

# What Web Users Copy to the Clipboard on a Website: A Case Study

Ilan Kirsh <sup>a</sup>

*The Academic College of Tel Aviv-Yaffo, Tel Aviv, Israel*  
kirsh@mta.ac.il

**Keywords:** Clipboard, Copy and Paste, Web Analytics, Web Usage Mining, Human-Computer Interaction (HCI), User Behavior, Personalization, Adaptive Websites, Text Simplification, Text Summarization, Plagiarism, Search Engine Optimization (SEO).

**Abstract:** The clipboard is a central tool in human-computer interaction. It is difficult to imagine a productive day-to-day interaction with computers, tablets, and smartphones, without copy and paste functionalities. This study analyzes real usage data from a commercial website in order to understand what types of textual content users copy from the website, for what purposes, and what can we use such user activity data for. This paper advocates treating clipboard copy operations as a bidirectional human-computer dialogue, in which the computer can gain knowledge about the users, their preferences, and their needs. Copy operations data may be useful in various applications. For example, users may copy to the clipboard words that make the text difficult to understand, in order to search for more information on the internet. Accordingly, word copying on a website may be used as an indicator in Complex Word Identification (CWI) and help in text simplification. Users may copy key sentences in order to use them in summaries or as citations, and accordingly, the frequency of copying full sentences by web users could be used as an indicator in text summarization. Ten different potential uses of copy operations data are described and discussed in this paper. These proposed uses and applications span over a wide range of areas, including web analytics, web personalization, adaptive websites, text simplification, text summarization, detection of plagiarism, and search engine optimization.

## 1 INTRODUCTION

This study examines the content copied to the clipboard by users of a commercial website and proposes possible uses of copy operations data, in a wide range of applications.

Web analytics tools are widely used in almost all industries to better understand the interests, preferences, needs, and actions of website users. Premium web analytics services track client-side user activity at a very low level and go as far as to record all user mouse movements. However, tracking what the users copy to their clipboards is currently not part of the web analytics toolbox.

This paper is organized as follows: Section 2 reviews related work. Section 3 introduces the company website examined in this study and the data that were collected and used in this research. Section 4 presents examples of textual contents that were copied from

the website and taxonomy of copy operations types and subtypes. It also discusses possible user motivations to copy each type of textual content. Section 5 discusses ten different directions for potential applications and uses that can benefit from copy operations data. Section 6 concludes the paper.

## 2 RELATED WORK

Various studies show that web analytics is effective in understanding how visitors use websites and in improving and optimizing websites and web applications. This effectiveness has been demonstrated for a wide range of industries, including, for example, online news (Tandoc, 2015), online learning (Luo et al., 2015), e-commerce (Hasan et al., 2009), and digital marketing (Chaffey and Patron, 2012; Järvinen and Karjaluoto, 2015). Web analytics concepts, principles, and methods are described in detail in various books (Kaushik, 2007; Kaushik, 2010; Dykes, 2014; Alhlou et al., 2016).

The mouse cursor position is often used to esti-

<sup>a</sup>  <https://orcid.org/0000-0003-0130-8691>

mate which areas of the website capture the user’s attention. Studies show a correlation between the position of the mouse cursor on the screen and the user’s eye gaze (Huang et al., 2012), with a higher correlation when the user clicks the mouse or moves it (Chen et al., 2001; Rodden and Fu, 2007). Mouse cursor position information has been found to be an effective indicator of user’s attention in various web applications, including in e-commerce (Schneider et al., 2017), web marketing (Tzafilkou et al., 2014), online surveys (Cepeda et al., 2018), task execution (Milisavljevic et al., 2018), text reading analysis (Kirsh and Joy, 2020b), and web search (Guo and Agichtein, 2008; Huang et al., 2012; Navalpakkam et al., 2013; Rodden and Fu, 2007).

The cumulative attention of all the visitors in different areas of a web page can be visualized by heatmaps (Špakov and Miniotas, 2007; Lamberti and Paravati, 2015; Lamberti et al., 2017). Attention heatmaps are also offered by many commercial web analytics services, which track, record, store, and visualize user mouse activity (Kirsh and Joy, 2020a).

