Web-Navigation Skill Assessment through Eye-Tracking Data

Patrik Hlavac[1[0000−0001−7624−339X]], Jakub Simko[1[0000−0003−0239−4237]], and Maria Bielikova[1[0000−0003−4105−3494]]

Faculty of Informatics and Information Technologies,
Slovak Technical University, 811 07 Bratislava, Slovakia
{patrik.hlavac, jakub.simko, maria.bielikova}@stuba.sk

Abstract. Eye-tracking data provide many new options in domain of user modeling. In our work we focus on the automatic detection of web-navigation skill from eye-tracking data. We strive to gain a comprehensive view on the impact of navigation skills on addressing specific user studies and overall interaction on the Web. We proposed an approach for estimating the web navigation skill, with support of self-evaluation questionnaire. We have conducted eye-tracking study with 123 participants. Dataset from this study serves as a base for exploration analysis. We pair different web-navigation behavior metrics with result score from our questionnaire in order to find differences between participant groups. The results of the classification show that some stimuli are more appropriate than others.

Keywords: web navigation · eye-tracking · web literacy.

1 Introduction

Aim of our research is to show that analyzing user's gaze can bring valuable information into user classification. In our paper we describe its relation to web-navigation skill.

The main goal is to automatically distinguish between two groups based on inferred gaze metrics. Groups differ in efficiency when performing navigational tasks.

We estimate one specific aspect of web-navigation skill, the web-navigation efficiency. Our proposed questionnaire focuses on user's success and efficiency during navigational tasks on the Web.

Short and automatic web-navigation skill assessment would be helpful UX tool with perspective usage during user studies but also in general user modeling. Our main motivation for starting to research a topic of web-navigation impact on user behavior was very limited existing knowledge about gaze relations in the field of User studies. Nowadays, User Experience (UX) research become an interesting domain because of growing popularity of UX testing worldwide. Thanks to the power of UX testing, it quickly became a massively used approach in industry. More and more companies started to use UX methods to improve their products and services. More and more companies that support UX testing and consulting were established. Even general public is aware of these possibilities. This area deserves more exploration and deeper understanding.

Whilst qualitative studies generally consist of the interaction of a participant with an environment with presence of moderator as an important mediator, quantitative studies are carried mostly without the moderator and thus without further analysis of a perception of the given participant. Given that in this approach we usually gather a large amounts of data from logs or questionnaires, the evaluation of quantitative studies is inferred from methods of mathematical analysis [14].

Quantitative studies allow to generalize results for a greater population. However, this group could be so diversified, that we could not compare or evaluate the results. Broad study was performed with 445 participants [15] using an application to record data and session by cameras. There was an indication that each user has different characteristics and properties that enter the process and therefore that detecting of individual impacts could be more important than quantity of participants.

Our work is structured as follows. Related work section contains different views on the Web Navigation definition and skill assessment. Our study goal is explained in section 3. Experimental methodology is described by scenario, participants, devices and acquired data, followed by results. We summarized the observed results in section Conclusions.

2 Related Work

2.1 Web Navigation Definition

Web navigation includes many areas that could be inspected as well as there exist many factors that should be taken into account while investigating.

The navigation term can describe user activity and also website element (that usually contains hyperlinks). When we explore the user navigation skill, we have to consider also navigational environment that includes different elements. And vice versa, when we measure the quality of website navigation, we should consider the navigational ability of each user.

Website navigation is typically based on visual perception, accompanied by the goal to get somewhere (most likely to meet information needs) and exploring ways of doing this.

Since most of the information regarding the various products, actions and situations from daily life has gradually shifted to a web-based platform, there is a need for definition of new concept.

Nowadays, we consider digital literacy as a persons ability to perform tasks effectively in the digital environment, with emphasis on the representation of information in a form usable for computers [13]. It covers a wide range of qualities, abilities and skills, such as reading and interpretation of the media, replicating and modifying data (from documents to pictures and music). It also deals with the electronic equipment such as computers, mobile phones and portable devices and software cooperating with them. It is very connected to the Information literacy (search, examination and handling of information) and media literacy (competences allowing to work with the media in various formats and genres) [1].

