The Study of Readability of Bibliographical Styles in Digital Environment Using Eye-tracking

DOI: http://dx.doi.org/10.12775/FT.2019.004

Abstract. Aim: The aim of the study was to gain knowledge about the way in which students use the electronic bibliography when looking for their positions. These investigations are designed to indicate the most intuitive bibliographic description style from the student’s level. It is also to contribute to the improvement of the aesthetics of the ways of publishing the subject bibliography presented in the didactic support services and in the electronic management systems of students and academic staff. Research method: The study in question was conducted using a mobile eye tracker, operating in infrared technology and 60 Hz frequency, on the group of students of Information Architecture. In order to achieve a balanced degree of complexity for different bibliographic styles, the experiment, apart from using the eye tracker, also included a questionnaire concerning detailed questions in the field of particular elements of the bibliographic description. The study used the Vancouver style, Traditional Polish agreement with the norm (PN-ISO 690: 2002 and PN ISO 690 2) mixed and the style based only on network resources. Results/Conclusions: These are pilot studies, and their results will be presented on the example of the USOS website, which is used at Nicolaus Copernicus University in Torun.
Introduction

As it is commonly known, bibliographic data are a description of literature referred to in academic publications. They serve two major purposes: firstly, they protect the author from being accused of plagiarism; secondly, they show the reader the source from which a given piece of information derives and make searching for this information easier.

One shall omit the role of the study of the state and trends in the development of the literature, including scientific literature based on bibliographical lists and publishing statistics. The purpose of the research on the development of scientific literature is quantitative characteristics, its structures (e.g. formal, thematic, language) and development trends. Due to the increasing significance of bibliographic lists, whose number is estimated at 9,386 by ZOTERO (the figure remains unconfirmed), it is essential to consider the issue of the actual usability of proposed solutions, understood as the source of searching for specific information. Aware of the volume of the data studied, the difficulty with the classification and the complexity of bibliometric problems, one should pay attention to the manner of conducting research into the usability of bibliography from a practical perspective.

It is obvious that bibliometrics touches upon a wide variety of problems; however, the primary goal of the pilot studies presented herein is to explore the search process for a given bibliographic style. The difference in the reception of bibliography in the physical and digital forms poses an additional problem. The process of the digitalisation of library collections is a subject of other studies, but it goes without saying that the present and subsequent generations of students are mostly oriented towards the use of digital collections. The reception and search analysis of a given set of items in bibliography may, and should, differ from a similar process conducted with the use of physical objects in the form of books and journals, particularly that "Visual search is a combination of two different types of behavior: goal-directed and exploratory search behavior, each using a different part of our brain".

The method of conducting an experiment with the use of eye-tracking technology, put forward in this paper, may be essential as it may show the advantages and disadvantages of a given style of creating bibliographic lists and indicate which elements should be effectively emphasised while drafting such lists to enhance their readability and legibility.

Tracking of eye movements is a relatively young research technique developed dynamically only from the mid-20th century, when effective digital methods were created allowing for precise measurements. What is significant here is the fact that the attention of researchers has always been occupied by the anatomy of

---

1 B. Stefaniak B., Badania bibliometryczne, informatyczne, naukometryczne, [in:] Informacja naukowa w Polsce: tradycja i nowoczesność, ed. by E. Ścibor, Olsztyn 1998, pp. 198-199.
4 J. R. Bergstrom, A. J. Schall, Eye tracking in user experience design, Amsterdam, Boston 2014, p. 34.
the human eye as a cognitive organ.\textsuperscript{5} It is worth noting that the process of seeing was the subject of research already in ancient times. Initially, it was associated with the anatomy and the process of image recognition. The technique of tracking eye movements and proprietary measuring tools were primarily the domain of doctors specializing in ophthalmology. Their main goal was to treat visual dysfunction, and the research was more of a university than clinical nature.