Unlike the mouse cursor position and the mouse activity, copy operations are hardly used to gain knowledge on users. A recent study proposed and demonstrated the visualization of cumulative user copy operations using heatmaps, similarly to the mouse cursor heatmaps (Kirsh and Joy, 2020a), but commercial web analytics services do not offer such heatmaps yet.

This study is the first to analyze web user copy operations of different types, to define a taxonomy of copy operations by types and subtypes, and to propose various possible uses of copy operations data. Initial work on two of the ten proposed uses have already started: Using copy operations in automatic text summarization (Kirsh and Joy, 2020c) and in automatic text simplification (Kirsh, 2020a). Further work is required to explore the other eight proposed uses and applications.

### 3 COPY OPERATIONS DATASET

Data for this study were collected by tracking and recording copy operations of users on a commercial website of a software company, ObjectDB Software. The examined website contains hundreds of documentation pages with technical information on the company’s products and tools, and about related subjects. These web pages are visited daily by thousands of visitors.

Many of the visits to the website are short. Software developers, who make up most of the traffic to

the website, use the website as a learning and knowledge source. They frequently arrive from search engines for short visits, after searching for specific technical solutions and code examples. When the desired code example is found, it is often copied to the user’s clipboard in order to be pasted and integrated into the user’s software project. Although most of the copy operations on this website are of code, this study explores also copy operations of text. Copy operations of images and other resources, however, are out of the scope of this study. Most of the uses that this paper discusses are more relevant to copy operations of text.

Figure 1 shows the architecture that was used to collect usage data. A reference to a *Tracking Script* is embedded in the relevant web pages. As a result, every request for one of these web pages returns a revised page that triggers an additional request to load the *Tracking Script* from the *Copy Tracking Server*. This script tracks clipboard copy operations (by listening to JavaScript events) and reports them to the *Collector* component in the server, which stores the data (following anonymization) in a dedicated database. The database contains the input dataset for this study, while the *Reporter* component supports retrieval of copy operations based on specific parameters (e.g. type, length, and language of the copied content).

Recording copy operations of web users is related to *session recording*, which is a common practice in modern web analytics of tracking and recording user activity on websites, including mouse movements and keystrokes. This practice raises questions related to personal data protection and user privacy, because of the risks of collecting sensitive personal data intentionally or unintentionally (Gilliam Haije, 2018). Session recording does not necessarily require prior user consent under personal data protection regulations, such as GDPR (under certain terms, as explained by the privacy and IT lawyer Arnoud Engelfriet (Gilliam Haije, 2018)). Unnecessary personal data should not be collected. If the collected data are completely anonymized, which is a standard practice in web analytics, then they are no longer considered personal data (e.g. according to GDPR).

The copy tracking system was run on 231 pages of the website. Usage data were collected for a period of several months, ending in March 2020. During this period, 654,399 page-views from 241,644 unique visitors (estimated), and 53,131 copy operations were recorded. This dataset of copy operations is the input for this study.

Obtaining data for web usage mining research is usually much more challenging than for web content mining research, and therefore, most of such studies

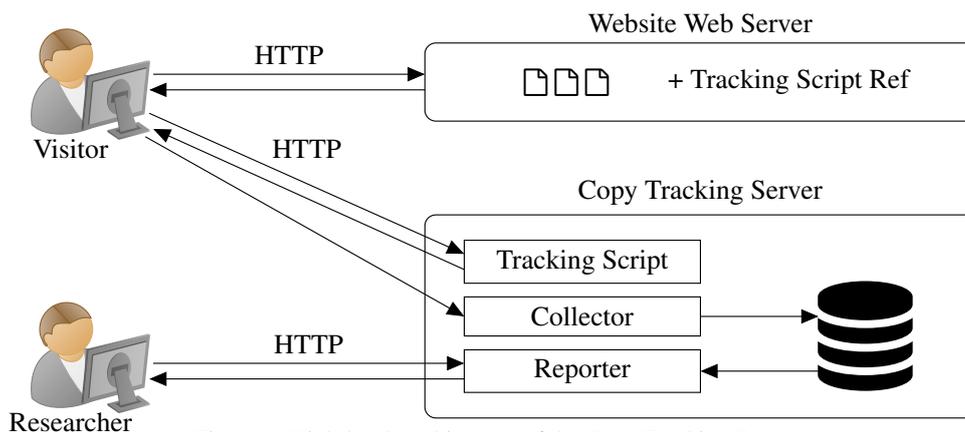


Figure 1: High-level Architecture of the Copy Tracking System.

focus on a single website. This study is no different, and further work is required in order to explore copy operations on other websites.

