A Usability Evaluation Method Applying AHP and Treemap Techniques

Toshiyuki Asahi, Teruya Ikegami, and Shin'ichi Fukuzumi

Common Platform Software Research Laboratories, NEC Corporation,
8916-47, Takayama-Cho, Ikoma, Nara 630-0101, Japan
t-asahi@bx.jp.nec.com

Abstract. This report proposes a visualization technique for checklist-based usability quantification methods. By applying the Treemap method, the hierarchical structure of checklists, weights of check items and evaluation results for target systems can be viewed at a glance. Effective support for usability analysis and its presentation tasks of usability evaluation results are expected. A prototype tool was implemented on a PC and experimental studies assuming actual usability evaluation tasks were conducted. The results indicate that the proposed method improves performance time of some typical tasks. Usability engineers gave higher subjective scores on the usefulness of the proposed method than that of printed table presentation.

Keywords: Usability quantification, checklist, Treemap, visualization, analytic hierarchy process, design tools.

1 Introduction

Usability quantification has been considered one of the key technologies for systematizing and spreading user-centered design and has been studied for some time [1]. Several methods such as 1) performance analysis/prediction, 2) subjective questionnaire, 3) checklist-based heuristic and 4) biological metric measurement/analysis have been developed as quantitative usability evaluation methods. Among those, checklist-based heuristics can provide early stage evaluation and synthetic rating even for software systems with many functions and can also pinpoint the user interface elements or interaction techniques to be improved when the checklist is designed carefully and goes into detail [2]. Therefore, this method is expected to fit for benchmarking and usability improvement of large-scale software applications. However, in addition to its relatively higher execution cost, it has a disadvantage in that it is not easy for evaluators or developers to grasp the evaluation result because of the huge number of check items and their complex (hierarchical) structure.

One of the solutions to this problem is to visualize the evaluation results properly. Yamada et al. [3] devised an explanation of the score distribution of target systems along with several check item categories by using segmented bar graphs. Although this method seemed to be successful in showing and comparing target systems' total scores,

J.A. Jacko (Ed.): Human-Computer Interaction, Part I, HCII 2009, LNCS 5610, pp. 195–203, 2009.

it was still hard to read out which and how check items contribute to (or lower) the synthetic scores since the detailed check items below the second hierarchy (88 items) are not presented on the graph.

In this paper, we propose to apply the well-known technique, Treemap, to visualize the evaluation result of checklist-based heuristics. One of the quantification techniques using checklist will be explained first; then the Treemap visualization techniques for that method will be demonstrated with the examples on the PC tool prototype. Section 4 describes the procedure and the results of the experimental study validating the effectiveness of the proposed method. Finally, we discuss the result and its implications for future study.

2 Checklist-Based Usability Quantification

Treemap visualization was applied on a checklist-based heuristic method developed by Ikegami et al. [4]. The checklist consists of 126 check items categorized into five sections and 18 sub-sections, which were extracted and arranged from various user interface (UI) guidelines, ISO standards [5][6] and consultation know-how.

Each check item is described from the viewpoints of UI components or system functions such as "Are titles attached to each window?" or "Are substitutive operations provided for double-click operations?" However, the usability score is expected to be given from user viewpoints (ex. learnability, memorability) when it is utilized in usability testing or benchmarking. Therefore, a weighting value was given to each checklist item for selected user viewpoints. Several techniques have been known for weight calculation such as expert opinion, task usage frequency or task importance, KANO model, entropy model, geometric mean of pair-wise comparison result in AHP (Analytic Hierarchy Process) and number of problems collected [7]. Ikegami et al. adopted the AHP method considering reliability and execution cost. A few usability experts applied a pair comparison method to give weighting value sets for every viewpoint. Thus, four weighting value sets, which correspond to "learnability," "memorability," "efficiency" and "low error rate" (selected by referring to [8]), were given to the check items, their sections and sub-sections (Fig. 1).

