

AL 25 JAAR OP ZOEK NAAR HET E-BOEK

Door Joost Kircz,
office@kra.nl

Onderstaand artikel is gebaseerd op een voordracht die gehouden werd op XML Holland 2010, het congres van de vereniging van Nederlandse SGML- en XML-gebruikers dat dit jaar voor de 25^{ste} keer werd georganiseerd. Het behandelt de ontwikkeling van de eerste stappen van computer netwerk en het gebruik van databases, tot wat er nog allemaal gedaan moet worden. Van belang is dat we zien dat vragen en analyses van 25 jaar geleden hoewel in veranderende vorm steeds weer opkomen. Dit geeft dan ook weer aan dat ons werk nooit af is en dit 25-jarig jubileum slechts een mijlpaal is, geen eindstation.

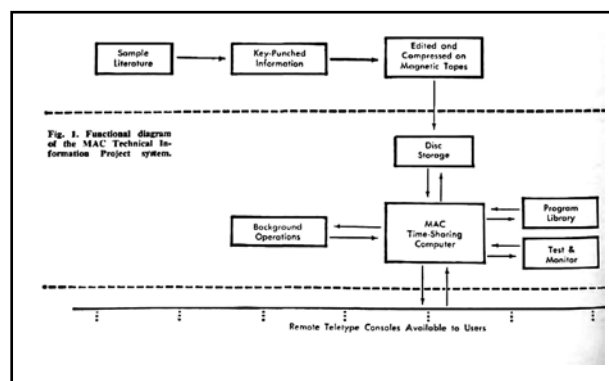
De eerste stappen naar SGML en standaardisatie

Een kwart eeuw is veel in een mensenleven en ik wil dan ook aan de hand van een stukje, deels eigen, geschiedenis de vraag proberen te beantwoorden “wat zijn e-boeken?”.

In de oertijd, we spreken van 1978, werd een ANSI-werkgroep ingesteld om te komen tot een eenduidige formaat voor tekstoverdracht en een annotatietaal die geschikt was voor verdere verwerking van tekst (1). In deze discussie waren er verschillende drijvende krachten. Enerzijds grote bedrijven zoals IBM en Boeing die hun documentatie eenduidig opgeslagen wilde hebben zodat handleidingen foutloos hergebruikt konden worden, anderzijds was er de wetenschappelijke wereld, met name CERN in Genève, waar men hard werkte aan de eerste versies van een elektronische bibliotheek, zodat wetenschappelijke artikelen probleemloos de nieuwe 27 km lange ring met vier werkplekken konden bewandelen om de verspreid werkende onderzoekers te helpen. In de uiteindelijke SGML ISO-standaard 8879 (2) speelden CERN-mensen zoals Anders Berglund dan ook een flinke rol. Een van de eerste SGML-leerboeken werd door CERN-medewerker Eric van Herwijnen geschreven (3).

In die tijd werkte ik als uitgever natuurkunde bij de Noord-Hollandse, een belangrijke wetenschappelijke uitgeverij die net door Elsevier Science was overgenomen. Alle principes die SGML wilde benoemen, namelijk een scheiding tussen inhoudelijke coherente eenheden en hun uiteindelijke representatie zijn vandaag de dag, meer nog dan men toen kon vermoeden, relevant.

De interesse in SGML bij uitgeverijen was een direct gevolg van de majeure overstap naar computer netwerk enerzijds en fundamenteel onderzoek naar bibliografische bestanden anderzijds. Figuur 1, een plaatje uit 1965 van het MIT technical information project illustreert dit (4).

