

Experimental Evaluation of Usability and Accessibility of Heading Elements

Takayuki Watanabe
Tokyo Woman's Christian University
2-6-1 Zenpukuzi, Suginami-ku
Tokyo, 167-8585, JAPAN

nabe@lab.twcu.ac.jp

ABSTRACT

Task completion times of sighted and blind users were measured with two kinds of Web sites: sites marked up appropriately with heading elements and sites with the same visual appearance but with no heading elements marked up. The experiment was carried out with user agents that could navigate through heading elements. The results showed that 1) task completion time was reduced by as much as one half with marked up heading elements, 2) the benefits of markup on task completion time were greater for blind users, and 3) the overall difference in response time between sighted and blind users diminished with sites that were appropriately marked up.

Categories and Subject Descriptors

H.5.4 [Hypertext/Hypermedia]: Navigation, H.5.1 [Multimedia Information Systems]: Hypertext navigation and maps, K.4.2 [Social Issues]: Assistive technologies for persons with disabilities, I.7.2 [Document Preparation]: Hypertext/hypermedia, Markup languages, H.5.2 [User Interfaces]: Evaluation/methodology, Standardization

General Terms

Measurement, Documentation, Performance, Experimentation, Human Factors, Standardization, Languages.

Keywords

Web, accessibility, structure, heading element, usability, blind

1. INTRODUCTION

Web content accessibility guidelines such as WCAG and JIS X 8341-3 require content developers to markup the content structure of Web pages appropriately. For example, WCAG 1.0 [4] requires "Checkpoint 3.5: Use header elements to convey document structure and use them according to specification. (Priority 2)." In Japan, JIS X 8341-3 [6], Japan Industrial Standard of "Guideline for older persons and persons with disabilities - information and communications equipment,

software and services - Part3: Web Content", requires "Clause 5.2 a) Web content shall define document structure using heading, paragraph, list, and other elements."

The requirement for structure markup is well explained in the WCAG 2.0 working draft [1]. Guideline 1.3 of WCAG 2.0 requires designers and developers to "Ensure that information and structure can be separated from presentation." The techniques to address the Success Criteria 1.3.1 of this guideline are listed in the document "Understanding of WCAG 2.0" [2]. The "How to Meet Success Criterion 1.3.1" section of the Understanding document lists the following HTML techniques to markup content structure: table elements, caption elements, label elements, fieldset and legend elements, list elements (ol, ul, and dl), h1-h6 elements, p elements, summary attributes, scope attributes, id and header attributes, and title attributes. Specific techniques are explained in the "Techniques for WCAG 2.0" document [3]. For example, "H42: Using h1-h6 to identify headings" of this document explains the objective, and provides examples of implementing this guideline.

Heading elements can be used to show the hierarchical structure of the content. They also show the topic for the section that follows the heading element. Sighted users can use heading elements as a visual cue for distinguishing between content blocks. Visually disabled users can also use them if the topic is properly marked up with heading elements and if their user agents, including Web browsers and assistive technologies, have functions that can make use of the marked up heading elements.

Web content in the real world, however, is generally not properly marked up. For instance, in Web sites for Japanese local governments such as Tokyo and Osaka, only 7 out of 10 Web sites are properly marked up with heading elements. Two out of 10 Web sites are not marked up with heading elements at all. As for Web sites of major Japanese private companies, only 1 out of 10 Web sites is properly marked up with heading elements. Five out of 10 Japanese Web sites are not marked up with heading elements at all [7]. Other examples are CNN¹ and the Washington Post². As of this writing, the CNN site used only one h2 element and the Washington Post site did not use any heading elements.

A further problem is that user agents generally do not have much capability to use content structure. For example, major Web browsers such as Internet Explorer and Firefox have no functions

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

W4A2007 -Technical Paper, May 07-08, 2007, Banff, Canada.
Co-Located with the 16th International World Wide Web Conference.
Copyright 2007 ACM 1-59593-590-8/06/0010 ...\$5.00.

¹ <http://us.cnn.com/>

² <http://www.washingtonpost.com/>

to navigate through heading elements or list heading elements to make a table of contents. Additionally, in the Japanese context, major Japanese screen readers do not have functions that utilize structure markup [8,9,10].

