LAYOUT-DRIVEN WEBPAGE READING

Koki Tsumura¹, Junko Shirogane², Hajime Iwata³ and Yoshiaki Fukazwa¹
¹Waseda University, Japan
²Tokyo Woman’s Christian University, Japan
³Kanagawa Institute of Technology, Japan

ABSTRACT
Most visually impaired people use a screen reader when using a computer. However, numerous problems are encountered when using screen reader. One of the problems is the reading order of a webpage. Current screen readers basically read aloud the webpage from top to bottom, left to right. Webpages have structures, and information is emphasized by color and size. Normally, when sighted people read a webpage, they decide the reading order by referring to nontext information such as color and structure. In this research, we propose a method to determine the reading order on the basis of the saliency of each element of the webpage and the structure information of the webpage, such as menus and layout patterns.

KEYWORDS
Screen Reader, Accessibility, Menu, Saliency, Webpage Layout

1. INTRODUCTION
According to the World Health Organization (Blindness and vision impairment, World Health Organization), 1 billion people are visually impaired worldwide. With the spread of Internet access, the use of the web by the visually impaired is increasing. According to a survey (Watanabe 2013), 96.3% of blind people use computers, mostly for web browsing. Many visually impaired people use a screen reader to navigate on computers, smartphones, and tablets. A screen reader basically reads aloud the text from top to bottom in the HTML file of a webpage visited by a user, which may cause unwanted parts of the page to be read aloud. According to a survey (Lazar 2007), screen-reader users waste an average of 30.4% of their time because of the reading order, images without alt text, etc.

In recent years, the use of various layout patterns in web design that use nontext information such as images, colors, and positions has increased. Layout patterns exist so that users can acquire necessary information as quickly as possible without becoming bored with the content of the webpage. Many designers use layout patterns to guide the eye movement of visitors of a webpage. Users who visit a webpage will skip certain parts of the layout pattern and read certain parts carefully. As a result, their eye movement and reading order will change accordingly. However, current screen readers do not read aloud the webpage in the same order that sighted people read.

To fill the usability gap between how sighted and blind users acquire information when visiting a webpage, we propose a method to determine the reading order of webpages by focusing on menus, saliency, and webpage layout patterns. Specifically, we first classify webpages into two major parts—menus and content—and create saliency rankings of child elements in each part. The reading order of webpages is then determined on the basis of the rankings and the detected layout pattern. We believe that this method will enable more efficient web browsing by visually impaired users, improve users’ understanding of each part of a webpage, and provide a reading order similar to that of sighted users.
2. RELATED WORKS

Extensive research has been conducted to fill the computer usability gap between sighted and visually impaired people. Some research has focused on improving screen-reader usability, and some has focused on improving voice user interface usability. In this chapter, some of these approaches are introduced and the difference between these approaches and the method proposed in the present work is discussed.

Authors of long web documents such as wiktionaries, manuals, tutorials, blogs, and scientific literature provide a table of contents (TOC) for users to access the desired content instantaneously. However, screen-reader users cannot exploit the TOC to the same extent as their sighted peers. Screen-reader users need to press a multitude of shortcuts to access the TOC from the beginning of the page as well as from an arbitrary section in the page, which they find tedious and cumbersome. To solve this problem, a method to identify and extract TOC hyperlinks from the web documents and then facilitate on-demand instant screen-reader access to the TOC from anywhere in the webpage has been proposed in previous works. The method has been proposed as a browser extension called iTOC (Lee 2020). Whenever a user presses a special shortcut, the TOC in the current webpage is identified from the document object model and can be accessed from a screen reader on-demand. However, this approach is limited to long documents that include a TOC inside and does not focus on changing the reading order of webpages.

ARIA (Accessible Rich Internet Applications) landmarks serve as reference points that can be conveniently and quickly accessed using keyboard shortcuts. If ARIA landmarks are properly implemented in web pages, screen reader users can jump to content they are interested in. However, in many cases, web developers use landmarks sporadically and inconsistently, and some webpages do not even have landmarks. To solve this problem, Aydin et al. proposed the SaIL system (Aydin 2020), which dynamically injects role landmark attributes into the HTML source using a deep learning model called SalGAN. However, ARIA landmarks are not commonly used, and their research is not providing method to determine reading order of webpages.

