

Project Proposal

Petr van Blokland

Automatic Page Layout

Finding the possibilities and ultimate challenges

Introduction

Deze notitie belicht noodzaak en mogelijk aanpak

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 in-

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.

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 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 stroken-

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 on-

0 van de relevante parameters en
1 hun onderlinge samenhang.

2
3 Daarnaast zijn met name
4 tabellen lastige typografische
5 bouwstenen, vooral als niet niet
6 duidelijk welke volume ze
7 moeten bevatten. Veel op-
8 maakprogramma's lopen hierop
9 vast.

10 Compositie van pagina's

11
12 De mate van complexiteit van
13 het componeren van één of
14 meerdere pagina's uit een gege-
15 ven volume aan strokenproeven
16 en beelden, is direct afhankelijk
17 van de structuur van het de in-
18 formatie en het medium waarin
19 moet worden afgebeeld.

20
21 Een tekst met grove structuur
22 (als er beelden of tabellen in de
23 tekst staan of veel hiërarchie in
24 koppen) is moeilijker te plaatsen
25 dan een homogene tekst. Die
26 gedraagt zich meer als een

27 aan een homogene tekst. Die
28 gedraagt zich meer als een
29 vloeistof.

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

36 Het soort van probleem is gere-
37 lateerd aan andere gebieden
38 zoals speltheorie en kunst-
39 matige intelligentie. In praktijk
40 komt het er op neer dat een op-
41 timale oplossing gezocht moet
42 worden in een snel vertakkende
43 boom. Het aantal vertakkingen
44 neemt exponentieel toe, waar-
45 door het aantoonbaar onmo-
46 gelijk is om ze allemaal af te
47 lopen. Net als bij het berekenen
48 van de beste zet in een schaak-
49 partij moet daarom externe con-
50 text informatie worden
toegevoegd om te zorgen dat de

0 om te zorgen dat de waarde van
1 takken kan worden berekend
2 zonder dat deze in detail zijn
3 geanalyseerd.

4 Toevoegen van document 5 informatie

6
7 Pas als de compositie van alle
8 kopij en beelden is afgerond
9 kan het document worden afge-
10 maakt met de informatie die
11 correspondeert met pagnering,
12 zoals paginanummering, in-

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.

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.

Randvoorwaarden van het systeem

Er zijn veel voorbeelden van derdelijke systemen die niet goed werken of waarbij het eisenpakket zodanig is vermin-

derd dat met eenvoudige algoritmes kan worden volstaan. De layout van een pagina met één kolom, zoals in tekstverwerkers of boeken, is betrekkelijk 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 schaakprogramma's trekken, dan is veel verbetering te behalen door het toevoegen van domeinkennis. Dit is een concept notitie, aanpassingen en uitbreidingen zijn nog nodig. Deze pagina's werden automatisch opgemaakt

Deze pagina's werden automatisch opgemaakt met PageBot (www.pagebot.io), een Open Source applicatie voor Drawbot. Buro Petr van Blokland + Claudia Mens

Rietveld 56

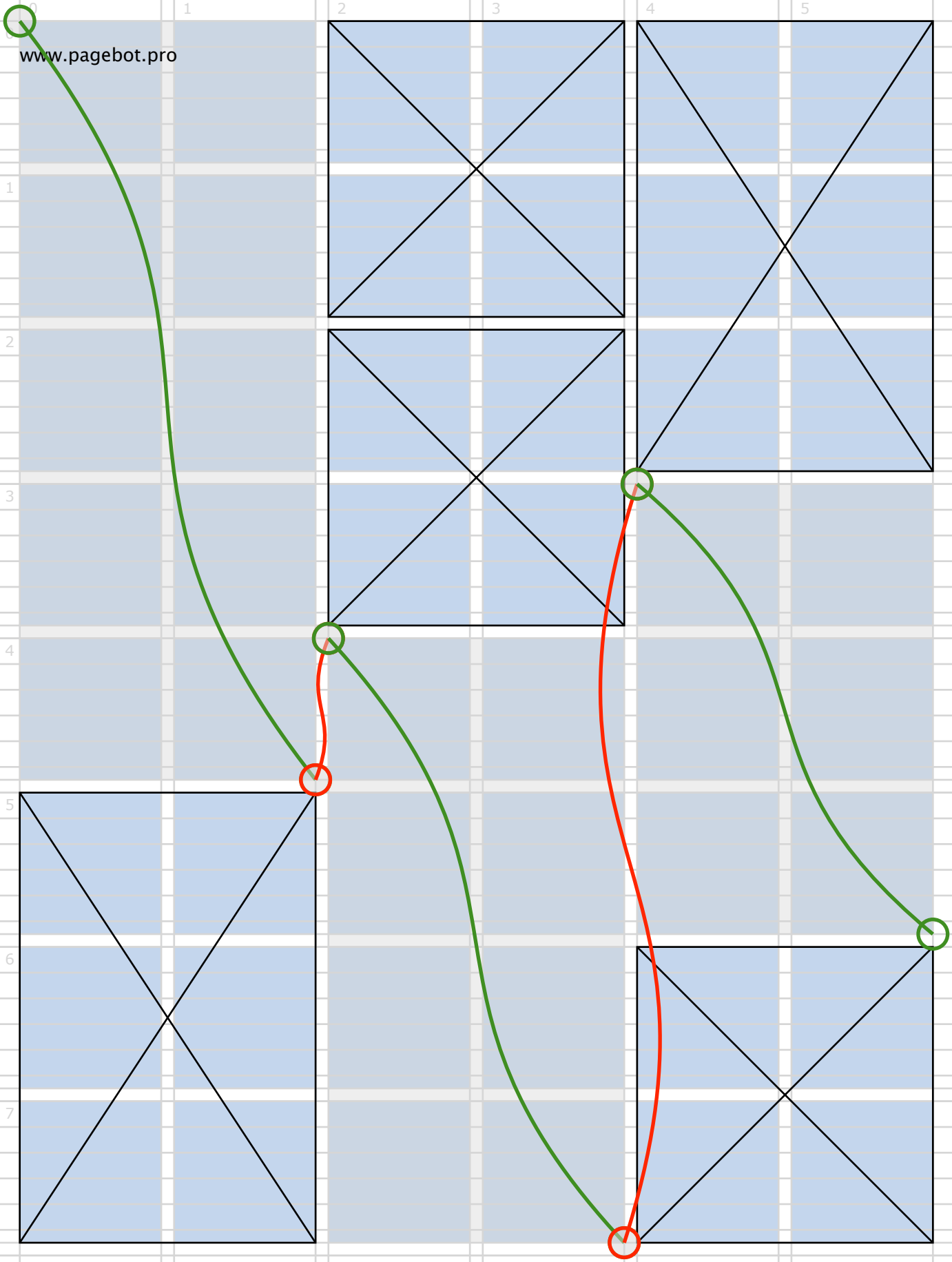
2611 LM Delft

@petrvanblokland

buro@petr.com

typetr.typenetwork.com

www.pagebot.io



www.pagebot.pro

###