## 4 WHAT USERS COPY TO THE CLIPBOARD: EXAMPLES AND A TAXONOMY

Users copy from websites different things for different reasons. This section presents real examples of copied textual content from the studied dataset of copy operations. The goal is to classify different types of copy operations and to understand possible user motivations for copying each of these types of content (although there is no attempt to cover every possible motivation). Each of the following subsections presents one type of copied content, with a further division into subtypes.

### 4.1 Copying Single Words

We start by examining copying of single text words. Table 1 presents the most frequently copied words (taking into account only copy operations of a single word). The examples of copied words are divided into three categories of word types.

The first category contains acronyms of technical terms. The three examples provided represent central concepts in the website domain knowledge. A possible reason for copying these terms into the clipboard is that users need more information about these concepts, so they copy the terms to search them on the internet. They might search these terms on their own or

as part of a longer (and more specific) search queries that contain additional words.

The second category contains real-life words relating to the specific examples used in the website tutorials. The tutorials explain how to develop simple “Guestbook” applications in several environments. Users can follow the step by step instructions in these tutorials to create their own versions of these applications in their IDEs. The words: “Guestbook”, “guest”, “GuestDao”, and “GuestListener” are the suggested names for the created projects and classes. Users copy these names from the tutorials to their clipboards and then paste them in the relevant dialog boxes in their IDEs.

The third category is complex words. These are regular but less frequent words, which may be less familiar for some non-native English speakers. A plausible explanation as to why users copy complex words rather than simple words, such as “of” and “the”, is that they need more information about these words and therefore paste them in online dictionaries or search engines, searching for more information or a translation. This is discussed and analyzed in a follow-up study (Kirsh, 2020a), which shows that copied words are more likely to be evaluated as complex than uncopied words and words that are copied more frequently are more likely to be evaluated as complex than words that are copied less frequently, by three distinct word complexity measures.

### 4.2 Copying Phrases and Clauses

In this subsection, we progress from single words to sequences of several words. Table 2 presents the most frequently copied phrases and clauses (which are not complete sentences).

Table 1: Examples of Frequently Copied Words.

| Type                | Examples (the numbers of copy operations are indicated in the parentheses)  |
|---------------------|---|
| Concepts (Acronyms) | JPQL (229), JPA (36), JDO (10)  |
| Tutorial Words      | Guestbook (70), guest (32), GuestDao (20), GuestListener (14)   |
| Complex Words       | criteria (36), transient (24), embeddable (21), embedded (20), persistence (17), composite (16), redundant (14), retrieved (14), explicit (11), cascaded (11) |

Table 2: Examples of Frequently Copied Phrases and Clauses.

| Type    | Examples (the numbers of copy operations are indicated in the parentheses)  |
|---------|---|
| Phrases | Composite Primary Key (86), JPA Criteria API Queries (53), Embedded Primary Key (39), JPA Criteria API (32), The Sequence Strategy (23), JPA Named Queries (21) |
| Clauses | This page covers the following topics (7),<br>A value will be automatically generated for that field (5)  |

The phrases category contains short sequences of two or three words. These are mainly technical terms and concepts, similar to the acronyms category in subsection 4.1. Users may copy them to the clipboard in order to search for more information on these subjects (possibly with additional search words, in order to refine and focus the search). They may also be copied in order to paste them later in texts that the users write, including summaries, documentation, answers to on-line questions, etc.

The clauses category contains longer sequences of words. Usually searching for longer sequences of words on search engines is less effective, because such search attempts may be too restrictive. The most frequently copied clause “This page covers the following topics” is very general and not related specifically to the subject of the website. It is more likely that users copy such clauses in order to use them in their own text (they may search them on the internet first, to verify that these clauses are commonly used). The frequency of these copy operations are relatively low, and the information that they provide may be of less value in the context of this study.

### 4.3 Copying Sentences

From phrases and clauses, we proceed to full sentences. Table 3 presents some of the frequently copied sentences and sequences of sentences.

