MANCHESTER 1824
The University of Manchester

The University of Manchester Research

Experiential transcoding: an EyeTracking approach

DOI:
10.1145/2461121.2461134

Link to publication record in Manchester Research Explorer

Citation for published version (APA):
Yesilada, Y., Harper, S., & Eraslan, S. (2013). Experiential transcoding: an EyeTracking approach. In *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility* (pp. 1-4). (W4A '13). Association for Computing Machinery. https://doi.org/10.1145/2461121.2461134

Published in:
Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [http://man.ac.uk/04Y6Bo] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.

Download date:09. Jun. 2022

Experiential Transcoding: An EyeTracking Approach

Yeliz Yesilada
Middle East Technical University Northern Cyprus Campus, Güzelyurt, Mersin 10 Turkey
yyeliz@metu.edu.tr

Simon Harper
[1]Schoool of Computer Science University of Manchester, Manchester UK
simon.harper@manchester.ac.uk

Sukru Eraslan[1,2]
[2]Middle East Technical University Northern Cyprus Campus, Güzelyurt, Mersin 10 Turkey
seraslan@metu.edu.tr

ABSTRACT

Transcoding web pages for ease of use for small screen device users and for disabled users have been researched extensively. However, there has been very little research on transcoding web pages based on understanding and predicting users' experiences. In this paper, we discuss the concept of experience-based transcoding, called "*experiential transcoding*", and present our initial work on identifying patterns in eye-tracking data to guide transcoding of web pages for improving the experience of blind and situationally impaired users.

Keywords

Transcoding, Eyetracking, UX, Mobile Web, Visual Disability, Situational Impairment

Categories and Subject Descriptors

H.1.2 [**User/Machine Systems**]: Human information processing.; H.5.4 [**Hypertext/Hypermedia**]: Navigation, User issues.

1. INTRODUCTION

The World Wide Web (web) is not anymore 'Desktop only'. It can accessed by different devices with different requirements and constraints. This enriches the web but brings a lot of burden on web designers. Typically, web pages are designed for visual interaction under specific conditions, which means it is difficult to access in alternative forms [8, 16]. One approach to address this problem is transcoding which reengineers web pages such that they are easy to access by small screen devices and by assistive technologies [2]. There has been significant research on transcoding but so far the focus has been on using the content as input to transcode web pages. In this communication paper, we expand on the concept of experience-based transcoding, called "*experiential transcoding*", which we first rationalised in in our previous SASWAT[1] project [4] and later contextualised [7]. Experiential transcoding differs from plain content transcoding because it is user centred and includes techniques that transcode web pages based on understanding and predicting users' experience. One approach to understand users' experience is to use eyetracking data. Eye tracking has widely been used to investigate cognitive processes for over 30 years [12], but comparatively speaking, monitoring people's eye movements during web use is a relatively new area [17]. While reading, the eye does not generally move smoothly but it makes continuous movements called saccades. Between the saccades fixations occur where the eye is relatively immobile and those fixations show the area where attention is being allocated. Scanpaths are repetitive sequences of fixations and saccades that occur upon a visual stimulus. In our previous work, we have investigated fixations and saccades to understand how people use visual elements of web pages [17], how they allocate their attention to dynamic content [11], and how they process simple and complex pages [3]. These studies generated enormous amount of data and in our current project, called eMine[2], we are developing an algorithm for identifying people's scanpaths and relate those scanpaths to visual elements of web pages. By relating scanpaths to the underlying source code of web pages, we are aiming to achieve "experiential transcoding" of web pages such that they are easier to access in constrained environments for both small screen device users and disabled users who use assistive technologies to access the web. In this paper, we present our preliminary work on identifying common scanpaths and explain how those scanpaths can be used for achieving experiential transcoding.

[1]http://wel.cs.manchester.ac.uk/research/saswat/

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. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
W4A2013 - Communication, May 13-15, 2013, Rio de Janeiro, Brazil. Co-Located with the 22nd International World Wide Web Conference.
Copyright 2013 ACM 978-1-4503-1844-0 ...$5.00.

2. RELATED WORK

Transcoding is the manipulation of content in various ways to make it more suitable for alternative presentation, such as on small screen devices and audio presentation. Transcoding can be achieved by using different approaches, such as by using heuristics, users' preferences, external annotations or via proxy [2]. Transcoding methods include adding a skip link, generating summaries of the content, generating thumbnail images of web pages, ranking and reordering the content, segmenting content into smaller fragments and removing irrelevant content [2]. Although existing research shows that transcoding techniques improve disabled and mobile web users' experience on the web [18], it is also clear that a good transcoding technique depends on a good understanding of structure, content and context of use [17]. Unfortunately, not many of these studies try to understand how web pages are used in reality to do transcoding. We believe eye-tracking studies provide a good way of achieving that.