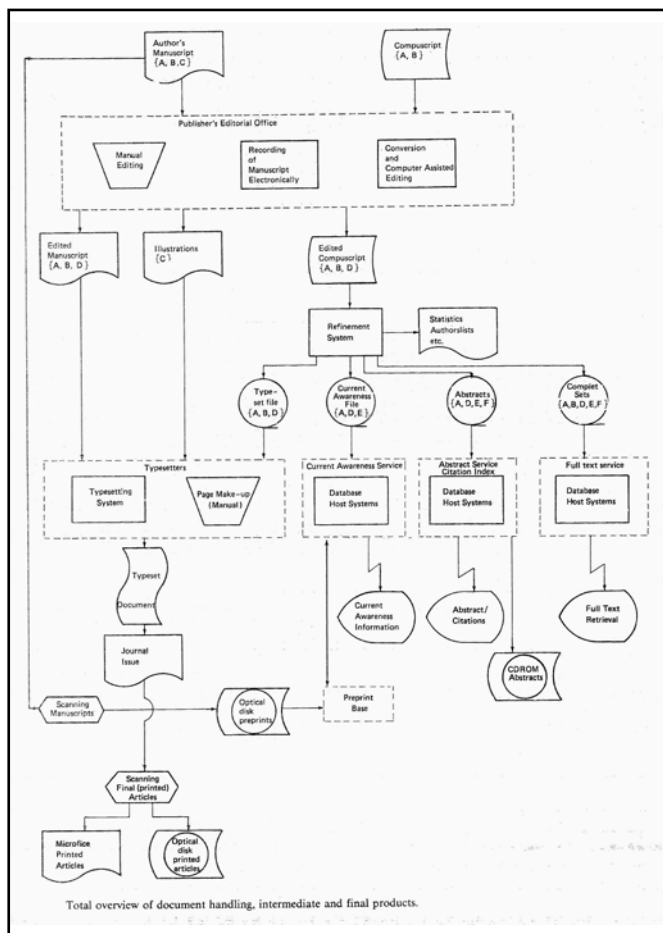


Figuur 1: MIT technical information project Schema

Een voorbeeld van het eerste is te vinden in een artikel van Ken Metzner uit 1975. Hij was toen Director of Publications van de *American Physical Society* (APS). Later zou hij bij Elsevier New York directeur worden. Het is een overzicht van een project dat twee jaar daarvoor van start ging (5). In dit artikel behandelt hij de bestaande publicatiestroomschema's en de veranderingen die optreden bij “full-text computer photo-composition”. Een belangrijk punt was dat door deze werkwijze bibliografische informatie, zogenaamde secundaire informatie in uitgeverstermen, vanzelf genereerd werd als bijproduct van het primaire zetten van de artikelen. Op deze manier konden de *abstract journals*, weekbladen op krantenpapier met alleen de samenvattingen van artikelen in een vakgebied, die toen nog bestonden moeiteloos en correct gemaakt worden. Ook konden de SPIN-tapes (van spin off) waarop deze records stonden direct geladen worden in bibliografische on-line databases, die in die tijd overal hun kop opstaken.

Een citaat wil ik u niet onthouden:

“it is an operation involving.....about 100 000 pages per year. (this total, incidentally, has remained roughly constant over the last 5 years. To talk about an information explosion is no longer appropriate. The dust from the explosion has settled already).”



Figuur 2: Documentverwerking, tussenstappen en eindproducten

In deze periode begon ook Donald Knuth met zijn TEX (6). Dit was het begin van een bittere schoolstrijd tussen de TEX- en later Latex-adepten en de semantici van de SGML-school die duidelijk inhoud van representatie wilden scheiden. Weer tien jaar later, publiceerde ik met mijn Elsevier collega Jan Bleeker een artikel over het gebruik van relationele databases in wetenschappelijk uitgeven (7). In figuur 2 is een schema uit dat artikel te zien waarin alle stappen van documentenverwerking die we daarbij onderkenden zijn afgebeeld.

Wij maakten een onderscheid tussen verschillende soorten metadata, verdeeld in groepen:

- Groep A: identificatie, dat wil zeggen alle bibliografische informatie van de auteur, alsmede titel en abstract;
- Groep B: tekststructuurinformatie inclusief een ondergroep met de metadata van de literatuurreferenties;
- Groep C: illustraties;
- Groep D: uitgeversinformatie, d.w.z. document nummer (nu DOI), tijdschriftnaam en -identificatie, omvang, type, data van insturen, behandeling en publicaties, productie informatie;
- Groep E: ontsluitingsgegevens, zoals trefwoorden, registratienummers, structuurformules e.d.;
- Groep F: uiteindelijke bibliografische informatie. Let wel: in dit schema werd duidelijk de produc-

tiegang van auteur tot gedrukt artikel gevolgd. Belangrijk is hierbij de scheiding tussen de auteursinformatie en de uitgeversinformatie, iets wat heden in de open access discussie relevant is.

Minder inkloppen

In die tijd kocht de Noord-Hollandse een PDP 11 en werden de koppen van de "letter" (dat zijn snelle korte artikelen) tijdschriften apart ingevoerd en op magneetband weggeschreven. Het doel was duidelijk. Eenvoudig onderzoek op de ESA-IRS bibliografische database had uitgewezen dat de abstract van een artikel bij de Noord-Hollandse ongeveer acht keer werd ingevoerd. Let wel, acht keer opnieuw ingeklopt, zoals dat in oude zetterijtaal heet. Met een magneetband onder de arm bleek mij dat deze diensten wel interesse hadden om niet meer op papier, maar op band aangeleverd te krijgen en dat voor de helft van hun kostprijs. Met ongeveer 5 klanten waren wij al uit de kosten.

Het voorstel om SGML *North-Holland Machine Searchable Heads (Normash)* te produceren werd dan ook door de directie met gejuich ontvangen. Het besluit werd toen genomen dit niet alleen voor natuurkunde te doen maar voor het hele bedrijf. De ene kant van de munt is dat het hierdoor ruim een decennium duurde voor wat nu Science-Direct heet echt op streek kwam. Dit kwam omdat alle zetterijen, en dat waren er tientallen, opeens dezelfde elektronische output moesten leveren. De harde les is dat verregaande automatisering niet zonder rigoureuze standaardisering kan. De andere kant van de munt is dat daardoor Elsevier uiteindelijk een excellente tijdschrift-DTD heeft.

Vanaf toen dus, inderdaad zo'n 25 jaar geleden ging het hard. Maar voor ik een sprong in de tijd maak wil ik van de gelegenheid gebruikmaken om nog even te wijzen hoe het er met standaardisatie toen bij stond. Michael Buckingham, directeur van Elsevier filiaal in Lancaster, lanceerde een unieke dienst voor tekstconversie. Voor de aardigheid laat ik even het begin van de eerste en het einde van de dertiende pagina zien van zijn aanbiedingsbrochure, totaal zijn er zo'n 625 regels (8).