Information literacy, as well as the Digital literacy, has ambiguous and broadly defined the problem area. Thus, a new, unique concept of Web literacy [3] has been earmarked.

Unlike computer literacy or digital literacy, Web literacy includes more detailed aspects of Web interaction. Many definitions divide Web literacy into several areas, including navigation. Some standards describe it and provide specific examples; other standards rely on a detailed and unambiguous definition that does not require examples.

We consider the web-navigation skill as the ability to effectively move across the Web.

2.2 Web Navigation as an Ability

We expect the navigation behavior (the users ability to intentionally move across the Web) is influenced mostly by two factors - the web layout and the user web-navigational skill. Navigation layout may also influence interaction and overall user satisfaction [9]. Navigation behavior is influenced by familiarity [6] (navigational expectations can help or distract if website uses elements of unexpected functionality) and the ability of keeping the path to correct destination.

Web navigation is a path traversed with or without main intention (goal). The interaction path then consists of all visited web-pages. When we look to a lower level, we can say the navigation even consists of all website elements and related interaction on each visited web-page element [11].

Web navigation without specific aim, but with intention could be depicted by typical task: *Get understand what is the site about.* In this case we are measuring the ability to traverse the most important parts of website by the most efficient way. Thus, the real ability is that we know where to get crucial information and how to navigate to the next information point.

Web navigation with intention should evoke the need of choosing the most correct and efficient way to reach the final website (destination): *Find specific product.*

2.3 Web Navigation as a Layout

Web navigation could be divided according the size of working environment into: wearable, mobile device (smartphones, tablets), standard screen (PC, notebook), we should consider even voice form of navigation for blind users. We can navigate through icons, texts (and text abbreviations), shortcuts, sounds (read links for sightless persons).

Navigation design includes specific differences on the level of context domains [7] (e.g. e-shop, e-banking, e-government), or even on the level of individual websites of the same context domain. Navigation layout is influenced by individual design approaches, which are based on constantly evolving technologies. Secondly, it is also influenced by the graphic style. In some practical cases, the navigation could be influenced by a framework or brand that serves the software (e.g. specific color, icons, actions).

2.4 Web Navigation as a Process

Web navigation has been evolving over the years as well as technical aspects of the Internet. At present, we can see some of the well-established standards of creating web navigation by navigational elements that will be familiar to users even when they first arrive at an unknown page. Effective web browsing is the foundation for good usability, and so ongoing approaches for creating navigation are constantly evolving and improving [12].

The way to the navigation goal could lead through many different paths. The chosen way can lead to assessment of user's navigation skill. This could be a very powerful method in the topic of skill estimation. As we discovered during work on skill estimation [8], users can be clustered by the path they chose during navigation to the target.

Every path has its own reasoning. Whether it is web-page influenced by website owner, that tried to lead customers to specific information (or additional advertisement), or whether it is well known shopping cart process, we can describe the navigational path.

2.5 Web Literacy Estimating

Most current assessments test users for overall computer usage, however they test web interaction only marginally and even less they test web-navigation behavior. Naturally, research experiments are specially designed for testing specific issues [7], therefore they are not applicable to various aspects. Despite the expectations, we did not find a great variety of proven and established Web literacy assessments. Existing tests for Information, Computer or Internet literacy go only partially into details that we can observe in Web literacy domain.

The results of well-known worldwide skills measurements are based mainly on the subjective opinions of the participants. They are calculated from the scale of how often student do some activity [10], but not how successful or effective he actually is. PISA 2012 tests were accompanied by computer-based tests focused on reading literacy [2]. Interaction was logged during work in specialized framework. Raw files were then treated to extract the navigation sequence and standard metrics (e.g. number of steps) was inferred.