Eye tracking has widely been used to investigate cognitive processes for over 20 years [12], but eye tracking during web use is

[2]http://emine.ncc.metu.edu.tr/

a relatively new area [9]. The most obvious applications of eye-tracking are in improving the standard design and layout of web pages, and evaluating their usability [13]. Studies have also examined the saliency of items on a page under varying conditions, how eye movements vary according to information scent and how looking for a menu is influenced by page complexity and prior expectations [10]. In our previous work, we have investigated fixations and saccades, which are considered as static metrics [14], to understand how people use visual elements of web pages [17], how they allocate their attention to dynamic content [11], how they process simple and complex pages [3]. Alternative dynamic metric for analysing eye tracking data is scanpath analysis [14]. Although there are a number of studies on scanpath analysis, these studies focus on "String-edit" algorithm which uses web pages as images [14] and none of these studies investigated the relationship between scanpaths, visual elements of web pages and the underlying source code. Without relating the scanpath to the underlying source code, it is very limited what you can achieve with the eye tracking data.

3. EYETRACKING APPROACH

Our aim is to address a gap in the literature and develop an algorithm to mine the eye tracking data to identify scanpaths and relate them to visual elements of web pages and the underlying source code, so that web pages can be transcoded to provide better user experience to disabled and mobile web users. This kind of transcoding will be an example of "experiential transcoding".

3.1 Segmentation

Eye tracking data analysis is typically performed by creating areas of interests on web pages (AoI). These are then used to understand how people's eyes traverse a web page. Typically these AoIs are specified manually by the person who analyses the eyetracking data. However, as we have suggested in our previous work [17], web pages are designed in such a way that they include visual elements to guide the user in traversing and reading the content. To automate the process of identifying the visual elements, we have extended and improved the well known web page segmentation algorithm called VIPS [1], and developed it on the Accessibility Tools Framework (ACTF) platform[3]. This extended algorithm takes a web page and uses both the visual presentation and the underlying source code to divide a web page into a number of segments, for example Figure 1 shows an automatically segmented web page. Here, segment B is the header, segment D is the menu, segments A, C and F are part of main content, and segment G is the footer.

3.2 Eyetracking in terms of Visual Elements

Once we automatically segment a web page to identify visual elements, we can take an eyetracking data and relate it to these segments. This would allow us to mine the interesting patterns in eyetracking that which would allow us to understand and predict how people would use a web page.

Figure 1 has been used in another eyetracking study [5] to investigate how people interact with dynamic content. Here we repurpose that eyetracking data and investigate it to see how we can use this eyetracking data set to understand and predict the users' experience. Twelve people participated in this eyetracking study who were students/university employees between the ages of 18-45 and Figure 2 shows the heatmap of the collected eyetracking data. The task on this page was to locate the latest news and click on the special offers. Out of these 12 participants, there were some

[3]http://www.eclipse.org/actf/

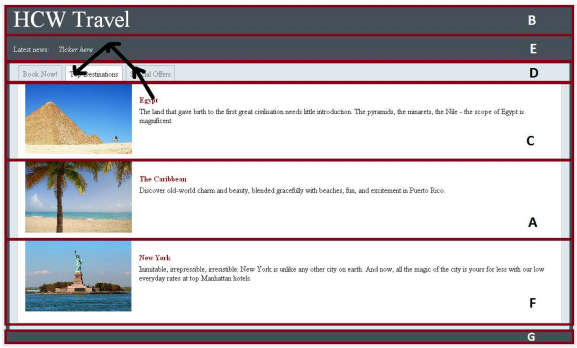

Figure 1: Example Segmented Web Page.

