From Print to Online Newspapers on Small Displays: A Layout Generation Approach Aimed at Preserving Entry Points

Sebastián Gallardo Díaz*
Universidad Técnica Federico Santa María
Valparaíso, Chile
sebastian.gallardod@usm.cl

Dorian Mazauric
Université Côte d'Azur, Inria
Sophia Antipolis, France
dorian.mazauric@inria.fr

Pierre Kornprobst
Université Côte d'Azur, Inria
Sophia Antipolis, France
pierre.kornprobst@inria.fr

ABSTRACT

Simply transposing the print newspapers into digital media can not be satisfactory because they were not designed for small displays. One key feature lost is the notion of entry points that are essential for navigation. By focusing on headlines as entry points, we show how to produce alternative layouts for small displays that preserve entry points quality (readability and usability) while optimizing aesthetics and style. Our approach consists in a relayouting approach implemented via a genetic-inspired approach. We tested it on realistic newspaper pages. For the case discussed here, we obtained more than 2000 different layouts where the font was increased by a factor of two. We show that the quality of headlines is globally much better with the new layouts than with the original layout. Future work will tend to generalize this promising approach, accounting for the complexity of real newspapers, with user experience quality as the primary goal.

CCS CONCEPTS

• **Human-centered computing → Accessibility design and evaluation methods**; • **Applied computing → Publishing**.

KEYWORDS

Newspapers, digital publishing, document transformation, content customization, layout generation, usability, accessibility, readability, aesthetics, navigation, algorithms

ACM Reference Format:
Sebastián Gallardo Díaz, Dorian Mazauric, and Pierre Kornprobst. 2022. From Print to Online Newspapers on Small Displays: A Layout Generation Approach Aimed at Preserving Entry Points. In *Proceedings of San Jose '22: ACM Symposium on Document Engineering (DocEng '22).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

*Postgraduate education financed by CONICYT (CONICYT-PFCHA/Magister Nacional/2020-22201427)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
DocEng '22, September 20–23, 2022, San Jose, CA, USA
© 2022 Association for Computing Machinery.
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

While there has been a lot of studies and debates around print versus online newspapers, print newspapers continue to be a point of reference in terms of quality and user experience [4]. This attachment we have for print can be explained mainly by the beauty of their design, which provides a strong, consistent, and appealing visual approach to enhance a story's impact and comprehension [3, 6] but also promotes reader engagement [12]. However, combining the design and aesthetics from the printed edition and the functionalities of the online edition remains a complex design challenge [9, 10]. In this sense, we need to go beyond the limits of the two main current modes of access to the digital press, which are: (1) access to an electronic version of the paper edition, on a reader or tablet, through digital kiosk applications (access to more or less enhanced PDF files, in general); (2) access to a web version, adaptable to various screen sizes thanks to "responsive design", which essentially gives us the possibility to read articles but missing out on the print experience.

The print experience is indeed very particular: Numerous behavioral studies have shown that a newspaper is not read linearly and exhaustively. In [5], the authors have highlighted the existence of *entry points* (e.g., photos, illustrations, headlines) that provide readers with the necessary elements to scan a page and select what they want to read. As such, these entry points are essential in the print user experience. Since online newspapers have been fundamentally a simple transposition of the print newspaper into digital media, one might think that the print experience should the same. Unfortunately, this simple reasoning does not work simply because the print newspapers are not designed to be read on small displays. To convince yourself, let us consider here the case of headlines which will be our focus in this paper. When the headlines become less readable (because their font is too small), they can no longer play their role as a hook to make us want to read a particular article (see Fig.1(a)). So there is a need for magnification. One classical option in kiosk applications is to zoom in on the content thanks to intuitive touch screen gestures (e.g., pinch-to-zoom). However, this type of magnification has a significant drawback as we quickly lose our positioning inside the page (Fig. 1(b)), which impacts reading comfort and navigation significantly.