Nu er een nieuwe oorlog op het gebied van besturings-systemen woedt is het goed voor ogen te houden dat een dienst als deze uit 1986 mogelijk nog allerminst uit de mode is. Gelukkig is de noodzaak van metalen en met name XML nu doorgedrongen. Om nog even een voorbeeld te geven van mijn oude werkgever. De DTD is het enige boek dat Elsevier ooit heeft uitgegeven met een auteursrechttekst die zegt: *This book may be reproduced and distributed in whole or in part in any medium, physical or electronic, so long as this copyright notice remains intact and unchanged in all copies.*

Het aardige is natuurlijk dat dit 444 pagina's dikke boek dat de Elsevier DTD5 familie code voor code beschrijft, gezet is in pdfTEX (9).

IRCS DISK FORMATS							
The following disk/tape formats can be read by IRCS for conversion to database form. Sample disks/tapes should be submitted with requests for a quotation. Magnetic tape (1/2 inch) should be recorded at 1600 bpi							
Description	Size	Sides	Density	Hard/Soft	TPI	Op.	System
ABC 800	5.25	D	D	S	48	Own	
ABC 830	5.25	S	D	S	48	Own	
ABC 832	5.25	D	D	S	96	CP/M	
ABC 832	5.25	D	D	S	96	Own	
ABC 838	8.00	D	D	S	96	Own	
ABS Systems	5.25	D	D	S	48	CP/M	
ACT Apricot	3.50	S	D	S	135	CP/M	
ACT Apricot	3.50	S	D	S	135	MS-DOS	
ACT Apricot	3.50	D	D	S	135	MS-DOS	
Action Discovery	8.00	S	S	S	48	CP/M	
AES 7200	8.00	S	D	H	48	Own	
AES AlphaPlus	5.25	D	S	H	96	Own	
AES MultiPlus	8.00	S	D	H	48	Own	
AES Plus	5.25	S	S	H	48	Own	
AI	8.00	D	D	S	48	CP/M	
AI	5.25	*	*	S	48	MS-DOS	
Ajile Professional	8.00	D	S	H	48	Own	
Xerox 860 US output format block	8.00	D	D	S	48	Own	
Xerox Star	8.00	S	S	S	48	CP/M	
Xycom Hard Hat	8.00	S	S	H	48	HDOS	
Zenith	5.25	S	D	S	48	-CP/M-	
Zenith Z-100	5.25	S	S	S	48	-CP/M-	
Zenith Z-100	5.25	S	D	S	96	-CP/M-	
Zenith Z-100	5.25	S	S	S	96	-CP/M-	
Zenith Z-100	8.00	D	D	S	48	MS-DOS	
Zenith Z-100	5.25	D	D	S	48	Z-DOS	
Zenith Z-100	5.25	D	D	S	96	Z-DOS	
Zenith Z-100	8.00	S	S	S	48	Z-DOS	
Zenith Z-100	5.25	S	S	S	48	-CP/M-	
Zilog MC22-20	8.00	S	S	S	48	CP/M	

Figuur 3: Conversiemogelijkheden in 1986.

Van zetten voor verspreiden naar vinden om te krijgen

Zoals boven aangegeven, was de drijvende kracht achter SGML hergebruik en duiding van vaste tekstdelen. De zin, "Controleer of het apparaat op 220 Volt staat", moet altijd foutloos op de goede plek staan. De kracht van SGML was dan ook dat meervoudige presentaties van dezelfde tekst mogelijk werden. Het is daarom dat de grote wetenschappelijke uitgevers ervoor kozen. De elektronische tijdschriftdepots vereisten dat een eens gezette en opgemaakte tekst identiek qua inhoud zowel op papier als scherm verscheen en dat ook indexen en abstract-tijdschriften automatisch gegeneerd konden worden. De ontwikkeling was en is nog steeds vooral productiegedreven.

Een van de meest komische, of zo je wilt wonderlijke resultaten bij mijn favoriete voorbeeld, de firma Elsevier, was dat met de ontwikkeling van Science Direct. Opeens werd de lay-out van alle bladen naar een homogeen formaat omgezet. Dit terwijl nu juist de essentie van de techniek is dat je een eenmaal gestandaardiseerde opgeslagen teksten onderling geheel anders kan vormgeven al naargelang het onderwerp of de markt dat waardeert. Er is geen enkele reden waarom een artikel in de topologie dezelfde lay-out zou moeten hebben als een artikel in de neurologie of de dierkunde. Hiermee vervalt ieder onderscheidend vermogen en wordt gesuggereerd dat de uitgeversnaam meer zegt dan de naam van een specialistisch vaktijdschrift.

Het tegendeel is eerder waar. De opmaak van een tekst is er voor om de boodschap van de schrijver zo helder mogelijk over te brengen aan de lezer. Vierhonderd jaar typografie en vormgeving waren niet voor niets. De rol van plaatjes en tegenwoordig ook geluid,

film of zelfs haptische elementen is per vakgebied en gebruikersgroep anders! Uiteraard geldt mijn kritiek ook al die andere uitgeverijen die denken dat hun missie verspreiding is en niet het faciliteren van communicatie.

Kort en goed, de stormachtige ontwikkeling van SGML en later XML in de uitgeverijen was en is vooral gedreven door logistiek en opslag. Pas bij de opkomst van mobiele apparaten van verschillende vormgeving zien we de opkomst van *style sheets* van allerlei soort die het mogelijk maken om documenten van vorm te laten veranderen als ze op verschillende soorten scherm of papier worden gerepresenteerd.