Some of the examples in Table 3 are single sentences and the others are multiple sentences. In both

cases, it is quite unlikely that users copy these sentences in order to use them in searches on the Internet, as such long strings are not effective for searching. It is much more likely that full sentences are copied in order to use them in other texts, e.g. as citations or in summaries. Possible uses for the copied sentences could be in documentations, summaries, presentations, blogs, websites, answers on forums (such as StackOverflow), or even private communications between colleagues who work on a project together. Copying full sentences is further demonstrated and discussed in a follow-up study (Kirsh and Joy, 2020c).

### 4.4 Copying Translated Text

The website under investigation contains only texts in English. Therefore, it was a surprise to find many hundreds of copy operations of text in other languages (mainly Spanish, but also French, Portuguese, and other languages). It seems that many users used Google Translate or other translation services to read translations of the pages, and subsequently copied translated text as well. Table 4 shows some examples of copied translated text, which are relatively short and could fit the limited space in the table.

Google Translate usage is usually transparent to websites and web applications, as ordinary requests are sent from the browser to the website server, and ordinary responses are sent back from the server to the client. Translations are done by communications between browsers and Google Translate, in which web-

Table 3: Examples of Frequently Copied Sentences.

| Type               | Examples (the numbers of copy operations are indicated in the parentheses)  |
|--------------------|---|
| Single Sentences   | <p>The JPA Criteria API provides an alternative way for defining JPA queries, which is mainly useful for building dynamic queries whose exact structure is only known at runtime. (24)</p> <p>Transient entity fields are fields that do not participate in persistence and their values are never stored in the database. (9)</p> <p>Changes to detached entity objects are not stored in the database unless modified detached objects are merged back into an EntityManager to become managed again. (7)</p> |
| Multiple Sentences | <p>The IDENTITY strategy also generates an automatic value during commit for every new entity object. The difference is that a separate identity generator is managed per type hierarchy, so generated values are unique only per type hierarchy. (63)</p> <p>The @Id annotation marks a field as a primary key field. When a primary key field is defined the primary key value is automatically injected into that field. (16)</p>  |

Table 4: Examples of Copied Translated Text.

| Type      | Examples (the numbers of copy operations are indicated in the parentheses)  |
|-----------|---|
| Phrases   | Requerimientos de plataforma (4), Fiabilidad y Estabilidad (3)  |
| Sentences | Adecuado para archivos de bases de datos que van desde kilobytes a terabytes. (1), Las relaciones son campos persistentes en clases persistentes que hacen referencia a otros objetos de entidad. (1) |

sites are not involved. However, client-side JavaScript copy events are aware of translated text in the browser (as shown to the users), and therefore, copying of translated text is tracked and recorded.

The copied translations include phrases and clauses (similar to the copied English phrases and clauses, shown in subsection 4.2), and sentences (similar to the copied English sentences, shown in subsection 4.3). Translated single words (as shown for English in subsection 4.1) have not been copied.

Three types of copied words have been described in subsection 4.1 (see Table 1). Words of the first two types, acronyms and names from the tutorials, might have been copied by Google Translate users, but since these words are never translated, they cannot be detected as copied translated text. Words of the third type, the complex words, are translated (and therefore could be detected if copied), but the translation to the user's preferred language eliminates the need to copy these words for the purpose of translation.

## 4.5 Copying Programming Code Elements and Fragments

All the types of copy operations that are discussed above are related to copying text, rather than code. Table 5 shows some of the more commonly copied code elements and fragments, which are short enough to be included as examples.

Code elements and fragments copied by users range from single words (query literals, class names, method names, files names, etc.), through lines of codes (e.g. queries), to larger fragments containing complete examples. This reflects the extent of assistance that the copying user needs. In some situations, only a single word is missing. In other situations, a complete example is needed. It is reasonable to expect that in most cases code elements and fragments that are copied on the website are then pasted in the users' IDEs and integrated into software projects. Single code words may also be used in further searches.