In this paper, we aim to find a solution to preserve the print experience online, i.e., enable this scanning and selection procedure that defines the print reading experience. We present a new approach to transforming a print newspaper layout to fit small displays while preserving entry points. This principle will be explored here, focusing on headlines. Our goal will be to guarantee that headlines should be readable (referring to the font size) and usable (referring to the number of lines to split the headline string into). Concerning

this last property, note that most newspaper headlines occupy one or two lines –sometimes three or four, which can be related to the easiness of reading the headline. The problem is that these two properties are *a priori* contradictory since increasing the font size makes headlines split into more lines (Fig. 1(c)).

In this paper, we propose a method that explores alternative shapes for articles to generate new layouts where headlines will be readable and usable. This brings a new kind of packing problem which is an NP-hard problem. To solve this layout generation problem, one possibility would be to consider constraint-based or template-based ones However, these approaches generally lead to a single optimal layout (or just a few). Thus a better option is to consider random search procedures like genetic algorithms (GA, e.g., [7]) which will bring two main advantaged to our problem: (i) many solutions can be obtained at once, (ii) more adapted to artistic tasks because they do not attempt to mimic or model any particular process by which solutions are created.

2 OUR APPROACH

2.1 Problem Definition and Fitness Function

Let us consider a newspaper page specified by a multi-column layout and denote by γ the width of the columns (Fig. 1(e)). This page is made of a set of N articles $\{\bar{\mathcal{A}}_i\})_{i=1..N}$ defining the original layout $\bar{\mathcal{L}}$ (Fig. 1(d)), which is denoted by the symbol $\bigoplus$:

$$\bar{\mathcal{L}} = \bigoplus_{i=1..N} \bar{\mathcal{A}}_i.$$

Note that bars over letters refer to the original layout. By construction, these articles define a partition of the page. Each article $\bar{\mathcal{A}}_i$ is defined by (i) its shape $\bar{\mathcal{S}}_i$, which has dimensions $(\bar{w}_i, \bar{h}_i)$, defining an aspect ratio $\bar{\lambda}_i = \bar{h}_i/\bar{w}_i$, (ii) its position here defined by the upper-left corner position $(\bar{x}_i, \bar{y}_i)$, and (iii) its content. In this paper we assume that the content will always consist in a headline and just body text for the main body.

Then let us introduce an essential notion for our paper, which is the notion of *allowed* versus *unwanted* shapes. Given a shape $\mathcal{S}$, let us first denote by the function $\lambda(\mathcal{S})$ the number of lines a headline is split into (concerning how this number of line is estimated, note that here we chose justified headlines in all cases, but any other choice would work). By definition, a shape $\mathcal{S}$ will be either *allowed* if $\lambda(\mathcal{A}) \leqslant \lambda_{max}$ or *unwanted* otherwise. As shown later, the number of unwanted shapes (UW) will be essential in evaluating a layout since it will provide information about the headline's usability.

Given these definitions, we can define our problem as follows: Consider that a reader wishes to increase the font size by a magnification factor of α (e.g., to compensate for a visual acuity limitation that prevents him or her from reading the headlines correctly), is the original layout satisfactory (i.e., no unwanted shapes) or does it exist alternative better layouts (with less unwanted shapes)?

Answering the first question is easy. If we simply increase font size using the original layout, most headline will naturally flow on many lines (see, e.g., Fig. 1(c) with six unwanted shapes out of nine). So we need to find alternative layouts which is a difficult problem. Indeed, one needs to discard unwanted shapes, but as soon as we change one shape, this generates a domino effect needing a complete relayout of the page by exploring alternative shapes for all articles (even if they were not *a priori* impacted by the magnification).

Let us now define more precisely the notion of alternative shapes. For an article $\bar{\mathcal{A}}_i$, we define N_i alternative shapes $\{\mathcal{S}_{i,j}\}_{j=1,...,N_i}$ that are the shapes having the same area as the initial one, widths being multipliers of γ (adjusting the height in a way that preserves the article's area), and of course fitting into the page. Note that we will use $j = 1$ for the original shape (i.e., $\mathcal{S}_{i,1} = \bar{\mathcal{S}}_i$). For example, Fig. 1(e) shows the five possible shapes for $\bar{\mathcal{A}}_6$ (including the original one). Note that the initial position should not be considered to build alternative shapes (they should only fit the page).

