

Visual walkthrough as a tool for utility assessment in a usability test

Kristiina Juurmaa
University of Helsinki
Institute of Behavioural Sciences
PO Box 9, FI-00014
kristiina.juurmaa@helsinki.fi

Janne Pitkänen
Aalto University School of Science
PO Box 15400, FI-00076
janne.p.pitkanen@aalto.fi

Sirpa Riihiahho,
Tuukka Kantola & Jani Mäkelä
Aalto University School of Science
PO Box 15400, FI-00076
firstname.lastname@aalto.fi

This paper presents a compact procedure for classifying the importance of elements in a user interface based on the visual walkthrough method. This method was used during a usability evaluation of an information service for healthcare professionals. Paper printouts were given to users who were asked to highlight the parts of the system they consider most important for them. This method proved to be a quick and useful way to understand which parts of complex user interfaces are the most important for users. In addition, heat maps were constructed based on these answers and they proved to be an easy way to visualise the results both for the evaluators and the different stakeholders. These heat maps could be formed right after the last test session, on the spot of the actual test session.

Visual walkthrough, heat map, discount usability, data visualisation, usability test.

1. INTRODUCTION

There is a remarkable gap between the theory and reality of usability testing. Nørgaard and Hornbæk (2006), for example, have described the specifics of this gap between academic research, *i.e.* what usability experts think they are doing, and how usability testing in real life is actually conducted. One major factor affecting this reality is time pressure. In real life projects, usability research is conducted at a fast pace and there is little time available for the analysis of results. Consequently, Nørgaard and Hornbæk (2006) conclude that new methods for rapid analysis should be developed and validated. Johannessen and Hornbæk (2013) also point out that there are only few methods for evaluating the utility of a system although plenty of methods are available for usability evaluation. This paper presents an approach dealing with both of these issues.

Currently, there are not many methods that allow rapid analysis of usability test results. Kjeldskov *et al.* (2004), for example, present an instant data analysis method that gives almost similar results as traditional video analysis with only 10% of the time required for video analysis. Instead of focusing only on the analysis phase of usability testing, we utilised a slightly modified testing method already in the phase of conducting the tests to streamline the analysis phase. In the spirit of discount usability (e.g. Nielsen 1989), this method is quick and

inexpensive since it does not require a functional prototype, but only printouts and coloured pencils. Additionally, it focuses on the utility of the system instead of mere usability.

Our preliminary interviews and usability inspections had indicated that the system suffered from considerable information overload, so one of our goals in the study was to filter the most relevant information from the service from the users' point of view. Additionally, we needed to present the results in an easily comprehensible and convincing manner to our customer. We had already planned a traditional usability test with interviews and questionnaires, and the participants for these tests were already recruited, so we had an opportunity to do an empirical study with some complementing method to get experiences on how well it meets these goals.

2. RELATED WORK

Our study relates to three issues in usability research, namely the visual walkthrough method for usability evaluation, the use of block, click and heat maps in visualising users' eye movements and foci, and the evaluation of utility as a part of usability evaluation. These are discussed briefly in the following subchapters.

2.1 Visual walkthrough

Nieminen and Koivunen (1995) introduce a visual walkthrough method to get information about users' perceptions and interpretations of the user interface and its components. The method is closely related to picture analysis of screen images described by Dehlholm (1992), as they both go through the screen several times getting deeper into details from the first impressions. Nieminen and Koivunen present the method as a complement to usability test so that it can be used before performing the test tasks or during the tasks as interesting components of the system come forth. If quantitative measures, such as performance time or number of errors, are of importance in the test, Riihiho (2000) recommends to conduct the visual walkthrough only after the test tasks to avoid biasing the results as the users may outline the system and its components differently if they concentrate on the user interface and its components before the tasks.