At the beginning of the 19\textsuperscript{th} century, the research on visual perception was carried out on the basis of the results of individual people as it was believed that specific principles could be introduced as universal on the basis of a single observation. Only color blindness or disordered visual axis, colloquially known as squint, were treated individually. However, advanced measurement tools were still insufficient. The contemporary researchers, such as Hermann Helmholtz or Ewald Hering\textsuperscript{6}, based their research mainly on substitute contrast called the afterimage. It is an optical phenomenon consisting in observing the focus of the selected object and the next reversal of sight, where the same shape \textit{v} will be observable in the complementary color. This mechanism is associated with the construction of the retina and the activity of color receptors when light falls on them.

In the second half of the 19\textsuperscript{th} century, the first precise measuring tools and the first basic definitions were created by Émile Javal and Ewald Hering, who were considered pioneers in the eye-tracking technology until today\textsuperscript{7}. Nowadays, the technology is mostly non-invasive and based on not only physical observation and measurement, but also on the measurement of electric potentials. What is meant here is electrooculography (EOG) “which relies on measurement of the skin’s electric potential differences, of electrodes placed around the eye”\textsuperscript{8}. “Video-assessed trackers utilize relatively inexpensive cameras and image processing hardware to compute the point of regard in real-time”\textsuperscript{9}. or “the corneal reflections of the light source (typically infra-red) is measured relatively to the location of the pupil center”\textsuperscript{10}. Not only does the proposed measurement method serve to study search processes nowadays, but it is also used to evaluate the architecture of information on websites as well as in neuroscience, psychology, computer science, marketing, advertising or industry\textsuperscript{11}.

Subject of Study

The study was conducted using the method of tracking eyeball movements with an EYETRIBE device\textsuperscript{12} that operates in the infrared technology at the frequency of...

\textsuperscript{5} N. J. Wade, Pioneers of Eye Movement Research, I-Perception, vol. 1: 2010, issue 2, pp. 33-68, \url{https://doi.org/10.1068/i0399}.

\textsuperscript{6} Ibidem, p. 50-51.

\textsuperscript{7} Ibidem, p. 51-55.

\textsuperscript{8} A. T. Duchowski, Eye Tracking Methodology. Theory and Practice, Cham 2017, p. 52.

\textsuperscript{9} Ibidem, p. 54.

\textsuperscript{10} Ibidem, p. 54.


60 Hz. The analysis was made with OGAMA\textsuperscript{13} software published under the General Public License, Version 5.0.1. The software is meant to register the measurements and comply with the device so that the recording of eyeball movements would deliver raw quantitative data that are analysed primarily with tests of significance (e.g. Student’s t-test) and with the analysis of variance. Additionally, there are created measurements allowing for the interpretation of how the elements shown on screen are perceived. Modern eye-tracking devices software enables immediate generation of the visualisation of these data and the automation of a considerable number of tasks, which used to be done manually. Baseline data comprise the identification of the points at which the user looked, the evaluation of the time spent on studying the material and the calculation of eye gaze coordinates (saccades). The most popular forms of the graphic representation of eye-tracking-derived data include: heat maps, fixation sequences and Areas of Interest (AOI) analysis\textsuperscript{14}.

Survey questionnaires and the University Study-Oriented System USOS\textsuperscript{15} meant that the student and the employee used to conduct the experiment. USOS is used on a daily basis by students as well as academic and administrative staff. Bibliographic lists were entered and appropriately modified for the subjects taught by the coordinator of course. Due to software limitations, the experiment was conducted with the use of static images prepared and formatted to the resolution of 1536 x 863 pixels with 32-bit colour depth saved in the PNG format so as to meet software requirements. This was the maximum value accepted by the software. Only the experimenter and the respondent were present while the study was conducted in conditions that would maximally minimise the possibility of the respondent’s distraction.

\textbf{III.1. Eye tracking experiment scheme.}

Source: Author’s own resources.


