"The Best of Both Worlds!": Integration of Web Page and Eye Tracking Data Driven Approaches for Automatic AOI Detection

DOI: 10.1145/3372497

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

Citation for published version (APA):
https://doi.org/10.1145/3372497

Published in:
ACM Transactions on the Web

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: 19. Aug. 2021
“The Best of Both Worlds!”
Integration of Web Page and Eye Tracking Data Driven Approaches for Automatic AOI Detection

SUUKRU ERASLAN, Middle East Technical University Northern Cyprus Campus
YELIZ YESILADA, Middle East Technical University Northern Cyprus Campus
SIMON HARPER, University of Manchester

Web pages are comprised of different kinds of elements (menus, adverts, etc). Segmenting pages into their elements has long been important in understanding how people experience those pages, and in making those experiences 'better'. Many approaches have been proposed which relate the resultant elements with the underlying source code, however, they do not consider users’ interactions. Another group of approaches analyses eye movements of users to discover areas that interest or attract them (i.e. areas of interest or AOIs). Although these approaches consider how users interact with web pages, they do not relate AOIs with the underlying source code. We propose a novel approach which integrates web page and eye tracking data driven approaches for automatic AOI detection. This approach segments an entire web page into its AOIs by considering users’ interactions and relates AOIs with the underlying source code. Based on the Adjusted Rand Index measure, our approach provides the most similar segmentation to the ground truth segmentation compared to its individual components.

CCS Concepts: • Human-centered computing → Human computer interaction (HCI); User studies; User interface programming; • Information systems → World Wide Web; Web interfaces.

Additional Key Words and Phrases: web page segmentation, visual element, visual block, segment, region of interest, ROI

ACM Reference Format:
Sukru Eraslan, Yeliz Yesilada, and Simon Harper. 0. "The Best of Both Worlds!": Integration of Web Page and Eye Tracking Data Driven Approaches for Automatic AOI Detection. ACM Trans. Web 0, 0, Article 0 (0), 30 pages. https://doi.org/10.1145/nmnnnn.nmnnnn

OPEN DATA
The implementation of our proposed approach with C programming language can be found in our external repository http://iam-data.cs.manchester.ac.uk/data_files/35. All the materials and data used for the evaluation of our approach can also be found in this repository.

1 INTRODUCTION
There are different kinds of visual elements on web pages. Some of core visual elements are the main contents, menus, headers and footers. These elements allow users to interact with web pages [53]. For example, when users access a particular website, they can use its main menu to navigate

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.

© 0 Association for Computing Machinery.
1559-1131/0/0-ART0 $15.00
https://doi.org/10.1145/nmnnnn.nmnnnn

ACM Trans. Web, Vol. 0, No. 0, Article 0. Publication date: 0.
within the website. These visual elements can be considered as areas of interest (AOIs) of web pages. Identifying these elements has long been important in understanding how people experience web pages, especially which elements people commonly use and in which order they use these elements [13], [14]. It has also been important in processing web pages for different purposes, such as re-organising or re-engineering web pages to improve user experience in constrained environments. For example, when a web page is accessed with a small-screen device, the page can be re-organised by dividing it into several sub-pages where each of these sub-pages is comprised with closely related content and appropriate for displaying on the screen of the device [1]. Moreover, some elements of web pages may be visually fragmented but their visual fragmentation may not be encoded in the source code. Hence, visually disabled people who use screen readers may not be able to access the desired information as screen readers follow the source code of web pages. Having the knowledge of visually fragmented elements would allow re-organising web pages to support accessibility for visually disabled users [3].

There has been extensive research to identify visual elements of web pages [51]. Many approaches have been proposed for segmenting web pages into their visual elements and relating these elements with the source code of the web pages, ranging from the use of heuristics to the use of machine learning models [51] (see Section 2). These approaches have widely been used in different fields. For example, they have been used to enhance not only small screen device interaction and web accessibility, but also web information retrieval, web page caching, web archiving, web page visual quality and aesthetics, etc. [51]. To the best of our knowledge, none of these approaches consider how users interact with web pages, and therefore their results may not represent the areas that are actually used by users.

Some eye tracking data driven approaches have also been used to automatically discover AOIs on web pages by analysing eye movements of users [12]. Once these AOIs have been discovered, they have been used for further analysis, such as identifying sequential patterns in terms of these AOIs [12]. Therefore, in contrast to web page driven approaches, these approaches consider how users interact with web pages (see Section 2). When users interact with web pages, their eyes become relatively immobile at certain points which are called fixations. Their eyes also make quick movements between fixations which are referred to as saccades. The series of fixations and saccades represent their scanpaths. Fig. 1 illustrates an example of a scanpath from a particular user on the home page of the Apple website. The circles show the fixations while the larger circles show the longer fixations. Even though eye tracking data driven approaches analyse eye movements of users to consider how users interact with web pages, they do not entirely segment web pages and relate AOIs with the source code of web pages. However, when we access AOIs in the source code of web pages, we can re-organise or re-engineer web pages. For instance, if we identify areas that attract users and remove other areas from web pages, then screen reader users can directly access their targets without spending unnecessary time on clutter [52].

In this paper, we propose a novel approach which integrates web page driven and eye tracking data driven approaches for automatic AOI detection (see Section 3). Therefore, our approach takes the best of both worlds. Specifically, our approach discovers areas attracting people and relates these areas with the underlying source code which is not supported by either only web page driven approaches or only eye tracking data driven approaches. This approach segments web pages into their AOIs by taking users’ interactions into consideration and relates the AOIs with the source code of the web pages. As a web page driven approach, we currently use the Vision-based Page Segmentation (VIPS) algorithm because it is the most popular algorithm in the web page segmentation field [2, 56]. As an eye tracking driven approach, we currently use the Hierarchical DBSCAN (HDBSCAN) algorithm because it deals with noisy fixations and partitions fixations into a number of stable clusters without specifying the number of clusters in advance [8]. As our
proposed approach has an open architecture, other suitable web page and eye tracking data driven approaches can also be integrated.

To investigate the validity of integrating web page and eye tracking data driven approaches, our proposed approach was compared with its individual components (see Section 4). We used an eye tracking dataset collected on a number of web pages [16]. We applied the VIPS algorithm on its own, the HDBSCAN algorithm on its own and our proposed approach to segment the pages into their AOIs. We also prepared the ground truth segmentation\(^1\) that was constructed by a group of users in two different ways. We then compared the similarity between the ground truth segmentation and the resulting segmentation of each of the VIPS algorithm, the HDBSCAN algorithm and our proposed approach. To compute the similarity between two segmentations, we used the Adjusted Rand Index (ARI) measure [4]. Based on the ARI values, our approach provides the most similar segmentation to the ground truth segmentation in comparison with its components (see Section 5.1). We also compared our approach with one of the recent state-of-art web page driven approaches called Block-o-Matic! [41] and our approach achieved better ARI values (see Section 5.1). We also investigated the effects of the number of participants in an eye tracking dataset on the performance of our approach and our approach performed well with different numbers of participants (see Section 5.2). We are planning to investigate other web page driven approaches and eye tracking data driven approaches for our integrated approach in the future (see Sections 6 and 7).

\(^1\)It is also known as gold standard segmentation [34]
2 RELATED WORK
In this section, we summarise and discuss the main features of two different kinds of approaches that have been used to segment web pages into their AOIs. Saliency maps which are typically computed with low-level image characteristics [24] can also be used to identify certain areas that may take visual attention without using eye tracking data. However, these maps may not always be consistent with the real visual attention [47].

2.1 Web Page Driven Approaches
A wide range of manual and automated approaches have already been proposed for segmenting web pages by using their structures, including their source code, CSS (Cascading Style Sheets), JavaScript, images, etc. These approaches use different features such as heuristics, DOM (Document Object Model), text densities, machine learning models, and so on. A detailed review of these approaches can be found in [51]. To summarise these approaches, here we give some examples: Ahmadi and Kong [1] propose a set of heuristics for identifying the main content of a particular web page and also the atomic and composite blocks in the main content. Wu et al. [49] propose another approach to differentiate content blocks from link blocks on a web page by using its DOM structure. Chen et al. [55] also propose a DOM-based approach which is comprised of three main steps: the detection of the high-level content blocks, the identification of explicit separators in these content blocks to divide them into smaller blocks, and the identification of implicit separators in these smaller blocks for further division. Kohlschütter and Nejdl [27] propose a different approach to segment web pages by focusing on text density. Furthermore, Bing et al. [4] propose a machine learning model for segmenting web pages based on some features, which are categorised as DOM structure features (e.g., structure similarity of the neighbouring segments), visual property features (e.g., background colour) and text content features (e.g., text similarity of the neighbouring segments).