3. LAYOUT PATTERN

Numerous layout patterns, including f-patterns, spotted patterns, layer-cake patterns, and zigzag patterns (Text Scanning Patterns: Eyetracking Evidence, Nielsen Norman Group), are intended to guide the eye movement of a user reading a webpage. For example, a layer-cake pattern is a pattern used by a user visiting a webpage to read each heading and subheading; when a heading relevant to the user’s interest appears, the user reads the accompanying body text below the heading. The left half of Fig 1 shows an example of a layer-cake pattern. A gaze plot is shown in circles, and the gazes are mainly plotted in the headings. Designers usually use layout patterns to prevent users from becoming bored and to convey information efficiently by locating elements in certain positions.

Zigzag patterns are commonly used patterns in modern webpages. A zigzag pattern is a layout pattern that alternates the placement of images and text on each horizontal row (Zigzag Image-Text Layouts Make Scanning Less Efficient). The right half of Fig 1 shows an example of a zigzag pattern. The image with a house is located on the left-hand side of the gray rectangular text block, and the image with leaves is located on the right-hand side of text block. A zigzag pattern is used to create a rhythm to maintain the reader’s interest in the content while also directing attention so that the user can quickly look through each element and grasp its importance. Zigzag patterns are often used in promotion pages and landing pages to communicate the features of the product and the benefits the product offers. Therefore, the images in a zigzag pattern are included for the purpose of adaptation to attract attention. However, in current screen readers, when a zigzag pattern is read aloud, the readings of these images are not skipped. To reduce redundant reading and to follow the scanning pattern of a sighted user when scanning a zigzag pattern, our proposed method skips the reading of the images.
4. DESIGN

Figure 2 presents an overview of a screen-reader browser extension that implements the proposed method. First, when a user visits a webpage, a URL is given to the saliency prediction system that uses Inagaki’s prediction model (Inagaki 2020) and a saliency ranking is generated by the system. The saliency ranking and DOM of the webpage are then passed to the read-order determinator. In the determinator, the menu and zigzag pattern are detected first; after these processes, the reading order of elements in the menu and content parts are determined according to the saliency ranking. In the end, the determinator determines the reading order of all the elements in the webpage and, whenever a user presses a predefined shortcut, it determines the next element to read and reads it aloud.

4.1 Menu Detection

For the sake of space, only the nav tag menu detection algorithm is explained here. In the proposed method, other algorithms are also used for menu detection. For parallel navigation, navigation elements are usually in the top one-fourth of the screen height; thus, in the first step, the nav tag elements that satisfy this condition are collected (Fig 3). After this step, each collected element in the same height is grouped to be marked as a single navigation element. In some cases, multiple nav tag elements exist at the same height; for them to be recognized as a menu, nav tag elements at the same height are grouped. For each group, the closest ancestor element is then collected and returned as a menu element. If no menu is detected by the algorithm, other algorithms are used to detect the menu.
4.2 Zigzag Pattern Detection

First, images larger than or equal to 0.2 times the screen width and less than or equal to 0.8 times the screen width are collected. In the next step, the collected images with the same or similar height text neighbor are collected as candidates (Fig 4). Collected images are then grouped by size and distance between each image and returned as zigzag pattern elements.

4.3 Menu Reading Order Determination

For each element in a detected menu, the reading order is determined according to a saliency ranking generated by a saliency generation system. Elements’ reading order is determined in order of saliency, and, before and after the reading aloud menu, “menu begin” and “menu end” are read aloud. The number of elements inside menu is also read after “reading of menu begin” in the same manner that existing screen readers read list elements.

4.4 Contents Reading Order Determination

For each element in the detected content, the reading order is determined according to saliency ranking generated by a saliency generation system. The reading order of elements is determined in order of saliency; however, for zigzag pattern images, reading orders are not set to skip reading alt text attributed to the image elements. After the reading-order determination steps, whenever the user presses a shortcut key, an element to be read is determined and read aloud by the system.
5. EVALUATION

5.1 Experimental Design

To evaluate our proposed method, we conducted an experiment. Twelve sighted people who usually use a computer in their daily life and work participated. Because of the COVID-19 situation, the experiment was conducted online using Zoom (video communication software). We grouped the 12 participants into 4 groups of 3 participants each and asked them to listen to a reading by the proposed method and to a reading by NonVisual Desktop Access (NV Access), which is a screen reader widely used by visually impaired people; the participants were then asked to answer some questions. In each session, they answered questions using hand signs, which we prepared, or answered via text. To ensure that each participant could not see other participants’ answers, the participants were asked to pin the host screen or to close their eyes when answering questions using hand signs.