3 Usability Visualization Using Treemap

In order to visualize the evaluation result of the checklist effectively, we think the major requirements are:

1. Giving the overall view of the checklist including its hierarchical structure, and
2. Displaying detailed information such as each check item's weight and checking result.

In other words, both structural and quantitative information, and both overall and detailed information should be displayed in one view in a form easy to understand. We assumed the Treemap method [9] developed at the University of Maryland would fill these requirements and tried to apply it to the checklist heuristics. The following parts of this section describe detailed techniques of Treemap implementation.

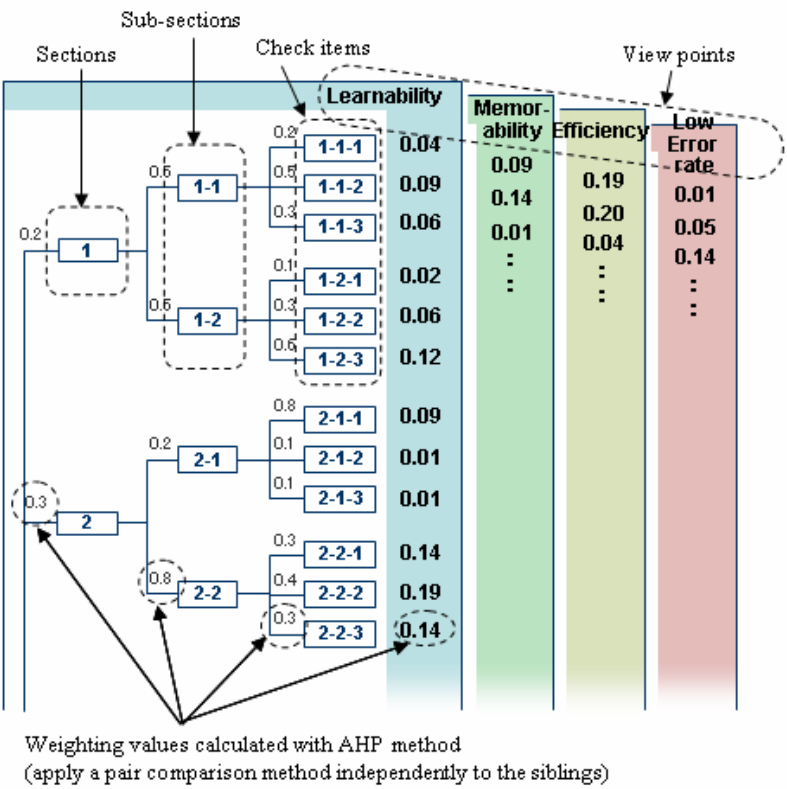
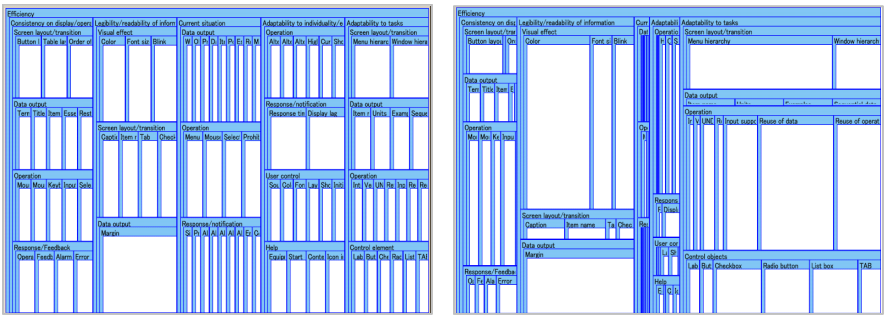


Fig. 1. Outline of checklist weighting method [4]

3.1 Visualize the Checklist

Treemap techniques were applied to the checklist introduced in Section 2. Figure 2 shows examples of checklist visualization, (a) without weights (assuming all sibling items or sections have equal weighting value), and (b) with the weight set given for the "efficiency" viewpoint. The "slice and dice" and "offset" techniques introduced in reference [9] were adapted to utilize the display area effectively and to present check item structure and titles clearly (full titles of check items appear by simple mouse-over operation). It is obvious that the structure and weight distribution can be viewed in a given display area even for a checklist consisting of more than 100 items. By providing four sets of weighting values and the corresponding map, each of which represents one of the four viewpoints, the entire checklists can be visualized. Although it is quite easy to merge them into one map by adding one more level of hierarchy, we chose to show them separately because the merged Treemap looked too busy and there seemed to be less need for viewing four viewpoints simultaneously.