Given this set of all possible shapes $\mathcal{S}_{i,i}$ for each article i, let us define the solution space, i.e., the combinations of shapes leading to good alternative layouts (i.e., no overlapping articles):

$$\mathcal{E} = \{\{\mathcal{S}_{i,j}\}_{i=1..N} \quad / \quad \exists \quad \mathcal{L} = \bigoplus_{i=1..N} \mathcal{A}_{i,j}\},$$

where $\mathcal{A}_{i,i}$ is derived from the original article $\bar{\mathcal{A}}_i$ and is defined by (i) the alternative shape $\mathcal{S}_{i,j}$, (ii) a new position $(x_{i,j}, y_{i,j})$, (iii) a content that is cut because of the magnification (normally concerns only body text for reasonable magnification factors). Note that we target a digital experience so that the full body text will still be accessible once an article is chosen.

Now the problem becomes to find the best layout, i.e., the best combination of shapes for each article, with a minimal number of unwanted shapes, thus leading to the best alternative layout $\mathcal{L} \in \mathcal{E}$. To do so, here we propose to maximize the fitness function E:

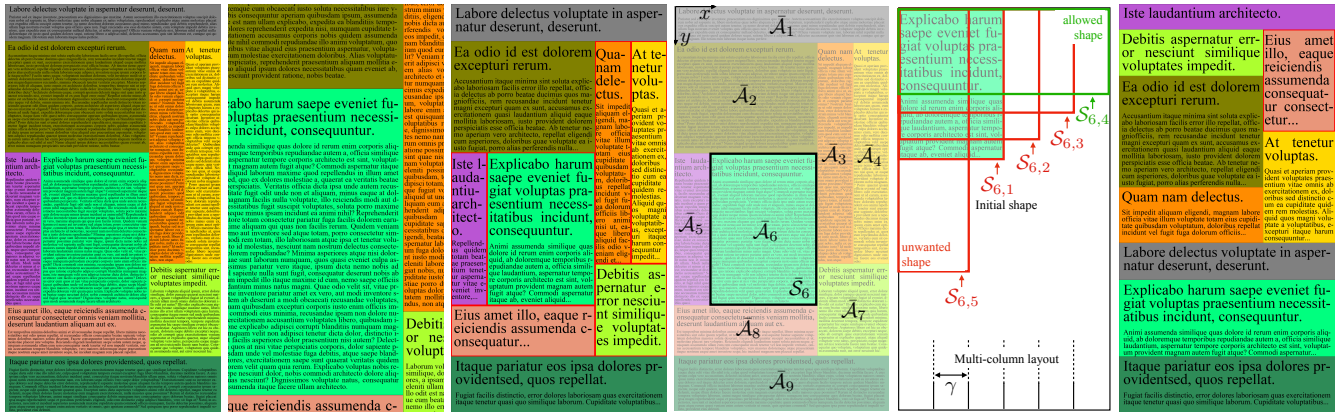
$$\max_{\mathcal{L}=\bigoplus \mathcal{A}_{i,j} \,/\, \{\mathcal{S}_{i,j}\} \in \mathcal{E}} E(\mathcal{L}) = H(\mathcal{L}) + G(\mathcal{L}) + F(\mathcal{L}),$$

where each of the three terms of have a specific role that we describe here concisely. The term $H(.)$ is related to the number of lines of each headings in chosen shape configurations. It penalizes strongly shapes that elicit headlines flowing on high number of lines ($> \lambda_{max}$). The term $G(.)$ is related to the style. This term can be interpreted as a fidelity attach term with respect to the original newspaper design. In this paper, for style, we have considered that the distribution of aspect ratio of shapes for our solution should respect what is statistically observed in an ensemble of pages from our target newspaper. Finally, the term $F(.)$ is related to the aesthetics of the layout. In this paper, it is based on three criteria measuring alignment, regularity and balance. We have defined them taking some inspiration from [2, 8].