Verification of literacy is significant for companies that provide education. In some cases, the assessment is just the last part of online educational process. Microsoft Digital Literacy [1] provides courses based on testing practical skills (Build Your First App; Creating an Internet Email Account; Writing a Great Resume; or newer Digital Literacy: Get Connected, Browse the Web, Search the

[1] https://www.microsoft.com/en-us/DigitalLiteracy

Web etc.). Assessments contain from 30 to 90 questions according to the level of tested literacy. The user usually has to choose some of prepared options. Northstar [2] provides much sophisticated tool, which is another assessment service, but provides a more attractive and verbally moderated tests. It tests the user from the basics of text processing, presentations, web or information literacy in a nine assessments. Its environment simulates a computer desktop, interactive captcha verification, completing a web form or searching the Web. Questions are answered by click or by explicit I dont know button. iSkills Assessment [3] contains many types of environment for Defining, Accessing (filtering with select boxes), Evaluating, Managing (sorting content in e-mail client), Integrating, Creating (statistical environment with graphs) or Communicating (composing e-mail).

Our initial effort is to automatize estimation of users web-navigation skill. There are many theoretically defined standards which help us to understand the issue in depth. However, there are almost no practical tests or questionnaires. While at the beginning of the millennium several attempts were made to estimate digital and Web literacy through questionnaires, they are no longer available today. In addition, with constant development and additional discoveries, what does the interaction on Web mean and what does it include - the number of such instruments to cover each area should have been created along with their definition. But specialized tools missed arise.

User experience research is full of different methods and modern approaches. On the one hand, in the UX industry, the time-limited website testing is the tool to understand the perception of the user. One of the pertaining method is the Three second rule, where after three seconds, user should be able to answer what he can expect from the website. This method could be also found as Five second rule [4] (with little adjustments).

On the other hand, UX research [5] shows that first few seconds of interaction with website are critical in terms of deciding to leave or to stay. There are even hints [5] that claim, that critical time that is needed to leave the website is decreasing, being now about 8 seconds.

How is it, when user has to orientate in 8 seconds? We assume that they will focus on the elements that they consider most important to gain most of information fast. Do experts choose the important elements better than novices?

3 Experiment: Influence of Web-navigation Skill on User Behavior

The way that user perceive the website through exploring multiple elements allows us to model his navigation. We have chosen the scenario tasks that help us to motivate the user to search for key information on the web through multiple

[2] https://www.digitalliteracyassessment.org/
[3] https://forms.ets.org/sf/iskills/rfi/
[4] https://fivesecondtest.com/
[5] https://www.nngroup.com/articles/how-long-do-users-stay-on-web-pages/

elements (thus navigate). Key information would be those, which user consider to be helpful for orientation on current homepage.

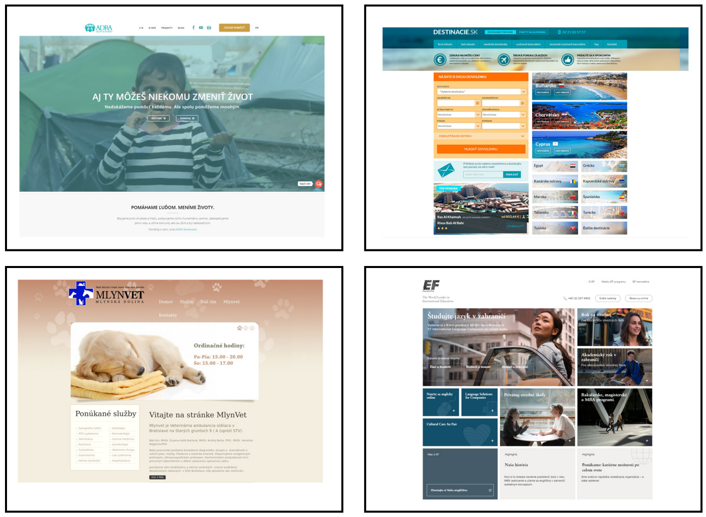

Fig. 1. Sample of four homepage websites among all 32. Website differs in domain, content and elements.