Table 5: Examples of Frequently Copied Code.

| Type      | Examples (the numbers of copy operations are indicated in the parentheses)  |
|-----------|---|
| Words     | CURRENT_DATE (243), CURRENT_TIMESTAMP (195), UPPER (112), LOWER (99), persistence.xml (93), EntityManager (80), SUBSTRING (57), Transient (54), orphanRemoval=true (47), YEAR{d'2011-12-31'} (41), @GeneratedValue (36), MEMBER OF (32), getSingleResult (25), PersistenceException (24), orphanRemoval (16). |
| Lines     | @NamedQuery(name="Country.findAll", query="SELECT c FROM Country c") (279)<br>SELECT c1, c2 FROM Country c1 INNER JOIN c1.neighbors c2 (158)  |
| Fragments | TypedQuery<Country> query =<br><br>em.createQuery("SELECT c FROM Country c", Country.class);<br><br>List<Country> results = query.getResultList(); (294)  |

## 5 POSSIBLE USES OF COPY OPERATIONS DATA

As shown in the previous sections, copy operations of website visitors can be tracked, stored, and analyzed. This potential source of web usage data could be useful in various applications. Each one of the following ten subsections proposes and discusses a different possible use of such data.

### 5.1 Understanding the Website Audience's Interest

Knowing the users and understanding how they use the website is a core element of user-centered design and a key to business success. Copy operations reveal valuable information about the audience of the website, which is not easily obtained by other means.

The frequencies of copying code indicate the importance of specific code fragments to the audience of the website. For example, Table 5 shows a high interest in the date and time literals, CURRENT\_DATE and CURRENT\_TIMESTAMP. Similarly, the frequencies of copying text indicate the importance of specific concepts. For example, Table 2 shows that certain types of primary keys (Composite Primary Key and Embedded Primary Key), and a certain type of queries (Criteria API Queries) are of particular interest to many users.

Existing web analytics tools, including premium services that track every user's mouse movement, fall short of providing this precious information. This study advocates the use of copy operations data as a new tool in the web analytics toolbox. As discussed in section 2, web analytics is effective in improving

and optimizing websites. Extending the boundaries of web analytics to include copy activity of users could make it even more effective.

### 5.2 Web Personalization and Adaptive Websites

The insights copy operations provide about user interest could also be used for personalization. Knowing the user's interest in real-time paves the way to commercial opportunities, such as presenting customized advertisements and special offers, as well as improving user experience by adjusting the website user interface, for example, by presenting new relevant links.

An adaptive website can also change the content that is presented to users based on their copy operations. In the context of a software company website with technical documentation, if a user copies only sample code fragments of a specific programming language, content that is relevant to that programming language should be presented where available, rather than content that is only relevant to other programming languages.

### 5.3 Text Simplification

Shardlow defined text simplification as "the process of modifying natural language to reduce its complexity and improve both readability and understandability" (Shardlow, 2014). In lexical text simplification, complex words are replaced with simpler words with similar meanings (Shardlow, 2014). This could be done manually or automatically. In any case, the identification of complex words is the first step.

Word complexity is subjective. Information about

which words users of a given website consider complex could be very helpful in improving the website content and making it more readable and understandable. Some users move the mouse cursor while reading to mark their reading position, so slowing down or stopping near words might indicate difficulties in reading or understanding (Kirsh, 2020b).

Table 1 shows examples of single word copy operations. The third category in that table contains regular words that have been copied by the users of the website. As discussed in subsection 4.1 and shown in a follow-up study (Kirsh, 2020a), these are relatively complex words, and apparently, some users have to search their meanings on search engines or online dictionaries. Therefore, simplifying these complex words might be particularly beneficial to the website's users.

## 5.4 Tooltips and Glossary

Similarly to complex words, users also copy professional and technical terms to the clipboard in order to search for more information about them, either on the website or externally on the internet. The "Concepts" category in Table 1 and the "Phrases" category in Table 2 contain terms that express concepts. Unlike complex words, these terms cannot be replaced with simple words in order to improve readability.

However, there may be other possible techniques to help readers. One option is to underline these terms and display tooltips when the mouse cursor hovers over them. Another option is to present a focused, local glossary of relevant terms beside the main text. Frequently copied terms may also be defined and explained on other pages of the website. However, most users arrive from search engines directly to specific pages, missing definitions found on other web pages, so they can benefit from employing such techniques. The frequencies of copy operations in the "Concepts" and "Phrases" categories could indicate where on the website such assistance is most needed.