In visual walkthrough, test users are asked to explain before they start using the system what they see, what kind of elements and units they recognise, how they understand the symbols and terminology, and what they expect will be found behind the functional elements. The results can be used to evaluate how visible and understandable the elements of the user interface are. (Nieminen & Koivunen 1995)

2.2 Block, click and heat maps

Choros and Muskala (2009) describe click and heat maps that show which areas of a website are most frequently clicked on by the users. The clicks are presented as crosses. Heat maps that summarise the amount of clicks with different colours are generated based on the mouse clicking data. To make the data easier to understand, Choros and Muskala (2009) also introduce block maps in which user interface elements forming a group are defined as blocks and the data of usage is presented for each of these blocks on top of the user interface. This block map technique can be used for usability evaluation of a website, and the technique encourages to restructure the layout of the website under examination based on actual use to improve user satisfaction. (Choros & Muskala 2009)

2.3 Utility evaluation

Several studies have brought up the problem that only few usability evaluation methods take utility into account (e.g. Johannessen & Hornbæk 2013), and leave only little room for assessing usefulness,

value and evolving use of systems (e.g. Sengers & Gaver 2006, Greenberg & Buxton 2008). Too many evaluation methods and also usability tests focus on micro level measures, such as task completion times and accuracy, instead of macro level measures, such as learning and cognitively complex problem solving (Hornbæk 2006). As Cockton (2006) states, the quality in use is a component of user experience during interaction, but "outcomes and lasting impacts endure beyond interaction". Therefore, Cockton (2006) recommends to use self-reporting methods to enable assessments in the real world instead of merely during the interaction.

3. MODIFIED VISUAL WALKTHROUGH AND HEAT MAPS

Inspired by the above mentioned methods, we made a mixture of them to serve our goals, i.e., to help us prioritise the elements in the system we were evaluating. The system was an information intensive web service for healthcare professionals including several databases and separate web sites, each of them offering a large amount of information and links to other services. Figure 1 shows the main page of the service.



Figure 1: Main page of the healthcare information system evaluated in the study.

We had already planned a traditional usability test using thinking aloud method with predefined test tasks to reveal the most important usability problems of the service. With such an information intensive service, we also needed ways to identify the most important elements of the user interface to give them more visibility and room, and also the least significant elements that could be removed, hidden behind links or otherwise given less visibility. For these goals, we decided to combine elements both from the visual walkthrough method and the block maps to produce explicit data for

quick analysis and to support the detection of most important components of the web pages.

As the system to be evaluated was an information service for healthcare professionals, the test users needed to have at least some experience with medical issues users but not necessarily with the evaluated system. Although our group of participants was statistically low (n=6) and presented only one profession of the variety of user groups, the users were good representatives of their profession, some of them still being novices on medical issues and some having experience on the profession for several decades. Many of the test participants used the service every day, as it is a central source for healthcare related information in Finland.

In the visual walkthrough, we utilised paper printouts of different parts of the service to collect feedback from the users. These printouts were presented one by one to the users. Since the main page was already familiar with all the users, it could be walked through before the actual test tasks as a warm-up task to the test. The printouts from the other parts of the service, however, were addressed only after the test tasks to avoid possible changes in users' behaviour by forcing them to focus on the screens in more detailed level than usual. Along with the printouts, the users were given three highlighter pens to colour code the elements on the page with the following colours:

- **Green** for elements they use often or are of interest to them,
- **Yellow** for elements they do not use often or have not used but might need at some point,
- **Red** for elements that they have never used or that they find useless.

Given these instructions, the users marked the printouts while explaining reasons for their colourings. In case an element was left uncoloured, we asked the user to colour it if possible. One of these printouts is presented in Figure 2 along with a user's markings.

We did not specifically ask the users to think aloud while making the markings, but most of them did so, and thereby, provided valuable information about the relevance of the components and the conventions of use by the professionals. The printouts with the users' markings summarised these ratings in a way that was easy to analyse and visualise. As the participants were already familiar with the service, the order of the walkthroughs and the test tasks did not seem to have an effect on the users' performance. Still, most of the walkthroughs were made only after the tasks to avoid possible bias in performance.



Figure 2: An example of a user's markings.

4. RESULTS AND ANALYSIS

After finishing the evaluations we went through all the printouts and used them to create summative heat maps. Instead of eye-fixations and gaze movements, the heat maps that we generated represented average importance ratings of certain elements in the printouts. We kept the colour codes the same as in the users' markings, although this is opposite to the general use of heat maps. A gradient from green, through yellow, to red was used to illustrate the approximate averages from the most important to the least significant ones. Elements that received both red and green ratings from the users were coded with red-green stripes to illustrate the mixed opinions. An example of our heat map is presented in Figure 3.