Through exploratory analysis we strive to gain deeper insight into human behavior during navigational tasks on the Web. We chose 32 screenshots [6] from different website homepages that are provided as testing stimuli for each participant with simple task "to understand what is the website about". We record gaze data during sessions that are further processed to transition matrices. Our approach consists of three main parts:

- collection of eye-tracking data from the first encounter with the homepage,
- collection of a web-navigation skill questionnaire and calculating results,
- classification of quantified data on participant behavior.

We expect that the division of Web literacy by other aspects will be the step to modelling the Web literacy more transparently and automatically. Thus, maybe this will be the "missing link" in modern user modelling and a primal approach to fast and easy web-navigation skill detection.

We approach with analysis of user orientation on the website. Dataset contains behavior from the orientation on the homepages of different websites (see Figure 1). We expect that users with similar skill would have similar behavior in terms of specific gaze metrics. We analyze only first 7 seconds of participant

[6] http://eyetracking.hlavac.sk/web-navigation/

interaction to obtain the information about what elements considers participant as the most important when perceiving new website.

3.1 Experiment Goals

The main goal is to find differences in eye-tracking data that allow us to classify users according their web-navigation skill. We have to go through the following steps:

- **To verify whether the questionnaire has a distinctive ability to estimate web-navigation skill.**
  The experiment was designed to compare the results of the proposed questionnaire with real behavior - more precisely, to find out whether the questionnaire can in specific cases significantly demonstrate a difference in metrics (duration of the task execution, number of tasks, SPI, gaze metrics).
- **To compare the behavior on different groups of participants on modern websites.**
  All participants perform four tasks on modern single-page websites. All four websites are of single-page structure, therefore links in the navigation only scroll the content and do not open new pages.
- **To analyze the user orientation on the homepages among groups.**
  Contains collecting the dataset from the orientation on the home page. These websites come from different domains (education, stores, services).

3.2 Questionnaire

In order to partially abolish the subjective assessment effect, we did not ask directly for web-navigation skill estimate. We rather asked about the situations associated with it. Using examples from existing questionnaires, we have formulated 19 statements, where the participant indicates on Likert scale 1-5 how much they agree with each (from 1 - definitely disagree, to 5 - definitely agree).

The questionnaire contains a participant identifier (usually 6 characters), filling time time-stamp, and integer values for 19 statement questions [7]. For example: “I often do not finish my online purchase because the page is too chaotic”, “I often get to a page where I can not orientate myself”, “It often happens that I do not find what I’m looking for at all.”

Questionnaire is provided as online form. It is anonymous, individual codes are used to pair with gaze data. Questionnaire is not limited by the time or requirements to be met. Questions are provided in random order to each participant.

[7] http://eyetracking.hlavac.sk/web-navigation/q_description.pdf

3.3 Apparatus

The testing room was equipped by 20 computers with eye-trackers and headphones, which enables multiple studies to be carried out at the same time with identical conditions. Participants were recorded by standalone Tobii X2-60 eye-tracker with 60 Hz sampling rate. 20 simultaneously running computers with OS Windows 8 was used in our setup [4]. The participants' display were 24 LCD. Referenced freedom of head movement was 30x15cm at the distance of 70cm.

3.4 Session Description

After welcoming and introducing study we performed eye-tracker calibration. Calibration of each participant was checked individually by a team member, only then the participants could start their session or re-calibrate. Recording part started by unrelated study with video stimuli, followed by our experiment: navigation tasks on live websites, getting familiar with static homepages and filling the questionnaire.

Scenario had the following purpose:

1. to collect reference eye-tracking data from the use of live websites (4 stimuli),
   - Task: to accomplish given task (find information, order product etc.). Task is finished by clicking on the correct link or choosing button "I do not know".
2. to collect eye-tracking data from the first encounter with the website homepage for short time (32 stimuli),
   - Task: familiarize with current website. 32 different website screenshots stimuli were presented in random order for 7 seconds. After each stimuli, participants were provided with form containing two questions ("what is the site about" and "what can you expect of it").The main role of two tasks after each screenshot was to motivate all participants to achieve a conscientious interaction. We limited the answers by the size of text input field. Only small number of responds contained information that participant was unable to answer (2,9% of responds).
3. to complete a web-navigation skill questionnaire (19 questions).