Even though various web page driven approaches are available, they also have considerable limitations. In particular, some of these approaches have very strong assumptions which cannot be generalised because web pages are designed in different ways and most of them are not properly formatted. Similarly, some proposed heuristics are limited to simple models, and therefore they cannot be generalised. As web technologies evolve very rapidly, some of the web page driven approaches can also be easily outdated. In addition, some of these approaches aim to detect specific parts of web pages, such as only the main content [26] or only the important parts of web pages based on Google’s PageRank algorithm [54]. These approaches can be useful for specific purposes but their application areas are limited. There are some other approaches to segment web pages based on their DOM structures that contain lots of information about the structure of web pages, but there are also lots of information encoded in the visual representation of web pages. Some web page segmentation approaches are also available which discover the elements of web pages based on their visual representation either substantially (e.g., [21]) or solely (e.g., [40]). However, the source code of web pages should also be taken into consideration jointly for a better understanding of their structures. The correlation between the resultant elements and the source code would allow further processing of web pages in a more efficient way as it provides direct access to the elements within the source code. Boi et al. [5] propose an algorithm to segment web pages into their content blocks by combining their visual representation with their source code. Their algorithm first analyses a given web page as an image to detect its content blocks and then uses the geometric features of these blocks to match them with the content blocks in the DOM structure.

The Vision-based Page Segmentation (VIPS) algorithm is the most popular approach among those web page driven approaches [7]. This algorithm takes a web page and all of its components (CSS, JavaScript, etc.) as an input and then analyses both its source code and its visual representation to
provide a tree of elements where the deeper levels have more and smaller elements. For example, the main menu may not be divided into its menu items in the second level but it may be divided into its items in the third level. The level to cut the tree, which is referred to as a segmentation level or granularity level, can be selected by researchers and practitioners based on their objectives. In particular, they can select the first level if they are interested in the main elements of web pages. However, this selection may not be straightforward for researchers and practitioners who are not sure about which level they should use for their studies. Although a study conducted by Akpınar and Yesilada [2] suggests the fifth level as the most successful level with ≈74% user satisfaction, it has limitations. Specifically, only the first five levels were evaluated and the fifth level was found as the most successful level. However, we do not know about further segmentation levels. In addition, varying levels within the tree might be more preferred by users instead of a strict level. For example, the main menu might not need to be further divided but the main content might need to be further divided.

There are also other approaches which use the combination of the source code and visual representation of a given web page with a different set of rules to generate a tree of elements. Block-o-Matic! is one of these approaches [41]. It follows a logical approach to cluster elements based on the Gestalt laws – Proximity, Similarity, Closure and Simplicity. Similar to the VIPS algorithm, Block-o-Matic! also provides a tree of elements and has a granularity parameter. There is also no specific value suggested for this parameter. Block-o-Matic! is not the first approach which considers the Gestalt laws for web page segmentation. An approach proposed by Yang et al. [50] also considers the Gestalt laws to segment web pages into their elements by using both of the source code and visual representation of web pages.

Web page driven approaches consider the data available in either the source code, the visual representation or both, but there is a lack of integration of web page data and user interaction data for the entire web page segmentation process. Hence, we aim to develop an algorithm for this kind of integration to entirely segment web pages into their elements based on the data available in the source code and visual representation of the pages and the data collected from users’ interactions with the pages. Furthermore, we also aim to relate the resultant elements with the underlying source code to support further processing of the web pages.

### 2.2 Eye Tracking Data Driven Approaches

Most of the approaches used for discovering AOIs based on eye tracking data are conventional clustering algorithms which can cluster fixations and each cluster represent a different AOI. Different clustering algorithms have already been used for discovering AOIs on web pages or other visual stimuli, such as K-Means [35], Affinity Propagation [22, 29], Mean-Shift [12, 42] and DBSCAN [9, 31]. However, none of these approaches relates AOIs with the source code of web pages. Besides, these approaches also have the following considerable problems.

Eye tracking dataset usually includes a large number of fixations but some of these fixations might be made involuntarily [11]. These fixations are referred to as noisy fixations and they potentially affect the process of clustering. Due to these noisy fixations, some fixations may not be located in the right clusters. In addition, these fixations may unnecessarily increase the coverage of some clusters. Some of the clustering algorithms, such as K-means [30] and Affinity Propagation [20], partition a given dataset into a number of clusters without considering noisy fixations.

The clustering algorithms typically have a number of different parameters, and these parameters affect the process of clustering. For example, the Mean-Shift algorithm has a parameter called bandwidth [10]. When the value of this parameter is increased, larger clusters are generated. Similarly, the DBSCAN algorithm has two main parameters: min_samples and epsilon (a distance parameter) [18]. When the min_samples parameter is set to a high value or the epsilon parameter...
is set to a low value, a higher density will be needed to create a cluster [43]. Unfortunately, the adjustment of these parameters is not straightforward.

It is difficult to estimate the number of clusters before applying the clustering algorithm because we typically do not know which parts of visual stimuli will attract the attention of users. However, some clustering algorithms need the number of clusters specified in advance, such as K-Means [30] and some hierarchical clustering algorithms [44]. If our estimation is not correct, the clustering algorithm may provide unnecessarily large or extremely small clusters.

A clustering algorithm should provide the same clusters for the same dataset with the same parameters. Specifically, it should not be similar to the K-Means algorithm which starts with random initialisation and possibly provides slightly different clusters for the same dataset with the same parameter. Because of this reason, the K-Means algorithm does not provide stable clusters. Therefore, if eye tracking data analysis is conducted based on certain clusters and these clusters are not stable, then the analysis may not reproducible, thus not reliable.

Fixations may not be consistently distributed over web pages. Hence, some clusters may have higher density in comparison to other clusters. If a clustering algorithm is sensitive to different densities, it may not work well with an eye tracking dataset that contains a number of clusters with different densities. To take this issue into consideration, Campello et al. [8] propose a different clustering algorithm, called Hierarchical DBSCAN (HDBSCAN), which is also based on density and provides a tree of clusters. The tree is then pruned at different levels to select the most stable varying density clusters [32]. Although the HDBSCAN clustering algorithm has not been applied to cluster fixations to discover AOIs on visual stimuli yet, it is a potential candidate because it deals with noisy fixations, provides the same results with the same dataset and the same parameters, does not need the number of clusters specified in advance and has a few reasonable parameters (min_cluster_size and min_samples).

2.3 Lessons Learnt

Web page driven and eye tracking driven approaches have been researched independently, however as can be seen from this review they can be considered as two lines of research with similar goals. Buscher et al. [6] aimed to develop a model for predicting visual attention that individual elements of web pages may receive. This model was trained by using some features of DOM elements with certain eye tracking features generated for each DOM element, which are mean fixation impact, viewing frequency and median time to the first fixation. In our study, we aim to segment an entire web page into their areas instead of determining which DOM element may receive visual attention. If we only focus on specific areas that may receive visual attention, then our work will be limited to the identification of particular areas. Some elements may contain important information but may not receive visual attention because of some distractive elements. We aim to provide a full and more generic web page segmentation. Therefore, our proposed approach firstly generates a tree of segments, and then prunes the tree with eye tracking data.

3 PROPOSED APPROACH

We propose an approach which integrates both a web page driven approach and an eye tracking data driven approach for segmenting web pages. By integrating these two kinds of approaches, we will be able to combine the advantages of both worlds because eye tracking data is rich in terms of users' interactions and web page data is also rich in terms of the structure and visual representation of web pages.

Our proposed approach is visualised in Fig. 2. It has two main inputs: a Web Page and a set of Fixations on that page. The Web Page is used as an input of a web page driven approach whereas the set of Fixations is used as an input of an eye tracking driven approach. The outputs of the web page
driven and eye tracking data driven approaches are used by an integration algorithm to provide the *Areas of Interest* on the *Web Page* as an output. Therefore, there are three main distinct components here: (1) Web Page Driven Approach, (2) Eye Tracking Data Driven Approach and (3) Integration Algorithm. These three components are explained below. Since our proposed approach is an open approach, other web page driven approaches which provide a tree of elements and eye tracking data driven approaches which label fixations with their cluster IDs can also be used within our approach.

### 3.1 Web Page Driven Approach

The web page driven approach provides a *Tree of Elements* as a result of its segmentation process which includes the width, height, top-left x and top-left y coordinates of each element. As a web page driven approach, we currently use the extended Vision Based Page Segmentation (VIPS) algorithm [2] as it provides a tree of elements and very popular in the literature [2] (see Section 2.1).

Let Fig. 3 to represent a web page with all its visual elements generated with the VIPS algorithm. As illustrated in Fig. 4, the element A is the root and it has three children which are the elements B, C and D. The element B has no child whereas the elements C and D have two children. The elements E and F are children of the element D whereas the elements G and H are children of the element C. The entire Tree of Elements is used by our integration algorithm.

