

Project Proposal

Petr van Blok-

land

Automatic Page Layout

1 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 1:

- 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 informatie-structuur 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. 2 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 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

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 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.

Daarnaast zijn met name tabellen lastige typografische bouwstenen, vooral als niet duidelijk welke volume ze moeten bevatten. Veel op-

0 moeten bevatten. Veel op-
1 maakprogramma's lopen hierop
2 vast.

3
4 **Compositie van pagina's** De mate van
5 complexiteit van het componeren van één
6 of meerdere pagina's uit een gegeven volu-
7 me aan strokenproeven en beelden, is direct
8 afhankelijk van de structuur van het de in-
9 formatie en het medium waarin moet wor-
10 den afgebeeld.

11
12 **2 Een tekst met grove structuur**
13 (als er beelden of tabellen in de
14 tekst staan of veel hiërarchie in
15 koppen) is moeilijker te plaatsen
16 dan een homogene tekst. Die
17 gedraagt zich meer als een
vloeistof.

18 **3**
19 Er zijn een aantal strategieën
20 mogelijk om het probleem op te
21 lossen. Niet duidelijk is welke
22 strategie in alle gevallen het
23 beste is of welke strategie past
op een bepaalde situatie.

24 **4**
25 Het soort van probleem is gere-
26 lateerd aan andere gebieden
27 zoals speltheorie en kunst-
28 matige intelligentie. In praktijk
29 komt het er op neer dat een op-

30 **5**
31 Een optimale oplossing gezocht
32 moet worden in een snel ver-
33 takkende boom. Het aantal ver-
34 takkingen neemt exponentieel
35 toe, waardoor het aantoonbaar
36 onmogelijk is om ze allemaal af
37 te lopen. Net als bij het bereke-
38 nen van de beste zet in een
39 schaakpartij moet daarom ex-
40 terne context informatie worden
41 toegevoegd om te zorgen dat de
42 waarde van takken kan worden
43 berekend zonder dat deze in de-
44 tail zijn geanalyseerd.

45 **Toevoegen van document informa-
46 tie** Pas als de compositie van alle kopij en
47 beelden is afgerond kan het document wor-
48 den afgemaakt met de informatie die cor-
49 respondeert met paginering, zoals pagina-
50 nummering, inhoudsopgave, beeld- en tref-
woordindexering en verwijzingen voor voet-
noten, literatuur en citaten. De uitdaging in
dit stadium is dat het benodigde volume

18 uitdaging in dit stadium is dat het benodig-
19 de volume voor deze informatie pas aan het
20 eind bekend is, terwijl tijdens de opmaak
21 wel al voldoende ruimte moet worden gere-
22 serveerd.

23
24 Het kan in extreme situaties
25 nodig zijn om via backtracking
26 de opmaak aan te passen als
27 blijkt dat de gereserveerde
28 ruimte niet toereikend is ge-
29 weest.

Randvoorwaarden van het systeem

30 Er zijn veel voorbeelden van deidelijke sys-
31 temen die niet goed werken of waarbij het
32 eisenpakket zodanig is verminderd dat met
33 eenvoudige algoritmes kan worden vol-
34

0 bestaan. De layout van een pagina met één
1 kolom, zoals in tekstverwerkers of boeken,
2 is betrekkelijk eenvoudig automatisch te be-
3 rekenen. Het wordt exponentieel complex als
4 er meerdere tekststromen tegelijk lopen,
5 zoals bij een tijdschrift of krant het geval is.

6 1 Als we de vergelijking met de ontwikkeling
7 van schaak-programma's trekken, dan is
8 veel verbetering te behalen door het toe-
9 voegen van domeinkennis. Dit is een con-
10 cept notitie, aanpassingen en uitbreidingen
11 zijn nog nodig. Deze pagina's werden auto-
12 matisch opgemaakt met Pagebot (www.pa-
13 gebot.io), een Open Source applicatie voor
14 Drawbot. Buro Petr van Blokland + Claudia
15 Mens

16 Rietveld 56

17 3 2611 LM Delft

18 @petrvanblokland

19 buro@petr.com

20 typetr.typenetwork.com

21 www.pagebot.io

22 www.pagebot.pro