## 5.5 Text Summarization

In the era of information explosion, text summarization is essential in bridging the gap between computer capabilities to store texts and human abilities to read them. A common approach to automatic text summarization is to compose a summary from key sentences, which are extracted from the original text. Various statistical metrics can be used to evaluate the importance of sentences in the original text. The most important sentences, based on the results of this evaluation, are selected and included in the summary (Sajjan

and Shinde, 2019).

However, copy operations data might provide a better indicator of the importance of sentences. Table 3 shows examples of sentences that are frequently copied by users. As discussed in subsection 4.3, users may copy sentences for various reasons, including for summaries, and it is reasonable to expect that more important sentences would be copied more frequently.

This novel approach is further developed and demonstrated in a follow-up study (Kirsh and Joy, 2020c), which shows that users tend to copy important key sentences more frequently and that a good summary can be built from the most frequently copied sentences.

Even if fully automatic text summarization is not used, copy operations data can help a human summarizer to make difficult decisions regarding what to include in a summary. Sentences copied by users more frequently are likely to be more important to the audience of the website, and therefore, should probably be included in summaries.

## 5.6 Reference Cards and Tip of the Day

Sentences that are copied frequently by users may also emphasize important points. A further examination of the examples of copied sentences in Table 3 reveals that there are two different types of frequently copied sentences.

The first type is summary sentences, such as key definitions and conclusions. For example: "Transient entity fields are fields that do not participate in persistence and their values are never stored in the database."

The second type highlights important details that have to be considered and may be less familiar, e.g. caveats and edge cases. For example: "Changes to detached entity objects are not stored in the database unless modified detached objects are merged back into an EntityManager to become managed again."

A summary can benefit from both types of key sentences. In addition, sentences of the second type may be selected (manually or automatically) and highlighted using reference cards (or cheat sheets) or as part of a Tip of the Day system.

## 5.7 Detecting Plagiarism

The vast majority of copy operations are legitimate, e.g. embedding short pieces of copied text as citations, with references to the source, is usually considered legal. On the other hand, some copy operations violate copyright rules, e.g. copying and publishing long sections of text, without explicit permission and

with no reference to the source, is illegal in most circumstances.

Copy operations of isolated words and phrases are usually less concerning in this context. Plagiarism is associated mainly with copying larger sections of text and code (as well as images and other resources, but this study focuses on copying textual content).

Simple Google searches of sentences and lines of code that were copied frequently by the website users reveal that many of them appear on other websites. Some occurrences, such as answers to questions on StackOverflow with proper references to the sources, are not only legitimate but even desirable. On the other hand, using large original sections of the website content, without any attribution to the source, are pure plagiarism.

This process of searching for unauthorized uses of the content of a website, based on copy operations data, can be either manual or automatic. An automatic plagiarism detection implementation can track suspicious copy operations, search for the copied text on the internet frequently, and alert the website owner when instances of plagiarism have been found.

## 5.8 Tracking User Progress in Tutorials

Table 1 shows several words that are frequently copied from the website tutorials. The “Guestbook”, “guest”, “GuestDao”, and “GuestListener” are the suggested names for the tutorial project and its classes. Users are expected to copy these names from the tutorial’s web pages and paste them in their IDEs. Similarly, users are expected to copy fragments of code from the tutorials and integrate them into their projects.

Tracking these copy operations (possibly in combination with other indicators) may be useful in analyzing user progress in tutorials. It could be used to detect possible breaking points, i.e. points at which many users abandon the tutorials. Some reduction in the number of copy operations throughout a tutorial, due to users’ decisions to quit the tutorial, is expected. However, if the copy operations data show an extreme drop at a certain point it may indicate an issue with that specific section of the tutorial and may require further investigation.

## 5.9 Understanding Language Translation Needs

As discussed in subsection 4.4, some users view websites through Google Translate. The translation provided by Google Translate is normally invisible in the website statistics and web analytics data. The website

is accessed ordinarily from the browser, and the translation is done by the browser and Google Translate. In this case study, the copy operations expose a community of users that use Google Translate to translate pages on the website to Spanish.

Examining copy operations data is a simple way to detect the level of usage of Google Translate on a website. This method has the additional benefit of detecting also specific interests of these users on the website (see subsection 5.1 above).

Deciding on investing in the translation of a website or specific web pages can take into account copy operations data of Google Translate users. That data is much more relevant for decisions on translation than the distribution of users by country (which is ordinarily available by web analytics services) since many non-native English speakers do not need translation and may even prefer reading the content in the original language.