3.5 Participants

123 participants, took part in this exploratory study. Very few participants were not able to pass the calibration process or we were not able to pair questionnaire score with few gaze recordings. The average age in the sample was 25.29 (SD = 7.16, min = 19, max = 55). Most of the sample were females (n = 95),

Participant groups came from different environments. Different groups participated:

- students of pedagogy on bachelor grade (N = 45),
- students of pedagogy on master grade (N = 35),
- high school teachers (N = 28),
- IT faculty doctoral students and researchers (N = 15).

3.6 Data

Eyetracking Data The recorded gaze interaction duration on each stimuli across all participants was lower than expected 7 seconds, but stable (M= 5.66s, SD = 0.18s). The duration of participant's gaze interaction across all stimuli varies much more (M = 6.30s, SD = 1.62s).

Three unexpected situations happened when stimuli was displayed considerably more than 7 seconds (12s, 25s, 30s) due to software error. These recordings were cropped to standard 7 seconds in further analysis.

For analysis were only used cleaned gaze data where eye-tracker was able to estimate the Gaze Point position on the screen.

We defined 54 different areas of interest (AOI), among all stimuli, from which we picked 4 (logo, site options, footer, top navigation) with the highest occurrence among stimuli. Their intersection resulted in 21 websites, that we further used for classification.

We worked with gaze data processed into AOI transition matrix. We generated unique matrix for every participant on each stimuli. We use their fixation points to calculate transitions between AOI. As the result we got relatively sparse matrices with count of transitions in each cell, see Figure 2.

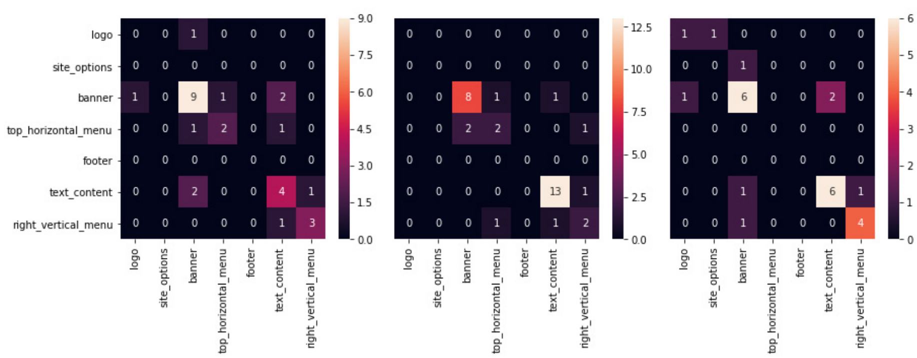

Fig. 2. Sample of three transitional matrices, for three different participants on the same stimuli. Number of transitions will become a classification feature.

Questionnaire Data Questions are not strongly correlating, as can be seen on Figure 3, so we should get comprehensive information about participant.

Although, overall success on four tasks was relatively high (M = 3.44, SD = 0.93). Success shown on participants ordered according questionnaire score differ for top 20% participants (M = 3.85, SD = 0.37) and bottom 20% (M = 3.5, SD = 0.95).