	Scanpath
S1:	ABCBDCACACBEBDEBE
S2:	CAEBEBDEBEBEDEDCECEDECACDC
S3:	CABFACDCDEDAEDFDADFAEDEDEDEDEDACDCDEDC
S4:	CACDCDCACFDEBECACDEBDEBEBEBEBECEDECECD
S5:	DACADEDEDEDEDEDC
S6:	DCDCABEDCD
S7:	BDCEBCDCDCDEDEDCD
S8:	CACDECECD
S9:	BCECDCDCDEDECDC
S10:	CACDADCBECDCB

Table 1: Sample Scanpaths on Figure 1

problems in eye-tracking recordings for two participants since they were distracted, therefore we eliminated their data from our further study. Table 1 shows the scanpaths of these participants in terms of the visual elements given in Figure 1. In order to also support the overlying of eyetracking data onto given visual elements on a web page, we extended our VIPS algorithm implementation on the ACTF platform. Once a web page is automatically segmented, one can export existing eyetracking data set on that particular page to see the scanpaths of those participants in terms of the visual elements that are automatically identified. As can be seen from Table 1 people have different strategies to complete the same task, and they follow slightly different paths to achieve the same task. For example, participant 3 and 4 followed a longer path in terms of the number of visual elements visited to complete the given task compared to participant 6 and 8.

3.3 Patterns in Scanpaths

In order to properly compare the scanpaths, their similarities and

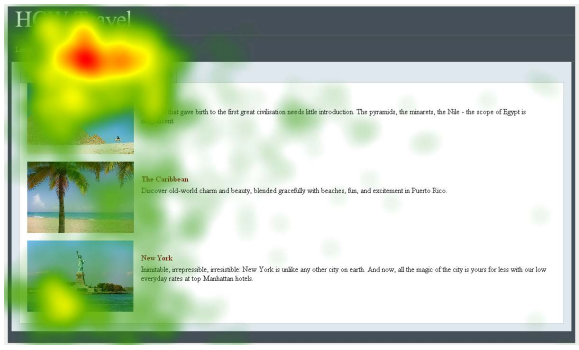

Figure 2: Heat Map of an Example Web Page.

differences, there exist algorithms in the literature [12]. Most of these studies are based on the algorithm known as "String-Edit" that compares two strings of data and finds out if they are similar or not. However, these algorithms mainly compare two scanpaths as given as string inputs and finds if they are similar. Furthermore, these algorithms mainly interpret input as images and there is no further understand of the underlying source code. However, in order to be able to use the identified patterns in transcoding, we need to be able to relate scanpaths to the underlying source code. Our approach, however, explained in the previous section allows us to generate scanpaths in terms of visual elements of web pages.

When we look at the literature, there is also not much research that takes more than two scanpaths and tries to understand how people traverse a web page, and tries to finds common patterns in scanpaths – Dotplots and T-Pattern are two examples [6]. However, when we look at these example algorithms, they all suffer from the following limitations:

- They are reductionist. When multiple scanpaths are compared, the existing algorithms are very reductionist. When the number of scanpaths increase, the common scanpaths get smaller and not very useful. For example, for the data set in Table 1, Dotplots algorithm generates "D" as the common scanpath. This is actually valid conclusion as all the participants were asked to locate the "special offers" which is part of the menu in visual element D. Recently, there has been also some work on improving Dotplots to address the issue of being reductionist but further work needs to be conducted on this issue.
- They tend to ignore the complexities of underlying cognitive processes: when one follows a path to achieve a task, there is a reasoning that affects their decision, and none of these algorithms capture that.
- All of these algorithms accept simple string as inputs. However, when we work with web pages it is not so straightforward to divide a web page into AoIs, for instance will the page be divided based on the visual elements? Will the page be divided according to the task? etc.
- Eye trackers record a lot of data. It is also important to have a systematic approach to eliminate the noisy data from the eyetracking data. None of the existing algorithms address the issue of how to identify noisy data.

Our current work aims to address these limitations and issues, however, in this paper we only focus on the first problem - being reductionist. We are developing an algorithm that takes a number of scanpaths and returns a pattern that is common in all scanpaths – that means we are trying to identify a route in terms of visual elements followed by our participants.

Algorithm 1 shows our proposed algorithm which takes a set of scanpaths and return a scanpath which is common in all the given scanpaths. If there is only one scanpath, it returns that one as the common scanpath, if there is more than one, then it tries to find the most similar two scanpaths in the given list. It does this by using the "Levenshtein Distance" which is a traditional string-edit algorithm [12]. Then it removes these two scanpaths from the given list of scanpaths and introduces their common scanpath to the list of scanpaths given originally. This continuos until there is only one scanpath. This process is illustrated in Figure 3. When we start the algorithm there are 10 scanpaths, then in the first iteration the most common scanpaths are S2 and S4, therefore we apply the Longest Common Subsequence (LCS) to these two scanpaths and generate a common one called S24. S2 and S4 are then removed from the list and S24 is introduced to the list of scanpaths.