![Fig. 3. An example of a web page segmented with the VIPS algorithm & an example of fixation clusters created with the HDBSCAN algorithm](image)

**Fig. 4. The tree of the elements in Fig. 3**

### 3.2 Eye Tracking Data Driven Approach

The eye tracking data driven approach provides a set of *Labelled Fixations* where the label of each fixation show its cluster id and the noisy fixations are ignored. Listing 1 shows an example set of *Labelled Fixations* including their x and y coordinates, and also cluster ids. As an eye tracking data driven approach, we currently use the HDBSCAN algorithm [8] because of its promising features (see Section 2.2). Assume that Fig. 3 also illustrates the four fixation clusters identified with the HDBSCAN algorithm.

<table>
<thead>
<tr>
<th>x-position</th>
<th>y-position</th>
<th>cluster-id</th>
</tr>
</thead>
<tbody>
<tr>
<td>105</td>
<td>130</td>
<td>0</td>
</tr>
<tr>
<td>547</td>
<td>320</td>
<td>2</td>
</tr>
<tr>
<td>203</td>
<td>429</td>
<td>1</td>
</tr>
<tr>
<td>429</td>
<td>680</td>
<td>3</td>
</tr>
<tr>
<td>. . .</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---


Listing 1. An example of *Labelled Fixations* (Format: x-position y-position cluster-id)
3.3 Integration Algorithm

The Labelled Fixations (see an example in Listing 1) are then used by the integration algorithm to identify the levels or points to prune the Tree of Elements (see an example in Fig. 4). The integration algorithm follows these rules:

- If an element contains more than one cluster, then it will be further divided. Therefore, the final elements do not contain more than one cluster, if it is not a leaf node as a leaf node cannot be further divided.

- If an element does not contain more than one cluster, then it will not be further divided.

When an element contains a certain amount of fixations of a particular cluster, it is considered that the element contains that cluster. Thus, if an element contains only a few fixations from a particular cluster and the number of these fixations is below a threshold value, then the fixations will be ignored. Let an element to contain all the fixations of the first cluster and only 4% of the fixations of the second cluster. If the threshold is set to 5%, then the fixations from the second cluster will be ignored and the further division will not be needed in that case. However, if the threshold is set to 3%, then the further division will be performed as the element will contain the fixations from more than one cluster. This threshold value allows our approach to prevent unnecessary divisions caused by borders of clusters. The final elements are referred to as the Areas of Interest or AOIs of the Web Page. The pseudocode of the integration algorithm is shown in Algorithm 1.

Algorithm 1: Integration Algorithm

```plaintext
input : Tree of Elements, Labelled Fixations, Threshold, AOIs
output: AOIs
/* AOIs is an empty list when the integration algorithm is first called. */
begin
R ← The root node of Tree of Elements
if R == NULL then
    return AOIs
else
    Use Labelled Fixations to count fixations of each cluster in R
    if R contains only one cluster based on Threshold or R is a leaf node then
        Add R to AOIs
        return AOIs
    else
        Run Integration Algorithm for each child of R
        return AOIs
    end
end
end
```

Let the Tree of Elements and Labelled Fixations shown in Fig. 3 are used by this integration algorithm. The algorithm starts from the root of the tree as shown in Fig. 5. Since the root includes all the fixations from the four clusters, the root should be further divided as shown in Fig. 6. The elements B and C do not contain fixations from more than one cluster, and therefore these elements will not be further divided. However, the element D includes all the fixations of two clusters, thus it should be further divided as shown in Fig. 7. After this step, there is no element which contains fixations from more than one cluster. Therefore, our approach finishes the adjustment of the segmentation levels and provides the elements B, C, E and F as the AOIs of the web page.
Fig. 5. The segmentation level adjustment - Step I

Fig. 6. The segmentation level adjustment - Step II

Fig. 7. The segmentation level adjustment - Step III
As a real-world example, Figures 8, 9 and 10 show a resulting segmentation made by the HDBSCAN algorithm, the VIPS algorithm and our approach (Threshold=0.05) respectively. Since our approach does not further divide an element when it contains fixations from only one or no fixation cluster, it is likely to provide a minimum number of desired AOIs based on users’ interactions.

Fig. 8. Fixation clusters created with the HDBSCAN algorithm on the home page of the Apple website

Fig. 9. The home page of the Apple website segmented with the VIPS algorithm
Our proposed approach was evaluated by using a dataset from an eye tracking study on six web pages [16]. In this evaluation, we compared the results of our approach with the results of its individual components (the VIPS and HDBSCAN algorithms) based on their similarities to the ground truth segmentation which were constructed with two different ways by a group of users. This comparison method has been commonly used in the web page segmentation field to assess the validity of the proposed approaches [4, 19, 28].

We also wanted to compare our approach with the state-of-art approaches. In our previous work, we followed a systematic way to find web page segmentation approaches with publicly available implementation, but we could find only a limited number of those approaches [17]. Not all web page segmentation approaches with publicly available implementation are suitable for comparison. For example, we tried to use the space-based decomposition approach proposed by Reinecke et al. which segments web pages based on their visual appearance and provides a tree of elements[40]. The decomposition approach was used to generate some features for predicting users’ initial impressions of website aesthetics. Therefore, the main focus of Reinecke et al. was not segmenting web pages. Furthermore, unfortunately there is no default level or suggested level to prune the tree of elements. As the number of leaves was used as a feature in their prediction model, we tried to use the leaves as the elements of the web pages but the leaves were very tiny items, such as one letter in a word. We believed that it would not be fair to select a specific level and then compare the approach with our approach. Therefore, we selected Block-o-Matic! as it was one of the recent state-of-art approaches in this field and its implementation was publicly available as a Google Chrome extension with a default parameter.

We also investigated how the number of participants in an eye tracking dataset affect the performance of our approach. Below we explain our evaluation in detail.

---

4https://github.com/rmardiko/vizweb
5http://www-poleia.lip6.fr/ sanojaa/BOM/
4.1 Eye Tracking Study

The eye tracking study used for the evaluation of our approach was originally conducted for evaluating a scanpath analysis algorithm which was designed to analyse eye movements of multiple users on a particular web page and identify their trending path in terms of the AOIs of the page [16]. Thus, the detailed description of the eye tracking study can be found in [16]. Here, we present a summary.

- **Participants:** The eye tracking study was carried out at the University of Manchester and Middle East Technical University Northern Cyprus Campus with 40 participants (20 female and 20 male) in 2013. These participants were mainly students and some of them were administrative and academic staff at these universities. Their ages ranged from 18 to 54.

- **Equipment and Materials:** Tobii T60 eye tracker was used to record eye movements of the participants [46]. It was integrated into a 17-inch monitor and the screen resolution of the monitor was set to 1280 x 1024. Fixations were detected with the Tobii Fixation Filter which was the implementation of Olsson [37]’s classification algorithm. The home pages of the following six websites which had various visual complexities (low, medium or high) were used: Apple (low), Babylon (low), AVG (medium), Yahoo (medium), Godaddy (high) and BBC (high). These web pages can be found in our external repository. The visual complexity of these web pages was assessed with the ViCRAM tool which computes a visual complexity score (VCS) for a given web page between 0 and 10 based on the number of occurrence of certain elements (top left corners, words and images) in its HTML Document Object Model (VCS < 3: low complexity, 3 ≤ VCS ≤ 7: medium complexity, 7 < VCS: high complexity) [33].

- **Procedure:** The participants were asked to complete searching and viewing tasks on the web pages in random order. There was no specific objective for the viewing tasks. During the viewing tasks, the participants spontaneously browsed the pages without clicking any links. For the searching tasks, the participants were asked to find specific information or items on the pages. For example, they were asked to find the link for watching the TV ads related to iPad mini on the home of the Apple website (see Fig. 1). Thirty seconds were given for the viewing tasks as used in other studies [38], and a maximum of two minutes were given for the searching tasks. For the viewing tasks, all the participants viewed all the pages for 30 seconds. However, for the searching tasks, when the participants completed their tasks on a particular page, they continued with other pages without waiting. After each session of the eye tracking study, the participants were also asked to draw what they remember about the structure of the web pages. Each participant was asked to draw one page from each complexity level (three pages in total). As there were 40 participants and they were asked to draw three different web pages out of six distinct pages, 20 drawings were expected for each page (i.e. \((40 \times 3)/6 = 20\)). Some of the participants could not provide any drawing for some pages, but each page had at least 17 drawings. These drawings were used to construct the ground truth segmentation for each page as explained in the following subsection.

4.2 Ground Truth Segmentation

Web page segmentation researchers typically provide a set of web pages and ask somebody to segment them in order to generate a ground truth segmentation for each page [4, 19, 28]. We created two types of ground truth segmentation for the same pages to evaluate our approach. For the creation of the first ground truth segmentation, we used the drawings of the participants from the eye tracking study explained above which represent what they remembered about the structure of the web pages, in other words how they segmented the pages in their minds. For the creation of
the second ground truth segmentation, we asked five people who were experienced in web design & development to segment the web pages based on their opinions.