To determine which reading method was more effective, we asked participants to listen to the menu reading, the zigzag pattern reading, and the reading determined using saliency separately. Otherwise, it would not be possible to determine which reading was most effective. In the experiments, webpages that include a zigzag pattern and a menu were used and we grouped participants into 4 groups (A to D), with 3 participants in each group. Table 1 shows the order in which each group listened to the readings. The participants in group A listened to proposed method’s reading first and the NVDA’s reading after for the menu reading experiment, whereas the participants in group D listened to the NVDA’s reading first and the proposed method’s reading after for the same experiment. The orders of the other experiments are also presented in Table 1. The participants participated in the menu reading experiment and subsequently participated in the saliency-sorted reading experiment and then the zigzag-sorted reading experiment. Questions presented in Table 2 were asked during or after the experiments. After each experiment, we asked participants to send comments.

To verify the effectiveness of proposed method’s menu reading, we asked participants to listen to the proposed method’s reading and the NVDA’s reading. To ensure the experiment design did not favor NVDA or the proposed method, 2 groups listened to the proposed method’s reading first and the other 2 groups listened to the NVDA’s reading first. After or while listening to each reading, participants were asked to answer questions 1–4.

To verify the effectiveness of the proposed method’s saliency-sorted reading order, we asked the participants to listen to the proposed method’s reading and the unchanged original reading. To ensure the experiment design did not favor the original or the proposed method, 2 groups listened to the proposed method first and the other 2 groups listened to the original first. After listening to each reading, participants were asked to answer questions 5 and 6.

To verify the effectiveness of the proposed method’s zigzag pattern reading, we asked the participants to listen to the proposed method’s reading and the NVDA’s reading. To ensure that the experimental design did not favor NVDA or the proposed method, 2 groups listened to the proposed method’s reading first and the other 2 groups listened to the NVDA’s reading first. After listening to each reading, participants were asked to answer questions 7 through 9.

Table 1. The order in which each group listened to the readings

| Group Name | Menu       | Saliency Sort | Zigzag Pattern |
|------------|------------|---------------|----------------|
| A          | Proposed, NVDA | Sorted, Original | Proposed, NVDA  |
| B          | Proposed, NVDA | Sorted, Original | NVDA, Proposed  |
| C          | NVDA, Proposed | Original, Sorted | Proposed, NVDA  |
| D          | NVDA, Proposed | Original, Sorted | NVDA, Proposed  |
In most cases, people

The reading of NVDA’s reading, which is insufficient for users to understand the menu. Our proposed method’s reading outperformed the NVDA’s reading in understanding a menu; the correct rate of guessing the menu from the proposed method’s reading was 100%, whereas that from the NVDA’s reading was only 8.33%. In addition, the participants preferred the proposed method’s reading.

For question 5, 8 participants felt that necessary information was read earlier in the proposed method. For question 6, 6 participants preferred to use the proposed method. From these results, the proposed method’s reading was not effective. There were few positive comments and numerous negative comments for the proposed method’s reading. The following comments are some excerpted participant comments that seem valuable for our future research:

- The reading of a menu should be skipped first and read whenever the user wants the screen reader to read aloud.
- In most cases, people don’t visit a webpage to register a service, so a screen reader should read aloud a menu in the original order.

| Question Number | Question                                                                 |
|-----------------|--------------------------------------------------------------------------|
| 1               | Raise your hand while the menu is being read                              |
| 2               | Could you understand where the menu started and where it ended? Choose from 5 choices. (1) Completely ununderstandable, (2) Ununderstandable, (3) Neither understandable nor ununderstandable, (4) Understandable, (5) Very understandable. |
| 3               | Which reading was easier to understand where the menu started and where it ended? |
| 4               | If you were blind, which reading would you prefer to use?                |
| 5               | Which read necessary information earlier?                                |
| 6               | If you were a visually impaired person, which reading would you prefer to use? |
| 7               | How many duplicate readings were there? (1) A very large amount, (2) A large amount, (3) Neither large nor few, (4) A Few, (5) None |
| 8               | Which reading had more redundant reading?                                |
| 9               | If you were blind, which reading would you prefer to use?                |

5.2 Results

Table 3 presents the correct answer rate to question 1. The participants were asked to guess the menu while a webpage was being read, and the correct answer rate was calculated; the answer rate is reported as a percentage calculated as the number of people who answered correctly divided by 12 (the total number of participants). Although every participant correctly answered where the reading of the menu started and where it finished for the proposed method’s reading, 1 participant did not correctly guess the start and finish of the menu in the NVDA’s reading. Table 4 shows the average answer score for question 2. Eleven participants scored 4 or 5 for the proposed method’s reading, corresponding to “understandable” or “very understandable”, respectively, whereas 6 participants scored 1 or 2, which correspond to “completely ununderstandable” and “ununderstandable”, respectively, for the NVDA’s reading.