Het hele doel van de exercitie uitgeven is echter niet alleen boeken en tijdschriften te verkopen, maar de inhoud van deze communicatieve objecten qua inhoud en vorm dusdanig te laten schrijven en redigeren dat zij aansluiten bij een communicatieve behoefte van de lezer, die kennis wil of moet nemen van de boodschap van de schrijver.

In de loop van de geschiedenis is er duidelijk een verschil ontstaan in de manier van vormgeving en opbouw van werken in verschillende vakgebieden. Hierbij is het begrip *Genre* van belang. Een genre is niet alleen een term voor het culturele veld, maar geldt voor alle vakgebieden. Juristen structureren hun verhalen nu eenmaal anders dan synthetisch organisch scheikundigen, en linguïsten doen het ook weer anders. Voor de humaniora heeft het geleid tot de belangrijke onderneming van de TEL, de *Text Encoding Initiative* (8).

Bij een correcte en uitvoerige annotatie is het mogelijk om niet alleen bibliografische records te genereren, maar ook gelijkaardige benoemde tekstvelden aan elkaar te verbinden. Dus bij een systematische aanpak is het niet alleen mogelijk uit een corpus artikelen alle achternamen van auteurs te verzamelen, maar ook alle tekstdelen die na het hoofdje "conclusie" komen. Op deze wijze is bladeren of browsen niet meer zoeken, maar tekstdelen vinden.

Atomen en moleculen

Het probleem waar we nu tegenaan lopen is net als bij de scheikunde. Eenheden zijn meer dan hun samenstellende delen. Als je een molecule uit elkaar sloopt in zijn constituerende atomen, dan houden we niets meer over van wat ooit de eigenschappen van die molecule waren. Om de losse atomen weer tot molecule te maken, moeten we het begrip chemische binding introduceren. Losse atomen vormen alleen een geheel nieuwe stof als er een bindende kracht is. Een molecuul bestaat uit de samenstellende atomen en de bindingskrachten. Hetzelfde geldt voor elektronische teksten. Elektronische teksten zijn, in principe, gestandaardiseerd tezamen beschikbaar. Om daar iets coherent mee te doen moeten de relaties tussen deze teksten ook gespecificeerd worden. De relatie(s) tussen tekst(delen) zijn net als chemische bindingen, zij maken iets nieuws vanuit de optelling van samenstellende delen.

Het zal iedereen duidelijk zijn dat als wij een min of meer homogeen corpus in een genre hebben, we graag gelijkaardige tekstdelen, hiermee bedoel ik tekstdelen die het zelfde onderwerp behandelen, naast of achter elkaar zien, met weglating van de rest. Dit betekent dat er naast de structuurmetadata zoals we die nu bijvoorbeeld uit DTD's kennen ook inhoudsmetadata nodig zijn. Dit dwingt om even een klein uitstapje te maken naar zoekmachines.

Er is een duidelijk verschil tussen enerzijds tekstueel materiaal dat intrinsiek een structuur kent, zoals een wetenschappelijke of professionele tekst en anderzijds literaire teksten en allerhande losse communicaties. Maar let wel, als we dieper kijken naar de opbouw van bijvoorbeeld krantenartikelen zien we ook structuur.

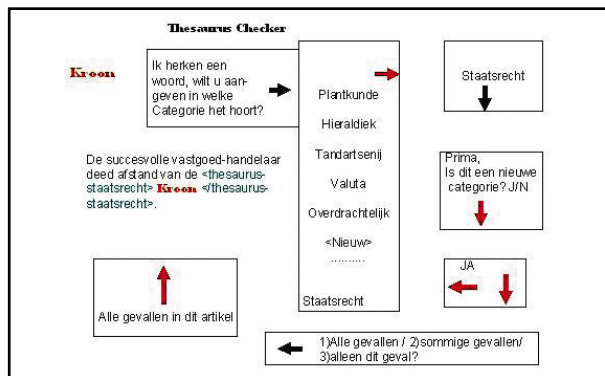
Dit wil zeggen dat als we in een oceaan vissen zoals naar webpagina's op het internet, zware statistische en probabilistische zoekmethoden de aangewezen weg zijn. Iedere structuur is meegenomen, hoe meer hoe beter, maar we kunnen er niet op wachten. Gaan we echter naar gebieden waar gecertificeerde en valideerde informatie regeert dan zijn semantische benoemingen cruciaal. Het dataveld *trefwoord* is dan volstrekt onvoldoende. Ondanks alle Booleaanse Gymnastiek kom je er niet altijd uit.

We komen dan op de noodzaak om de volgende zaken uit te werken:

- A. Gestructureerde trefwoordlijsten of thesauri per vakgebied. In modern jargon heet dat nu *ontologie*. In feite heb je een waaier ontologieën nodig. Een woord heeft immers verschillende betekenissen. Neem het woord Kroon, dit kan een heraldische, een staatsrechtelijke, een plantkundige, een overdrachtelijke, een tandkundige, een monetaire en nog zo'n paar betekenissen hebben. Zoeken op Kroon heeft dus geen zin. Voorstel: de omgekeerde spellingchecker. Hier stopt de checker als een woord wel herkend wordt en vraagt dan tot welke categorie het behoort. Dit is uitgewerkt in figuur 4.
- B. Typografische structuur is mede een uitdrukking van een betoogtrant. Een verhaal, zeker in wetenschap of onderwijs, is een betoog, wil iets uitleggen, overbrengen, beargumenteren.