## 5.10 Search Engine Optimization

Copy operations data can be useful in Search Engine Optimization (SEO), i.e. in making a website perform better in search engine results, and consequently in increasing the traffic to the website (which is usually desired for commercial websites). Copy operations indicate which topics are more popular among users of the website. Investing in new content on these subjects may be a cost-effective way to increase the traffic to the website.

As part of an SEO process, new pages can be created for phrases that users frequently copy for the purpose of further searches. A new dedicated page can provide additional content about the subject of a frequently copied phrase, with a title containing that phrase, as well as other optimizations that help search engines to establish the relevancy of the new page to the phrase.

This could attract more visitors from search engines. It may even be possible to catch a rebound of users that leave the website with a phrase in their clipboard to search for more information on the internet. These users may find themselves after that search back on the same website, possibly on a page that was created in this SEO process.

## 6 CONCLUSIONS AND FURTHER WORK

This study explored what users of a website copy to their clipboard. Copy operations of users of the website were recorded and analyzed, and different pat-

terns were identified. Accordingly, copy operations were classified into types and subtypes, and possible motivations for each of the copying types were discussed.

Copy operations data may be valuable in various situations. This paper proposes ten potential applications and uses that can benefit from that data. Additional work is needed in order to explore these proposed directions further and to develop and evaluate the usefulness of these potential applications.

Most of the proposed applications focus on specific types of copy operations. Therefore, automatic classification of copy operations types (possibly based on the taxonomy that this paper defines), may be needed as the first step in most of these applications. An automatic classifier may also help in estimating the distribution of copy operations among the different types, and the frequency of copy operations of each type.

Work on two of these applications, the use of copy operations in text summarization and the use of copy operations in Complex Word Identification (CWI) and text simplification, has already started with promising initial results, as discussed above. Further work is needed in order to explore the other eight proposed uses and to extend the initial work on the two first applications.

Although it is reasonable to expect that the results of this research are not unique for the selected website and can be extrapolated to other websites, due to the study focusing only on one website, further work on other websites is needed in order to establish and generalize these results.

## REFERENCES

- Alhlou, F., Asif, S., and Fettman, E. (2016). *Google Analytics Breakthrough: From Zero to Business Impact*. John Wiley & Sons, USA.
- Cepeda, C., Rodrigues, J., Dias, M. C., Oliveira, D., Rindlisbacher, D., Cheetham, M., and Gamboa, H. (2018). Mouse tracking measures and movement patterns with application for online surveys. In Holzinger, A., Kieseberg, P., Tjoa, A. M., and Weippl, E., editors, *Machine Learning and Knowledge Extraction*, pages 28–42, Cham. Springer International Publishing.
- Chaffey, D. and Patron, M. (2012). From web analytics to digital marketing optimization: Increasing the commercial value of digital analytics. *Journal of Direct, Data and Digital Marketing Practice*, 14:30–45.
- Chen, M. C., Anderson, J. R., and Sohn, M. H. (2001). What can a mouse cursor tell us more? correlation of eye/mouse movements on web browsing. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '01, page 281–282, New York, NY, USA. Association for Computing Machinery.
- Dykes, B. (2014). *Web Analytics Kick Start Guide: A Primer on the Fundamentals of Digital Analytics*. Peachpit, Pearson Education, USA.
- Gilliam Haije, E. (2018). Are session recording tools a risk to internet privacy.
- Guo, Q. and Agichtein, E. (2008). Exploring mouse movements for inferring query intent. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, page 707–708, New York, NY, USA. Association for Computing Machinery.
- Hasan, L., Morris, A., and Proberts, S. (2009). Using google analytics to evaluate the usability of e-commerce sites. In *Proceedings of the 1st International Conference on Human Centered Design*, pages 697–706, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Huang, J., White, R., and Buscher, G. (2012). User see, user point: Gaze and cursor alignment in web search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, page 1341–1350, New York, NY, USA. Association for Computing Machinery.
- Järvinen, J. and Karjaluo, H. (2015). The use of web analytics for digital marketing performance measurement. *Industrial Marketing Management*, 50:117 – 127.
- Kaushik, A. (2007). *Web Analytics: An Hour a Day*. SYBEX Inc., USA.
- Kaushik, A. (2010). *Web Analytics 2.0*. SYBEX Inc., USA.
- Kirsh, I. (2020a). Automatic complex word identification using implicit feedback from user copy operations. In *Proceedings of the 21st International Conference on Web Information Systems Engineering (WISE 2020)*, *Lecture Notes in Computer Science*, forthcoming, Cham. Springer International Publishing.
- Kirsh, I. (2020b). 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, forthcoming, New York, NY, USA. Association for Computing Machinery.
- Kirsh, I. and Joy, M. (2020a). 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, pages 543–546, Cham. Springer International Publishing.
- Kirsh, I. and Joy, M. (2020b). Exploring Pointer Assisted Reading (PAR): Using mouse movements to analyze web users' reading behaviors and patterns. In *Proceedings of the 22nd HCI International Conference (HCII 2020)*, *Lecture Notes in Computer Science*, Cham. Springer International Publishing.
- Kirsh, I. and Joy, M. (2020c). An HCI approach to extractive text summarization: Selecting key sentences based on user copy operations. In *Proceedings of the 22nd HCI International Conference (HCII 2020)*, *Communications in Computer and Information Science*, Cham. Springer International Publishing.