For all ten scanpaths in Table 1, our algorithm returns "CDED"

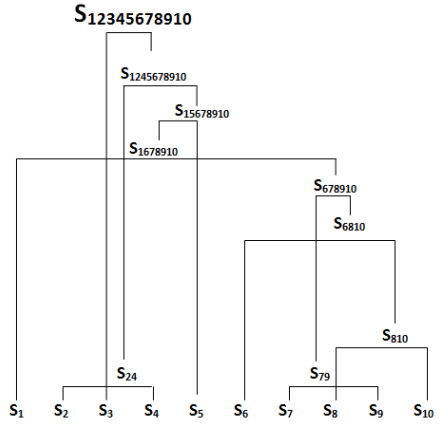

Figure 3: eMine Algorithm applied on scanpaths in Table 1.

as the common scanpath which is illustrated in Figure 1 by arrows. It shows that people start to look at the first item of the content, then they looked at the menu, the latest news part and then they looked at the menu.

Algorithm 1 Find Common Scanpath

Input: Scanpath List
Output: Scanpath

```
if the size of Scanpath List is equal to 1 then
    return the scanpath in Scanpath List
end if
while the size of Scanpath List is not equal to 1 do
    Find the two most similar scanpaths in Scanpath List with Levenshtein Distance
    Find the common scanpath by using Longest Common Subsequence
    Remove the similar scanpaths from the Scanpath List
    Add the common scanpath to the Scanpath List
end while
return the scanpath in Scanpath List
```

3.4 Initial Informative Validation

If we revisit the task that the participants were asked to complete in the original study (see Section 3.2), we see that the path generated by our algorithm is quite logical [5]. The path very nicely matches with what people asked to complete. In the literature there is also an approach that looks at the probability of AoIs following each other in a set of scanpaths, called a transition matrix 2 [15]. When we generate such a transition matrix, we also see that it supports this path. In a transition matrix each cell has three rows where the second row shows the row probabilities and the third row illustrates the column probabilities. Row probabilities allow identifying the next AoI of the particular AoI and Column probabilities allow identifying the previous AoI of the particular AoI. For example, according to the transition matrix, if the people look AoI 'D', it is most likely that they will look AoI 'E' (46.16 %) after AoI 'D' and they have already looked AoI 'C' (42.6 %) just before AoI 'D'. However, further studies with more participants and pages need to be conducted to validate the proposed approach.

3.5 Experiential Transcoding

Algorithm 3 has also been developed on the ACTF platform, given a web page and a set of eyetracking data, it generates a com-

	A	B	C	D	E	F
A	0 0.0% 0.0%	3 15.0% 15.0%	11 55.01% 23.92%	3 15.0% 5.56%	3 15.0% 6.53%	0 0.0% 0.0%
B	0 0.0% 0.0%	0 0.0% 0.0%	3 14.29% 6.53%	5 23.81% 9.26%	12 57.15% 26.09%	1 4.77% 25.0%
C	12 25.54% 63.16%	4 8.52% 20.0%	0 0.00% 0.00%	23 48.94% 42.6%	7 14.9% 15.22%	1 2.13% 25.0%
D	5 9.62% 26.32%	0 0.0% 0.0%	21 40.39% 45.66%	0 0.0% 0.0%	24 46.16% 42.18%	2 3.85% 50.0%
E	0 0.0% 0.0%	13 28.89% 65.0%	11 24.45% 23.92%	21 46.67% 38.89%	0 0.0% 0.0%	0 0.0% 0.0%
F	2 50.0% 10.53%	0 0.0% 0.0%	0 0.0% 0.0%	2 50.0% 3.71%	0 0.0% 0.0%	0 0.0% 0.0%

Table 2: Transition Matrix

mon scanpath in terms of visual elements of a web page. This common scanpath can be used to present the content on that page in alternative forms. For instance, one can only show the visual elements in the common scanpath or guide the reader through that scanpath which would behave like a macro. These techniques show how such a common scanpath can be used to transcode web pages. This is a very specific example of "experiential transcoding".

4. DISCUSSION

Our work on identifying a common scanpath in terms of visual elements of web pages is still under development, however we have presented it here as an example to "experiential transcoding". Further work is needed to investigate the following:

What are the possible experiential transcoding techniques? Our work presented here showed a concrete example of how eyetracking data can potentially be used to drive experiential transcoding, but further studies need to be conducted to investigate the techniques for experiential transcoding.