Modularisatie versus coherentie

Dit brengt mij op onderzoek dat in de jaren '90 is gestart met een artikel van mijn hand waarbij ik stel dat de argumentationele structuur van een wetenschappelijk artikel zou kunnen helpen bij het zoeken van het *soort* informatie dat je zoekt (11, 12, 13). Al snel kwamen we er achter dat een taalkundige analyse van argumentationele structuren wel mooi is, maar dat je in teksten die in essentie geschreven zijn als unieke zelfstandige essays niet veel verder komt. In een verzameling werken moet er meer structuur zijn om vergelijkingen tussen werken te kunnen maken. Dit leidt dan onomstotelijk tot de discussievraag: wat



Figuur 4: De omgekeerde spellchecker

is de coherentie van een werk: artikel, document, roman, serie, enz.?

Kun je een werk uit losse bouwstenen, die je met de juiste metadata en hun onderlinge relaties beschrijft, tot een goed leesbaar, spannend en overtuigend nieuw document opbouwen? Je komt dan op de notie van modularisatie van informatie kortom terug, maar wel een spa dieper dan in het schema dat ik eerder liet zien (14).

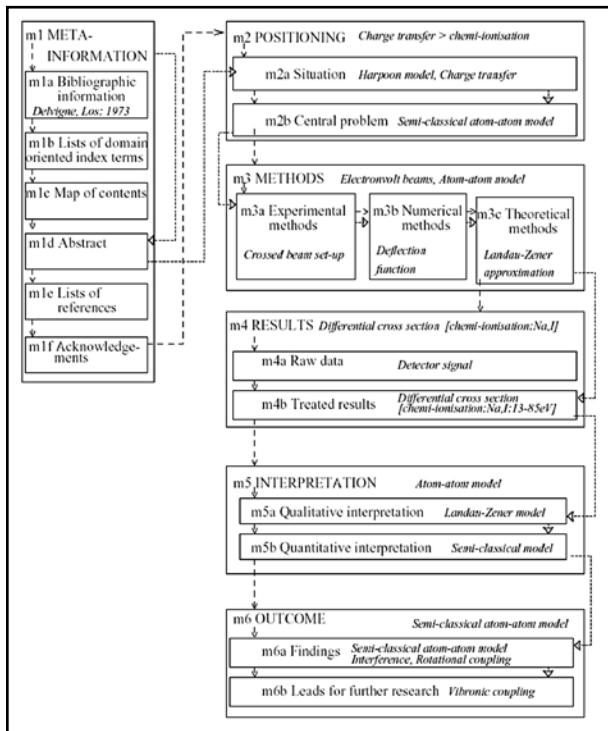
Voor wetenschappelijke artikelen in de molecuulfysica heeft Frederique Harmsze dit uitputtend en prachtig in haar proefschrift uitgewerkt (15, 16). Zij komt met instructies voor auteurs in dat vakgebied, hoe een modulair artikel geschreven zou kunnen worden op dusdanige manier dat hergebruik van modules zinvol is om nutteloze herhalingen te voorkomen. *Een module wordt gedefinieerd als een uniek gekarakteriseerd, op zichzelf staande representatie van een conceptuele informatie-eenheid bedoeld om die informatie te communiceren.*

In haar analyse komen twee zaken aan de orde die voor de verdere ontwikkeling van XML voor elektronische artikelen en boeken essentieel zijn:

1. Een hyperlink is een onaf object. Behalve dat we nog steeds geen punt-punt koppelingen hebben waardoor je wel heen kunt verwijzen maar nooit meer goed terugkomt, zegt deze koppeling niet meer dan het oude "see also". Nou en! denk je dan. In dit werk wordt de fundamentele asymmetrie van *verwijzen naar* en *verwezen worden* uitgewerkt. Ik haal jou aan als bewijs van mijn stelling, maar jij kunt soms niet meer zeggen dan dat een ander jou om voor de ander moverende redenen heeft aangehaald, maar soms ook: "kijk mijn ideeën worden daar bevestigd".
2. Er is een verschil tussen organisatorische-logistische verbanden die de structuur van een werk aangeven en argumentationele verbanden die de gang van de redentatie weergeven.

In figuur 6 is een voorbeeld zien van een aantal van die organisatorische verbanden, zoals:

- Proximity: broertjes, neven etc.
- Range based: van micro naar algemeen.



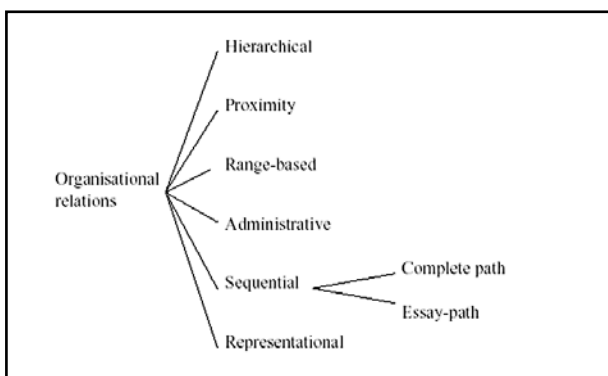
Figuur 5: Voorbeeld van modulaair wetenschappelijk artikel.

- Administrative: om aan te geven wat we nu hebben.
- Representational: denk aan formules, grafiek, tabel, histogram.
- Sequential: dit geeft aan dat je op verschillende manieren door het verhaal heen kan; volgorde of alleen de tekstdelen.

