket ger dokument med minimal SGML-markering.



Skärmbild av ett dokument validerat med CheckMark. Högst upp till höger är en indikator som visar hur stor del av filen som validerats.

*Understödda hårdvaruplattformar/systemkrav:*

Macintosh II och Macintosh SE/30 med minst 4MB RAM.

*Prisbild:*

Software Exoterica ger f n inga högskolerabatter. CheckMark kostar \$495.

## Write-It

*Yard Software Systems  
Avonbridge House, Bath Road,  
Chippenham, Wiltshire SN15 2BB  
Tel: + (0249) 656194 Fax: + (0249) 655723*

Write-It är en strukturerad SGML-editor som arbetar mot Mark-It (version 3.0 och högre). SGML-dokument kan inom editorn navigeras enligt deras struktur. Write-It tillåter interaktiv validering medan man skriver.

*Understödda hårdvaruplattformar/systemkrav:*

IBM PC/AT, 640K, MS-DOS 2.11 eller högre

*Prisbild:*

Ingår med Mark-It under MS-DOS.

*Anmärkning:*

Mera utförlig information beträffande Write-It saknades i skrivande stund.

## Verktyg för versionshantering

När material skickas ut på remiss och när flera författare samarbetar behövs verktyg som understöder versionshantering. Till PC och Macintosh finns t ex programmet DocuComp, som nämns nedan, och på Unix-sidan finns RCS, som tas upp på nästa sida. Gemensamt för DocuComp och RCS är att man kan jämföra två olika versioner av en text till en ny version.

### DocuComp

*Advanced Software ,Inc.  
1095 East Duane Avenue, Suite 100  
Sunnyvale, California 94086-9648*

*Leverantör i Sverige:*

*Eurodex Sverige AB  
Tel: (08) 98 97 60 Fax: (08) 751 43 92*

Programmet DocuComp — tillgängligt för såväl PC som för Macintosh — upptäcker hur dokument förändrats i samband med revisioner. Skillnaderna som upptäcks är text som tillkommit, raderats, ersatts, samt flyttning av textblock (och ev ändringar inom sådant block). Jämförelsen kan presenteras på ett antal olika sätt, där den mest åskådliga formen är genom ett fönster som visar samma avsnitt i respektive fil.

*Understödda hårdvaruplattformar/systemkrav:*

IBM PC och Macintosh

*Prisbild:*

Ca 2500 SEK

## RCS

RCS står för *Revision Control System* och är ett versionhanteringssystem utvecklat under Unix av Walter F. Tichy vid Purdue University. Följande beskrivning avser RCS med versionsnummer 1.2 daterad 890227.

RCS automatiserar lagringen, identifieringen, åtkomsten, loggningen samt sammanslagningen av olika versioner av dokument. Systemet hanterar upprepade revisioner i dokument och är i synnerhet lämpat för texter som ändras ofta eller på periodisk basis, som program, dokumentation o s v.

Gränssnittet till programet är mycket enkelt, nybörjaren behöver endast använda två kommandon: `ci` och `co`. `Ci` står för *check in*, och används för att arkivera innehållet i en RCS-fil. `Co` (*check out*) används för att hämta ut filer ur arkivet för uppdatering och redigering.

### *Egenskaper i RCS*

- Varje revision kan tilldelas ett symboliskt namn och t ex markeras som preliminär eller stabil. RCS identifierar automatiskt alla förändringar med namn, nummer, tidpunkt o s v.
- RCS lagrar revisioner på ett kompakt sätt. Valfri version kan återskapas och hämtas på basis av bl a versionsnummer, symboliskt namn, datum, författare. Om man vill kan intermediära versioner raderas utan att historiken mellan de sparade versionerna korrumpas.
- Systemet loggar automatiskt alla ändringar mellan versioner. Förutom dokumentet lagras författarens namn, tid och datum för check-in, samt en notering beträffande utfört arbete. Noteringen utförs vid check-in när man har arbetet färskt i minnet.
- När två eller fler författare önskar ändra samma version av ett dokument kommer systemet att hindra en förändring utförd av andra från att förstöra ens egna revisioner. Det är dessutom möjligt att ha flera parallella versioner vilka alla redigeras och som senare sätts samman till ett ny version. RCS upptäcker då överlappande förändringar.

Mer om RCS finns att läsa t ex i [Tichy82] och i online-dokumentationen till systemet.

## Ett prototypramverk för dokumenthantering

Ett ramverk för dokumenthantering håller på att byggas upp inom det aktuella projektet, med en maskinpark bestående av Macintosh, Sun SparcStation och PC-kompatibla datorer. Mellan dessa plattformar är det tänkt att man systematiskt skall kunna ta emot, redigera, och exportera dokument. Denna process kan uppdelas i ett antal faser, av vilka dock samtliga inte nödvändigtvis behöver genomlöpas.