Research Group

A total of 30 respondents participated in the experiment – 14 women and 16 men, all fully eligible students of the 3rd year of the first cycle (Bachelor) programme of Information Architecture at Nicolaus Copernicus University in Toruń. All the students declared that they reached for academic literature and bibliographic lists to study and write their diploma (Bachelor) dissertations. They also admitted that they used the USOS regularly, up to several times a week.

The average age of the respondents was 22.4 years with women being 22.25 and men being 22.57 years old. Only two students declared in the survey to have completed higher education, which accounts for 6.67% of respondents versus the remaining 93.33% with secondary education.

Eight respondents admitted to having visual impairments: six indicated nearsightedness (myopia) with five being -0.5 to -1.5 dioptres and one being -5 dioptres, whereas two persons declared to suffer from astigmatism and one reported to have undergone an eye surgery, but he was not able to specify the date and type of the procedure.

Device calibration failed in the case of three persons. In two cases, a probable cause could be heavy makeup, which is mentioned in expert literature as a factor that can affect the correct functioning of measurement devices. In one case, the cause of lack of calibration remained unknown; no technical issues appeared and no illnesses were reported that could prevent the eye-tracker from working properly.

The experiment was conducted in line with the schedule specified in Table 1. The research problem and measurement tools were introduced to the respondents at the first stage of the study. After this introduction, the experimenter completed the statistical questionnaire. Some respondents confirmed they were acquainted with the eye-tracking technology and had participated in other studies in which the device was used.

The next stage was to calibrate the device prior to which the respondent was asked to limit all movements. The very calibration process consisted in the observation of nine measurement points generated by the application algorithm. As mentioned above, the calibration failed in three cases and the process had to be repeated for several students.

The experiment began with the display of black screen during which the experimenter asked the respondent a question. After obtaining confirmation from the student that they understood the nature of the research problem, the question was accepted and the screen with dedicated content was displayed for 20 seconds, followed by another black screen during the display of which the respondent was supposed to provide an answer.
Table 1. Experiment scheme.

<table>
<thead>
<tr>
<th>STAGE</th>
<th>EXPERIMENT SCHEME</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>Introducing the research problem and measurement device to the respondent</td>
</tr>
<tr>
<td>II.</td>
<td>Statistical questionnaire completed by the experimenter</td>
</tr>
<tr>
<td>III.</td>
<td>Device calibration</td>
</tr>
<tr>
<td>IV.</td>
<td>EXPERIMENT</td>
</tr>
<tr>
<td>1.</td>
<td>Question reading</td>
</tr>
<tr>
<td>2.</td>
<td>Single display – duration: 20 seconds</td>
</tr>
<tr>
<td>3.</td>
<td>Answer</td>
</tr>
</tbody>
</table>

Questions in the questionnaire were divided into four groups, arranged according to a particular bibliographic style. The styles included: the Vancouver Style, the Traditional Polish Style compliant with PN-ISO 690: 2002 and PN-ISO 690-2, a combination of the two above-mentioned styles (the mixed style) and the style based solely on online resources. Five questions were asked in reference to the Vancouver and combined styles, whereas four questions were asked in relation to the remaining two styles. The selection of bibliographic styles resulted from their popularity in Poland. The main tasks in the search process included locating and providing data in respect of the year of publication, author, title and number of publications, including the number of online publications.

Results

In the case of the evaluation of the relevance of answers per student the results of pilot studies fluctuated from 72 to 100%; therefore, statistically-wise, irrespective of the bibliographic style analysed, each student provided three wrong answers on average. However, as the experiment progressed, the number of correct answers increased producing the following results: 26 wrong answers for the Vancouver Style, 11 incorrect answers for the Traditional Polish Style, 13 wrong answers for the mixed style and one incorrect answer for the style based on online resources. In a conversation held after the measurement, the respondents indicated that at the initial stage they had a problem with understanding the task. The second reason why they made errors despite having the text displayed in the full-screen mode on a 17-inch screen was the accommodation of the eye to a barely legible font in a relatively small size. Apart from the bibliographic set containing solely links, determining the correct number of electronic sources for practically all bibliographic styles posed a problem to the students.

