

Project Proposal

Petr van Blokland

Automatic Page Layout

Finding the possibilities and ultimate challenges

Introduction

Deze notitie belicht noodzaak en mogelijk 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 een online source - en waar 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 eisen die aan handmatige opmaak worden gesteld.

Fasering

Het opmaken van een pagina vanuit ruwe kopij en meta-informatie is te verdelen 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

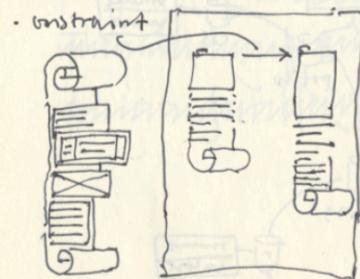
Om de software te maken die een strokenproef kan genereren is typografische kennis nodig.

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. 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 du-

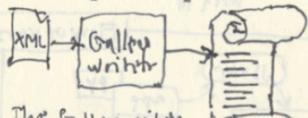


Automatic composition of components in local strategies.

AUTOPAGER

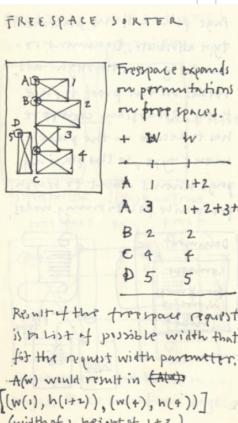
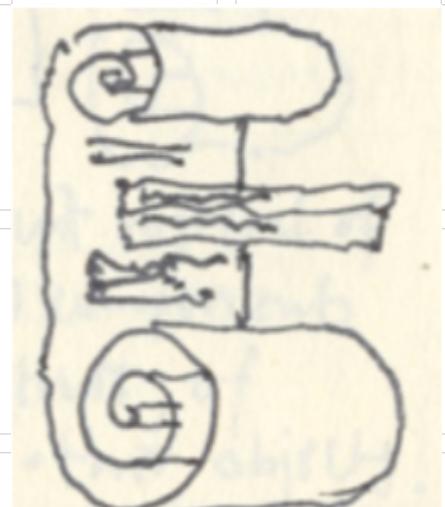
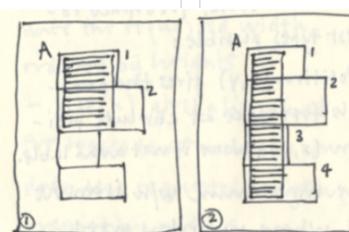
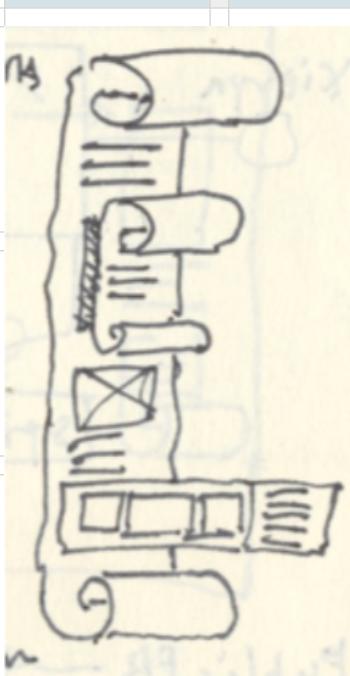
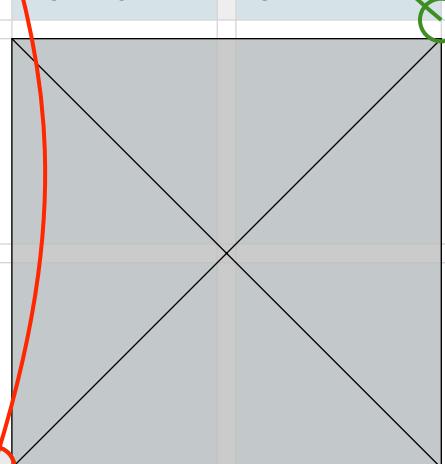


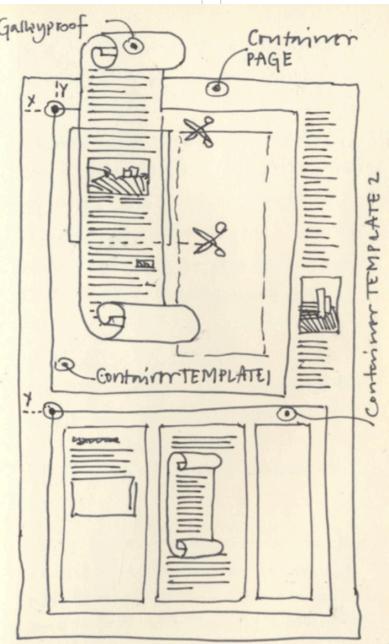
Step 1 > [Notes 85]



The gallery writer is XML builder and Xmp+Builder, but instead of writing it, it writes all components into a gallery structure, of Gallery kind and other objects.

#??#

0	1	2	3	4	5
Met name tabellen zijn lastige typografische bouwstenen, als niet duidelijk is welk volume ze zullen bevatten 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 stroken- 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. Veel opmaakprogramma's lopen hierop vast.				bepaalde situatie. Het soort van probleem is gerelateerd aan andere gebieden zoals speltheorie en kunstmatige intelligentie. In praktijk komt het er op	
		Compositie van pagina's De mate van complexiteit van het componeren van één of meerdere pagina's uit een gegeven volume aan strokenproeven en beelden, is direct afhankelijk van de structuur van het de 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. Er zijn een aantal strategieën mogelijk om het probleem op te lossen. Niet duidelijk is welke strategie in alle gevallen het beste is of welke strategie past op een bepaalde situat		neer dat een optimale oplossing gezocht moet worden in een snel vertakkende boom. Het aantal vertakkingen neemt exponentieel toe, waardoor het aantoonbaar onmogelijk is om ze allemaal af te lopen. Net als bij het berekenen van de beste zet in een schaakpartij moet daarom externe context informatie worden toegevoegd om te zorgen dat de	
					#??#

0	de waarde van takken kan worden berekend zonder dat deze in detail zijn geanalyseerd.			
1	Toevoegen van document informatie			
2	Pas als de compositie van alle kopij en beelden is afgerekend kan het document worden afgemaakt met de informatie die correspondeert met paginering, zoals paginanummering, inhoudsopgave, beeld- en trefwoordindexering en verwijzingen voor voetnoten, literatuur en citaten. De uitdaging in dit stadium is dat het benodigde volume voor deze informatie pas aan het eind bekend is, terwijl tijdens de opmaak wel al voldoende ruimte moet worden gereserveerd.			
3	Het kan in extreme situaties nodig zijn om via backtracking de opmaak aan te passen als blijkt dat de gereserveerde ruimte niet toereikend is geweest.			
4	Randvoorwaarden van het systeem			
5				
6	Er zijn veel voorbeelden van derdelijke systemen die niet goed werken of waarbij het eisenpakket zodanig is verminderd dat met eenvoudige algoritmes kan worden volstaan. De layout van een pagina met één kolom, zoals in tekstverwerkers of boeken, is relatieve eenvoudig automatisch te berekenen. Het wordt exponentieel complex als er meerdere tekststromen tegelijk lopen, zoals bij een tijdschrift of krant het geval is. Als we de vergelijking met de ontwikkeling van schaak-programma's trekken, dan is veel verbetering te behalen door het toevoegen van domeinkennis.			
7	Dit is een concept notitie, aanpassingen en uitbreidingen zijn nog nodig. Deze pagina's werden automa-			
8				#??#