Other than eyetracking data, what could be the possible driving input for experiential transcoding? For experiential transcoding, we need to be able to understand and predict users' needs and requirements, in our work, we explain how we propose to use eyetracking data for that but further research needs to be conducted to investigate the kind of inputs that can be used to guide experiential transcoding, for instance, can we use log data on servers? Can we use users' interaction pattern for prediction and transcoding?

How can we generalise experiential transcoding? By nature experiential transcoding is going to be specific to specific user's needs and requirements but further research need to be conducted to investigate the ways and techniques to generalise experiential transcoding across users.

5. SUMMARY

This paper introduced the concept of experience-based transcoding which we call 'experiential transcoding', and presented our preliminary work on identifying patterns in eye-tracking data to guide transcoding of web pages for improving users' experience in constraint environments.

6. ACKNOWLEDGEMENTS

The project is supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) with the grant number 109E251. As such the authors would like to thank to (TÜBİTAK) for their continued support.

7. REFERENCES

[1] Elgin Akpinar and Yeliz Yesilada. Vision based page segmentation: Extended and improved algorithm. Technical report, Middle East Technical University Northern Cyprus Campus, 2012.

[2] Chieko Asakawa and Hironobu Takagi. Transcoding. In Simon Harper and Yeliz Yesilada, editors, *Web Accessibility: A Foundation for Research*, Human-Computer Interaction Series, chapter 14, pages 231–260. Springer, 2008.

[3] Andy Brown, Caroline Jay, and Simon Harper. Audio presentation of auto-suggest lists. In *W4A '09*, pages 58–61. ACM, 2009.

[4] Andy Brown, Caroline Jay, and Simon Harper. Audio access to calendars. In *W4A '10*, pages 1–10. ACM, April 2010.

[5] Andy Brown, Caroline Jay, and Simon Harper. Tailored presentation of dynamic web content for audio browsers. *International Journal of Human-Computer Studies*, 70(3):179 – 196, 2012.

[6] Joseph H. Goldberg and Jonathan I. Helfman. Scanpath clustering and aggregation. In *ETRA '10*, ETRA '10, pages 227–234. ACM, 2010.

[7] Simon Harper. Deep accessibility: Adapting interfaces to suit our senses. *Invited Talk - Technical Superior Insitute, LaSIGE, Lisbon, Portugal*, 2012.

[8] Simon Harper and Yeliz Yesilada, editors. *Web Accessibility: A Foundation for Research*. Springer, 2008.

[9] Sheree Josephson and Michael E. Holmes. Age differences in visual search for information on web pages. In *ETRA '04*, pages 62–62. ACM, 2004.

[10] J. D. McCarthy, M. A. Sasse, and J. Riegelsberger. Could I Have the Menu Please? An Eyetracking Study of Design Conventions. In *HCI2003*, pages 401–414, UK, 2003.

[11] Eleni Michailidou, Simon Harper, and Sean Bechhofer. Investigating sighted users' browsing behaviour to assist web accessibility. In *Assets '08*, pages 121–128. ACM, 2008.

[12] K. Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, pages 372–422, 1998.

[13] Mark Russell. Using eye-tracking data to understand first impressions of a website. *Usability News 7.1*, 7(1), 2005.

[14] H. Takeuchi and Y. Habuchi. A quantitative method for analyzing scan path data obtained by eye tracker. In *Computational Intelligence and Data Mining*, 2007.

[15] Julia M. West, Anne R. Haake, Evelyn P. Rozanski, and Keith S. Karn. eyepatterns: software for identifying patterns and similarities across fixation sequences. In *ETRA '06*, ETRA '06, pages 149–154. ACM, 2006.

[16] Yeliz Yesilada, Alan Chuter, and Shawn Lawton Henry. Shared Web Experiences: Barriers Common to Mobile Device Users and People with Disabilities. W3C, 2008. `http://www.w3.org/WAI/mobile/experiences`.

[17] Yeliz Yesilada, Caroline Jay, Robert Stevens, and Simon Harper. Validating the use and role of visual elements of web pages in navigation with an eye-tracking study. In *WWW '08*, Beijing, China, 2008.

[18] Yeliz Yesilada, Robert Stevens, Simon Harper, and Carole Goble. Evaluating DANTE: Semantic transcoding for visually disabled users. *ACM Trans. Comput.-Hum. Interact.*, 14(3):14, 2007.