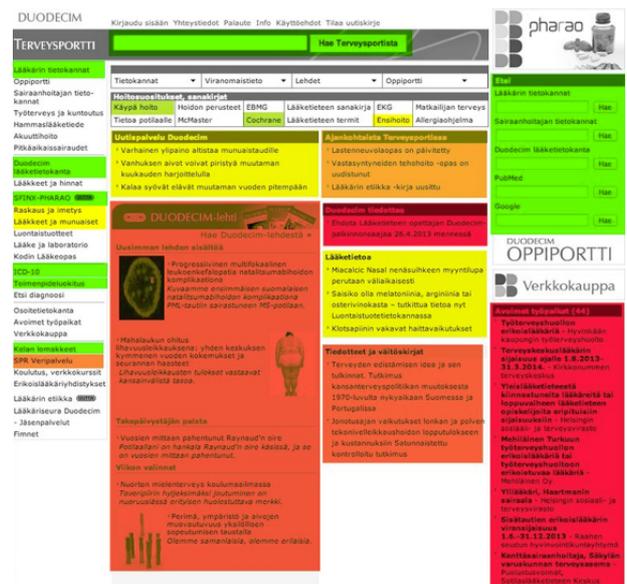


Figure 3: A heat map with average importance ratings.

Heat maps are commonly used also in eye tracking studies to visualise users' fixation patterns, *i.e.*, which parts the users look at and how long they look at each spot (e.g. Cutrell & Guan 2007). The use of eye-tracking has become quite general in usability testing, as well, and eye-gaze plots are commonly used to visualise the results in these studies (e.g. Eger *et al.* 2007, Freeman 2011). The visualisation method is thereby familiar to many customers. The heat maps and eye-gaze plots of eye tracking studies are usually generated with specific software analysing the data. The heat maps in our study were, however, generated manually, and did not require specific hardware or software.

Unlike the heat maps of eye-tracking studies, our heat maps did not show where the users looked at, but how the users valued the elements in the user interface, and how useful they considered the elements to be. Although eye-fixations may reflect the subjective salience of certain parts of a screen to the user in a given situation, it is also possible to fixate on a given target without understanding its meaning and even without being conscious of this fixation. While fixations are often utilised as a proxy for user's attention (e.g. Cuttrel & Guan 2007), fixations alone do not convey how well the user actually understands and values certain user interface elements. Therefore, our heat maps focused on the value instead of eye-fixations to filter the relevant parts from the service, and also assess the utility of the service and its user interface elements.

Our heat maps were used to analyse the current situation of the service as well as a way to convey findings to the stakeholders. Communicating findings with heat maps is effective because they are easily understood and enable an efficient way to summarise large amount of information. The simple procedure with printed screen shots also enable to quickly summarise the results after the last test session, as the experiences with the test users are still fresh in mind. In this case, it took us approximately ten minutes to generate a heat map from six users after a consistent criterion for forming an aggregated heat map was established.

Figure 4 illustrates an example of restructuring one of the pages providing a search utility to each specialised part of the portal. The results of visual walkthrough were applied as follows when the new layout of the web page was considered:

- **A** Search element (green) was moved into the middle and the topmost position within the content section of the page.
- **B** Introductions of the database contents and purpose (yellow) was kept in the middle and placed right after the search.

- **C** Database licensing information (red) placed at the bottom with mandatory parts only.
- **D** Listing of the contents (orange) was removed and put behind a link next to the search element.

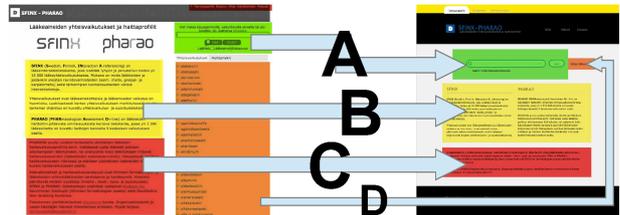


Figure 4: Restructuring of a web page including elements for (A) search, (B) introductions, (C) licensing information, and (D) contents listing.

The heat maps revealed both completely new information about the relevance and support of some user interface elements the stakeholders and fortified many of their intuitions about the service. Although some of the issues that our study revealed were at some level already known to the stakeholders, the heat maps transformed these issues into something more concrete. According to the stakeholders, the heat maps enabled us to communicate the pervasiveness and severity of the information overload, and also shake up the customer's image of their service which was slightly burdened by design decisions made a long time ago.