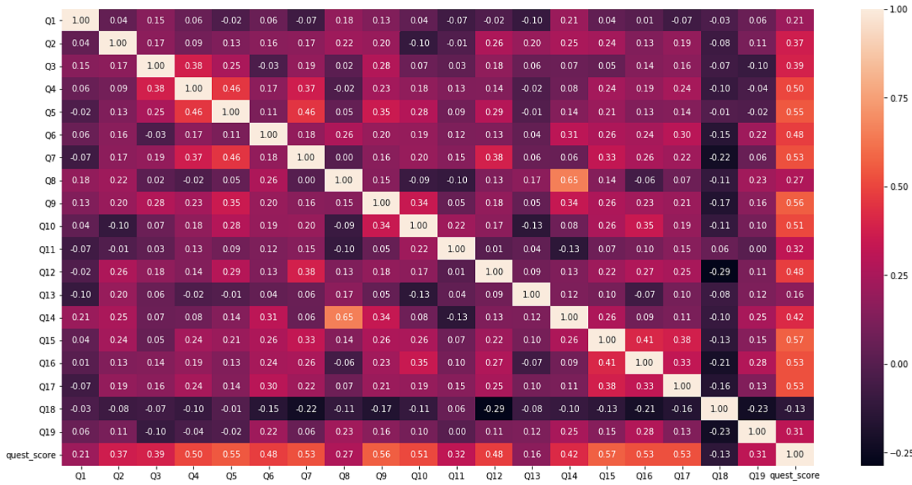

Fig. 3. Correlation matrix of 19 tasks based on 158 respondents

3.7 Normalization

Each cell on Figure 2 represents behavior, whether participant has transition between two specific areas. Each matrix was converted to 1-dimensional array and used for Support Vector Classification, therefore each matrix cell became a classification feature. These data are referred as *Raw* in Table 1.

In order to determine further relations, we used data normalization in two more testings. *Norm1* is used for labeling arrays where values are scaled to 0-1 but in each array solely - so this normalization is related for single participant. *Norm2* is used for labeling arrays where values are scaled to 0-1, but maximum value is calculated among all participants.

For classification was used SVC with k-fold cross-validation with 20 splits.

3.8 Results

From the wide range of participants (117) we picked extreme samples (20 best and 20 worst according to their web-navigation score). These were labeled by two classes for binary classification ("less skillful", "more skillful" participant). For every stimuli we got evenly balanced 40 instances.

Results of the classification for each stimuli separately can be seen in Table 1. Each row contains classification accuracy with standard deviation. Only few websites (stimuli 6, 8, 11, 12, 14) support the classification to our two classes. Almost in all cases, standard normalization approaches were not useful for classification.

For further analysis we see possible improvements in understanding the differences between stimuli that are useful for classification and that which are not useful, and therefore not suitable. Because even though they have similar

Table 1. Comparing classification results on each stimuli. Normalized data are not more useful for this type of task.

Stimuli	Raw (std)	Norm1 (std)	Norm2 (std)
0	0.42 (±0.40)	0.72 (±0.33)	0.42 (±0.43)
1	0.47 (±0.37)	0.03 (±0.11)	0.15 (±0.28)
2	0.28 (±0.37)	0.15 (±0.28)	0.03 (±0.11)
3	0.45 (±0.42)	0.47 (±0.37)	0.42 (±0.43)
4	0.25 (±0.34)	0.17 (±0.33)	0.25 (±0.34)
5	0.47 (±0.37)	0.35 (±0.36)	0.15 (±0.28)
6	0.60 (±0.41)	0.42 (±0.40)	0.20 (±0.37)
7	0.50 (±0.39)	0.55 (±0.35)	0.17 (±0.29)
8	0.70 (±0.33)	0.68 (±0.36)	0.30 (±0.43)
9	0.30 (±0.40)	0.28 (±0.40)	0.10 (±0.25)
10	0.45 (±0.42)	0.05 (±0.15)	0.15 (±0.28)
11	0.72 (±0.37)	0.60 (±0.41)	0.15 (±0.28)
12	0.60 (±0.41)	0.30 (±0.43)	0.25 (±0.40)
13	0.38 (±0.38)	0.45 (±0.42)	0.38 (±0.47)
14	0.72 (±0.33)	0.47 (±0.40)	0.10 (±0.20)
15	0.28 (±0.29)	0.38 (±0.38)	0.17 (±0.33)
16	0.55 (±0.38)	0.20 (±0.24)	0.10 (±0.20)
17	0.45 (±0.44)	0.35 (±0.39)	0.33 (±0.43)
18	0.50 (±0.42)	0.38 (±0.41)	0.17 (±0.29)
19	0.35 (±0.39)	0.35 (±0.42)	0.15 (±0.28)
20	0.40 (±0.34)	0.07 (±0.18)	0.17 (±0.33)

features and website elements, they have various usage and purpose. Possible solution for improving results would be to infer more general metrics, that could be used across all stimuli and will enlarge the testing sample.