2.2 Implementation as a GA-inspired approach

2.2.1 Method Overview. We use a random procedure inspired by GA. First, we need to define the basic data structure (a chromosome) representing, in our case, a single generated layout. We use a structure that defines a specific order or permutation of the articles that will be packed using a heuristic. Here, we use Bottom Left Heuristic [1], which packs each article as low as possible in a left-justified position. So, each chromosome contains two lists: (i) a list of which shape to choose for each article and (ii) a permutation list that defines the order in which the articles will be packed in the page.

Given a page (defining a set of articles and multi-column layout) and a magnification factor, the algorithm works as follows: (1) Generate and classify alternative shapes for each article as allowed



(a) (b) (c) UW=6 (d) (e) (f) UW=1

Figure 1: From print to online newspapers on small displays: (a) Print newspaper page where each article has been colored to highlight the structure. On small displays, headlines become too small to be readable and usable for navigation. Magnification is needed. (b) Pinch-zoom result with a magnification factor of two: Illustrates the common local/global navigation difficulty encountered when reading newspapers via digital kiosk applications. (c) Increasing the font with a factor two, keeping the original layout: Articles with unwanted shapes (UW with $\lambda_{max} = 3$, see Sec. 2) are boxed in red; (d) Original page with articles numbered. (e) Alternative shapes for article $\bar{\mathcal{A}}_6$ using the multicolumn grid (suggested at the bottom), where colors indicate whether the shapes are allowed (in green) or unwanted (in red). For article $\bar{\mathcal{A}}_6$ the only allowed shape is the one occupying all page width. (f) Best result of our approach showing how the page in (a) has been transformed to preserve headlines quality.

or unwanted; (2) Generate M random chromosomes, which means choosing random allowed shapes for each article (or unwanted ones if there are no allowed shapes), and a random permutation; (3) Build the layout and compute the fitness value for each chromosome (if the solution does not lead to an admissible layout, fitness is set to 0); (4) Then, in each one of the N generations, apply crossover with probability p_c (see Sec. 2.2.2), mutation with probability p_m (see Sec. 2.2.3), and select the best M generated solutions.

2.2.2 Crossover. The crossover procedure is similar to the mechanism proposed by [11], adapted to the chromosome proposed in this work. The idea is, starting from two parent chromosomes, perform two random cuts in both (that means, define two points in the chromosome array), and copy the first parent information to the child (new chromosome) only for the articles between these two points, and copy the second parent information for the rest of the articles in the child. We also build a second child with the analog process: copy the information of the second parent only between these two points and from the first one for the rest of the articles.

2.2.3 Mutation. The method can change the shape to another one (chosen randomly) of the same article, with probability p_{ms}. In this case, with probability p_{uw}, change the shape to another one from the *unwanted* list. In this scenario, the algorithm chooses which shape using a *roulette wheel selection*, a random procedure where shapes that need more lines to fit the content have less probability of being chosen. On the other hand, with probability $1 - p_{uw}$, change the shape to a randomly chosen article from the *allowed* list. Also, with probability p_{or}, change the permutation array: generate two random numbers p, q, and reverse the order of the sub-array between indices p and q.

3 RESULTS

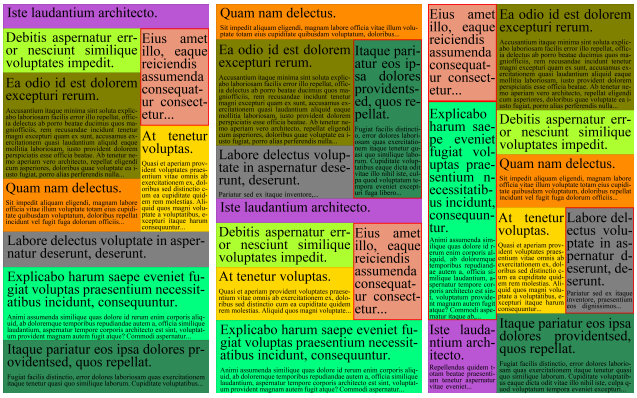
We tested our approach on realistic examples inspired by authentic newspapers. Given pages from a target newspaper (here front pages from *New York Times*), we used them in two ways: (i) to define the distribution of aspect ratios appearing in the style term (for the proportion measure). (ii) to define our input: Starting from one page, we replaced original articles' content with a heading and a body text generated with a lorem-ipsum generator. We assigned to the headlines a random number of lines between 1 and $\lambda_{max} = 3$ so that the original layout had only allowed shapes (before magnification was applied).