The first question concerning the Vancouver Style was: “Indicate which company published *E-learning w szkolnictwie wyższym* [E-learning in Higher Education].” The respondent was supposed to name the publisher. 66.7% of respondents replied correctly as shown in Figure 2 presenting the summary heat map of correct answers16.

---

16 Statistical image generated on the basis of the time in which the pupil focused in one position in relation to a given point on screen for all respondents who provided the correct answer, J. R. Bergstrom, A. J. Schall, op. cit.
III. 2. Heat map of correct answers
Source: Author’s own resources.

In the case of incorrect answers, it can be seen that “e-learning” was the keyword on which the respondents focused. Despite concentrating on the correct areas containing the word searched, the students had a problem with specifying the reply. Due to the fact that it was the first search attempt and that it was the first contact with this type of measurement for most respondents, it can be assumed that the answer to the above question was difficult to provide and may have resulted unreliable.

Another factor that might affect the results may be the fact that the bibliographic list comprised as many as six publications on e-learning, so the user should have searched for the entire phrase “e-learning w szkolnictwie wyższym” [E-learning in Higher Education].

III. 3. Heat map of incorrect answers.
Source: Author’s own resources.
The next question had a similar degree of complexity. The respondent was supposed to provide information about the year of publication. In the case of this test, the challenge was simpler because the respondent could notice that this publication was listed two lines below the previous one had they not searched for keywords in the former task. The number of correct answers for this question was higher and equalled 85.1%.

Owing to the fact that the respondent’s task was to locate the title of publication on the basis of the author’s name, the third question from this set was the least complex. 100% correct answers were provided because the list of references was arranged in the alphabetical order.

The penultimate task was to find the author, the prompt being the name of the publishing house. The second largest percentage of wrong answers provided in the entire experiment concerned this question – at the level of about 48.2%. Consequently, almost every second respondent answered this question incorrectly or did not answer it at all. In conversations held after the experiment, the students who had not replied to the question explained that they had not expected an intentional question about the same publication. Therefore, it may be assumed that the respondents relied on the existing answer key rather than the actual search process during the experiment. In the case of several correct answers, it is worth paying attention to the number of subsequent fixations (gaze points) and saccades (eye movement between fixations lasting from 20 to 40 milliseconds). It is presented in Figure 4 in which it can be seen that the information about the publishing house was listed already in point 6.

Ill. 4. Sequence of fixations and saccades for an example of a correct answer.
Source: Author’s own resources.

---

17 U. Garczark-Bąk, op. cit., p. 58.
The analysis of saccades and fixations for incorrect replies is more difficult to conduct due to the fact that their number varies and depends on individual predispositions. As shown in Figure 5, it is also more chaotic.

III. 5. Sequence of fixations and saccades for an example of an incorrect answer. Source: Author’s own resources.

The last question concerned the number of electronic sources. For a start, the respondent had to locate the HTTP tag[^18], a specific unspaced record using non-standard symbols or constant underlining with a changed font colour. For the purposes of the experiment, linking – and, consequently, the underlining of digital sources – was removed and standardised with the formatting of the remainder of the text while maintaining the correct structure. This made it difficult to count electronic items correctly. The correct answer was given by 15 respondents; nevertheless, in most cases (with the exception of three students) the limit of errors did not exceed two. Thus, finding digital sources was not a major challenge. The percentage of correct or almost correct answers can be partially ascribed to the knowledge and continuous use of the HTTP as against physical sources.

The next style in reference to which four questions were asked in the experiment was the traditional Polish bibliographic description, compliant with standards PN-ISO 690: 2002 and PN-ISO 690-2. The analysis of test results revealed a considerably smaller number of incorrect answers, at about 10%. Only two wrong answers were provided for the first task of finding the title on the basis of the name of the publishing house. One more incorrect answer was given to the next question concerning the e-mentor publishing house. Despite the fact of it being an electronic source, the location of the answers did not pose much difficulty. The worst

results were produced for the more difficult task of indicating the title of the journal, not the paper, on the basis of the names of two editors.

