

# Visualizing Web Users' Attention to Text With Selection Heatmaps

Ilan Kirsh<sup>1</sup>[0000-0003-0130-8691]

The Academic College of Tel Aviv-Yaffo, Tel Aviv, Israel, [kirsh@mta.ac.il](mailto:kirsh@mta.ac.il)

**Abstract.** Web analytics tools provide useful information about the interaction of users with websites, and particularly, on what captures the attention of web visitors on websites. User attention to areas of web pages can be visualized using heatmaps. Two types of attention indicators are commonly used in web analytics heatmaps: visibility duration of page sections in the browser's viewport and mouse activity on areas and elements of web pages. This work introduces a new type of user attention heatmap, which visualizes the frequency of text selection operations on websites. Selection is the first step in the process of copying text to the clipboard, but it is also used to highlight important points while reading, similarly to highlighting words on a notebook with a marker pen. As demonstrated and discussed in this paper, selection heatmaps provide interesting perspectives on user attention to paragraphs, sentences, and words on websites, and this could be useful in web analytics.

**Keywords:** Web Analytics · Visualization · Heatmaps · Text Selection

## 1 Introduction

Web analytics aims to shed light on how a website is used by its visitors, and particularly, what captures the users' attention [1]. In web analytics heatmaps, page areas that attract more attention are displayed with 'hot' background colors, such as shades of red, and areas that attract less attention with 'cold' background colors, such as shades of blue. Two types of attention indicators are commonly used in commercial web analytics: visibility time (page areas become visible and invisible as users scroll the page) and mouse activity (clicks and movements on page areas) [2, 3]. A recent paper proposed to track and use copy operations of text to the clipboard by web users as a new source of data for web page heatmaps [3]. The usefulness and importance of heatmaps in web analytics are demonstrated and discussed in previous works [2, 3].

This work introduces a new type of heatmap, which visualizes user attention to text as reflected by selection (or highlighting) operations. Text selection is often considered as merely the first step in copying text to the clipboard. However, analysis shows that the vast majority of text selection operations in this study's dataset are not followed by copying. Text selection can be used by online readers to highlight points while reading, similarly to highlighting words

in a book with a marker pen. Accordingly, selection heatmaps can extend the existing web analytics toolbox and may provide new perspectives on web user attention to elements of online text.

## 2 Implementation

Figure 1 illustrates the selection heatmaps implementation. At the top, we can see a standard HTTP client-server communication between a browser (on the left side) and a web server (on the right side).

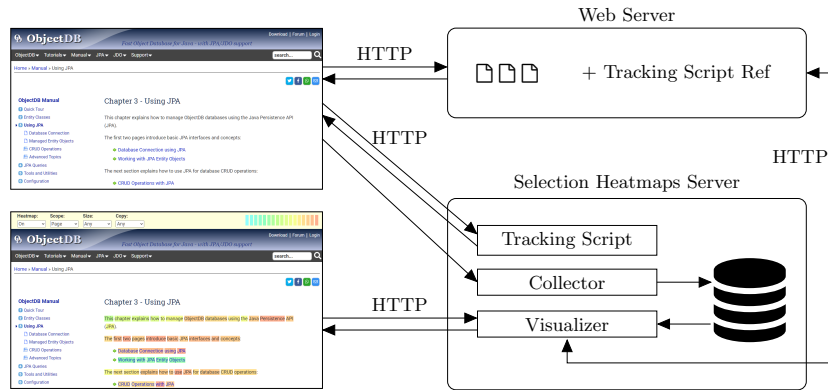


Fig. 1. Architecture of the Selection Heatmaps Implementation

To track text selection operations, a reference to a *Tracking Script* is embedded in the website pages. As a result, when a web page is loaded it triggers a request to load the Tracking Script from the Heatmaps Server. Once loaded, the Tracking Script records text selection operations and reports them to the *Collector* component in the Heatmaps Server, which stores the data in a dedicated database. Following the common practice of web analytics, and to protect user privacy, all the collected data are anonymized.

User actions, such as clicking and moving the mouse, scrolling the page, and copying text to the clipboard, can be tracked by listening to JavaScript events in the browser. Text selection is different, as it does not trigger any JavaScript event. It is possible, however, to get information from JavaScript on which text is selected, if any, at any given point in time. Therefore, the implementation uses a timer to routinely check for text selections, at a rate of once every tenth of a second (which only adds a negligible performance overhead). This poses an additional challenge, as a single selection operation can be detected multiple times with different text strings, while a user extends the selection (and sometimes also shrinks it). Therefore, detected selections are filtered: selections that are substrings of the preceding or the succeeding selection are discarded.