For the GA, we use $N = 100$ generations with a population of $M = 100\,000$ chromosomes. We used $p_c = 0.4$, $p_m = 0.9$, $p_{ms} = 0.9$, $p_{uw} = 0.1$, and $p_{or} = 0.9$.

Our input here is shown in Fig. 1(a). We considered a magnification factor $\alpha = 2$. We obtained 2440 acceptable layouts out of the 100 000 chromosomes. On a standard laptop, the GA took about 10 minutes to run and obtain all these solutions for one page.

Three alternative layouts are shown in Fig. 2. They are ranked in decreasing order with respect to the fitness function (Fig. 3(a)). The best layout found by our method (Fig. 2(a)) is much better in terms of number of unwanted shapes (UW=1) than using the original layout (Fig. 1(c), UW=6). More generally, if we look at all solutions, we observe that our method generates many other alternative layouts with few unwanted shapes (a majority with one or two unwanted shapes, see Fig. 3(b)).

Figure 3(c) shows the style term values (here proportion) w.r.t. layout rank. We can observe its impact by looking at the layout #2440 (Fig. 2(c)) which has the highest proportion measure. When compared to layouts #1 and #222 (Fig. 2(a)–(b)), layout #2440 has an aspect ratio distribution closer to our sample set of pages, which



(a) UW=1 (b) UW=2 (c) UW=3

Figure 2: Layouts generated by our approach corresponding to the original layout shown in Fig. 1(a) with $\alpha = 2.0$: (a) Best layout (layout #1), (b) layout #222, (c) layout #2440.

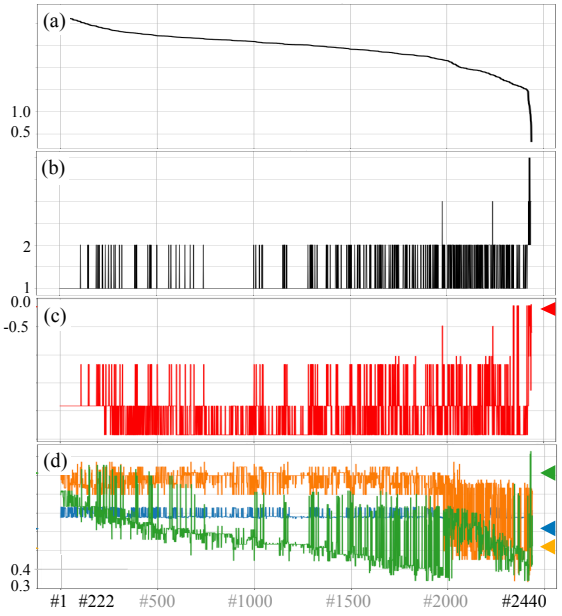


Figure 3: Properties of solutions. In each panel the x-axis represents the solutions sorted in decreasing order of the fitness function (i.e., layout #1 is the best). (a) Fitness function, (b) number of unwanted shapes, (c) style measure, namely proportion (red), (d) aesthetic measures, namely alignment (blue), regularity (orange), and balance (green). For (c) and (d), colored triangles on the right-hand side show values corresponding to the original layout magnified (Fig. 1(c)).

is achieved by using narrow blocks instead of wide ones. However, this solution is the worst given our parameters because there is a trade-off between style and the other terms. In our case, we chose to penalize the number of unwanted shapes more since our focus is on the quality of entry points.