Ill. 6. Heat map of incorrect answers.
Source: Author’s own resources.

In the case of this task, 18% of incorrect answers were provided -five wrong replies. As depicted in Figure 5, which shows the heat map of all incorrect searches, the respondents did not know the structure of the scientific journal. Seemingly, the respondents treated the academic journal in the same manner as a book.

The last question about electronic sources had a similar degree of difficulty as the one concerning the Vancouver Style; nevertheless, one of the respondents neither provided any reply after 20 seconds, nor was he able to give the reason for failing to answer this query.

The third set of questions concerned the mixed style. Although this part comprised five questions, similar to those pertaining to the Vancouver Style, the number of incorrect answers was significantly higher and amounted to as much as 30%. This high percentage resulted from one question that required much focus. The first task was to indicate the title on the basis of the name of the publishing house. As with the previous styles studied in the experiment, the amount of incorrect answers equalled 18.5%. The second question required the respondent to specify the number of publications by a given author. In this case, only two wrong answers were given. With a higher degree of difficulty, the third question turned out to pose a relatively considerable problem as the respondents had to determine the number of authors on the basis of a publication in English. The difficulty of the task consisted in the fact that the only prompts given were the surnames and initials; moreover, the authors were of Asian descent. What made the task even more challenging was the fact that the publication was not the only one written in a foreign language.
The study showed that the students had a very big problem with determining the number of authors of whom there were 11. This answer was given by fewer than 15% of respondents. The reason for this may be the fact that students were not familiar with Asian names that rarely appear on bibliographic lists. Another cause may be the fact that the students seldom reach for foreign publications. The focus map depicted in Figure 6 shows further fixations whose number equals 21 for the time over 133 milliseconds according to the data imported into the database.

The concentration in fixations number 18 and 21 results from the fact that there is a double name that might have misled the respondents. In the case of a single surname, the limit of error increases the number of correct answers to nearly 45%.

III. 7. Heat map for all correct answers.
Source: Author’s own resources.

Two remaining tasks concerned searching for the year of publication on the basis of a specific title and the number of electronic sources listed in the mixed style. Both questions were answered correctly.

The smallest number of incorrect replies was given for the set of questions pertaining the style based on online resources. Only one wrong answer was provided for the pool of all respondents and all questions. Most correct replies were given for the last question asked in all sets, that is the number of digital sources in a given bibliographic style. The limit of errors of the two sources was exceeded merely in four cases.

Conclusion

The ease with which the recipient, reader or student uses a given bibliographic style depends primarily on the degree of their familiarity with a bibliographic style and the frequency with which they refer to it. The pilot studies show that restric-
tive standards for creating bibliographic lists can pose a problem in the process of searching for one piece of information. Bibliographic lists based on digital materials seem to be more intuitive for contemporary students. In this respect, particular attention should be paid to students’ habit of using materials on a global scale. The problem is significant because the number of Open Access journals is constantly growing, so they are given as internet sources in PDF format. Web sources are a separate subset of bibliographic styles, so it’s a worth considering to create a new bibliographic style or develop DOI-based (digital object identifier) instead of URLs.

Thus, it can be concluded that the restrictiveness of bibliographic description should depend on the place of publication. The aim of the pilot research was first of all to check the readability and legibility of four popular bibliographic styles used on the didactic support portals which showed the students’ high skills in searching the electronic sources. While all publications should be fully compliant with the styles of bibliographic descriptions, tools such as closed e-learning portals or university or school management systems should be more user-friendly, which in turn should encourage users to extend their knowledge and explore a given topic rather than scare them away with its strictness. This theorem becomes particularly important because, as it is commonly known, the style, form and size of a given bibliographic list is usually imposed by the academic teacher.