(a) No weight (b) With a set of weights for "Efficiency"

Fig. 2. Checklist visualization

3.2 Visualize the Checking Result

Two large-scale Web application systems (tentatively named “A” and “B”) were evaluated along with the checklist, and the result was displayed as shown in Fig. 3. The preceding study [10] and Ikegame proposed to do checklist evaluation with yes (meaning the target user interface satisfies the corresponding check item) or no judgment for every check item in order to minimize the effect of individual difference. In the example of Fig. 3, colored areas (yellow for A, red for B) mean “yes” and black areas represent “no” judgment. The total scores, which are the summations of colored areas, are displayed as the bar chart below the map.

Intuitively, it is quite easy to read out which check items have bigger influence on the total score, or why system B gets a higher score than A from the viewpoint of

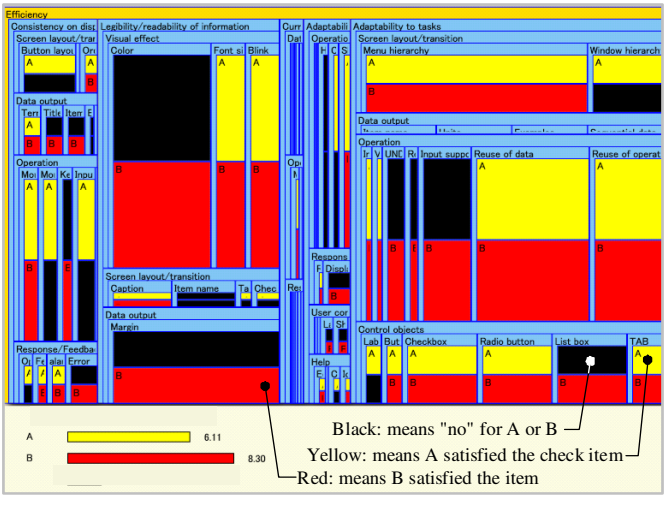


Fig. 3. Visualization of checking result with a bar chart

efficiency, for instance. On the other hand, it is hard to explore the checking result for the low-weighted items. (However, this problem has been partly solved by implementing the zooming function, which enables the zooming of any node (rectangle) out to the base rectangle area with its subsidiary nodes.)

4 Experimental Study

In the previous sections, Treemap visualization for a checklist-based quantification method was proposed. The simulated map seems to be useful for usability analysis tasks and for reporting their results. In this section, we try to validate its usefulness experimentally assuming actual data analysis tasks in usability consultation.

4.1 Experimental Design

We assumed two kinds of situations in which Treemap visualization would be helpful: 1) usability engineers analyze the checking result, and roughly estimate the usability level of the target systems or detect where they should be changed for effective usability improvement, and 2) usability engineers or consultants explain the analysis result to developers of targeted systems or clients of the consultation. The tasks in the experiment were designed assuming situation (1).

Outline. The Treemap on the PC screen (such as Fig. 2) and the table forms printed on paper sheets, both of which represent the checklist items, their weighting values, and checking results for two systems (A, B), were displayed to subjects. Each subject executed tasks assuming situation (1). Task completion time was measured by an experimenter along with the correctness of the answers. Subjective questionnaires were applied after completing all tasks.