**The First Ground Truth Segmentation:** Fig. 11 shows the drawing of one participant for the home page of the Apple website based on what s/he remembers about the structure of the page after his/her eye tracking session. This method is known as recall. According to Johnson [25], visual cues support users to orient themselves. The recall method allowed us to recognise the areas which really took the attention of the participants in the study. Therefore, with this method, we could identify the AOIs that supported the participants to recognise where they were.

The drawings of the participants were analysed by three researchers (including one external researcher) who had worked in the field of web page segmentation to create the first ground truth segmentation of the pages. Each researcher identified the AOIs from each drawing for each page. When the researchers could not match the drawn areas with any element of the pages, they discarded them. All of the AOIs identified by the researchers were then combined to construct the ground truth segmentation. Specifically, we considered all the AOIs drawn by the participants because if a specific area is taken into consideration by a particular user, the area will possibly be taken into consideration by other users. In case of nested AOIs, smaller AOIs were taken into consideration as they were more specific and more detailed. As an example, the ground truth segmentation of the home page of the Apple website is illustrated in Fig.12.

**The Second Ground Truth Segmentation:** Some well-known studies in the literature use three to seven external assessors or volunteers to assess the validity of their web page segmentation approaches or create ground truth segmentation. For example, Cai et al. [7] asked five assessors to evaluate the results of their web page segmentation approach, Kreuzer et al. [28] and Kohlschütter et al. [26] used three and seven external volunteers respectively to create their ground truth segmentation. Sanoja and Gançarski [41] and Bing et al. [4] also recruited external volunteers for the creation of their ground truth segmentation, but they did not clearly state how many volunteers...
they used. A further detailed discussion about the creation of ground truth segmentation can be found in [28].

In order to evaluate our approach with two types of ground truth segmentation, we conducted a separate user study for the creation of the second ground truth segmentation for the same pages. We recruited five people who had experience in web design & development and did not participate in the eye tracking study whose dataset was used in the evaluation of our approach. Thus, we ensured that we do not have subjective volunteers and we have a sufficient number of volunteers to create new ground truth segmentation [36]. We asked our participants to draw the AOIs of the pages based on their opinions, and then combined all the AOIs drawn by the participants to create the ground truth segmentation for each page. Similar to the creation of the first ground truth segmentation, smaller AOIs were considered in case of nested AOIs. The details of this user study are given below.

- **Participants:** One female and four males participated in the study. Three of them were aged between 25-34, one of them was aged between 18-24, and another one was aged between 35-54. All of the participants were daily web users and spent more than four hours per day on the web, apart from one of them who claimed that s/he spent 3-4 hours per day on the web. They all had at least a bachelor’s degree in computer science, information systems, and/or web design & development. Three of them worked in web design & development for more than four years whereas others worked for 1-2 years. The mean of the number of projects that they worked was 9.8 with the standard deviation 6.42, and three of the participants were actively working in web design & development at the time that we conducted the study.

- **Materials:** The screen-shots of a number of web pages including the home pages of the Apple, Babylon, AVG, Yahoo, Godaddy and BBC websites were given to the participants. These screen-shots are available at our external repository (see Open Data section).
participants were then asked to draw the AOIs of the web pages based on their opinions. Figure 13 shows how the home page of the Apple website was segmented into its AOIs by one of the participants.

![Apple website segmentation](image)

---

**Procedure:** The participants firstly read the information sheet which explained the objective of the study and their rights, and then they signed the consent form. After that, they filled in a short questionnaire about their basic demographic information (e.g., educational background) and their work experience in web design & development (e.g., the number of projects they worked related to web design & development). The information sheet, consent form, questionnaire of the study are also available at our external repository (see Open Data Section). Once the participants were ready, they were asked to draw the AOIs directly as a new layer on the top of a screen-shot of each web page in random order.

**4.3 Methodology**

In web page segmentation, two AOIs can be located under the same block by one algorithm, however these two AOIs can be located under different blocks by another algorithm. Moreover, one algorithm may tend to construct larger AOIs whereas another algorithm may tend to create smaller AOIs. In particular, one algorithm may provide AOIs which cover other AOIs created by another algorithm. Hence, web page segmentation can be considered as a clustering of atomic units which cannot be further divided, thus the cluster of atomic units can be considered as an AOI [56].

In this paper, we refer to all visible leaf nodes and images on web pages as atomic units. Other remaining units without any tags under the parents of the leaf nodes are also referred to as atomic units. For example, if the line in Listing 2 is available on a web page, “Aa”, “Bb” and “Cc” are all
considered as atomic units. All horizontal and vertical separators (e.g., |) are ignored here as they are only used to visually separate items. In addition, icons on images are assumed to be parts of the images. As an example, Fig. 14 shows the atomic units of the home page of the Apple website.

Listing 2. An example of atomic units

```html
<div>
  Aa <a href="bb.html">Bb</a> <a href="cc.html">Cc</a>
</div>
```

Fig. 14. The atomic units of the home page of the Apple website

By using the atomic units of the web pages, we could compare how our approach and other approaches clustered the atomic units on the web pages. We expect that the clusters provided by our approach are more similar to the clusters in the ground truth segmentation.

To compute the similarity between the clusters of atomic units (AOIs) generated by the approaches and the clusters of atomic units (AOIs) in each of the ground truth segmentation, we computed the Adjusted Rand Index (ARI) in a pair-wise manner for each page. ARI is a similarity measure of two clusterings and it is increasingly used to assess the success of proposed segmentation algorithms [4, 56]. In this context, Rand Index is a similarity measure between two different segmentations which is computed by considering all pairs of atomic units and counting the pairs which are located under the same or different AOIs in the resulting segmentation and the ground truth segmentation [4]. The value of Rand Index can be between 0 and 1 where 0 means that the segmentations are completely different from each other and 1 means that the segmentations are exactly the same [4]. Adjusted Rand Index is a corrected-for-chance version of the Rand Index and it is equal to 0 on average for random segmentation and 1 when two segmentations are exactly the same [4]. To compute ARI for two different segmentations of a particular web page called A and B, a contingency table is first created to count common atomic units between each pair of segments in A and B as
illustrated in Table 1, and then the contingency table is used for ARI computation with Equation 1 [23].

Table 1. Contingency table for ARI computation for two different segmentations called A and B where A and B include x and y segments respectively and \( n_{ij} \) represents the number of common atomic units in \( A_i \) and \( B_j \)

<table>
<thead>
<tr>
<th></th>
<th>( B_1 )</th>
<th>( B_2 )</th>
<th>( B_3 )</th>
<th>( \cdots )</th>
<th>( B_y )</th>
<th>( \text{Total} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_1 )</td>
<td>( n_{11} )</td>
<td>( n_{12} )</td>
<td>( n_{13} )</td>
<td>( \cdots )</td>
<td>( n_{1y} )</td>
<td>( a_1 )</td>
</tr>
<tr>
<td>( A_2 )</td>
<td>( n_{21} )</td>
<td>( n_{22} )</td>
<td>( n_{23} )</td>
<td>( \cdots )</td>
<td>( n_{2y} )</td>
<td>( a_2 )</td>
</tr>
<tr>
<td>( A_3 )</td>
<td>( n_{31} )</td>
<td>( n_{32} )</td>
<td>( n_{33} )</td>
<td>( \cdots )</td>
<td>( n_{3y} )</td>
<td>( a_3 )</td>
</tr>
<tr>
<td>( \vdots )</td>
<td>( \vdots )</td>
<td>( \vdots )</td>
<td>( \vdots )</td>
<td>( \cdots )</td>
<td>( \vdots )</td>
<td>( \vdots )</td>
</tr>
<tr>
<td>( A_x )</td>
<td>( n_{x1} )</td>
<td>( n_{x2} )</td>
<td>( n_{x3} )</td>
<td>( \cdots )</td>
<td>( n_{xy} )</td>
<td>( a_x )</td>
</tr>
<tr>
<td>( \text{Total} )</td>
<td>( b_1 )</td>
<td>( b_2 )</td>
<td>( b_3 )</td>
<td>( \cdots )</td>
<td>( b_y )</td>
<td></td>
</tr>
</tbody>
</table>

\[
ARI(A, B) = \frac{\sum_{ij} \binom{n_{ij}}{2} - \sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}}{\frac{1}{2} \sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2} - \sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}}
\] (1)

We expect that the segmentation provided by our proposed approach is more similar to each of the ground truth segmentation compared to its individual components (the VIPS and HDBSCAN algorithms) and Block-o-Matic! which one of the recent state-of-art web page segmentation approaches, and therefore we expect to have higher ARI values with our approach.