5. REQUIREMENTS FOR USE

The visual walkthrough method does not require a functional prototype or a finished system, as printouts or sketches of the user interface are enough for the walkthrough. The visualisation of the aggregated results as heat maps does not require specific tools either, but can be done with almost any photo editor or even with coloured pencils.

The use of the visual walkthrough method to prioritise the elements of a user interface does anyhow have some requirements on the test users' part. In our study, the test users were mostly experts on the domain – all having some experience on medical issues although some were still students – and most of them were already very familiar with the studied service. Therefore, it seemed rather effortless for the users to mark the relevant parts of the service, and to rule out the parts they considered useless. On that account, we recommend that the test users in visual walkthrough with utility assessment would be experts on the domain, and familiar with the tasks that the service is intended to assist, especially if

the system is for occupational use. This way, the test users have experience and knowledge on which to base their assessments of the usefulness and value of various user interface elements. With walk-up-and-use systems or other systems that are intended for anyone, the requirements are naturally less strict as the system and its relevant components need to be intuitive and visible for everyone. To avoid bias in users' performance, the use of visual walkthrough with utility rating is, nevertheless, recommended to be used only after the corresponding tasks if the method is used as a part of a usability test.

On the moderators' part, the method requires that all the participants are given similar instructions on the method. Therefore, the test moderators need to make sure the criteria for choosing colours is consistent between subjects, and the instructions and answers to potential questions are consistent. For example, it is possible that even an experienced user does not know the meaning of all of the elements, and wants to leave some parts without markings. On the other hand, a common comment from the test participants was roughly: "I don't need this element but maybe it's good to have it there". The moderators need to take a consistent stance (within-study and between-subjects) on how to instruct the test users on these sort of situations: whether to force the users to make an opinion or let them leave some parts unmarked, and whether to focus on the personal needs of the test users or also try to incorporate test user's impressions about an approximate user's hypothetical expectations and needs.

Concerning the analysis, it is essential to resolve how to present

- missing information (e.g. if some participant has not marked some specific element at all)
- the deviation of responses, and thus
- reliability of an aggregated colouring based on only a few test users markings.

6. PROS AND CONS

The method of colouring the elements in printouts is simple and inexpensive as it does not require a fully functional prototype or specific software or equipment. It is also fast, since the analysis of results can be done right after the tests in quite a straightforward procedure as long as a few basic criteria for combining the responses into a heat map has been formed. Although the users are not given much time to study the system, this method relies on the expertise that the users already have on the domain area, and the experience and

knowledge they have on the issues that are relevant to their work. In a way, making the markings gives the users a chance to self-report their experiences with the system – a prospect recommended for example by Cockton (2006) to evaluate the evolving use of various systems.

The method also enables a convincing and intuitive way to communicate the results with heat maps. Using a common representation for data gathering, analysis and further communication is both economic and fosters intuitive understanding of the data. The results are grounded in the same format of presentation as the actual service. We found that the heat maps were a great tool for forming and communicating a higher level general picture of the relevance of various users interface elements. This intuitiveness, however, can also have a downside. Without the understanding of how the data was gathered and how the heat maps were generated, their seeming intuitiveness can cause biased perceptions about the actual needs of the user group.

7. CONCLUSIONS AND DISCUSSION

When observing usability practitioners at their work, Nørgaard and Hornbæk (2006) identified a lack of systematic analysis of observations immediately after test sessions. Since our heat maps can be generated quite rapidly, on the spot, they could provide a common ground for the practitioners to form a mutual understanding of the data. More importantly, this could be done while observations from the test sessions are still fresh in memory.

Especially, if the method is used with expert users, it also gives the users an opportunity to assess the usefulness of various user interface elements, and thereby lays ground for the evaluation of the utility of the system as a whole. As the users comment various elements, the moderator has a chance to ask for clarifying questions on what sort of tasks the users normally do when entering the specific screen, and how relevant these tasks are in the users' work.

Even though this method provides an efficient way to summarise the results, they provide just one point of view to the use of the system without detailed instructions on how to proceed with the development work. In our study, the heat maps functioned most effectively as a thought provoker. The website we tested suffered from information overload, the method of colouring user interface elements helped in filtering the most relevant parts, and the heat maps helped in communicating the results to the stakeholders. Essentially, the heat maps did not tell about the usability of the system but more about the amount of irrelevant information and the utility of various elements.