Table 1. Experimental tasks (task set TA)

	Task description	Meaning of tasks
a	From the "learnability" viewpoint, what is the most significant category?	Understand which category is supposed significant in each viewpoint.
b	Select all viewpoints in which system A's total score is higher than that of B.	Grasp target systems' usability feature roughly.
c	From the "low error rate" viewpoint, what is the most important check item?	Understand which check item is most influential on the final score.
d	From the "low error rate" viewpoint, what is the least important check item?	Estimate which check item is not so critical.
e	From the "memorability" viewpoint, which system's score is higher?	Compare overall scores among target systems.
f	From the "memorability" viewpoint, which check item should be improved to raise system B's score most effectively?	Identify where to improve for raising usability score most effectively.
g	Count up the number of check items in sub-category "data output," where only B is OK.	Compare system usability for certain (restricted) usability aspect.

Participants. Nine usability engineers with usability testing or consulting experience ranging from 20-40 years old participated in the experiment. They were asked to do tasks supposing they were in a consulting situation.

Tasks. Seven simple tasks supposed to be typical in actual usability analysis work were selected as shown in Table 1. In addition to the seven tasks (task set "TA"), similar tasks (doing the same thing using another data viewpoint) a' - g' were also prepared as task set "TB." Each subject tried TA first, then tried TB. Half of them used table data for TA and Treemap data for TB, and the other half executed TA with the Treemap data and TB with the tables. This order is considered to minimize the training effect possibly appearing in the performance time. Just after completing TA and TB, all subjects were asked to check the seven subjective questionnaires. All subjects could complete all tasks and questionnaire responses in 40-60 min without any serious problem.

4.2 Results

Figure 4 shows mean and standard deviation of task completion time and number of correctly executed tasks. In five of seven tasks (except for d and e), completion times on the map seem shorter than those on the table data when simply comparing with mean value. ANOVA shows significant difference in task c ($t=2.67$, $P<0.02$) and in task f ($t=4.94$, $P<0.01$).

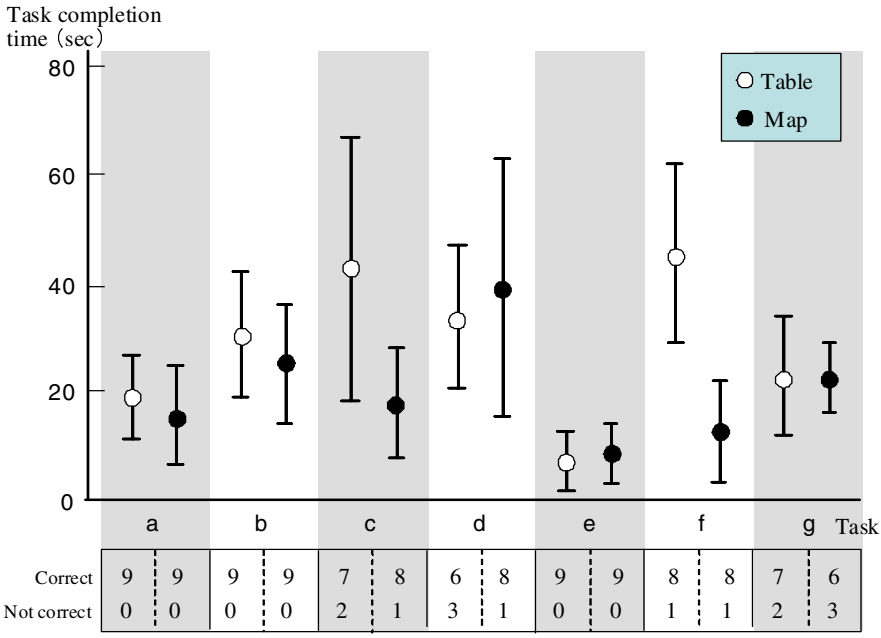


Fig. 4. Task completion times and numbers of tasks correctly executed or not

Figure 5 shows the result of subjective rating for seven questionnaires. As for the overall impression, participants (usability engineers) gave higher scores to table form for "simplicity," to map form for "comfortableness," and the same score for "easiness to understand." As for the readability of check item weights, higher scores were given to Treemap. (As for the check item structure, the same scores were given.) Also, participants tended to feel the Treemap was more useful for both usability analysis and presentation tasks.