Although relevant guidelines require the markup of content structure, in practice, relatively little Web site content is appropriately marked up. Even when sites are marked up, user agents tend not to have the functions needed to take advantage of the mark up. Why is this? What prevents the effective use of content structure in the real world? This author assumes that this unfortunate state of affairs continues to hold, among the other reasons, because there is insufficient evidence to show that well marked up content structure improves usability and accessibility. Thus, the primary objective of the current paper is to demonstrate how content that is well marked up improves usability and accessibility. More specifically, the current paper aims to investigate the following questions:

- Q1. How much (if at all) is the task completion time reduced by marking up heading elements?
- Q2. Do the benefits to (shorter) task completion time of marked up heading elements differ between sighted and blind users?
- Q3. Is the difference in task completion time between sighted and blind users decreased when structure markup is both available and utilized by user agents?

In this paper, experiments were carried out with two kinds of Web sites: one type of site was appropriately marked up with heading elements, while the other type was not marked up with heading elements but had the same visual appearance. Task completion time was measured for both type of site. Task completion times were collected on these sites for both sighted and blind users.

2. METHODS

2.1 Components of Web Accessibility

As shown in “Essential Components of Web accessibility” [5], various components affect Web accessibility. They are Web content, user agents, authoring tool, evaluation tool, guidelines, Web techniques, content developers, tool developers, and users. Evaluations of Web accessibility must consider these components. In order to examine the effect of structure markup, more specifically heading elements, all other components must be accessible. Thus, the experiment reported below utilized user agents (Web browsers and screen readers) with enough capability to utilize marked up Web structure, skilled users who know how to use the user agents’ functions, and accessible content (aside from the issue of heading elements which was the experimental manipulation used in the study).

2.2 Subjects

There were two groups of subjects: sighted and blind subjects. There were 16 sighted subjects and 4 blind subjects in the study.

The sighted subjects were all female university students between 20 and 22 years of age. They reported using computers and the Web almost everyday. Every experiment was carried out using a PC provided by the experimenter.

The blind subjects were all male ranging between 30 and 34 years of age. Each of them had either been born blind or lost his sight soon after birth. All of them had used JAWS for several years

prior to participating in the study and they reported using the Web almost everyday. Experiments were carried out either on the subject’s laptop computer, or on a computer provided by the experimenter.

2.3 User Agents

Internet Explorer and Firefox, at the time of this writing, did not have functions that took advantage of heading elements. Therefore, for sighted subjects, the “Document Map³” extension of Firefox was used.

As shown in Figure 1a, which is shown in the last page, the Document Map displays the content’s heading structure in the sidebar of Firefox. A user can move to any heading element of the content by clicking that heading in the sidebar, allowing rapid navigation between content elements. In case where content is not marked up with heading elements, the Document Map displays nothing (Figure 1b, shown in the last page).

The experiment with blind subjects was carried out using Internet Explore and the Japanese version of either JAWS 6.2 or 7.1. JAWS has a lot of functions that utilize heading elements. JAWS reads aloud how many heading elements exist in the content when loading the page. JAWS has heading navigation functions. Users can also skip to the next or previous heading element of the specified level, h1 to h6, by pressing a key.

2.4 Experimental Design

Prior to the experiment, a pilot study was carried out. In the pilot study, a strong learning effect was found, making it difficult to compare the two types of sites (with markup vs. without markup). In addition, large individual differences in task completion time were found between the subjects.

The experiment was mixed design study with one within-subjects factor of site type (marked up or not marked up/structured) and one between-subjects factor of subject (participant) type (sighted vs. blind). Two structured and two unstructured sites were used, yielding four sites in total.

In order to control for order effects, a half of the subjects started the experiment with the structured sites first (Order 1) and the other half started the experiment with the unstructured sites first (Order 2). The structured sites were marked up with heading elements, whereas the unstructured sites were not marked up with any heading elements.

2.5 Web sites

The experiment was carried out by instructing the subject to accomplish the given tasks by using the Web site. Figure 1 (attached at the end of this paper) shows part of one site A: Figure 1a shows the structured version of the site (Ao), and Figure 1b shows the unstructured version (Ax) of the same site. CSS was used to make the visual appearance of both versions of the site look the same. Except for not marking up heading elements, the sites were all accessible.

