Focus Paragraph Detection for Online Zero-Effort Queries

Lessons learned from Eye-Tracking Data

Christin Seifert
Technische Universität
Dresden, Germany
christin.seifert2@tu-dresden.de

Annett Mitschick
Technische Universität
Dresden, Germany
annett.mitschick@tu-dresden.de

Jörg Schlötterer
University of Passau,
Germany
joerg.schloetterer@uni-passau.de

Raimund Dachselt
Technische Universität
Dresden, Germany
dachselt@acm.org

ABSTRACT

In order to realize zero-effort retrieval in a web-context, it is crucial to identify the part of the web page the user is focusing on. In this paper, we investigate the identification of focus paragraphs in web pages. Starting from a naive baseline for paragraph and focus paragraph detection, we conducted an eye-tracking study to evaluate the most promising features. We found that single features (mouse position, paragraph position, mouse activity) are less predictive for gaze which confirms findings from other studies. The results indicate that an algorithm for focus paragraph detection needs to incorporate a weighted combination of those features as well as additional features, e. g. semantic context derived from the user's web history.

Keywords

focus paragraph detection; zero-effort queries; eye tracking

1. INTRODUCTION

Just-in-time retrieval [8], also called zero-effort queries [1] is the idea of retrieving new, relevant information with minimal user effort, ideally in a fully automatic way. Zero-effort retrieval requires i) identification of relevant user context and ii) query construction based on this context. Our goal is to generate zero-effort queries in a web-based setting, i. e., when browsing arbitrary web pages. Conceptually, the ideal relevant textual user context is a paragraph, because each paragraph represents one main idea and the concepts within the paragraph are related to each other and to the main idea (topical coherence) [2]. In particular, on pages that contain diverse topics (in different paragraphs), such as news pages, it is desirable to find the topic (and hence the paragraph) the user is actually interested in.

CHIIR '17, March 07 - 11, 2017, Oslo, Norway

ISBN 978-1-4503-4677-1/17/03...$15.00

DOI: http://dx.doi.org/10.1145/3020165.3022138

This paper investigates the identification of the relevant user context, i. e., the *detection of focus paragraphs* in web pages. We hereby assume that the query construction given a focus paragraph is an independent and/or already optimal process [9]. In this paper we present i) an algorithm for *paragraph detection*, splitting a web page into content-bearing textual parts solely based on structural properties, ii) the results of a first user study on the accuracy of a naive baseline algorithm for *focus paragraph detection*, and iii) the results of an eye-tracking study on the effect of selected layout and interaction features on focus paragraph detection.

2. RELATED WORK

Paragraph detection, i. e., automatic splitting of larger documents into smaller, semantically meaningful parts, has been investigated before [5, 12, 3], but to the best of our knowledge not in the context of real-time processing of general web pages. Web page paragraph detection is especially challenging for two reasons: First, markup tags are used to convey both, semantics and layout information[1]. Second, the storage capabilities and processing power within web browsers are limited, making real-time paragraph detection based on content [5] or visual building blocks [3] intractable. Lagun and Agichtein [7] suggested manually engineered segmentation (i. e. hard-coded) for popular web pages which tend to share the same layout, such as Google search results, and supervised automatic classification based on selected training data for less frequent pages. However, our scenario of zero-effort retrieval on the web requires the paragraph detection to be generally applicable to all kinds of web pages and feasible online.

Previous work on focus detection investigated the correlation between mouse interactions (movements, clicks, text selections) on a web page and gaze with the goal of estimating the reading time spent on single paragraphs in the context of adaptive hypermedia systems. Hauger et al. [4] could predict the focus paragraph with approx. 79 % accuracy a-posteriori. Mouse clicks and text selections were identified as highly indicative for gaze position. Additionally, the authors found that the mean value of vertical eye

[1]Even though semantic elements (like article, section, ...) have been introduced with HTML5 (https://www.w3.org/TR/html5), their use is neither enforced nor guaranteed.

position on web pages is not centered, but slightly shifted towards the top (mean value at pixel position 473 of 996).