Ingewikkelder zijn de argumentationale relaties. In figuur 7 geef ik hier een voorbeeld van zonder verder in te gaan in de details. In het artikel van Kircz en Harmsze (16) wordt hierop in gegaan.

Dit soort structureringwerk wordt nu onder meer voortgezet door Anita de Waard, die zich vooral bezighoudt met de analyse van taalstructuren in de moleculaire biologie, die aangeven wanneer een auteur bijvoorbeeld een stelling, een voorbeeld, een resultaat of een conclusie beschrijft (17,18).

Ingewikkelder wordt het ook als je tekst verbindt met een bestand van plaatjes en/of data. Je moet heel goed



Figuur 6: Voorbeeld van organisatorische verbanden.

weten hoe deze data en plaatjes gedefinieerd zijn. De koppelingen van databanken zoals in de genetica met wetenschappelijke reportages is dan nu ook een *hot item*.

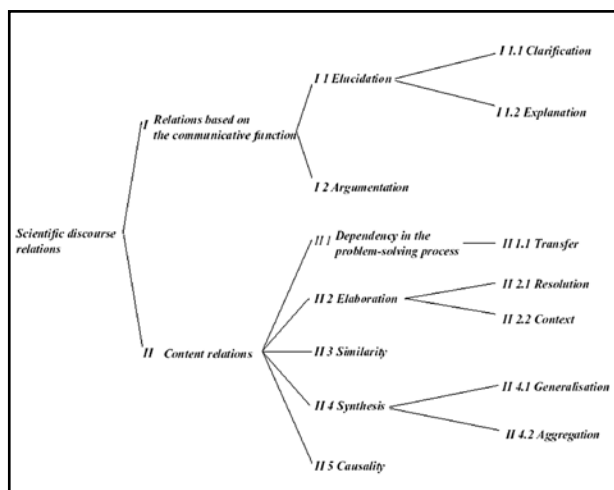
Maar eigenlijk zijn plaatjes veel lastiger. Hoe vaak immers is er geen enkele heldere relatie tussen tekst en plaatje. In de tekst staat: zie plaatje, of er staat niets en de illustratie spreekt kennelijk voor zichzelf. Vaak heeft het plaatje een ondertekening, maar dat verwijst maar al te vaak expliciet of impliciet terug naar de tekst. Een illustratieondertekening is een brug tussen tekst en beeld en is daarmee in feite een zelfstandige tekst die volledige informatie geeft, inclusief zaken als de techniek waarmee het gemaakt is en een volledige beschrijving wat de auteur wil weergeven. In die zin krijgt de beschrijving van een afbeelding die opgeslagen wordt voor hergebruik, in de database die de afbeelding bewaart de vorm van een uitgebreide verzameling metadata, te vergelijken met Iconoclast in de kunstgeschiedenis (19).

En nu het multimediale lesboek

Bovenstaande analyses en discussies vinden eigenlijk plaats binnen het digitaal depot. Maar er is ook nog zoets als lezen. Lezen is een bijzondere en unieke fysiologische activiteit die totaal anders is dan luisteren. In al het mediageschetter over dat de boekindustrie moet kijken naar de muziekindustrie wordt een essentieel element vergeten. Alle geluidsapparaten leiden tot luchttrillingen die het oor bereiken, die dat verder passief registreert onafhankelijk van de voorgeschiedenis. Je kunt dus op de fiets telefoneren en naar muziek luisteren. Leesapparaten echter zijn allemaal anders. Lezen is een actief proces waarbij symbolische tekens, ook wel letters en cijfers genaamd, geïnterpreteerd moeten worden. Je kunt alleen lezen op de fiets als je achterop zit (20).

Deze notie brengt ons tot de werkelijke uitdaging van het elektronische boek en met name het leerboek. Het is geen probleem meer om een reproductie van een willekeurige pagina in pixels om te zetten. Het is ook geen probleem meer om de lopende tekst in Unicode op te slaan en dat later te mengen met plaatjes, tabellen en nu ook film, animaties, of spelletjes. Maar was 400 jaar boekgeschiedenis dan eigenlijk iets dat alleen met papier te maken had? Niets is minder waar. De kern van het probleem is hoe je gegeven een techniek de boodschap van schrijver naar lezer overbrengt en dat geldt nu nog.

Op de Hogeschool van Amsterdam hebben wij een onderzoeksproject lopen onder de naam Amsterdam E-boekenstad (21). In dat project proberen wij met name leerboeken te onderzoeken. Uiteraard gaan we mee in de hype dat alles on-line beschikbaar is, dat de iPad een mooi apparaat is en je alles kunt kopiëren en doorsturen. Maar onze eigenlijke vraag is weer niet de logistiek, distributie en opslag, maar hoe ziet, om het formeel te zeggen, een digitale leeromgeving eruit. Hoe mengen wij voorbeelden, animaties, casussen



Plaatje 7: Voorbeeld van argumentatieve relaties.

in de vorm van spel, beeld, geluid en ja, daar zijn we weer, tekst, op dusdanige wijze dat de juiste techniek voor de juiste communicatie wordt gebruikt.

In eerste instantie richten wij ons op *e-inkt*-lezers en hebben daarvoor een flinke stapel Irex 1000 A4 lezers gekocht. We hebben 30 studenten zo'n lezer gegeven, 30 hadden een laptop en 30 hadden het papieren boek. Het enthousiasme was even groot als de verwachtingen. Toch..... papier won!