To compare the results of our approach with the results of its individual components and Block-o-Matic!, we firstly applied them to all the pages for the viewing and searching tasks. Due to the problems in the source code of the saved version of the home page of the Godaddy website, we could not identify its atomic units properly, and therefore we excluded the page from the evaluation.

To compute ARI, we used the ARI function provided by Python scikit-learn library\(^6\) which accepts two lists of labels for the same data points (i.e. atomic units in our case) in the same order. Therefore, we generated a list of labels for each page for each approach and each ground truth segmentation.

An overview of our evaluation methodology is illustrated in Figure 15.

When the VIPS algorithm was applied, we took the elements from the first level to the maximum available level (i.e. the level of the deepest leaf element in the segmentation trees of all the pages). The maximum available level was 12 from the BBC page and therefore we considered each level from one to 12 separately for all the pages. When we take the elements from a specific level, we take the elements with the specified level and other leaf elements from lower levels. After that, each atomic unit was associated with one of the AOIs on each page at each level. The same label was given to the atomic units which were located under the same AOI whereas different labels were given to the atomic units which were located under different AOIs. Therefore, we had a list of labels for each page at each level with the VIPS algorithm.

We also discovered the fixation clusters on each page for the viewing and searching tasks with the HDBSCAN algorithm by setting its \texttt{min_cluster_size} parameter to the number of participants. By default, its \texttt{min_samples} parameter was also set to the same value. The reason of setting the \texttt{min_cluster_size} to the total number of participants is to find the clusters which include at least one fixation from all the participants, and the clusters which contain at least the same number of fixations as the clusters which include at least one fixation from all the participants. It means that we considered all the participants in this evaluation to be objective as much as possible because outlier/noise detection could be considered as a subjective process, even though there are some outliers.

\(^6\)\text{http://scikit-learn.org/stable/modules/generated/sklearn.metrics.adjusted_rand_score.html}
techniques which help for outlier/noise detection. Since we could not use the eye tracking data of two participants because of the problems in their recordings, the min_cluster_size parameter was set to 38, and therefore min_samples parameter was also set to 38 by default.

When the clusters were generated by the HDBSCAN algorithm, we labelled the atomic units by considering the following rules: (1) if an atomic unit is not included by any cluster, it has its own label; (2) if an atomic unit includes fixations from more than one cluster, it is related to the cluster with the highest number of fixations in the atomic unit; (3) if two atomic units are associated with the same cluster, they are labelled with the same label; (4) if two atomic units are associated with different clusters, they are labelled with different labels. Thus, we also had a list of labels for each page for the viewing and searching tasks with the HDBSCAN algorithm.

In our integrated approach, we used the trees of elements generated by the VIPS algorithm and the fixation clusters generated by the HDBSCAN algorithm. In order to not bias the evaluation results, the threshold value of our approach was set to 0.05 because a five percent chance of not being true has widely been used in statistics [48]. Hence, if an element contains at least 5% of the fixations of a particular cluster, it means that the element contains that cluster. Once the AOIs were generated, we again associated each atomic unit with one of the AOIs on each web page. Therefore, we also had a list of labels for each page for the viewing and searching tasks with our approach.

To segment the web pages into their AOIs with Block-o-Matic!, we used its default granularity parameter 0.30. Each atomic unit was then associated with the AOIs where the same label was given to the atomic units under the same AOI. Hence, we also had a list of labels for each page.

To sum up, Table 2 lists all the parameters used in the evaluation of our proposed integrated approach along with their values. We used these values to be objective as much as possible. For example, we used the default parameter of Block-o-Matic! which was set by its developers, not us. We considered all the possible granularity levels of the VIPS algorithm in our case instead of choosing a specific granularity level. Table 2 also shows a short summary of the significance of each approach for the evaluation.

We also associated each atomic unit with the one of the AOIs in each of the ground truth segmentation to have a list of labels for each page. We then applied the ARI function to the list.
of labels generated by each approach with the list of labels created by each of the ground truth segmentation in a pair-wise manner for each page.

In addition to the comparison between our approach with its individual components and one of the recent state-of-art approaches, we investigated whether or not our approach performs well with a lower number of participants in an eye tracking dataset. We re-run our approach with randomly chosen 25%, 50% and 75% of the participants from the eye tracking dataset for the viewing and searching tasks and computed the similarity between the resulting segmentation with both of the ground truth segmentation based on ARI. We then compared the ARI values achieved by our approach with 25% (10 participants), 50% (19 participants), 75% (29 participants) and 100% (38 participants) of the participants.

5 RESULTS
In this section, we present the results of our comparison for both the viewing and searching tasks. We also present our analysis of the effects of the number of participants on the performance of our approach.

5.1 Comparison
Since both the VIPS algorithm and Block-o-Matic! do not use eye tracking data for segmenting web pages, it provided the same results for both types of tasks. Table 3 and Table 4 show the mean, median, standard deviation, minimum and maximum of the ARI values for each approach on all the pages for the viewing and searching tasks with the first and second ground truth segmentation respectively. Furthermore, Table 5 and Table 6 show the ARI values achieved by the VIPS algorithm, the HDBSCAN algorithm, our approach and Block-o-Matic! for each page for the viewing and searching tasks with the first and second ground truth segmentation respectively. The best values of the viewing tasks are shown in bold whereas the best values of the searching tasks are shown in italic. If a particular value is considered as the best value for both the viewing and searching tasks, it is shown in bold italic.

Viewing Tasks: With both of the ground truth segmentation, our approach achieved the highest mean (0.78, 0.85), median (0.80, 0.88), minimum (0.53, 0.68), and maximum (0.94, 0.98) ARI values. It also achieved the lowest standard deviation with the second ground truth segmentation (0.11) and the closest standard deviation (0.16) to the lowest standard deviation achieved by Block-o-Matic! with the first ground truth segmentation (0.14). Therefore, overall, our approach provided the most similar segmentation to the ground truth segmentation in comparison with its individual components and Block-o-Matic! for the viewing tasks.

When we look at the ARI values achieved by the VIPS algorithm, the HDBSCAN algorithm, our approach and Block-o-Matic! for each page for the viewing tasks with the first ground truth
Table 3. The mean, median, standard deviation, minimum and maximum of the ARI values for each segmentation approach for the viewing and searching tasks

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Parameter</th>
<th>Mean</th>
<th>Median</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extended VIPS</td>
<td>Level=1</td>
<td>0.21</td>
<td>0.13</td>
<td>0.18</td>
<td>0.04</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>Level=2</td>
<td>0.28</td>
<td>0.22</td>
<td>0.19</td>
<td>0.10</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>Level=3</td>
<td>0.54</td>
<td>0.57</td>
<td>0.26</td>
<td>0.11</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>Level=4</td>
<td>0.60</td>
<td>0.61</td>
<td>0.19</td>
<td>0.34</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>Level=5</td>
<td>0.53</td>
<td>0.46</td>
<td>0.24</td>
<td>0.23</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>Level=6</td>
<td>0.47</td>
<td>0.46</td>
<td>0.36</td>
<td>0.06</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>Level=7</td>
<td>0.33</td>
<td>0.17</td>
<td>0.34</td>
<td>0.04</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>Level=8</td>
<td>0.30</td>
<td>0.17</td>
<td>0.30</td>
<td>0.04</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>Level=9</td>
<td>0.27</td>
<td>0.17</td>
<td>0.24</td>
<td>0.04</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>Level=10</td>
<td>0.18</td>
<td>0.14</td>
<td>0.17</td>
<td>0.04</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>Level=11</td>
<td>0.15</td>
<td>0.08</td>
<td>0.18</td>
<td>0.02</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>Level=12</td>
<td>0.15</td>
<td>0.08</td>
<td>0.18</td>
<td>0.02</td>
<td>0.46</td>
</tr>
<tr>
<td>HDBSCAN - Viewing</td>
<td>Min Cluster Size=38</td>
<td>0.22</td>
<td>0.20</td>
<td>0.20</td>
<td>-0.01</td>
<td>0.49</td>
</tr>
<tr>
<td>HDBSCAN - Searching</td>
<td>Min Cluster Size=38</td>
<td>0.15</td>
<td>0.11</td>
<td>0.17</td>
<td>-0.05</td>
<td>0.37</td>
</tr>
<tr>
<td>Our Integrated Approach - Viewing</td>
<td>Threshold=0.05</td>
<td>0.78</td>
<td>0.80</td>
<td>0.36</td>
<td>0.53</td>
<td>0.94</td>
</tr>
<tr>
<td>Our Integrated Approach - Searching</td>
<td>Threshold=0.05</td>
<td>0.61</td>
<td>0.62</td>
<td>0.29</td>
<td>0.18</td>
<td>0.96</td>
</tr>
<tr>
<td>Block-o-Matic!</td>
<td>Granularity=0.3</td>
<td>0.42</td>
<td>0.42</td>
<td>0.14</td>
<td>0.26</td>
<td>0.63</td>
</tr>
</tbody>
</table>