The relationship between cursor and gaze has also been investigated for search engine result pages [6]. The authors approximated the x and y position of the gaze separately with a linear model, and achieved a RMSE of approx. 125 pixel in each direction. Their linear model combined the dwell time, the time since a movement, the x-coordinate of the cursor position, and the most likely x-coordinate of the gaze based on future cursor positions (analogous for the y-coordinate). Both authors (as well as [7]) use features that are not applicable because they are not suitable for online processing (future mouse position) or for zero-effort querying (mouse click[2]). Further, a text selection already provides a very specific user context, rendering the paragraph detection unnecessary. While our application scenario is different, we conclude from related work, that i) mouse pointer is a semi-informative feature, ii) frequent mouse usage has to be identified, iii) vertical position seems to be highly indicative.

3. INITIAL SITUATION

In this section we describe the initial situation for our approach, presenting an algorithm for paragraph detection, an initial (naive) solution to determine the focus paragraph, and the results of a first user evaluation.

In order to determine the focus paragraph in a web page, it is necessary i) to split the page into paragraphs and remove non-content elements and ii) to select the paragraph in focus. For the fist step, we utilize the DOM structure of the web page[3]. First, we create a set of candidate paragraphs, composed of nodes that have a textual child node of at least 40 characters. Only the topmost node in a subtree satisfying this criterion is considered, since it already includes all child elements. Next, neighbouring candidates are combined, if they are only separated by text nodes (i. e., there are no other nodes, such as <p> or
 in between). Finally, the candidate set is filtered by visible paragraphs, which are either combined or contain at least 100 characters and a dot.

A **naive baseline** algorithm for focus paragraph detection is to select the topmost (completely visible) paragraph as focus paragraph and update the selected paragraph only on scroll-events. We used this baseline for two reasons: First, the topmost paragraph is the most likely paragraph to be looked at after following a hyperlink. Second, the topmost area of the page is already excluded by the paragraph detection algorithm if it merely contains navigation elements or advertisements.

In a first study with 77 users[4] we investigated the accuracy of our zero-effort query generation process including the accuracy of the paragraph and focus paragraph detection [10]. The prototype implemented the **naive baseline** for focus paragraph detection. We asked users to correct the focus paragraph if they considered it to be incorrect. This was achieved by simply clicking inside the correct paragraph. In 35% of all cases the focus paragraph was adapted, resulting in an accuracy of 65% for the focus paragraph detection. This accuracy is highly biased, because i) we did not ask whether the paragraph is correct in general, but asked users to correct wrong paragraphs (which took manual effort), ii) we visually outlined the detected paragraph, which might draw the focus of attention, and iii) the detection was embedded in a study with a different focus.

4. IMPROVING FOCUS DETECTION

Although the first study showed an accuracy of the naive baseline algorithm of at least 65%, the study had several limitations as described above. In this section we describe our approach of improving the focus paragraph detection for the context of web-based zero-effort queries using selected features based on related work and the results of an eye-tracking study.

4.1 Feature Selection

We chose to investigate four layout and interaction features in detail (cf. Figure 1): The **paragraph's size** $a = w \cdot h$ as a layout factor should account for the fact that larger areas on the screen have a higher probability of being looked at. The **paragraph's position** d (vertical and horizontal) on the page is included, because it has been identified as predictive for gaze on web pages in a different scenario [4]. We also included the **position of the mouse pointer** relative to the paragraph m, which has shown to be of predictive value if users move their mouse frequently [6]. Thus, we also included a **mouse pointer activity** threshold, indicating whether the mouse is used as aid for cognition [4].

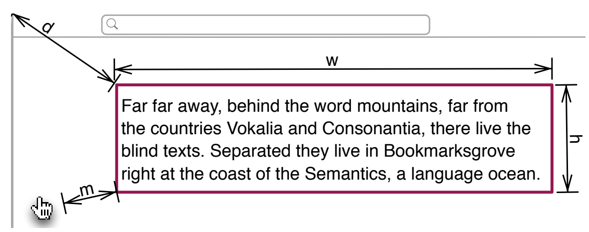

Figure 1: Features for focus paragraph detection

4.2 Eye-Tracking Study

To investigate the applicability of the selected features (paragraph size and position, mouse pointer position and activity) we conducted a user study using an eye-tracker to log the visual focus of users while reading web pages. The initial question was: which feature(s) could improve the accuracy of focus paragraph detection, and to what extend?