The common theme of the four sites was “Summer Dish Recipes⁴.” The sites had four h2 level heading elements which

³ <https://addons.mozilla.org/firefox/475/>

⁴ The experiment was carried out with Japanese Web pages. Content of the experimental sites were translated into English in this paper.

showed the category name of recipes: ingredients, purpose of the dishes, cooking time, and calories. As shown in Figure 2, these categories were divided into subcategories of h3 levels and sometimes included h4 level subcategories that were nested beneath h3 items.

- h1: Summer Dish Recipes
 - h2: Select from ingredients
 - h3: Vegetable ----- 4 h4 level subcategories
 - h3: Fish and shellfish -- 3 h4 level subcategories
 - h3: Meat ----- 3 h4 level subcategories
 - h3: etc ----- 2 h4 level subcategories
 - h2: Select by purpose
 - h3: Main dish
 - h3: Sub dish ----- 3 h4 level subcategories
 - h3: a la carte
 - h2: Select by cooking time
 - h3: 5 min.
 - h3: 10 min.
 - h3: 11 to 20 min.
 - h3: Over 21 min.
 - h2: Select by calories
 - h3: less than 100 kcal
 - h3: 100 kcal range
 - h3: 200 kcal range
 - h3: 300 kcal range
 - h3: 400 kcal range
 - h3: 500 kcal range

Figure 2. Hierarchical structure of site A

The site had 27 links to other recipe pages, which showed details of the recipe such as ingredients, cooking time and calories. The links were classified into the appropriate subcategories of the h3 or h4 level. As shown in Table 1, the differences among the four sites differed in terms of the ordering of h2 level categories. Thus, subjects could easily grasp the mental image of the site but could not use the memory of a previous site to navigate because the order of categories was different between the sites.

Table 1. Category order for four sites

	Site A	Site B	Site C	Site D
1 st h2	ingredients	calories	cooking time	purpose of the dishes
2 nd h2	purpose of the dishes	cooking time	calories	ingredients
3 rd h2	cooking time	purpose of the dishes	ingredients	calories
4 th h2	calories	ingredients	purpose of the dishes	cooking time

In the experiment, subjects investigated the top page (site A, B, C, or D) to find the candidate recipes and, by selecting a link in the top page, then went to a recipe page to look for more details.

For the unstructured sites, the category and subcategory names of the top page and any content in every recipe pages were not marked up with heading elements.

2.6 Tasks

The same four tasks (described below) were carried out for each site. However, some task details changed according to the arrangement of content on the sites, as noted below.

- Task 1: Yes or no question. A subject was required to answer if a particular category existed in the h3 level of the top page. (Except for site C, the answer was yes. While the answer to the question for site D was also yes, the actual location of the answer was nested within the h3 category at the h4 level.)
- Task 2: A subject was required to study 8 or 9 recipes which were nested within the h3 or h4 levels and then choose one of the recipes that matched a set of target conditions. This task required subjects to go to the recipe pages many times.
- Task 3: A subject was required to choose three recipes which matched a set of conditions posted in the question for the task. For instance, “Find three recipes that have fewer calories, and where the cooking time is between 11 to 20 minutes.” The three recipes to be found for this task were categorized in the same h3 level subcategories except for site C in which recipes were categorized in two h4 level subcategories. This task also required subjects to refer to the recipe pages.
- Task 4: A subject was required to count number of recipes that matched each of two conditions and then indicate which of the conditions was matched by more recipes. In this task subjects generally had to look at h4 level subcategories associated with different h2 level categories. (However, for site D, this task did not require looking at h4 level subcategories.)

2.7 Data Recording

Every screen encountered by subjects was recorded either by video camera or else by using screen-recording software. Twelve out of the 16 sighted subjects and 2 of the 4 blind subjects were also recorded with key logging software. The experimenter also took notes during the session so that they could be referred to later in interpreting the results.

2.8 Procedure

For both the sighted and blind groups, half of the subjects saw the structured site first and the other half saw the unstructured site first.