Table 4. The mean, median, standard deviation, minimum and maximum of the ARI values for each segmentation approach for the viewing and searching tasks with the second ground truth segmentation

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Parameter</th>
<th>Mean</th>
<th>Median</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extended VIPS</td>
<td>Level=1</td>
<td>0.22</td>
<td>0.13</td>
<td>0.18</td>
<td>0.04</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>Level=2</td>
<td>0.31</td>
<td>0.22</td>
<td>0.25</td>
<td>0.09</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>Level=3</td>
<td>0.58</td>
<td>0.60</td>
<td>0.30</td>
<td>0.12</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>Level=4</td>
<td>0.68</td>
<td>0.74</td>
<td>0.24</td>
<td>0.35</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>Level=5</td>
<td>0.64</td>
<td>0.67</td>
<td>0.22</td>
<td>0.38</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>Level=6</td>
<td>0.60</td>
<td>0.52</td>
<td>0.28</td>
<td>0.24</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>Level=7</td>
<td>0.45</td>
<td>0.49</td>
<td>0.29</td>
<td>0.16</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>Level=8</td>
<td>0.42</td>
<td>0.49</td>
<td>0.26</td>
<td>0.12</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>Level=9</td>
<td>0.38</td>
<td>0.49</td>
<td>0.22</td>
<td>0.09</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>Level=10</td>
<td>0.29</td>
<td>0.21</td>
<td>0.20</td>
<td>0.09</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>Level=11</td>
<td>0.27</td>
<td>0.21</td>
<td>0.23</td>
<td>0.03</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>Level=12</td>
<td>0.27</td>
<td>0.21</td>
<td>0.23</td>
<td>0.03</td>
<td>0.52</td>
</tr>
<tr>
<td>HDBSCAN - Viewing</td>
<td>Min Cluster Size=38</td>
<td>0.21</td>
<td>0.17</td>
<td>0.11</td>
<td>0.10</td>
<td>0.39</td>
</tr>
<tr>
<td>HDBSCAN - Searching</td>
<td>Min Cluster Size=38</td>
<td>0.26</td>
<td>0.25</td>
<td>0.12</td>
<td>0.11</td>
<td>0.38</td>
</tr>
<tr>
<td>Our Integrated Approach - Viewing</td>
<td>Threshold=0.05</td>
<td>0.85</td>
<td>0.88</td>
<td>0.11</td>
<td>0.68</td>
<td>0.98</td>
</tr>
<tr>
<td>Our Integrated Approach - Searching</td>
<td>Threshold=0.05</td>
<td>0.64</td>
<td>0.60</td>
<td>0.31</td>
<td>0.18</td>
<td>0.94</td>
</tr>
<tr>
<td>Block-o-Matic!</td>
<td>Granularity=0.3</td>
<td>0.46</td>
<td>0.45</td>
<td>0.14</td>
<td>0.26</td>
<td>0.59</td>
</tr>
</tbody>
</table>

segmentation (see Table 5) and the second ground truth segmentation (see Table 6), we can see that the ARI values of the HDBSCAN algorithm and Block-o-Matic! were always considerably lower than the ARI values of our approach. However, on a few of the pages, the ARI values of the VIPS algorithm at varying segmentation levels were slightly higher than the ARI values of our approach. For example, with the first ground truth segmentation, the ARI value of the VIPS algorithm was slightly higher than the ARI value of our approach on the home page of the BBC website when its segmentation level was set to six (VIPS: 0.92, Our Integrated Approach: 0.90). However, none of the levels of the VIPS algorithm achieved better results compared to our approach.
Table 5. ARI values for each web page with each segmentation approach for the viewing and searching tasks with the first ground truth segmentation

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Parameter</th>
<th>Apple</th>
<th>Babylon</th>
<th>AVG</th>
<th>Yahoo</th>
<th>BBC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Level=1</td>
<td>0.25</td>
<td>0.51</td>
<td>0.12</td>
<td>0.04</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Level=2</td>
<td>0.52</td>
<td>0.45</td>
<td>0.22</td>
<td>0.10</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Level=3</td>
<td>0.75</td>
<td>0.57</td>
<td>0.75</td>
<td>0.54</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>Level=4</td>
<td>0.80</td>
<td>0.50</td>
<td>0.61</td>
<td>0.74</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>Level=5</td>
<td>0.45</td>
<td>0.46</td>
<td>0.23</td>
<td>0.86</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>Level=6</td>
<td>0.17</td>
<td>0.46</td>
<td>0.06</td>
<td>0.73</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>Level=7</td>
<td>0.17</td>
<td>0.46</td>
<td>0.04</td>
<td>0.12</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>Level=8</td>
<td>0.17</td>
<td>0.46</td>
<td>0.04</td>
<td>0.09</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>Level=9</td>
<td>0.17</td>
<td>0.46</td>
<td>0.04</td>
<td>0.08</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>Level=10</td>
<td>0.17</td>
<td>0.46</td>
<td>0.04</td>
<td>0.08</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>Level=11</td>
<td>0.17</td>
<td>0.46</td>
<td>0.04</td>
<td>0.08</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Level=12</td>
<td>0.17</td>
<td>0.46</td>
<td>0.04</td>
<td>0.08</td>
<td>0.02</td>
</tr>
<tr>
<td>Extended VIPS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDBSCAN - Viewing</td>
<td>Min Cluster Size=38</td>
<td>-0.01</td>
<td>0.33</td>
<td>0.49</td>
<td>0.20</td>
<td>0.09</td>
</tr>
<tr>
<td>HDBSCAN - Searching</td>
<td>Min Cluster Size=38</td>
<td>-0.05</td>
<td>0.27</td>
<td>0.11</td>
<td>0.03</td>
<td>0.37</td>
</tr>
<tr>
<td>Our Integrated Approach - Viewing</td>
<td>Threshold=0.05</td>
<td>0.80</td>
<td>0.53</td>
<td>0.94</td>
<td>0.72</td>
<td>0.90</td>
</tr>
<tr>
<td>Our Integrated Approach - Searching</td>
<td>Threshold=0.05</td>
<td>0.74</td>
<td>0.57</td>
<td>0.18</td>
<td>0.62</td>
<td>0.96</td>
</tr>
<tr>
<td>Block-o-Matic!</td>
<td>Granularity=0.3</td>
<td>0.34</td>
<td>0.45</td>
<td>0.26</td>
<td>0.63</td>
<td>0.42</td>
</tr>
</tbody>
</table>

Table 6. ARI values for each web page with each segmentation approach for the viewing and searching tasks with the second ground truth segmentation

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Parameter</th>
<th>Apple</th>
<th>Babylon</th>
<th>AVG</th>
<th>Yahoo</th>
<th>BBC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Level=1</td>
<td>0.33</td>
<td>0.47</td>
<td>0.12</td>
<td>0.04</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>Level=2</td>
<td>0.67</td>
<td>0.45</td>
<td>0.22</td>
<td>0.09</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Level=3</td>
<td>0.92</td>
<td>0.60</td>
<td>0.75</td>
<td>0.51</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Level=4</td>
<td>0.98</td>
<td>0.53</td>
<td>0.74</td>
<td>0.80</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>Level=5</td>
<td>0.72</td>
<td>0.49</td>
<td>0.38</td>
<td>0.94</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>Level=6</td>
<td>0.52</td>
<td>0.49</td>
<td>0.24</td>
<td>0.80</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>Level=7</td>
<td>0.52</td>
<td>0.49</td>
<td>0.21</td>
<td>0.16</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>Level=8</td>
<td>0.52</td>
<td>0.49</td>
<td>0.21</td>
<td>0.12</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>Level=9</td>
<td>0.52</td>
<td>0.49</td>
<td>0.21</td>
<td>0.09</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>Level=10</td>
<td>0.52</td>
<td>0.49</td>
<td>0.21</td>
<td>0.09</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>Level=11</td>
<td>0.52</td>
<td>0.49</td>
<td>0.21</td>
<td>0.09</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Level=12</td>
<td>0.52</td>
<td>0.49</td>
<td>0.21</td>
<td>0.09</td>
<td>0.03</td>
</tr>
<tr>
<td>Extended VIPS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDBSCAN - Viewing</td>
<td>Min Cluster Size=38</td>
<td>0.14</td>
<td>0.17</td>
<td>0.39</td>
<td>0.24</td>
<td>0.10</td>
</tr>
<tr>
<td>HDBSCAN - Searching</td>
<td>Min Cluster Size=38</td>
<td>0.38</td>
<td>0.25</td>
<td>0.18</td>
<td>0.11</td>
<td>0.38</td>
</tr>
<tr>
<td>Our Integrated Approach - Viewing</td>
<td>Threshold=0.05</td>
<td>0.98</td>
<td>0.68</td>
<td>0.90</td>
<td>0.80</td>
<td>0.88</td>
</tr>
<tr>
<td>Our Integrated Approach - Searching</td>
<td>Threshold=0.05</td>
<td>0.91</td>
<td>0.60</td>
<td>0.18</td>
<td>0.59</td>
<td>0.94</td>
</tr>
<tr>
<td>Block-o-Matic!</td>
<td>Granularity=0.3</td>
<td>0.58</td>
<td>0.45</td>
<td>0.26</td>
<td>0.59</td>
<td>0.41</td>
</tr>
</tbody>
</table>

on the majority of the pages. In particular, the VIPS algorithm achieved better results compared to our approach at most two out of five pages at a specific level (see the sixth level with the first ground truth segmentation). No approach has been proposed to adjust the segmentation level of the VIPS algorithm to achieve the most accurate segmentation. When researchers use the VIPS algorithm, they typically use a particular segmentation level, such as the fifth level [16].