Web analysts can examine the web pages with the selection heatmaps by visiting the website through the *Visualizer* component, which functions as a proxy server. The Visualizer retrieves the original web pages from the web server and converts them to heatmap pages by adding background colors and shades. The colors that this implementation uses are shown on the right side of the heatmap toolbar at the top of the web page (see the bottom browser image in Figure 1), from the coldest color on the left to the hottest color on the right.

### 3 Demonstration

The selection heatmaps implementation was examined using pages of the ObjectDB website<sup>1</sup>. Web usage data were collected for six months, ending in June 2020. In total, 783,028 text selection operations were collected over 1,295,221 page views. Only a fraction of the selection operations were followed by copying text to the clipboard (there were only 109,525 copy operations).

Figure 2 demonstrates selection heatmaps for two lists. The explanations on the right side of these paragraphs attracted more selections than the class and method names on the left side. Some words, and noticeably the word ‘exactly’ (in both of its appearances), gained more selections than the other words. The similarity between the painting of the lines in these two paragraphs indicates repeating patterns. Possibly, users highlighted the explanations, and particularly the word ‘exactly’, to emphasize specific words to themselves while reading.

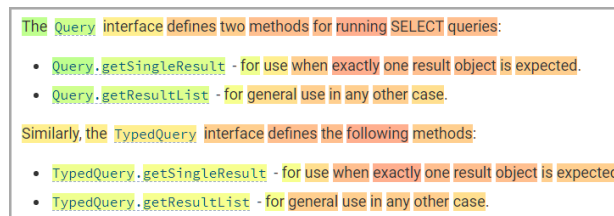


Fig. 2. A Selection Heatmap

Figure 3 presents heatmaps with the primary and secondary headers of two web pages. These examples demonstrate another pattern: words that are less frequent for the specific topic, or more unique, were selected more frequently. Common words such as ‘Obtaining’, ‘Using’, and ‘Access’ were selected less frequently than words that are specifically related to the context. The entire website is about ‘JPA’ (and database), thus ‘Entity Fields’ and ‘Connection’ are much more specific to these pages than ‘JPA’, and indeed, they were selected more frequently. Similar behavior is also identified when moving from the page and the primary header level down to page sections and secondary headers. In the ‘JPA

<sup>1</sup> [www.objectdb.com](http://www.objectdb.com)

Entity Fields’ page, the word ‘Fields’ is very frequent, and not directly linked to any particular section of the page. Accordingly, the words ‘Transient’, ‘Persistent’, etc. (which are specifically related to certain sections of the page) were selected by users more frequently. This user behavior of selecting specifically related terms more frequently might be linked to the concept of Inverse Document Frequency (IDF) in Information Retrieval (IR). Terms that are generally less frequent in the context provide more information, and therefore, highlighting such terms might be more beneficial while reading.

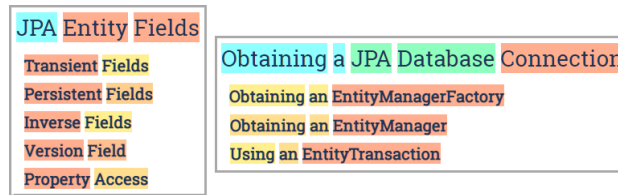


Fig. 3. Selection Heatmaps for Headers

## 4 Conclusions and Future Work

This paper demonstrates a new type of user attention heatmap, the selection heatmap, which visualizes the frequency of text selection operations across web pages. Selection heatmaps may provide new web analytics perspectives regarding user attention to paragraphs, sentences, and words on websites. The examples in this demo highlight interesting user behaviors with respect to text selection. Quantitative research will be needed to evaluate whether these examples represent significant, rather than causal patterns. Future work should also explore particular potential uses: the use of selection heatmaps as an alternative for conventional mouse movement heatmaps, which are used in commercial web analytics to visualize user attention; automatic identification of key sentences, for example for automatic text summarization; and automatic identification of important keywords in textual content, possibly as a complementary metric for Inverse Document Frequency (IDF) in Information Retrieval (IR).

## References

1. Kaushik, A.: Web Analytics 2.0. SYBEX Inc., USA (2010)
2. Kirsh, I.: Using mouse movement heatmaps to visualize user attention to words. In: Proceedings of the 11th Nordic Conference on Human-Computer Interaction (NordiCHI 2020), Tallinn, Estonia. pp. 117:1–5. Association for Computing Machinery, New York, NY, USA (10 2020). <https://doi.org/10.1145/3419249.3421250>

3. Kirsh, I., Joy, M.: A different web analytics perspective through copy to clipboard heatmaps. In: Proceedings of the 20th International Conference on Web Engineering (ICWE 2020), Lecture Notes in Computer Science, vol 12128. pp. 543–546. Springer International Publishing, Cham (06 2020). [https://doi.org/10.1007/978-3-030-50578-3\\_41](https://doi.org/10.1007/978-3-030-50578-3_41)