4.2.1 Procedure

The user study was conducted in a university lab using a Tobii Pro TX300 eye-tracker (providing a 23″ monitor). Twelve participants volunteered (2 female; age between 21 and 51; mostly students and staff from the local university), four needed a vision aid and used it during the experiment. Each session lasted around 30 minutes. First, the participants had to fill out a questionnaire about their web browsing and retrieval experiences and skills, and their knowledge

[2]A click typically corresponds to following a hyperlink, thus changing the page content. We regard clicks to the non-linked space as feature of user control, explicitly marking the paragraph as focused and hence not applicable to the automatic detection.

[3]The source code is available as part of the c4 framework https://github.com/eexcess/c4#paragraphdetection

[4]A detailed description of the study setup is available at https://github.com/schloett/p2q

and interest in 8 predefined topics such as geography, history, politics, etc. From these topics we later picked 2 or 3 for the session tasks (depending on the user's rating, preferring topics where the user indicated high interest and low or medium prior knowledge). The participants were given basic information about the setup of the study (accompanied by an individual calibration of the eye tracker), but no further details about the purpose. So all subjects knew that their eye movements were tracked and logged, but did not know what exactly we were interested in. Afterwards, the participants were given three questions within the scope of the selected topic and a Wikipedia article[5] in a full-screen browser window on the Tobii monitor (full-HD). The article directly provided the answers to the questions or links to other relevant articles. The task was to read the relevant Wikipedia articles (without time limit) until one felt to be able to answer the questions. We were able to do at least 2 iterations (topics) with each subject.

We tracked and logged coordinates and time stamp of gaze and mouse pointer positions on the screen with 50 Hz, the URL and content of the visited web page including position and layout of all detected paragraphs and detected focus paragraphs (naive algorithm), as well as scrolling and mouse click activities (cf. Figure 2 for some examples). The subjects were instructed not to use the browser's full text search functionality to highlight keywords, so that they were forced to read larger parts of the text to find relevant information.

4.2.2 Results

In order to investigate the relevance and impact of layout and interaction features for focus paragraph detection, we analyzed the logged data as follows: First of all, we evaluated the performance of the naive focus paragraph detection approach by counting the number of gaze events which were targeted at a point within the bounding box of a focus paragraph candidate. To compensate inaccuracies of the eye-tracker[6] we extended the area of the bounding box by 10 pixels in each direction. Furthermore we smoothed the samples using a weighted average filter and removed short fixations (<100 ms), which are usually a result of noise but are especially dispensable in our case as we are interested in sequences of focused reading (at least 100 ms up to 500 ms fixation duration [11]). We learned that an average of 36.63 % (SD 12.93 %) of the gaze events targeted to the screen did not hit a paragraph candidate at all. This is due to the fact that elements like images, headlines, formulas, table of contents, etc. do not provide sufficient textual content to construct the search context for a query formulation, and are thus ignored by the algorithm (as described above). Taking this into account, we only considered gaze events within the set of paragraph candidates. This resulted in an average success rate of 40.71 % (SD 18.27 %) for the naive baseline algorithm. We found that the lowest accuracy (9.5 %) was achieved when the subject was reading thoroughly through large parts of text without any mouse interaction or scrolling, which would trigger the reallocation of the focus. The best result (87.9 %) was achieved when the subject used anchor links within the document to directly jump to a paragraph and read it.

[5]Please note that all participants shared the same native language, in which all reading material was provided.

[6]According to Tobii the TX300 has an average accuracy of 0.4° and a precision of 0.14° [13].