For the proposed method’s reading, participants submitted both positive and negative comments. Two participants wrote that it was understandable to hear the “menu begin” keyword before the start of the reading menu segment and that it was also understandable to hear the “menu end” keyword after the end of the reading menu segment. However, 1 participant mentioned that there should be more time before reading “the number of links in the menu”, and some participants mentioned that a more detailed menu explanation should be given because people who do not usually use the Internet might not understand what a menu is.

For the NVDA’s reading, participants submitted numerous negative comments. Five participants mentioned that they had no idea when the reading of the menu started and ended. Also, some participants mentioned that they could predict the menu when “link” was read multiple times and when “navigation” was read. However, as the results in Table 3 show, most of the participants failed to guess the answer; thus, even if they made predictions, the predictions were incorrect in most cases.

The results for questions 3 and 4 are shown in Table 5. For both questions, all of the participants preferred the proposed method’s reading.

From the results, the NVDA’s reading of a menu is insufficient for users to understand the menu. Our proposed method’s reading outperformed the NVDA’s reading in understanding a menu; the correct rate of guessing the menu from the proposed method’s reading was 100%, whereas that from the NVDA’s reading was only 8.33%. In addition, the participants preferred the proposed method’s reading.

For question 5, 8 participants felt that necessary information was read earlier in the proposed method. For question 6, 6 participants preferred to use the proposed method. From these results, the proposed method’s reading was not effective. There were few positive comments and numerous negative comments for the proposed method’s reading. The following comments are some excerpted participant comments that seem valuable for our future research:

- The reading of a menu should be skipped first and read whenever the user wants the screen reader to read aloud.
- In most cases, people don’t visit a webpage to register a service, so a screen reader should read aloud a menu in the original order.

| Question Number | Question                                                                 |
|-----------------|--------------------------------------------------------------------------|
| 1               | Raise your hand while the menu is being read                              |
| 2               | Could you understand where the menu started and where it ended? Choose from 5 choices. (1) Completely ununderstandable, (2) Ununderstandable, (3) Neither understandable nor ununderstandable, (4) Understandable, (5) Very understandable. |
| 3               | Which reading was easier to understand where the menu started and where it ended? |
| 4               | If you were blind, which reading would you prefer to use?                |
| 5               | Which read necessary information earlier?                                |
| 6               | If you were a visually impaired person, which reading would you prefer to use? |
| 7               | How many duplicate readings were there? (1) A very large amount, (2) A large amount, (3) Neither large nor few, (4) A Few, (5) None |
| 8               | Which reading had more redundant reading?                                |
| 9               | If you were blind, which reading would you prefer to use?                |
- When I visit a webpage, I first read the contents and rarely use a menu, so a screen-reader user might want to read like this.

For question 7, as presented in Table 7, 8 participants chose 4 or 5 as their choice for the proposed method’s reading, whereas 9 participants scored 1 or 2 for the NVDA’s reading. No negative comments were submitted for the proposed method’s reading; however, several participants submitted negative comments for the NVDA’s reading. Examples include multiple readings of the “image” keyword, which hindered understanding, many duplicate readings, and inability to imagine a webpage from the reading.

The results for questions 8 and 9 are presented in Table 8. For question 8, all participants felt that there were more redundant readings in the NVDA’s reading. For question 9, 11 participants preferred the proposed method’s reading.

In summary, compared with the NVDA’s menu readings, the proposed method’s menu readings outperformed with respect to understanding; in addition, all of the participants preferred to use the proposed method rather than the NVDA. Eleven participants preferred to use the proposed method’s zigzag pattern readings, and all of the participants felt that the proposed method’s reading was less redundant. However, saliency-sorted reading did not perform well.

