

# Dag van het Document

## Verslag informatiedag ITI-TNO

Kees van der Laan

24 Februari 1992

### 1 Achtergrond

Het waren er twee: de eigenlijke dag met lezingen en een beurs, en de volgende ochtend met een workshop. De organisatie lag in handen van ITI-TNO, met Joop van Gent, als coördinator.

### 2 Lezingen

De leidraad van de lezingen was de opslag en retrieval van documenten, en algemener hoe gaan wij om met de zwellende informatiestromen. Joop van Gent schetste de context. Een steeds nijper wordend probleem tekent zich af: het wordt ons allemaal teveel, wij worden overspoeld door de berg van papier en de minder zichtbare maar oh-zo dwangmatige elektronische informatie. Kunnen wij het met behulp van computers hanteerbaar houden? Of moeten wij maar rücksichtslos opslaan en terugzoeken wanneer nodig, dit na selectie? De automatische bibliothecaris? Of ontstaat er het beroep van IT-keurder?

Godfrey Smart sprak zeer meeslepend onder de titel Document Management and Retrieval.

Hilde van der Tocht deed weer eens uit de doeken hoe het productieproces bij Elsevier Science Publishers verloopt, onder de titel: Van manuscript tot elektronisch product.

David Evans ging er puur taal-technisch tegenaan. Dit was meer een stand van zaken m.b.t. de research. Het draaide om het 'partoon' herkennen, het automatisch herkennen van inhoudelijke aspecten uitgaande van de syntaxis en de semantiek. Ambitieuus.

Gerard Kempen sprak over intelligente auteursomgevingen. Zijn groep houdt zich bezig met de ontwikkeling van hulpmiddelen zoals: Norma—ter ondersteuning van huisstijlen, normalisatie en consistentie, Corrie—spellingschecker met kennis van woord- en zinstructuur.

In Libraries without walls, werd gerapporteerd over het CMU/DEC project om bibliotheken via de (huis)computer te raadplegen. Interessant ook vanwege de consequenties t.a.v. toegang: geen tijds- noch

afwezigheidsrestricties. Dit laatste geldt zowel t.a.v. personeel als het materiaal zelf.

Ikzelf gospelde een ander maal over SGML en T<sub>E</sub>X, m.n. over de relatie tussen beiden. Benadrukt werd dat de transformatie van wiskunde en tabellen in zijn algemeenheid een probleemgebied is, en dat voorlopig auteurs naast schrijver ook programmeur moeten zijn voor een perfecte opmaak.<sup>1</sup>

Bronkhorst sprak over de verwerkingsachterstand bij ministeries, en vooral over het hoe om te gaan met de menselijke aspecten bij de invoering van kantoorautomatisering.

De dag werd afgesloten met een forumdiscussie, helaas was deze wat flauw. De bitterballen en de drankjes waren daarentegen pittig en des te lekkerder.

#### 2.1 Indruk

Mijn persoonlijke mening is dat er niet gestreefd moet worden naar het 'alles' bewaren, niet ons door de angst—het zou weleens belangrijk kunnen worden—laten meeslepen.

Wij moeten de computer—de gebruikers ervan dus—leren, het niet-relevante (automatisch) te vergeten. Dit roept een schijnbaar selectieprobleem op. De context van betrokkenen en hun onderlinge verhoudingen is een graadmeter voor relevantie. Ik vond het vreemd dat het wetenschappelijke model van samenvattingen en trefwoorden, c.q. betreffende/kenmerk bij brieven, en het hele referee mechanisme, onderbelicht bleven. Misschien omdat voor algemene documenten er geen classificatiesysteem bestaat?

### 3 De beurs

De beurs was zeer goed bezet met een 20-tal stands. Wij (NTG) hadden ons bescheiden PR-materiaal in eenvoudige mapjes op een tafel gelegd.

#### Stands/bedrijfjes:

- CELEX  
Centre for Lexical information. Lexicografische

<sup>1</sup> Afgezien van de rol van uitgevers.

database voor Nederlands, Engels en Duits. Zij hebben een nieuwsbrief (gratis).

- **Cognitech**  
Language checking tools for Dutch.
- **Customized Publishing**  
VSR en Océ. Zij hanteren Wordperfect en SGML.
- **Druide**  
Document Retrieval Using Information from Document Elements.
- **Profglot**  
A multi-lingual natural processor, geprogrammeerd in prolog.
- **SPIN**  
Human-Computer Communication using natural language. Een groot project met druide als onderdeel.
- **Spirit**  
Een hulpmiddel voor 'schrijvers.' Het bevat een tekstverwerker, planningshulpmiddel, en een opslaggereedschap (database?) voor document on-

derdelen.

#### 4 Conclusie

Ik vond Corrie en Norma aardig en realistisch.<sup>2</sup> Een ding is vooral duidelijk geworden: er is veel gaande rondom computer-assisted writing, vooral vanuit de taalkundige hoek. Een weelde aan nieuwe informatie werd over het voetlicht gebracht. De echte IT is aan het opstaan. Iets voor de NTG, een beurs tijdens de bijeenkomsten?

ITI-TNO verdient een pluim met de organisatie van dit gebeuren en het vlekkeloos laten verlopen ervan. Een soort proceedings, met extended abstracts staat er aan te komen.

Document $\TeX$ nisch gesproken is het al heel gewoon dat vrij *complex* kopij elektronisch overgestuurd en verwerkt wordt. Mijn verhaal was een bewijs daarvoor. Hoezo uitwisseling van documenten? Just do it!

---

<sup>2</sup>Theo beïnvloedt mij toch meer dan ik dacht.