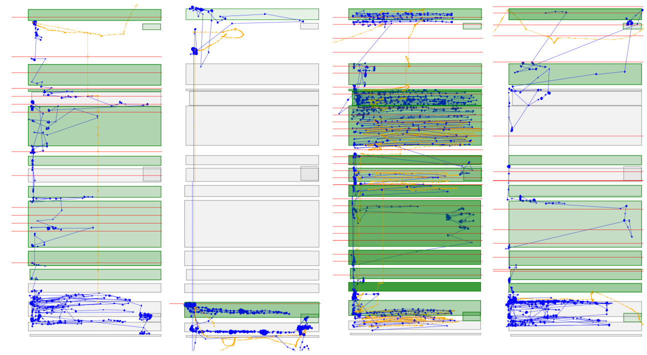

Figure 2: Four participants reading the same web page: gaze hits (blue), assumed focus paragraphs (green), vertical scrolling position (red) and mouse activity (orange)

Next, we investigated the influence of the paragraph's size. We calculated the ratio between the size of the actual read paragraph (fixation duration of at least 3 seconds) and the size of the largest visible paragraph on the screen. The average ratio was 0.56 (SD = 0.35), which means that the mean size of a focused paragraph was about a half of the size of the largest visible paragraph. Only an average of 33 % of the gaze events hit the largest visible paragraph in each case.

After that we focused on the influence of the vertical position of the paragraph on the screen[7]. Regarding the overall vertical distribution of gaze hits on the screen we found that gaze is primarily directed to the upper half of the screen (mean value at 402 of 1080 pixels from the top). We used the frequency values (see Figure 3) to weight the visible paragraphs according to their vertical position when rerunning the focus paragraph detection based on the logs. In this experiment we achieved an average success rate of 41.59 % (SD 16.65 %), which is only slightly better than the baseline algorithm. According to expectations, this approach performed better in cases of thoroughly reading of long text passages.

Furthermore, we investigated the relation between gaze events and simultaneous mouse pointer positions (as one of the feature candidates). As to be seen in Figure 3 the mouse pointer is mainly placed near the vertical center (mean value at 472 pixels of 1080). We calculated a matching between mouse hovering in a vertical distance to the gaze target of under 100 pixels[8] in 38 % of all cases. According to the assumption that the activity of the mouse pointer is an indication for its usage as an aid for cognition [4], we focused on the samples showing high mouse pointer activity. In the selected subset we found a matching in 60 % of the cases.

4.3 Discussion

The results of the study are quite in line with our expectations, regarding the inferior performance of the naive baseline algorithm compared to the results of the first user study. This is mainly due to the limitations of the first study (cf. Section 3) and the difference in measuring the accuracy: while the first study achieved a binary measure through the (indirect) approval or (direct) rejection of the users without

[7]As the majority of the Wikipedia articles used in this study appear in a single-column layout we did not have sufficient data to evaluate the influence of the horizontal position.

[8]The average height of a detected paragraph was 100 pixels.

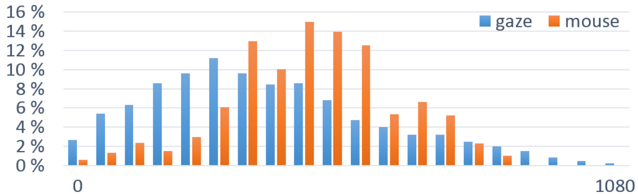

Figure 3: Histogram of the vertical distribution (0=top, 1080=bottom) of gaze (blue) and mouse pointer (red) positions

accounting for dwell time, the eye-tracking study measured gaze hits per paragraph in relation to time. The results confirm former studies [4, 6] concerning the deviation between gaze and mouse positioning, and that the mouse pointer is a semi-informative feature. The approach to use an activity threshold for the mouse pointer increases the accuracy significantly (up to 60 %) but was only applicable in 5 of 25 runs (20 %). The results showed that the relative size of a paragraph on its own is not an informative indicator for it being read. Furthermore, the weighting of the paragraphs according to their vertical positions based on the vertical distribution of gaze hits improved the accuracy only slightly, but seems to be promising in combination with navigational behavior of the users. This leads to the assumption that focus paragraph detection solely based on layout features is not feasible. Instead, future research should be focused on the weighted combination of layout, interaction and semantic features such as content-related background knowledge, browsing and query history, user profiles, etc.