It should be pointed out that the above studies concerned a very narrow group of students and only four styles. Therefore, other popular bibliographic description systems require studies and analyzes. What is more, the technology of tracking eye movements seems to be a good direction, which may allow for a better understanding of the problem of searching the bibliographic lists and their improvement.

References

Bergstrom, Jennifer Romano, and Andrew Jonathan Schall. 2014. Eye tracking in user experience design. Amsterdam, Boston: Morgan Kaufmann.


Mazurek, Jolanta. 2011. „Czytelnik, użytkownik, klient, odbiorca w bibliotece cyfrowej… czyli KTO?” In Polskie Biblioteki Cyfrowe 2009 : materiały z konferencji
Stefaniak, Barbara. 1998. „Badania bibliometryczne, informometryczne, naukometryczne” In Informacja naukowa w Polsce: tradycja i nowoczesność, ed. by E. Ścibor, 198-199. Olsztyn: Wydawnictwo WSP.


Received: 6 VII 2019
Accepted: 27 IX 2019
Badanie wzrokowe czytelności stylów bibliograficznych na przykładzie akademickich systemów zarządzania

DOI: http://dx.doi.org/10.12775/FT.2019.004

łowa kluczowa: śledzenie wzroku (eye tracking); bibliografia; czytelność; style bibliograficzne

treszczenie. Cel: Celem badania było zdobycie wiedzy na temat tego, w jaki sposób studenci korzystają z bibliografii elektronicznej, szukając swoich stanowisk. Badania te mają na celu wskazanie najbardziej intuicyjnego stylu opisu bibliograficznego z poziomu studenta, a także przycznienie się do poprawy estetyki sposobów publikowania bibliografii tematycznej stosowanej w ramach usług wsparcia dydaktycznego i elektronicznych systemów zarządzania studentami i personel akademicki. Metoda badawcza: Omawiane badanie zostało przeprowadzone przy użyciu mobilnego urządzenia do śledzenia wzroku, działającego w technologii podczerwieni i częstotliwości 60 Hz, na grupie studentów kierunku „Architektura informacji”. Aby osiągnąć zrównoważony stopień złożoności dla różnych stylów bibliograficznych, eksperyment oprócz korzystania z narzędzia do śledzenia wzroku obejmował także kwestionariusz dotyczący szczegółowych pytań w zakresie poszczególnych elementów opisu bibliograficznego. W badaniu zastosowano mieszany styl Vancouver, tradycyjny polski styl zgodny z normą (PN-ISO 690: 2002 i PN ISO 690 2) oraz taki oparty wyłącznie na zasobach sieciowych. Wyniki/wnioski: Są to badania pilotażowe, a ich wyniki zostaną przedstawione na przykładzie strony internetowej USOS, z której korzysta Uniwersytet Mikołaja Kopernika w Toruniu.
Adam Szalach
Nikolaus-Kopernikus-Universität in Thorn, Polen. Fakultät für Philosophie und Sozialwissenschaften
e-mail: aszalach@doktorant.umk.pl
ORCID ID: 0000-0001-8040-001X


Visuelle Lesbarkeitsprüfung der bibliografischen Stile am Beispiel der akademischen Verwaltungssysteme

DOI: http://dx.doi.org/10.12775/FT.2019.004

Dieser Text wird unter der Creative Commons-Lizenz Namensnennung-Keine Bearbeitungen 4.0 International (CC BY-ND 4.0) veröffentlicht.

Stichworte: Blickerfassung (eye tracking); Bibliografie; Lesbarkeit; bibliografische Stile

Norm (PN-ISO 690: 2002 und PN ISO 690 2) sowie eines, der sich ausschließlich auf die Netzbestände stützte. **Ergebnisse/Schlussfolgerungen:** Es sind Pilotuntersuchungen, und ihre Ergebnisse werden am Beispiel der USOS-Webseite der Nikolaus-Kopernikus-Universität dargestellt.