Table 3. Correct answer rate for question 1

| Question                  | Proposed Correct Rate | NVDA Correct Rate |
|---------------------------|-----------------------|-------------------|
| Raise your hand while menu is being read | 100% (12/12) | 8.33% (1/12) |

Table 4. Average answer score for question 2

| Question                                      | Proposed Average Answer Score | NVDA Average Answer Score |
|-----------------------------------------------|--------------------------------|----------------------------|
| Could you understand where the menu started and where it ended? Choose from 5 choices. | 3.83 (11 participants answered 4 or 5) | 2.42 (6 participants answered 1 or 2) |

Table 5. Answers for questions 3 and 4

| Question                                      | Proposed | NVDA |
|-----------------------------------------------|----------|------|
| Which reading was easier to understand where the menu started and where it ended? | 100%     | 0%   |
| If you were blind, which reading would you prefer to use? | 100% | 0% |

Table 6. Answers for questions 5 and 6

| Question                                      | Proposed | Original |
|-----------------------------------------------|----------|----------|
| Which read necessary information earlier?     | 66.7%    | 33.3%    |
| If you were blind, which reading would you prefer to use? | 50.0% | 50.0% |

Table 7. Average answer score for question 7

| Question                                      | Proposed Average Answer Score | NVDA Average Answer Score |
|-----------------------------------------------|--------------------------------|----------------------------|
| How many duplicate readings were there?       | 3.08 (8 participants answered 4 or 5) | 1.83 (9 participants answered 1 or 2) |

Table 8. Answers for questions 8 and 9

| Question                                      | Proposed | NVDA |
|-----------------------------------------------|----------|------|
| Which reading had more redundant reading?     | 0%       | 100% |
| If you were blind, which reading would you prefer to use? | 91.7% | 8.83% (1/12) |
6. CONCLUSION

This paper introduced a new method to determine a reading order of a webpage using saliency, layout pattern, and structure information to convey the intention of the webpage’s designer and convey information efficiently. A user study with 12 sighted participants demonstrated the potential of the method in substantially improving usability. We are planning to conduct an experiment with visually impaired people who are screen-reader users. The current method is implemented as a Chrome extension, and the voice implemented in the current method differs from that in NVDA, which is the comparison target; in addition, there are some shortcut keys for participants to remember, which might also influence the experimental results. To solve these problems, we will implement a new system that determines reading order of a given webpage and generates a sorted webpage according to the reading order. Each participant will be asked to perform some tasks using a screen reader on two webpages: an unchanged plain webpage and a webpage produced using our method. The number of shortcut presses will be counted, and the completion time will be measured for each task.

REFERENCES

Ali Selman Aydin, Shrin Feiz, Vikas Ashok, and I V Ramakrishnan. Sail: Saliency-driven injection of aria landmarks. International Conference on Intelligent User Interfaces, pp. 111–115, 2020.
Blindness and vision impairment, World Health Organization, [online] Available at https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment [Accessed 18 Dec. 2021].
Inagaki et al., Visualization Method of Important Regions by Combination of Webpage Structures and Saliency Maps. ICSOFT 2020. pp 204-211.
Lazar et al., 2007. What frustrates screen reader users on the web: A study of 100 blind users. INTERNATIONAL JOURNAL OF HUMAN-COMPUTER INTERACTION, Vol. 22, No. 3, pp 247-269.
Lee et al., 2020. iTOC: Enabling Efficient Non-Visual Interaction with Long Web Documents. Proceedings of IEEE International Conference on Systems, Man and Cybernetics. Toronto, ON, Canada, pp. 3799-3806.
NV Access, [online] Available at https://www.nvaccess.org/ [Accessed 18 Dec. 2021].
Screen reader user survey #9 results, WebAIM, [online] Available at https://webaim.org/projects/screenreadersurvey9/#landmarks [Accessed 18 Dec. 2021].
Text Scanning Patterns: Eyetracking Evidence, Nielsen Norman Group, [online] Available at https://www.nngroup.com/articles/text-scanning-patterns-eyetracking/ [Accessed 18 Dec. 2021].
Watanabe et al., 2014. A Survey on the Use of Cell Phones, Smartphones, and Tablets among People with Visual Impairments 2013. IEICE technical report. Welfare Information technology, Vol. 113, No. 481, pp 25-30.
Zigzag Image-Text Layouts Make Scanning Less Efficient, Nielsen Norman Group, [online] Available at https://www.nngroup.com/articles/zigzag-page-layout/ [Accessed 18 Dec. 2021].