De strijd tussen de dragertechnieken, laptop of tablet met backlit LCD schermen of een vorm van *e-inkt* met stilstaand beeld wordt nu gestreden op het niveau van het gadget. Wie heeft de meeste applicaties, zoals de keukenmachine die ook slagroom kan kloppen, terwijl voor dat laatste een garde of zelfs een vork al voldoende is (22). Onze studenten van Interactieve Media vonden de *e*-lezers maar niks. Alles wat ik wil weten staat toch al op het net, blufte ze. Maar ook bleek dat ze niet lezen. Dat wil zeggen gewoon een uurtje of wat verdrinken in een verhaal. Je hebt nu eenmaal mensen die de snuffjes aan hun nieuwe auto belangrijker vinden dan de wegligging. Alle resultaten en discussies staan overigens vrij toegankelijk op de projectwebsite en -wiki (21).

Lezen

Lezen van het scherm is een fysiologisch proces en, zoals al gezegd, in tegenstelling met muziek is er geen universeel overdrachtmiddel zoals lucht. Sinds de ontwikkeling van de kathodestraalbuis is er wel onderzoek gedaan naar de kwaliteit van het lezen, maar eigenlijk is dit nog vrij onontgonnen terrein in relatie met productontwikkeling (23). De kwaliteit van *e*-lezers wordt gemeten naar de consumentenmarkt. Er worden steeds meer *e*-boeken verkocht. Dat wil zeggen platte elektronische versies van een boek. Dat is ook geen kunst: nu alle boeken gemaakt worden met tekstverwerkers op de computer zijn dit een soort spin-offs net als de abstracts in het begin van mijn verhaal, een moeiteloos bijproduct. De discussie die gevoerd wordt gaat hoofdzakelijk over de handel en

de rechten. Zoals we ook met het Google Books-programma zien.

De werkelijke mogelijkheden liggen elders. Het gaat er nu om combinaties te maken van technieken en om die combinaties te maken zijn er strakke standaarden en metadateringen nodig. Het is leuk te bedenken dat het na, de mislukkingen van de eerste generatie Hypertext-romans, de spellen zijn die het concept hebben overgenomen. Immers, in *adventure games* kan de lezer/speler zelf een pad door het verhaal kiezen. Zoiets kan niet op papier, ondanks de heroïsch poging van Danielewski's Kaartenhuis (24).

Zoals zo vaak zijn de automatiseerders ver vooruit op de toepassingen. Maar net als in de bouw is meer dan een schetsontwerp nodig voor je de aannemer aan het werk zet. Voor de ontwikkeling van XML en de discussie over semantisch publiceren, gaat het er om per doelgroep en per genre structuren te ontwikkelen. Voor dermatologie hebben we kleurechtheid nodig, voor *Horror in de houtzagerij* is dat helemaal niet nodig, maar wel spectaculair rood. Wetboeken, en overigens de meeste grafieken, kunnen zonder kleur.

Wij zien echter een aanbodgedreven behoefte aan kleur ontstaan. Eeuwenlang ontwikkelde de wetenschap zich zonder inktjetvullingen. Dit wil zeggen dat in de metadatering het duidelijk moet zijn of kleur vereist is, of het ook in kleur kan of dat het alleen maar leuk is om 2 lijnen in een grafiek elk een andere kleur te geven.

Dit wil zeggen dat de boodschap zoals de auteur en de redacteur die definiëren, een maximum en een minimum representatie moet kennen. Via style sheets kunnen dan keuzes gemaakt afhankelijk van het beschikbare representatiemedium.

Ook staat bij het elektronisch boek de notie van een "werk" hoog op de agenda. Een mededeling van een auteur is over het algemeen een afgerond verhaal of betoog, ook al kan het einde open zijn. Vandaar ook dat wij in ons onderzoek naar modulair wetenschappelijk publiceren, de zelfstandigheid van een module als communicatief object centraal stellen. In het geval van een roman is de module de hele roman, niet meer en niet minder.

Naast de eenheid van een werk kan er enorm veel extra worden toegevoegd. Maar dat zijn de voetnoten die in zichzelf weer korte modules zijn. Het spanningsveld bij het elektronische boek is de mate van gesloten- of openheid.

Wij moeten dan ook een onderscheid maken tussen boeken die eigenlijk geen boeken zijn, maar de vorm van een boek hebben gekregen omdat dat de heersende techniek was, zoals het telefoonboek, het spoorboekje of een encyclopedie, en boeken die als zodanig passen in een papieren of *e*-papieren vorm en als compleet verhaal geconsumeerd worden. De toekomst van het *e*-boek, zeker in het onderwijs, zal bestaan uit centrale pedagogische modules ingebed in

verzamelings multimedievoorbelden, zie-ooks, en uitwerkingen.

Dit brengt mij dan weer terug bij mijn grootse zorg, de hyperlink. Al eerder heb ik aangegeven dat de koppeling ten principale asymmetrisch is. Maar niet alleen dat, als wij echt e-boeken willen maken dan moet de plek waar vanuit een, al dan niet on-line, depot een object hergebruikt wordt in een boek wel precies op de goede plaats landen. Kortom, een koppeling moet een punt-tot-punt-koppeling zijn en niet van een plek in een tekst naar een andere tekst als zodanig, zoals je dat nu altijd ziet.