To sum up, the overall results suggest that our approach performs better than its individual components for the viewing tasks in terms of providing the most similar segmentation to the ground truth segmentation.
Searching Tasks: Similar to the results for the viewing tasks, with the first ground truth segmentation, our approach achieved the highest mean (0.61), median (0.62) and maximum (0.96) ARI values. However, it could not achieve the highest minimum ARI value and the lowest standard deviation as its ARI value for the AVG page dramatically decreased (see Table 5). Figure 16 shows how the fixation clusters were distributed on the AVG page where the participants were asked to locate the link to download a free trial of AVG Internet Security 2013 and then the link to download AVG Antivirus Free 2013. As they required to locate the items illustrated by solid lines, they mainly focused on the upper half of the page and no fixation cluster was detected on the lower half of the page. Therefore, the lower menu was not further divided into their items by our approach as illustrated by dashed lines. Since the lower menu was divided into their items in the ground truth segmentation as illustrated by dotted lines, the ARI value of our approach considerably decreased on the AVG page. However, the third level of the VIPS algorithm segmented the lower menu into its items as the ground truth segmentation, and therefore it achieved considerably higher ARI value compared to our approach on the AVG page for the searching tasks. This case also existed in the second ground truth segmentation. Even though our approach achieved slightly better results in the second ground truth segmentation compared to the first ground truth segmentation, it could not achieve the best results in the second ground truth segmentation. Specifically, the VIPS algorithm performed slightly better than our approach with its level 4 (see Table 4).

Fig. 16. Fixation clusters created with the HDBSCAN algorithm on the home page of the AVG website

According to the ARI values achieved by the VIPS algorithm, the HDBSCAN algorithm, our integrated approach and Block-o-Matic! on each page for the searching tasks with the first ground truth segmentation (see Table 5) and the second ground truth segmentation (see Table 6), similar to the viewing tasks, the ARI values of the HDBSCAN algorithm were always considerably lower than the ARI values of our approach for the searching tasks. Although Block-o-Matic! achieved higher ARI values on the AVG and Yahoo pages with the first ground segmentation approach and a higher ARI value on the AVG page with the second ground truth segmentation in comparison with our approach, it could not perform better than our approach on the majority of the pages. The ARI values of the VIPS algorithm at varying segmentation levels were again slightly higher than
the ARI values of our approach. However, none of the levels of the VIPS algorithm achieved better results compared to our approach on the majority of the pages, except its fourth level in both of the ground truth segmentation.

The overall results of the searching tasks slightly differ from the overall results of the viewing tasks. The highest mean and median ARI values were achieved by our approach in the first ground segmentation, but not in the second ground segmentation. However, if the results of the AVG web page are excluded, then our proposed approach will also achieve the highest mean and median ARI values in the second ground truth segmentation. When we consider the overall results by taking the status of the AVG page into account, we can still suggest that our proposed approach is able to provide the most similar segmentation to the ground truth segmentation overall for the searching tasks.

5.2 Effects of the Number of Participants

Table 7 shows the mean, median, standard deviation, minimum and maximum of the ARI values achieved by our approach with randomly chosen 25%, 50%, 75% and 100% of the participants from the eye tracking study on all the pages for the viewing and searching tasks with the first and second ground truth segmentation where the best values are shown in bold.

For the searching tasks, our approach achieved the highest mean, median, minimum and maximum values with 75% and 100% of the participants with the first and second ground truth segmentation where the lowest standard deviation was achieved with 25% of the participants. For the viewing tasks, our approach achieved the highest mean, minimum and maximum values and the lowest standard deviation with 100% of the participants in the first ground truth segmentation where the highest median value was achieved with 75% of the participants. Even though it achieved the highest mean and median values in the second ground segmentation for the viewing tasks, it could not achieve the highest minimum and maximum values and the lowest standard deviation with 100% of the participants.

Table 7. The mean, median, standard deviation, minimum and maximum of the ARI values achieved by our approach with randomly chosen 25%, 50%, 75% and 100% of the participants from the eye tracking study for the viewing and searching tasks with the first and second ground truth segmentation.

<table>
<thead>
<tr>
<th>Task</th>
<th>Ground Truth</th>
<th>% of Participants</th>
<th>Mean</th>
<th>Median</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viewing</td>
<td>1</td>
<td>25</td>
<td>0.76</td>
<td>0.90</td>
<td>0.25</td>
<td>0.40</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>0.82</td>
<td>0.86</td>
<td>0.14</td>
<td>0.60</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>75</td>
<td>0.80</td>
<td>0.90</td>
<td>0.19</td>
<td>0.39</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.85</td>
<td>0.88</td>
<td>0.11</td>
<td>0.68</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>0.72</td>
<td>0.80</td>
<td>0.26</td>
<td>0.35</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>0.77</td>
<td>0.80</td>
<td>0.13</td>
<td>0.37</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>75</td>
<td>0.78</td>
<td>0.78</td>
<td>0.18</td>
<td>0.57</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.78</td>
<td>0.80</td>
<td>0.16</td>
<td>0.53</td>
<td>0.94</td>
</tr>
<tr>
<td>Searching</td>
<td>1</td>
<td>25</td>
<td>0.51</td>
<td>0.59</td>
<td>0.27</td>
<td>0.18</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>0.63</td>
<td>0.60</td>
<td>0.31</td>
<td>0.18</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>75</td>
<td>0.64</td>
<td>0.60</td>
<td>0.31</td>
<td>0.18</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.64</td>
<td>0.60</td>
<td>0.31</td>
<td>0.18</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>0.48</td>
<td>0.57</td>
<td>0.22</td>
<td>0.18</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50</td>
<td>0.60</td>
<td>0.57</td>
<td>0.28</td>
<td>0.18</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>75</td>
<td>0.61</td>
<td>0.62</td>
<td>0.29</td>
<td>0.18</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100</td>
<td>0.61</td>
<td>0.62</td>
<td>0.29</td>
<td>0.18</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Although our approach tends to provide slightly better results with a higher number of participants, the results suggest that our approach also performs well with a lower number of participants.
The standard deviation of the mean ARI values for the viewing tasks are 0.04 and 0.03 with the first and second ground segmentation respectively whereas the standard deviation of the mean ARI values for the searching tasks are 0.07 and 0.06 with the first and second ground segmentation respectively.

6 DISCUSSION

We have developed a new approach for segmenting web pages into their AOIs. This approach is integration of web page driven and eye tracking data driven approaches. It differs from existing approaches as it entirely segments web pages into their AOIs by both relating the AOIs with the source code of the web pages and considering users' interactions. Therefore, it is a novel contribution to both web page segmentation and eye tracking fields. In particular, when researchers conduct eye tracking studies on web pages, they can use our approach to automatically detect AOIs and use them in their analysis. Since the AOIs will be related to the source code of the web pages, they can directly use them to process the web pages. Specifically, they can re-order the AOIs or remove some of them to make the web pages more accessible in constrained environments [52]. For example, when there are many different AOIs on a particular web page, the most commonly used AOIs can be located at the beginning of the page such that small screen device users can directly access them without scrolling a lot.

We have also evaluated our approach by comparing its results with the results of its individual components in the same condition objectively. The eye tracking data used for this evaluation was collected for a completely different objective. According to our evaluation, our approach provides more similar segmentation to the ground truth segmentation constructed by a group of users. Specifically, it achieved very similar segmentation to the ground truth segmentation, such as 94% on the AVG page, 90% on the BBC page, and 80% on the Apple page in the first ground truth segmentation and 90% on the AVG page, 88% on the BBC page, and 98% on the Apple page in the second ground truth segmentation for the viewing tasks. Our approach identifies fixation clusters and uses them to prune a segmentation tree generated by the VIPS algorithm. The ARI values achieved by the HDBSCAN algorithm which only uses eye tracking data are considerably lower than the ARI values achieved by our approach. We also compared our approach with Block-o-Matic!, one of the recent state-of-art approaches in the field of web page segmentation, and the comparison showed that our approach also performed better.