Figure 3(d) shows the different terms of the aesthetic term values (here alignment, regularity, and balance) w.r.t. layout rank. Alignment value does not vary significantly across solutions and it is always better than the input. The reason is that we use a multi-column design to generate and position the shapes: horizontal alignment is mostly constant and we have small variations coming from the vertical alignment. A similar observation can be made for regularity. However, balance is worse sometimes because the algorithm prefers to drop that measure to focus on entry points and the other aesthetic criteria.

4 CONCLUSION

In this paper, we have highlighted that entry points are essential features for readers to scan pages, and thus, one main flaw of simply transposing print into digital media is that the reading experience cannot be as enjoyable and comfortable just because entry points do not play their role correctly anymore. A typical and easy-to-understand example of entry points that are impacted by magnification are headlines which have been our focus here. Interestingly, preserving headline quality (readability and usability) led us to a nontrivial layout generation problem where aesthetics and style were also taken into account. Results are promising. This work opens an entirely new avenue of research to tackle the complexity of newspapers, and make them accessible online to everyone.

ACKNOWLEDGMENTS

We are thankful to B. Génuit (Demain Un Autre Jour) who provided expertise that assisted the research.

REFERENCES

[1] Bernard Chazelle. 1983. The Bottom-Left Bin-Packing Heuristic: An Efficient Implementation. *IEEE Trans. Comput.* C-32, 8 (Jan. 1983), 697–707. https://doi.org/10.1109/TC.1983.1676307

[2] John G. Byrne David Chek Ling Ngo, Lian Seng Teo. 2000. A Mathematical Theory of Interface Aesthetics. *Visual Mathematics* 10, 8 (2000). http://eudml.org/doc/256723

[3] Javier Errea and Gestalten (Eds.). 2018. *Newspaper Design: Editorial Design from the World's Best Newsrooms.* gestalten.

[4] Leopoldina Fortunati, Sakari Taipale, and Manuela Farinosi. 2015. Print and online newspapers as material artefacts. *Journalism* 16, 6 (2015), 830–846. https://doi.org/10.1177/1464884914545439

[5] Mario R. Garcia and Pegie Stark. 1991. *Eyes on the News.* Poynter Inst Media Studies.

[6] Damien Gautier and Claire Gautier. 2017. *Design, Typography etc.: A Handbook.* niggli.

[7] Joe Geigel. 2002. Automatic Page Layout Using Genetic Algorithms for Electronic Albuming. *Proceedings of SPIE - The International Society for Optical Engineering* (08 2002). https://doi.org/10.1117/12.411879

[8] Steven J. Harrington, J. Fernando Naveda, Rhys Price Jones, Paul Roetling, and Nishant Thakkar. 2004. Aesthetic Measures for Automated Document Layout. In *Proceedings of the 2004 ACM Symposium on Document Engineering (DocEng '04).* Association for Computing Machinery, New York, NY, USA, 109–111. https://doi.org/10.1145/1030397.1030419

[9] Barry A. Hollander, Dean M. Krugman, Tom Reichert, and J. Adam Avant. 2011. The E-Reader as Replacement for the Print Newspaper. *Pub Res Q* 27 (2011), 126–134. https://doi.org/10.1007/s12109-011-9205-8

[10] Carina Ihlström, Maria Åkesson, and Stig Nordqvist. 2004. From Print to Web to e-paper-the challenge of designing the e-newspaper. *Proceedings of ICCC 8th International Conference on electronic Publishing* (2004), 249–260.

[11] Manuel Iori, Silvano Martello, and Michele Monaci. 2003. *Metaheuristic Algorithms for the Strip Packing Problem.* Springer US, Boston, MA, 159–179. https://doi.org/10.1007/978-1-4613-0233-9_7

[12] Ricki Morell. 2016. Design as a Driving Force for Audience Engagement. (2016). https://nieman.harvard.edu/articles/design-as-a-driving-force-for-audience-engagement/ Nieman Reports.