After collecting demographic information, the experimental session began with training on how to perform the task. Subjects who were recorded with the key logging software were also asked to press the X key (the W key for JAWS user) before starting the task, and to switch to the answer sheet by pressing the ALT + Tab key as soon as they found the answer to the question. These key events were used to analyze the task completion time.

The instruction described above was followed by a practice session. During the practice session, subjects were exposed to the features and interface that they used subsequently in the experimental tasks.

After the practice session, the site A was displayed to the subject. Subjects carried out the assigned tasks on each of the four tasks. After the 4th task seen by each subject was completed, 7-item questionnaires were administrated which evaluated 1) how easy the structure of the content was to understand, 2) how easy it was to operate the tools provided, and 3) how easy it was to find the requested information. After completing the questionnaire, a closing interview was carried out to get further feedback about thoughts that subjects had about the tasks and the way they carried them out.

The same procedures were carried out for the site B, C, and D.

3. RESULTS

In this paper, last two sites, site C and D, of four sites were used for analysis because the subjects were more accustomed to the site and tasks in these sites.

3.1 Analysis of Task Completion Time

After the experiment, subject actions such as task start and task completion, and in many cases, scroll, jump, click, key press, and page (window) name were recorded with a timestamp by experimenters. When possible this data was extracted from the output of the key logging software.

Task completion time was calculated as the time difference between the start of the task and the end of the task. Some tasks were evaluated as “not completed” because the answer was wrong.

3.2 Sighted Subjects

Two-way ANOVA was carried out to test if task completion time differed between structured and unstructured sites. In this analysis, two within-subject factors were considered: structured vs. unstructured and task ordering (first, second, etc.). The analysis showed that task completion time was affected by the structure factor, $F(1, 13)=10.286$, $p=0.007$. Task was also significant factor, $F(1.570, 20.416)=54.743$, $p=0.000$. There was no interaction between the structure factor and the type of task ($F<1$).

Figures 3a and 3b show the average task completion time for sighted subjects under two orderings. Standard deviation is also shown in these Figures. These Figures suggests that task completion time of structured site is shorter than that of unstructured site regardless of site order or kind of tasks.

The reduction in task completion time due to structuring of the site was evaluated as the ratio of the time for the structured site (Co or Do) to the time for the unstructured site (Cx or Dx). This “structure speed-up” ratio was calculated separately for each subject. Table 2 shows the average ratio of each task.

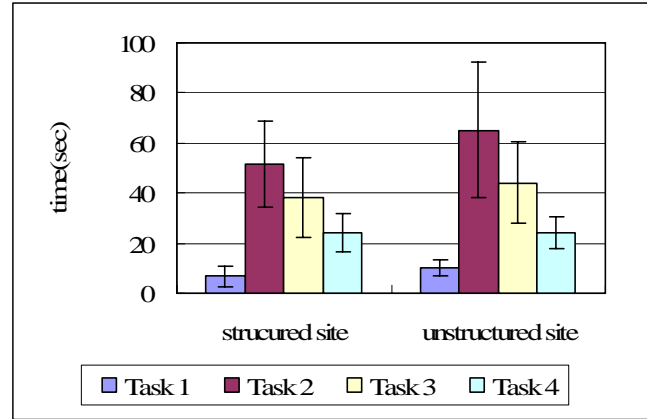


Figure 3a. Average task completion time and standard deviation for sighted subjects in Ordering 1 (users first accessed structured and then unstructured site)

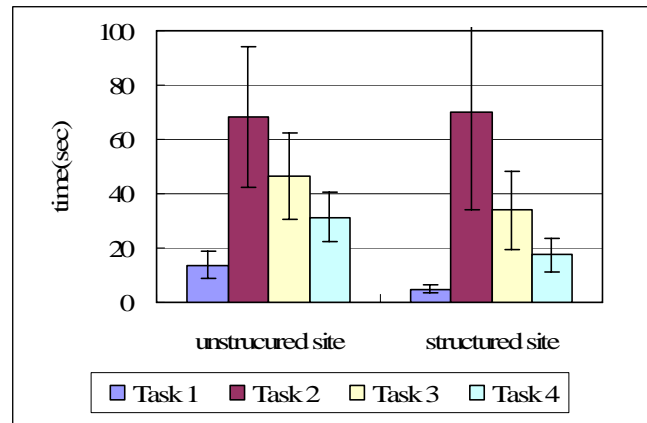


Figure 3b. Average task completion time and standard deviation for sighted subjects in Ordering 2 (users first accessed unstructured and then structured site)