4 Conclusions

In this paper, we provide new approach for estimating user web-navigation skill. We strive to determine user web-navigation skill from his interaction represented by eye-tracking metrics. We needed to take several steps to conduct this analysis. We performed exhaustive study on variety of participants in order to compare their web-navigation skill. Skill estimation was assessed by specialized self-evaluation questionnaire. Distribution of the questionnaire results suggests a realistic representation, additionally confirmed by the success in control tasks.

The overall results show that it is necessary to infer more user-related features to obtain better classification accuracy. The results of the transition matrix analysis revealed a need for sophisticated normalization of features among users that enter the classification. However, as we conducted the user study on multiple website homepages, we can see the results vary with each tested stimuli. For our classification we chose stimuli with four most common features to maintain a large number of stimuli. Equally important is to understand all the website elements relations and differences among stimuli.

Acknowledgement

This work was partially supported by the Slovak Research and Development Agency under the contracts No. APVV-15-0508, APVV SK-IL-RD-18-0004, grants No. VG 1/0725/19, VG 1/0409/17. The authors would like to thank for financial contribution from the STU Grant scheme for Support of Young Researchers.

References

1. Ala-Mutka, K.: Mapping digital competence: Towards a conceptual understanding. Sevilla: Institute for Prospective Technological Studies (2011)
2. Bardini, C.: Computer-based assessment of mathematics in pisa 2012. In: Assessing Mathematical Literacy, pp. 173–188. Springer (2015)
3. Belshaw, D..e.a.: Working towards a framework to understand the skills, competencies and literacies necessary to be a webmaker. https://wiki.mozilla.org/ Learning/WebLiteracyStandard/Legacy/WebLiteraciesWhitePaper (2013)
4. Bielikova, M., Konopka, M., Simko, J., Moro, R., Tvarozek, J., Hlavac, P., Kuric, E.: Eye-tracking en masse: Group user studies, lab infrastructure, and practices. Journal of Eye Movement Research **11**(3), 6 (2018)
5. Canada Microsoft: Attention spans. Tech. rep., Microsoft (2015), http://dl.motamem.org/microsoft-attention-spans-research-report.pdf
6. Eraslan, S., Yesilada, Y.: Patterns in eyetracking scanpaths and the affecting factors. J. Web Eng. **14**(5&6), 363–385 (2015)
7. Juvina, I., Van Oostendorp, H.: Individual differences and behavioral metrics involved in modeling web navigation. Universal Access in the Information Society **4**(3), 258–269 (2006)
8. Kubanyi, J., Hlavac, P., Simko, J., Bielikova, M.: Towards automated web navigation and search skill assessment: An eye-tracking study on the skill differences. In: 2018 13th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP). pp. 49–54. IEEE (2018)
9. Murano, P., Lomas, T.J.: Menu Positioning on Web Pages. Does it Matter? Tech. Rep. 4 (2015), www.ijacsa.thesai.org
10. OECD: Educational Research and Innovation Innovating Education and Educating for Innovation The Power of Digital Technologies and Skills: The Power of Digital Technologies and Skills. OECD Publishing (2016), https://books.google.sk/books?id=ylYhDQAAQBAJ
11. van Oostendorp, H., Aggarwal, S.: Modeling and supporting web-navigation. Journal of Interaction Science **3**(1), 3 (2015)
12. Pilgrim, C.J.: Website navigation tools: a decade of design trends 2002 to 2011. In: Proceedings of the Thirteenth Australasian User Interface Conference-Volume 126. pp. 3–10. Australian Computer Society, Inc. (2012)
13. Roberts, N.J.: Toward a literate future: Pairing graphic novels and traditional texts in the high school classroom. Ph.D. thesis, Colorado State University. Libraries (2012)
14. Rohrer, C.: When to use which user-experience research methods. Nielsen Norman Group (2014)
15. Şengel, E.: Usability level of a university web site. Procedia-Social and Behavioral Sciences **106**, 3246–3252 (2013)