Dit betekent dat de 15^{de} eeuwse prachtige uitvinding van paginering nu geheel overboord gaat. Een plek in een tekst wordt nu een dynamische plek. Een e-boek is immers in staat materiaal zo nodig te importeren, zelfs bij een roman. Bitjes tellen van af de start is dus geen optie.

Het is duidelijk, wij zijn op weg, we hebben fundamentele stappen gezet en daarom weten we dat we er nog lang niet zijn. Wij zien elkaar dus nog wel de komende decennia.

Referenties

1. Zie SGML Users' Group History op <http://xml.coverpages.org/sgmlhist0.html>
2. C. F. Goldfarb, editor, *The standard generalized Markup Language (ISO 8879)*. International Organisation for Standardization, Geneva 1986.
3. Eric van Herwijnen. *Practical SGML* Kluwer Academic Publishers, 1990.
4. *The MIT Technical Information Project*. Physics Today, March 1965, p 28-36. <http://ptonline.aip.org/getpdf/servlet/GetPDFServlet?filetype=pdf&id=PHTOAD000018000003000028000001&idtype=cvips>
5. A.W. Kenneth Metzner. *The multiple use and other benefits of computerized publishing*. IEEE Transactions on professional communication. Vol. PC-18, no3. September 1975, p. 274-278.
6. D.E. Knuth, *TEX and Metafont*. Am.Math Soc./Digital Press. Bedford, MA, 1979.
7. Joost G. Kircz en Jan Bleeker. *The use of relational databases for electronic and conventional scientific publishing*. Journal of Information Science 13 (1987) 75-89. <http://jis.sagepub.com/content/13/2/75.full.pdf+html>
8. IRCS Text processing Services for Electronic Publishing. IRCS, Lancaster UK, 1986
9. Bill Bernickus, Jos Migchielsen, Simon Pepping & Rob Schrauwen. *Tag by Tag. The Elsevier DTD 5 family of XML DTDs*. Version 1.0 March 2005, Elsevier 200.
10. The Text Encoding Initiative (TEI) , <http://www.tei-c.org/index.xml>.
11. Joost G. Kircz. Rhetorical structure of scientific articles: the case for argumentational analysis in information retrieval. Journal of Documentation, 47, 4 (December), 1991, 354-372. Ook via: <http://www.kra.nl/Website/Artikelen/Rhet-structure.pdf>
12. Sillince, J.A.A. Argumentation-based indexing for information retrieval from learned articles. Journal of Documentation, 48(4), 1992, 387-405.
13. Sillince, J.A.A. Literature searching with unclear objectives: a new approach using argumentation. Online Review, 16(6), 1992, 391-410.
14. Joost G. Kircz. Modularity: the next form of scientific information presentation?, Journal of Documentation. vol.54. No. 2. March 1998. pp. 210-235. Ook via: <http://www.kra.nl/Website/Artikelen/Jdoc98.htm>
15. Frédérique Harmsze. PhD Thesis, February 9, 2000. A modular structure for scientific articles in an electronic environment (HTML & PDF). PDF te vinden via: <http://dare.uva.nl/record/78293> (253 pp).
16. J.G. Kircz and F.A.P. Harmsze. Modular scenarios in the electronic age. Conferentie Informatiewetenschap 2000. Doelen, Rotterdam 5 april 2000. In: P. van der Vet en P. de Bra (eds.) *CS-Report 00-20. Proceedings Conferentie Informatiewetenschap 2000*. De Doelen Utrecht (sic), 5 april 2000. pp. 31-43. Ook via: <http://www.kra.nl/Website/Artikelen/mod2k.html>
17. Anita de Waard A. & Joost Kircz (2008). Modeling scientific discourse - shifting perspectives and persistent issues, ELPUB2008. *Open Scholarship: Authority, Community, and Sustainability in the Age of Web 2.0 - Proceedings of the 12th International Conference on Electronic Publishing held in Toronto, Canada 25-27 June 2008*. Edited by: Leslie Chan and Susanna Mornati. ISBN 978-0-7727-6315-0, 2008, pp. 234-245.
18. de Waard, A. and Pandermaat, H. (2010). A Classification of Research Verbs to Facilitate Discourse Segment Identification in Biological Text, Proceedings of the Interdisciplinary Workshop on Verbs. The Identification and Representation of Verb Features, Pisa, Italy, November 4-5 2010. On <http://elsatglabs.com/labs/anita/>
19. <http://www.iconclass.nl/>
20. Maryanne Wolf. *Proust and the Squid. The story and science of the reading brain*. Icon Books 2008.
21. www.e-boekenstad.nl en www.e-boekenstad.wikispaces.com
22. Joost Kircz. E-gadget or E-reader? To be published Logos: Forum of the World Book Community 21:1-2 (2010). Ook via: http://www.kra.nl/Website/Artikelen/Kircz_Egadget_or_Ebook_v4-final.pdf
23. Terje Hillesund. *Digital reading spaces: How expert readers handle books, the Web and electronic paper*, First Monday, Volume 15, Number 4 - 5 April 2010. <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2762/2504>
24. Mark. Z. Danielewski. *Het Kaartenhuis*. Cargo 2001.

◀ ■ ■ ■ Joost Kircz is hij part-time projectleider van het programma elektronisch uitgeven het Kenniscentrum DMCI HvA, Domein Media, Creatie en Informatie - Hogeschool van Amsterdam (www.create-it.hva.nl). Daarnaast is hij eigenaar van KRA Publishing Research (www.kra.nl), ■ ■ ■ ▶