

Project Proposal

Petr van Brokland

Automatic

Page Layout

Finding the possibilities and ultimate challenges

Introduction

Deze notitie belicht noodzaak en mogelijke aanpak van een actueel probleem waaraan weinig aandacht wordt besteed en dan bijna altijd op basis van foute uitgangspunten. Waar in de traditionele manier van werken met opmaakprogrammatuur zoals Quark XPress en InDesign altijd een menselijke beslissing de definitieve opmaak van een pagina bepaalt, zijn er steeds meer situaties waarin dat geen optie is. Doordat steeds meer pagina's worden gegenereerd met inhoud die uit een database komt - of van

de selectie van de informatie direct wordt bepaald door eigenschappen van de lezer, moet de layout van de pagina's automatisch worden berekend. Er bestaat op het moment vreemd genoeg geen digitaal gereedschap dat enerzijds voldoende flexibel is om in alle mogelijk technieken en soorten layouts te gebruiken, te koppelen is met een grote verscheidenheid aan informatiebronnen, en anderzijds voldoet aan de typografische kwaliteit die met handmatige opmaak wel kan worden bereikt.

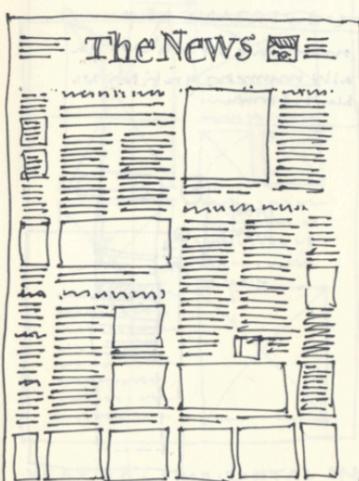
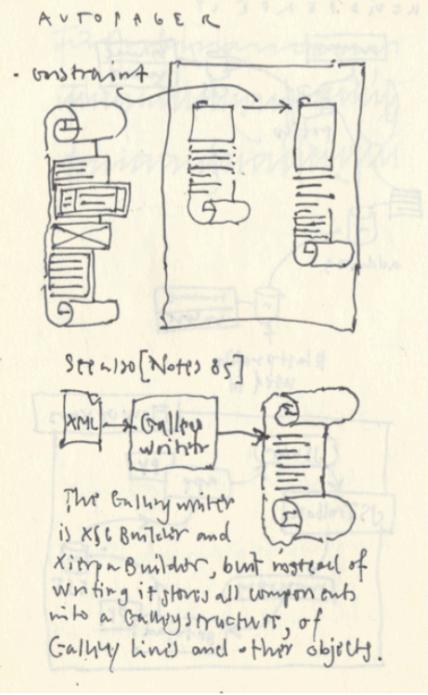
Fasering Het opmaken van een pagina vanuit ruwe kopij en meta-informatie is te verdelen¹ in een aantal fasen:

- Transformatie vanuit Markdown via XML;
- Zetten van de strokenproef;
- Compositie van pagina's;
- Toevoegen van document informatie.

XML transformatie Het transformeren van XML documenten naar een informatiestructuur is met standaardgereedschappen zoals Python-Markdown makkelijk uit te voeren.

Zetten van de strokenproef Om de software te maken die een strokenproef kan genereren is typografische kennis nodig². In vrijwel alle automatische opmaakprogramma's is dit een onderbelicht gebied. Dat komt doordat de bouwers van dergelijke software niet of nauwelijks op de hoogte zijn van de relevante parameters en hun onderlinge samenhang. Ook kunnen verschillen in culturele tradities een rol spelen. USA-typografie is niet hetzelfde als Europese.

In vrijwel alle automatische opmaakprogramma's is dit een onderbelicht gebied. Dat komt doordat de bouwers van dergelijke software niet



Automatic composition of layout in local strategies.