### Import av dokument

Det första steget vid import av dokument är att skapa en fysisk kopia som är i en *välkänd* maskinläsbar form. Dokument erhålls i princip på två sätt: som en fil eller i pappersburen form. Den förra kan behövas översättas till ett annat format, t ex mellan skilda textbehandlare, eller konverteras från ett medium till ett annat, säg mellan en MS-DOS-formaterad 5<sup>1</sup>/<sub>4</sub>" diskett och en 3<sup>1</sup>/<sub>2</sub>" Macintoshdiskett. En annan möjlighet är att dokumentet skickas via elektronisk post, och överförs till en lokal arbetsstation.

Pappersburen information skannas lämpligen in med OCR-programvara innan den bearbetas vidare, analogt med fallet ovan. Eventuellt behövs även program med stavningskontroll för att upptäcka fel som uppkommit genom OCR-konverteringen samt program för att hantera eventuella bilder i dokumentet.

Slutligen kan import av dokument följas av ytterligare en konvertering av dokumentet till en flyttbar intermediär form, t ex SGML, eller annan önskad maskinläsbar representation. Automatisk översättning till SGML är möjlig med olika "smarta" metoder (såsom mönsterigenkänning) men kan inte utföras fullt ut utan beslut från någon operatör på grund av tolkningsmöjligheter vid konverteringen.

#### *Realisering inom projektet*

Här har vi en Macintosh-baserad lösning som består av en UMAX 300 färgskanner försedd med arkmatore och OCR-programmet OmniPage för texthantering; programmet Adobe Photoshop används för bildhanteringen. OCR-programvaran känner igen de vanligaste typsnitten i grader mellan 8 och 72 punkter. Om så önskas bevarar programmet även textattribut som fet stil och kursivering. Inläst text kan sparas under ett antal olika textbehandlarformat eller som ASCII. För översättning till mer komplexa format används XGML Translator.

På Sun SparcStation avser vi att använda ICA Toolset från Chameleon-konsortiet.

### Remissbearbetning

Ett specialfall av import är remissen. Om dokument varit ute på remiss (i maskinläsbar form) utförs lämpligen en jämförelse mot utskickade original för att identifiera de ändrade passagera, eventuellt genereras nya versioner som inkluderar ändringar från annat håll. Versionshanteringen kan också gälla jämförelser mellan äldre och nyare versioner av samma text.

#### *Realisering inom projektet*

Programmet DocuComp — tillgängligt för såväl PC som för Macintosh — används för att upptäcka hur dokument förändrats i samband med revisioner. Skillnaderna som upptäcks är text som tillkommit, raderats, ersatts, samt flyttning av textblock (och ev ändringar inom sådant block). Jämförelsen kan presenteras på ett antal olika sätt, där den

mest åskådliga formen är genom ett fönster som visar samma avsnitt i respektive fil. På Sun-sidan är den nuvarande lösningen att arbeta med RCS.

## **Initial analys och redigering**

I detta skede är dokumentet översatt till önskad maskinläsbar form, och i fråga om remisser är de redigeringar som utförts identifierade. Materialet är alltså färdigt att bearbetas ur innehållsmässig synvinkel.

I fallet att dokumenten är SGML-kodade sker redigeringen lämpligen med verktyg vilka understödjer denna standard.

### *Realisering inom projektet*

Det är viktigt att ha SGML-baserade editorer på samtliga plattformar. På Macintoshsidan används editorn XGML CheckMark, och på PC programmet Write-It. Till Sun kommer editorn Grif. Alla dessa editorer utför validering av dokument mot godtycklig dokumenttyp. Dessutom finns parsern Mark-It och XGML Validator för batchkontroll på PC.

## **Komplex analys och presentation**

En mer komplex bearbetning av dokument kan innebära översättning av dokument för presentation eller hantering på ett annat medium. Andra medier kan t ex vara idébehandlare eller hypertextsystem. För att få överblick är det önskvärt med navigeringsmöjligheter liksom återgivning på papper.

### *Realisering inom projektet*

För att navigera genom stora dokument finns DynaText som kombinerar bl a idébehandlarfunktioner med fulltextindex. I skrivande stund är programvaran tillgänglig endast på Sun men "bläddraren" (DynaText Browser) kommer i en framtid att finnas till de andra två plattformarna. Man kommer då att skapa elektroniska böcker på Sun och kunna läsa dem på valfri plattform.

DynaText läser dokument kodade i (minimal) SGML varför denna intermediära form måste skapas för dokument kodade på annat sätt. För detta ändamål används XGML Translator som även är tillämplig för översättning till andra format. Ett möjligt alternativ är ICA Toolset.

För sedvanlig papperspresentation finns ett antal möjligheter. Från SGML-kodade dokument finns redan verktyg för batchformatering med MS Word (med hjälp av XGML Translator). ICA Toolset har konverteringsrutiner till bl a troff och L<sup>A</sup>T<sub>E</sub>X, medan Grif kan generera PostScript-kod själv. DynaText kan skriva ut papperskopior av sina elektroniska böcker.

Slutligen kan rutiner skrivas för översättning till valfritt format ur SGML-kodningen, t ex med Mark-It, XGML Translator eller Unix-verktyg. Här blir det aktuellt att skapa rutiner för konvertering till och från hypertextsystem.