4.3 Discussion of Experimental Result

From this experiment only, we could not reach a clear conclusion because the number of subjects was not sufficient for strict statistical analysis. One of the reasons was that participants were screened according to their experience as usability engineers. However, some implications or tendencies about the usefulness of the proposed method can be extracted.

The experimental results indicate the Treemap presentation should:

…be useful in data analysis tasks for usability consultation. We are paying attention to the result that significant task performance improvement was seen in tasks c and f. These tasks are to "select the most important check item (c)" and "pinpoint the check item for improving overall score most effectively." In many cases, usability engineers need to examine heavily weighted check items prior to others for improving target system usability effectively and promptly. Tasks c and f were designed with the intention of checking the adaptability to this requirement. Effectiveness and usefulness of the Treemap method are highly expected since significant improvement was observed in both tasks.

…not be suitable for examining low-weight check items. Although it was not statistically significant, performance time on tables was higher than that on Treemap (approximately 10% shorter mean time) in task d (select the lowest weight check item). When there are a lot of check items, sometimes weights are set smaller than 1/1000. In Treemap representation, where the weights are displayed as areas of rectangles, it is often hard to read out these areas exactly. In real situations, there are not many cases that require elaborate examination of low weight check items, and Treemap may not support such tasks sufficiently. (If engineers are accustomed to using the zooming function, the task completion time will be greatly improved, though.)

…not have clear advantages in "rough examination" tasks. As for tasks a, b, e and g, significant difference in task completion time was not observed, contrary to the authors' expectations. These tasks include roughly comparing target system characteristics, such as selecting viewpoints in which system A's score is higher. In these tasks, participants did not have to search for items by comparing weighting values, and they could complete the task easily just by reading numerical values in the tables.

…be useful by usability engineers. Participants gave higher subjective scores on Treemap presentation for two questionnaires about usefulness. Taking into consideration that every participant was a beginner in using Treemap, this indicates their expectations of Treemap are considerably high. More experience with Treemap and additional software functions that support usability analysis tasks will raise the subjective rating further.