- Lamberti, F. and Paravati, G. (2015). Vdhm: Viewport-dom based heat maps as a tool for visually aggregating web users' interaction data from mobile and heterogeneous devices. In *Proceedings of the 2015 IEEE International Conference on Mobile Services, MS '15*, page 33–40, USA. IEEE Computer Society.
- Lamberti, F., Paravati, G., Gatteschi, V., and Cannavò, A. (2017). Supporting web analytics by aggregating user interaction data from heterogeneous devices using viewport-dom-based heat maps. *IEEE Transactions on Industrial Informatics*, 13:1989 – 1999.
- Luo, H., Rocco, S., and Schaad, C. (2015). Using google analytics to understand online learning: A case study of a graduate-level online course. In *Proceedings of the 2015 International Conference of Educational Innovation through Technology, EITT '15*, page 264–268, USA. IEEE Computer Society.
- Milisavljevic, A., Hamard, K., Petermann, C., Gosselin, B., Doré-Mazars, K., and Mancas, M. (2018). Eye and mouse coordination during task: From behaviour to prediction. In *International Conference on Human Computer Interaction Theory and Applications*, pages 86–93, Setúbal, Portugal. SciTePress.
- Navalpakkam, V., Jentzsch, L., Sayres, R., Ravi, S., Ahmed, A., and Smola, A. (2013). Measurement and modeling of eye-mouse behavior in the presence of nonlinear page layouts. In *Proceedings of the 22nd International Conference on World Wide Web, WWW '13*, page 953–964, New York, NY, USA. Association for Computing Machinery.
- Rodden, K. and Fu, X. (2007). Exploring how mouse movements relate to eye movements on web search results pages. In *Proceedings of ACM SIGIR 2007 Workshop on Web Information Seeking and Interaction*, pages 29–32, New York, NY, USA. Association for Computing Machinery.
- Sajjan, R. and Shinde, M. (2019). A detail survey on automatic text summarization. *International Journal of Computer Sciences and Engineering*, 7:991–998.
- Schneider, J., Weinmann, M., vom Brocke, J., and Schneider, C. (2017). Identifying preferences through mouse cursor movements – preliminary evidence. In *Proceedings of the 25th European Conference on Information Systems (ECIS)*, pages 2546–2556, Guimarães, Portugal. Research-in-Progress Papers.
- Shardlow, M. (2014). A survey of automated text simplification. *International Journal of Advanced Computer Science and Applications(IJACSA), Special Issue on Natural Language Processing 2014*, 4(1):58–70.
- Tandoc, E. C. J. (2015). Why web analytics click. *Journalism Studies*, 16(6):782–799.
- Tzafilkou, K., Protogeros, N., and Yakinthos, C. (2014). Mouse tracking for web marketing: Enhancing user experience in web application software by measuring self-efficacy and hesitation levels. *International Journal on Strategic Innovative Marketing*, 01.
- Špakov, O. and Miniotas, D. (2007). Visualization of eye gaze data using heat maps. *Elektronika ir Elektrotechnika - Medicine Technology*, 115.