Table 2. Average “structure speed-up” ratios for sighted subjects

	speed-up ratio ± SD
Task 1	0.49 ± 0.22
Task 2	0.90 ± 0.28
Task 3	0.80 ± 0.21
Task 4	0.81 ± 0.35

3.3 Blind Subjects

Two-way ANOVA was carried out in the same way as in the analysis of sighted subjects, yielding the same pattern of results as found for the sighted subjects. The analysis showed that task completion time was affected by the structure factor, $F(1, 3)=33.371$, $p=0.010$. Task was also significant factor, $F(3, 9)=10.368$, $p=0.003^5$. There was no interaction between the two factors ($F<1$).

Figures 4a and 4b show the average task completion time for blind subjects for the two orderings. Standard deviation could not be calculated because each data was the average of two samples. Thus, the sample means should be interpreted cautiously. These Figures suggest that task completion time of structured site is shorter than that of unstructured site regardless of site order or kind of tasks. These Figures also suggests that the structuring speed-up ratio for task 4 is larger than that of other tasks.

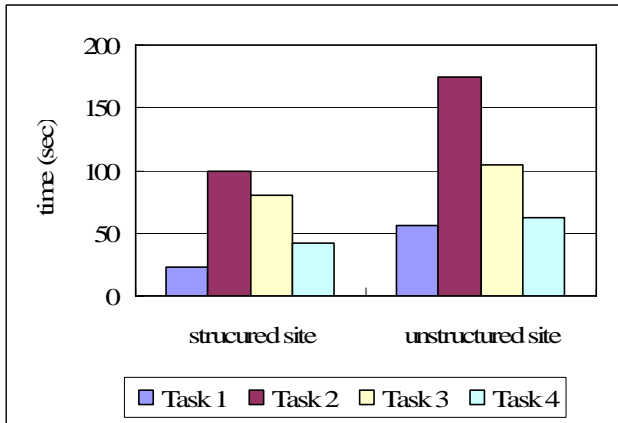


Figure 4a. Average task completion time and standard deviation for blind subjects in Ordering 1 (users first accessed structured and then unstructured site)

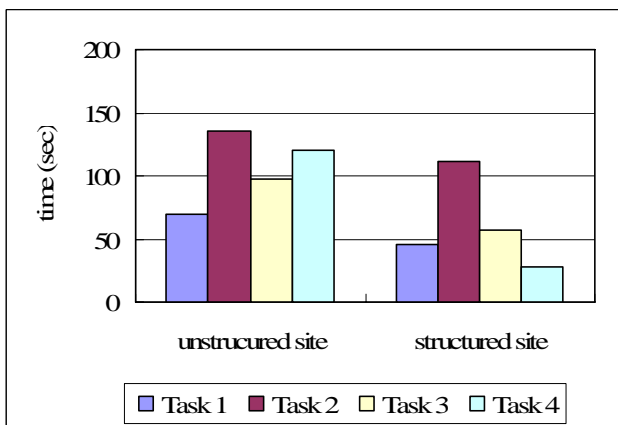


Figure 4b. Average task completion time and standard deviation for blind subjects in Ordering 2 (users first accessed unstructured and then structured site)

The structuring speed up ratio of task completion time of structured site to unstructured site was calculated in the same way as for sighted subjects. Table 3 shows the average ratio of each task.

Table 3. Average structure speed-up ratio for blind subjects

	speed-up ratio \pm SD
Task 1	0.56 \pm 0.23
Task 2	0.83 \pm 0.54
Task 3	0.71 \pm 0.20
Task 4	0.48 \pm 0.37

3.4 Differences between Sighted and Blind Subjects

To examine if task completion time was different between sighted subjects and blind subjects, two-way ANOVA of task completion time with two factors, structure (structured or not) and the type of subject (the sighted or the blind) was carried out for each task. In this analysis, structured or not was a within-subject factor, while sighted or blind was a between-subject factor. As summarized in Table 4, for every task, interaction between these two factors was significant and both factors were found to affect the task completion time.