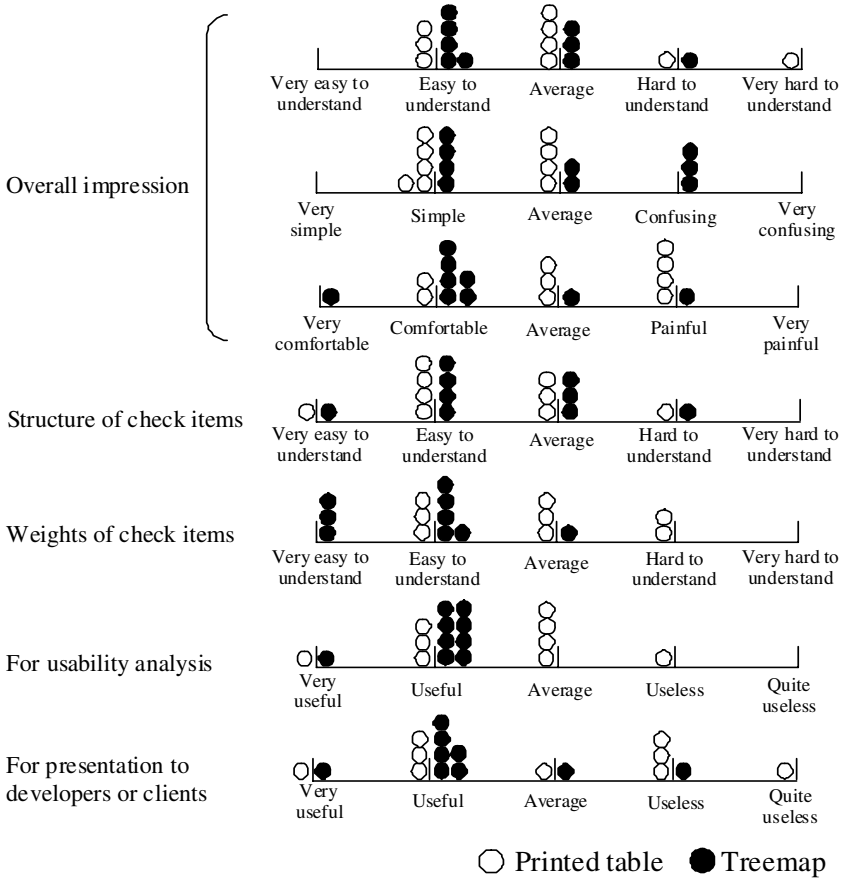


Fig. 5. Result of subjective questionnaires

5 Concluding Remarks and Future Work

In this report, a method applying the Treemap visualization to checklist-based heuristics was proposed. In some tasks that were intended to simulate usability analysis, improvement of task completion time was observed. Subjective ratings were higher than table form presentation on the usefulness for both data analysis and presentation tasks. Users' (usability engineers') experience and additional software functions will raise this score further.

We think the following three issues should be overcome for making the checklist-based quantification method practical and widespread.

1. Provide theoretical and reliable bases to the quantification method.
2. Enable objective evaluation by eliminating/minimizing the score difference caused by individual skills or impressions.
3. Provide practical and useful tools for evaluation and analysis tasks.

As for issue (2), Ikegami has claimed it will be achieved to some extent by designing check items and their terms elaborately and tuning them iteratively through experimental studies [4].

The proposed method was intended to contribute to resolving issue (3) and some effect was shown in the experimental studies. Of course, we will need many more functions and tools to support real usability analysis tasks. We are considering a tool for weight assignment by using the Treemap as a data input tool [11].

In order to create breakthroughs on issue (1), we need to present scientific bases of quantification, but it is hard to accomplish with short-term research. Although preceding studies [2][3][4][10] have tried to add reliability by adopting well-used guidelines or regulations, they have not become widespread as established methods. We think we need to ensure the validity of the method by developing or refining user cognitive/behavioral models for the checklist-based quantification as with the GOMS and KLM models for performance prediction with CogTool. Both scientific research and practical field activities should be merged harmoniously to develop reliable and systematic methodologies.

References

1. Hirasawa, N.: Usability Engineering for Software Development. IPSJ Magazine 44(2), 136–144 (2003)
2. Smith, S.L., Mosier, J.N.: A Design Evaluation Checklist for User-System Interface Software, Technical Report ESD-TR-84-358, MITRE, MA (1984)
3. Yamada, S., Tsuchiya, K.: A Study of Usability Evaluation of PC – Discussion of Model on PCs. The Japanese Journal of Ergonomics 32, 350–351 (1996)
4. Ikegame, T., Okada, H.: Toward Quantification of Usability. NEC Technical Journal 3(2), 53–56 (2008)
5. ISO 9241-10: Ergonomic requirements for office work with VDTs –Dialog principles (1996)
6. ISO 9241-12: Ergonomic requirements for office work with VDTs –Presentation of information (1998)
7. Ham, D.-H., Heo, J., Fossick, P., Wong, W., Park, S.-H., Song, C., Bradley, M.: Model-based approaches to quantifying the usability of mobile phones. In: Jacko, J.A. (ed.) HCI 2007. LNCS, vol. 4551, pp. 288–297. Springer, Heidelberg (2007)
8. Nielsen, J.: Usability Engineering. Academic Press, London (1993)
9. Johnson, B., Shneiderman, B.: Tree-Maps: A Space-Filling Approach to the Visualization of Hierarchical Information Structures. In: Proc. of the 2nd Conference on Visualization 1991, pp. 284–291 (1991)
10. Kato, S., Horie, K., Ogawa, K., Kimura, S.: A Human Interface Design Checklist and Its Effectiveness. Transaction of Information Processing Society of Japan 36(1), 61–69 (1995)
11. Asahi, T., Turo, D., Shneiderman, B.: Using Treemaps to Visualize the Analytic Hierarchy Process. Information Systems Research 6(4), 357–375 (1995)