In some cases, the VIPS algorithm performed better than our approach. This situation could be caused by the fixation clusters generated by the HDBSCAN algorithm. Some AOIs provided by our approach were larger than the AOIs in the ground truth segmentation because the fixation clusters did not allow further division. If the HDBSCAN algorithm had generated smaller fixation clusters, the further division would have been allowed. The current adjustment of the parameters of the HDBSCAN algorithm could be considered as strict, but the results would be more representative as all the participants were considered. By adjusting the parameters of the HDBSCAN algorithm, we could remove possible outliers, but we might decrease the representativeness of the results. The adjustment of the HDBSCAN parameters should also be further explored. Since the values of the parameters of the HDBSCAN algorithm potentially affect the results of our approach, we can test the HDBSCAN algorithm in the future by assigning different values to its parameters for finding the most appropriate values for their parameters in order to achieve the highest possible similarity to the ground truth segmentation.

As our integration algorithm works with the tree of elements provided by a web page driven approach (currently the VIPS algorithm), it does not divide the leaf elements in the tree. For example, on the Apple page, the menu could not be further divided into its menu items because the VIPS algorithm did not divide the main menu. In contrast, in the ground truth segmentation, the main...
menu was divided into their items. However, our integration algorithm is designed to work with
different kinds of web pages and eye tracking data driven approaches as long as their outputs are
appropriate. Therefore, another web page driven approach may allow to further divide the menu
on the Apple page.

The threshold value of the proposed approach was set to 0.05 in the evaluation. However, different
threshold values can also be explored in the future. Specifically, we can investigate whether the
threshold value should be correlated with the noise properties of the eye tracker and the spacing of
the elements.

Currently, our approach is a decision mechanism for identifying the levels to prune the tree of
elements provided by the VIPS algorithm based on the fixation clusters generated by the HDBSCAN
algorithm. However, other web page and eye tracking data driven approaches can also be used with
our approach in the future to investigate whether it is still possible to achieve higher similarities
to the ground truth segmentation in comparison with its individual components. For example,
as a web page driven approach, Block-o-Matic! can be used instead of the VIPS algorithm [41].
Although Block-o-Matic! is different from the VIPS algorithm, its segmentation process is also
based on the source code and visual representation of web pages. Therefore, we expect that our
approach can also work well with Block-o-Matic! and achieve the most similar segmentation to
the ground truth segmentation in comparison with its individual components. It is also possible to
use the whole DOM tree of a web page as a web page driven approach in our integrated approach.
However, not all information encoded in the visual representation of a web page is captured by its
DOM tree. Since the VIPS algorithm also considers the visual representation of web pages and the
visual attention of people are potentially affected by the visual representation of web pages, we
presume that the tree of elements generated by the VIPS algorithm is more suitable than the whole
DOM tree for our approach. However, further experiments can be conducted to investigate how
our approach works with the whole DOM tree.

Our current ground truth segmentation could also affect the evaluation of our approach, even
though we used two types of ground truth segmentation. We used the drawings of the participants
which show what they remembered about the structure of the web pages in the creation of the first
ground truth segmentation whereas we asked five people who had experience in web design &
development to segment web pages based on their opinions in the creation of the second ground
truth segmentation. For the creation of both the first and second ground truth segmentation, we did
not focus on specific tasks and this could cause our approach to perform better with the viewing
tasks where the participants did not require to answer specific questions. Since the first ground truth
segmentation was created with the drawings of the people who participated in the eye tracking
study whose dataset was used as an input of our integrated approach in the evaluation, it could
be argued that the ground truth segmentation and the resulting segmentation of our integrated
approach were unfairly correlated and potentially affected the evaluation. However, as presented
in Section 5.1, the evaluation results with the first ground segmentation are mainly consistent with
the evaluation results with the second ground segmentation created by different people.

In our study, we used an eye tracking dataset of 40 users. The number of users required for eye
tracking studies is still a controversial issue. In particular, Pernice and Nielsen [39] suggest that 39
users are required for a stable heat map which visualises which parts of visual stimuli receive more
attention. They also suggest when researchers want to conduct qualitative analysis to gain some
insights about how users interact with visual stimuli by watching their eye movements, 6 users
would be sufficient for them. Furthermore, Eraslan et al. [15] [17] raised the possibility to achieve
75% similarity to the results of 65 users with 27 users for searching tasks and 34 users for viewing
tasks when Scanpath Trend Analysis (STA) is conducted to identify the trending path of multiple
users on a web page in terms of its AOIs. The dataset we used to evaluate our approach here was
constructed with a higher number of users compared to the suggested number of users. Even though our experiments show that our approach performs slightly better with a higher number of participants, our approach also performs well with a lower number participants. For example, the standard deviation of the mean ARI values achieved with randomly chosen 25%, 50%, 75% and 100% of the participants for the viewing tasks was 0.03 and 0.04 with the first and second ground truth segmentation respectively. However, further experiments can be conducted to investigate the effects of the number of participants in depth.

The evaluation was not without limitations. One of these limitations was a limited number of web pages as our dataset was from an eye tracking study conducted on six web pages. However, this case was not specific to our dataset. Eye tracking studies are usually conducted with a small number of visual stimuli because a large number of visual stimuli potentially causes boredom and tiredness, thus decreasing the reliability of collected data. We also tried to find other public datasets collected on real web pages (not on screenshots as our approach also uses the source code of web pages), but we could not find any suitable dataset for our evaluation. Although the web pages used in the evaluation were the home pages of several popular websites and they might not be representative of the internal pages, they had different levels of visual complexities and they allowed us to evaluate our approach with the web pages with different levels of visual complexities. However, it would be worthwhile to conduct this evaluation with more and different kinds of web pages in the future. When we have a large number of web pages, we will also be able to conduct statistical analysis to compare our approach with its individual components and the state-of-art approaches. In this study, we could not apply statistical comparison tests due to a small number of web pages as small sample sizes tend to cause a failure to reach statistical significance due to low statistical power [45]. In our evaluation, we used both the viewing and searching tasks to investigate whether our proposed approach performs well regardless of the type of tasks. To draw stronger conclusions, the effects of different types of tasks should also be investigated. For example, synthesis tasks which require the combination of multiple facts to reach a new piece of information should also be investigated. The reason is that eye tracking studies are conducted with different types of tasks and we aim to explore the suitability of our approach for these tasks. This will allow us to demonstrate the most suitable task type(s) for our approach. Some participants may not be able to complete their tasks successfully. However, we included all the participants in our evaluation, even though they were not successful, as their data also showed real interaction.

In addition, instead of comparing the results of our approach with the results of its individual components and Block-o-Matic!, the results of our approach can also be compared with the results of other web page driven and eye tracking data driven approaches in the future to explore whether our approach still provides the most similar segmentation to the ground truth segmentation.

Our proposed approach is designed to automatically detect AOIs for a particular eye tracking study. Therefore, researchers should have eye tracking data for web pages to segment them. However, our approach provides an opportunity to generate ground truth segmentation based on both web page data and eye tracking data, and this ground truth segmentation can be used to train a machine learning model with certain features of the source code of web pages and these features can then be used to predict AOIs of web pages. Since there is a lack of approaches to generate ground truth segmentation based on both the source code of web pages and users’ interactions, our approach would be beneficial for the development of such model to predict AOIs of web pages.

Our proposed approach currently works with web pages with a minimal amount of dynamic content. Further studies should be conducted to process web pages with dynamic content (e.g. popups and menus appearing when we hover our mouse over them). One of the approaches that could be investigated is the division of eye tracking data into specific time periods based on appearance changes on web pages by preserving previous data on unchanged areas. Assume that a
A web page consists of four areas and one of these areas is changed after a particular period of time. In that case, we may keep both the previous and new data for the unchanged areas, but consider only the new data for the changing area. This approach could provide different segmentation for each appearance/instance of a particular web page, as this new appearance/instance of a web page may affect users’ perception about its segmentation.

7 CONCLUDING REMARKS

Different web page driven approaches have already been proposed for segmenting web pages into their AOIs and relating the AOIs with the source code of the web pages, but these approaches do not consider how users interact with web pages. Some eye tracking data driven approaches have also been used to identify AOIs on web pages. These approaches consider how users interact with web pages, but do not relate AOIs with the source code of web pages. In this paper, we propose an approach to integrate web page driven and eye tracking driven approaches for discovering AOIs and present its evaluation. Based on our evaluation, this approach performs better than its individual components in terms of providing the most similar segmentation to the ground truth segmentation constructed by a number of users. It also outperforms than Block-o-Matic! which was one of the recent state-of-art approaches in the field of web page segmentation. As we discussed in the Discussion section, the evaluation is not without limitations and could have been done in different ways. Therefore, we are planning to further evaluate our approach with different conditions in the future.

REFERENCES