The presented study has some limitations that should be further discussed. First of all, the number of subjects was too small to find similarities between users and derive adequate features in terms of the analysis of typical reading behavior. Second, the subjects knew that their gaze was tracked and logged, which means that we cannot ensure that they acted normally, i. e., performed normal reading behavior, although all of them stated that they did not feel affected by the presence of an eye-tracker. Due to the complexity of the given tasks it is very likely that the participants forgot about the tracking after some minutes. On the other hand, this complexity and a certain pressure to succeed might be the reason that several subjects tended to skim the page to spot relevant keywords instead of attentive reading. We are not sure if this adds to a bias and to what extend.

However, our aim was to get first results as an initial feedback to further improve focus paragraph detection. The insights we gained during the analysis led to promising features that should be taken into consideration.

5. CONCLUSION AND FUTURE WORK

In this paper we presented an approach for paragraph and focus paragraph detection for web-based zero-effort queries. We investigated the influence of layout and interaction features (namely paragraph size and position, and mouse pointer position and activity) with the help of an eye-tracking study. Our analysis showed that the pre-selected features were little informative in general, especially when considered in isolation, but might be effective in combination.

In the experiment we used a Wikipedia test corpus as a proxy for informational pages containing images, navigation elements and a sufficiently complex page structure. While we estimate the approach to be applicable to similar pages, future experiments might be necessary to investigate the applicability of the results to web pages of different types.

To further improve detection accuracy we will i) elaborate on weighting factors and parameters of features in combination, ii) investigate additional features, primarily semantic content features (from browsing/query history, etc.), and iii) investigate personalized focus paragraph detection by providing an initial model to all users, allowing for interactive correction of the detection and integrating the feedback into the model, e. g. to adapt the weights for different features.

6. REFERENCES

[1] J. Allan, B. Croft, A. Moffat, and M. Sanderson. Frontiers, Challenges, and Opportunities for Information Retrieval: Report from SWIRL 2012. *SIGIR Forum*, 46(1):2–32, 2012.

[2] G. Brown and Y. George. *Discourse Analysis.* Cambridge University Press, 1983.

[3] D. Cai, S. Yu, J.-R. Wen, and W.-Y. Ma. Extracting content structure for web pages based on visual representation. In *Proc. Asia-Pacific Web Conference*, pages 406–417, 2003.

[4] D. Hauger, A. Paramythis, and S. Weibelzahl. Using browser interaction data to determine page reading behavior. In *UMAP'11*, pages 147–158, 2011.

[5] M. A. Hearst. Multi-paragraph segmentation of expository text. In *Proc. Annual Meeting on Association for Computational Linguistics*, pages 9–16, 1994.

[6] J. Huang, R. White, and G. Buscher. User see, user point: Gaze and cursor alignment in web search. In *Proc. Conference on Human Factors in Computing Systems*, SIGCHI, pages 1341–1350, 2012.

[7] D. Lagun and E. Agichtein. Inferring searcher attention by jointly modeling user interactions and content salience. In *Proc. Conference on Research and Development in Information Retrieval*, SIGIR '15, pages 483–492, New York, NY, USA, 2015. ACM.

[8] B. J. Rhodes. *Just-In-Time Information Retrieval.* PhD thesis, Massachusetts Institute of Technology, 2000.

[9] J. Schlötterer, C. Seifert, and M. Granitzer. Supporting Web Surfers in Finding Related Material in Digital Library Repositories. In *Proc. Theory and Practice of Digital Libraries*, TPDL, 2016.

[10] C. Seifert, W. Bailer, T. Orgel, L. Gantner, R. Kern, H. Ziak, A. Petit, J. Schlötterer, S. Zwicklbauer, and M. Granitzer. Ubiquitous access to digital cultural heritage. *ACM Journal on Computing and Cultural Heritage (JOCCH)*, 2017. to appear.

[11] S. C. Sereno and K. Rayner. Measuring word recognition in reading: eye movements and event-related potentials. *Trends in Cognitive Sciences*, 7(11):489 – 493, 2003.

[12] C. Sporleder and M. Lapata. Automatic paragraph identification: A study across languages and domains. In *Proc. Conference on Empirical Methods in NLP*, EMNLP, pages 72–79, 2004.

[13] T. Technology. Accuracy and precision test report tx300. Technical report, Tobii Technology AB, 2013.