## **Export av dokument**

När ett dokument är klart för export återstår att utföra översättning till önskat format och kodning. Denna process är i princip trivial, eftersom dokumentet är väldefinierat och korrekt kodat. Rent praktiskt finns fortfarande problem såsom integration av bilder från olika källor eller översättning mellan format vilka inte har samma återgivningsmöjligheter. Lösningar till sådana problem är en viktig aspekt i det kommande arbetet.

## Referenser

Produktblad från tillverkare/leverantör samt

- [Broady90] Broady, Donald: "*Datorstöd för kunskapsarbete. Om hypertexttillämpningar i utbildningssammanhang*", Projektplan, IPLab, NADA, KTH, oktober 1990.
- [Goldfarb90] Goldfarb, Charles F.: "*The SGML Handbook*," Oxford University Press, 1990, ISBN 0-19-853737-9.
- [Herwijnen90] Herwijnen, Eric van: "*Practical SGML*," Kluwer Academic Publishers, Dordrecht, 1990, ISBN 0-7923-0635-X.
- [IBM90] IBM Corporation: "*Solutions for CALS Technical Publishing*," Document GC34-5153-01, Second Edition, November 1990.
- [Mamrak90] Mamrak, S.A., O'Connell, C.S. and Barnes, J.A.: "*The Integrated Chameleon Architecture: A Software Toolset to Support Data Translation*," Technical Research Report OSU-CISRC-11/90-TR37, November 1990.
- [Mark-It89] The Mark-It Manual, *Version 2.0*, Sobemap S.A./Yard Software Systems, January 1989.
- [SGML86] Information processing — Text and office systems — Standard Generalized Markup Language (SGML), ISO 8879-1986 (E), October 1986.
- [Tichy82] Tichy, Walter F., "*Design, Implementation, and Evaluation of a Revision Control System*, In: "Proceedings of the 6th International Conference on Software Engineering, IEEE, Tokyo, September 1982, ISSN 0270-5257.
- [XGML90] XGML Translator, XTRAN Programmer's Manual, Version 1.0, Software Exoterica Corporation, June 1990.

Det finns en 320-sidig bibliografi ("*Bibliography on SGML*") över SGML-relaterade artiklar och böcker som kan beställas från Queen's University. Bibliografin innehåller ca 1400 referenser. För mera information och aktuellt pris kontakta:

*Heather Ball*  
*Office of the Associate to the Vice-Principal(Resources)*  
*Richardson Hall, Room 215*  
*Queen's University*  
*Kingston, Ontario CANADA K7L 3N6*

*Tel: + (613) 545-2030      Fax: + (613) 545-6602*

*E-Mail: heather@qucis.queensu.ca*



## Appendix

Nedan sammanställs i tabellform programvaror kontra hårdvaruplattformar. Med IBM avses här IBM-kompatibla datorer med MS-DOS, PC-DOS eller OS/2 som operativsystem. Motsvarande avser Macintosh här Apple Macintosh datorer under AOS. För utförligare information, se under respektive produkt eller hör med leverantören.

Teckenförklaring: • = Finns, ° = Möjlig, — = Saknas. Alternativet "Möjlig" innebär att produkten kan tillhandahållas på begäran eller måste anpassas, är under utveckling, alternativt kan köras på aktuell plattform under förutsättning att man bygger ut denna med mjuk- och/eller hårdvara (t ex genom instickskort eller genom att byta operativsystem).

	IBM	Macintosh	Sun	Övrig
FastTAG	•	—	•	•
Glammar	°	°	•	°
ICA Toolset	—	—	•	—
Mark-It	•	—	°	•
NIST NCSL OSE SGML parser	°	°	•	°
SGML Translator DCF Edition	—	—	—	•
TextTagger Workstation Edition	•	—	—	—
XGML Translator	•	•	•	•

Tabell 1: Översättningsprogramvaror kontra understödda hårdvaruplattformar.

	IBM	Macintosh	Sun	Övrig
DynaText Indexer & Browser	°	°	•	—
Grif	—	—	•	•
SGML TextWrite OS/2 Edition	•	—	—	—
SGML TextWrite Tools OS/2 Edition	•	—	—	—
SoftQuad Publishing Software	°	°	•	•
SoftQuad Author/Editor	•	•	•	•
SoftQuad BuildRules/RulesBuilder	•	•	•	•
Write-It	•	—	—	—
Xerox DocuBuild	°	—	°	•
XGML CheckMark	—	•	—	—

Tabell 2: Editorer och presentationsverktyg kontra understödda hårdvaruplattformar.

### Addendum 920309

Alla NADA-rapporter ska framöver i princip göras tillgängliga för ftp-hämtning via nätet. Jag konstaterar att även denna lilla skrift (ca ett halvår gammal) har hunnit åldras; nya kommersiella produkter från Arbortext, Exoterica Corporation, Electronic Book Technologies m fl har dykt upp. Det är roligt att konstatera att utvecklingen i området går så fort. Vidare finns en public domain SGML-parser kallad sgmls som, utifrån en första version skapad av Charles Goldfarb, numera utvecklas och uppdateras av James Clark. Se comp.text.sgml!