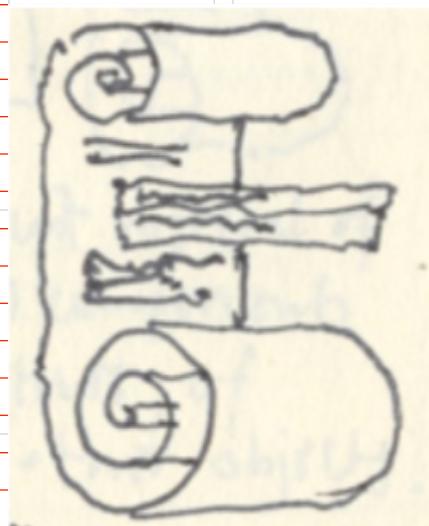
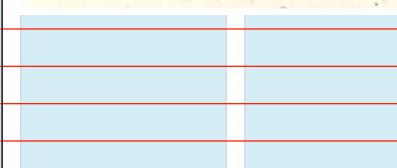
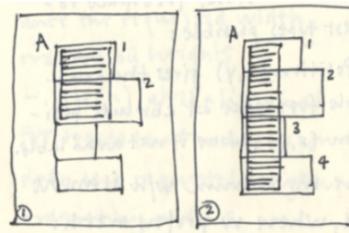
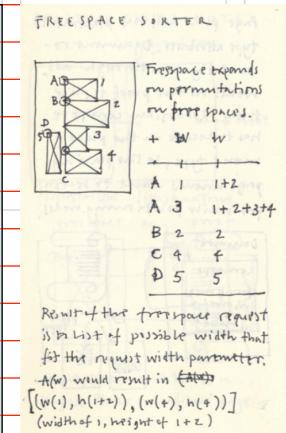
Europese. In vrijwel alle automatische opmaakprogramma's is dit een onderbelicht gebied. Dat komt doordat de bouwers van dergelijke software niet of nauwelijks op de hoogte zijn van de relevante parameters en hun onderlinge samenhang. Ook kunnen verschillen in culturele tradities een rol spelen. USA-typografie is niet hetzelfde als Europese. Met name tabellen zijn lastige typografische bouwstenen, als niet duidelijk is welk volume ze zullen bevat en tot welke marges de inhoud van hun cellen kan schalen. Veel automatische opmaakprogramma's lopen hierop vast.

Zetten van de strokenproef Om de software te maken die een strokenproef kan genereren is typografische kennis nodig. In vrijwel alle automa-

0 0 proef kan genereren is typografische kennis nodig. In vrijwel alle automatische opmaakprogramma's is dit een onderbelicht gebied. Dat komt doordat de bouwers van dergelijke software niet of nauwelijks op de hoogte zijn van de relevante parameters en hun onderlinge samenhang. Daarnaast zijn met name tabellen lastige typografische bouwstenen, vooral als niet niet duidelijk welke volume ze moeten bevatten.

12 2 Veel opmaakprogramma's lopen hierop vast.

15 **Compositie van pagina's** De mate van complexiteit van het componeren van één of meerdere pagina's uit een gegeven volume aan stukkenproeven en beelden, is direct afhankelijk van de structuur van het informatie en het medium waarin moet worden afgebeeld. Een tekst met grove structuur (als er beelden of tabellen in de tekst staan of veel hiërarchie in koppen) is moeilijker te plaatsen dan een homogene tekst. Die gedraagt zich meer als een vloeistof.



18 zijn geanalyseerd.

19 **Toevoegen van document informatie** Pas als de compositie van alle 20 kopij en beelden is afgerekend kan het 21 document worden afgemaakt met de 22 informatie die correspondeert met 23 paginering, zoals paginanummering, 24 inhoudsopgave, beeld- en trefwoor- 25 dindexering en verwijzingen voor 26 voetnoten, literatuur en citaten. De 27 uitdaging in dit stadium is dat het 28 benodigde volume voor deze infor- 29 matie pas aan het eind bekend is, 30 terwijl tijdens de opmaak wel al vol- 31 doende ruimte moet worden gereserveerd. Het kan in extreme situaties 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50

30 5 vloeistof.
31 Er zijn een aantal strategieën moge-
32 lijk om het probleem op te lossen.
33 Niet duidelijk is welke strategie in
34 alle gevallen het beste is of welke
35 strategie past op een bepaalde situatie.

36 Het soort van probleem is gerela-
37 teerd aan andere gebieden zoals
38 speltheorie en kunstmatige intelli-
39 gentie. In praktijk komt het er op
40 neer dat een optimale oplossing ge-
41 zocht moet worden in een snel ver-
42 takkende boom. Het aantal vertak-
43 kingen neemt exponentieel toe,
44 waardoor het aantoonbaar onmoge-
45 lijk is om ze allemaal af te lopen. Net
46 als bij het berekenen van de beste
47 zet in een schaakpartij moet daarom
48 externe context informatie worden
49 toegevoegd om te zorgen dat de
50 waarde van takken kan worden be-
rekend zonder dat deze in detail zijn