There is also a downside to these quick-and-dirty but credible looking heat maps. Usability professionals need to be fully aware of how the data has been gathered and what it means. The combination of impressive visual heat maps and quick – at worst unsystematic – analysis of results might lead to deceptive impressions of the users' needs, which in turn might lead to even worse design decisions. Impressive visual heat maps may lead to overemphasising of this information over other findings and results of the usability studies. Thus, the effective intuitiveness of visual heat maps is a characteristic that needs to be acknowledged and used with care by the usability researchers. In our experience, these heat maps are at their best when used as tools for drawing attention and starting a constructive discussion on improving the usability and utility of the studied system.

8. REFERENCES

- Choros, K. and Muskala, M. (2009) Block map technique for the usability evaluation of a website. In: Nguyen, N.T., Kowalczyk, R. and Chen, S-M. (eds.) *Computational Collective Intelligence. Semantic Web, Social Networks and Multiagent Systems*. Springer Berlin Heidelberg. 743-751.
- Cockton, G. (2006) Designing worth is worth designing. In: *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles* (NordiCHI '06). ACM, New York, NY, USA. 165-174.
- Cutrell, E. and Guan, Z. (2007) What are you looking for? An eye-tracking study of information usage in web search. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '07). ACM, New York, NY, USA. 407-416.
- Dehlholm, F. (1992) Picture analysis of screen images. In: Leponiemi, J. (ed.) *Precedings of the NordDATA '92 conference*, Tampere, Finland, 15-18 June 1992. Pitky, Tampere. 353-359.
- Eger, N., Ball, L.J., Stevens, R. and Dodd, J. (2007) Cueing retrospective verbal reports in usability testing through eye-movement replay. In: *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...but not as we know it - Volume 1* (BCS-HCI '07). British Computer Society, Swinton, UK, UK. 129-137.
- Freeman, B. (2011) Triggered think-aloud protocol: Using eye tracking to improve usability test moderation. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11). ACM, New York, NY, USA. 1171-1174.
- Greenberg, S. and Buxton, B. (2008) Usability evaluation considered harmful (some of the time). In: *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (CHI '08). ACM, New York, NY, USA. 111-120.
- Hornbæk, K. (2006) Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies*, 64(2). 79-102.
- Johannessen, G. and Hornbæk, K. (2013) Must evaluation methods be about usability? Devising and assessing the utility inspection method. *Behaviour & Information Technology*, DOI= 10.1080/0144929X.2012.751708, Published online: 23 Jan 2013
- Kjeldskov, J., Skov, M.B. and Stage, J. (2004) Instant data analysis: conducting usability evaluations in a day. In: *Proceedings of the third Nordic conference on Human-computer interaction* (NordiCHI '04). ACM, New York, NY, USA. 233-240.
- Nielsen, J. (1989) Usability engineering at a discount. In: *Proceedings of the third international conference on human-computer interaction on Designing and using human-computer interfaces and knowledge based systems*. Gavriel Salvendy and Michael J. Smith (Eds.). Elsevier Science Inc., New York, NY, USA. 394-401.
- Nieminen, M.H.T. and Koivunen, M-R. (1995) Visual Walkthrough. In: Allen, G., Wilkinson, J. and Wright, P. (eds.) *HCI'95, People and Computers, Adjunct Proceedings*. University of Huddersfield, UK. 86-89.
- Nørgaard, M. and Hornbæk, K. (2006) What do usability evaluators do in practice? An explorative study of think-aloud testing. In: *Proceedings of the 6th conference on Designing Interactive systems*. ACM, New York, NY, USA. 209-218.
- Riihiaho, S. (2000) Experiences with usability evaluation methods. Licentiate's thesis. Helsinki University of Technology, Finland. Available at: http://www.soberit.hut.fi/~sri/Riihiaho_thesis.pdf.
- Sengers, P. and Gaver, B. (2006) Staying open to interpretation: Engaging multiple meanings in design and evaluation. In: *Proceedings of the 6th conference on Designing Interactive systems* (DIS '06). ACM, New York, NY, USA. 99-108.