Table 4. Summary of two-way ANOVA of task completion time for 4 tasks

	structured or not	subject (blind or sighted)	interaction
Task 1	$F(1,17)=53.20$, $p=0.000$	$F(1,17)=153.3$, $p=0.000$	$F(1,17)=21.72$, $p=0.000$
Task 2	$F(1,17)=8.86$, $p=0.008$	$F(1,17)=30.62$, $p=0.000$	$F(1,17)=1.80$, $p=0.050$
Task 3	$F(1,18)=24.70$, $p=0.000$	$F(1,18)=17.25$, $p=0.001$	$F(1,18)=7.46$, $p=0.014$
Task 4	$F(1,18)=21.23$, $p=0.000$	$F(1,18)=45.48$, $p=0.000$	$F(1,18)=13.02$, $p=0.002$

Figure 5 shows the ratio of average task completion time for blind subjects over that of sighted subjects. Standard deviation was calculated using the standard deviation of a set of task completion time for blind subjects and that of sighted subjects. One can see task 1 takes more time for blind subjects than for sighted subjects. Task 1 was a Yes or No question and was quite easy for sighted subjects. Except for task 1, blind subjects needed roughly twice the time required by sighted subjects when the site was structured. As can be seen in the figure, the difference between the structured and unstructured for task completion time tended to be greatest for task 4.

⁵ Analysis of the blind subjects included one task completion time where answer of that task was wrong.

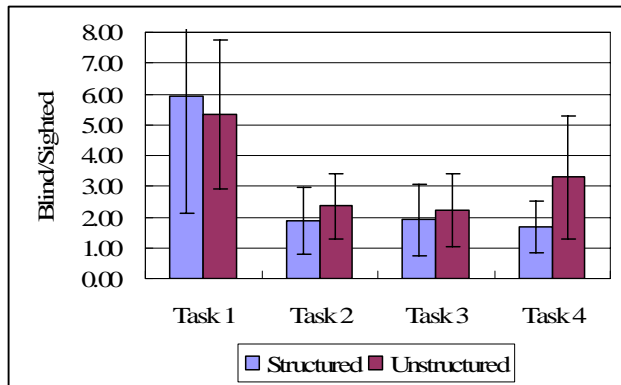


Figure 5. Ratio of average task completion time of blind subjects over that of sighted subjects

4. CONCLUSIONS

In this paper, an experiment was carried out which measured task completion time for 4 tasks. 16 sighted people and 4 blind people participated in the study. In the following subsections, the results are interpreted with respect to the questions presented earlier in the Introduction.

- Q1. How much (if at all) is the task completion time reduced by marking up heading elements?
- Q2. Do the benefits to (shorter) task completion time of marked up heading elements differ between sighted and blind people?
- Q3. Is the difference in task completion time between sighted and blind people decreased when structure markup is both available and utilized by user agents?

4.1 Reduction of Task Completion Time

Table 2 and Table 3 shows the speed-up attribute to using a structured site, for sighted and blind subjects, respectively. The response time for sighted subjects for the structured sites ranged between 50% and 90% of the required time for the unstructured sites (depending on the task). The corresponding ratios for the blind subjects ranged between 50% and 80% (depending on the task). Based on the present results, task completion time is reduced about 10 to 50 % when heading elements are properly marked up.

The impact of structuring can also be seen how it affected individual subjects. In task 1, one sighted subject spent only 1 to 2 seconds to navigate through the structured site, while she spent several seconds to navigate through the unstructured site. For task 1, some of the blind subjects spent about 20 to 40 seconds to navigate through the structured site, while they typically spent from 60 to 70 seconds to navigate through the unstructured site. Task 2 required additional time to think as well as navigate. In this task, some sighted subjects spent up to around ten seconds to think and about 25 seconds to navigate for the structured site. For the unstructured site, they spent roughly the same amount of time to navigate but somewhat longer to think.

As shown by the ANOVA results presented in subsection 3.2, task completion time is strongly affected by tasks. Therefore, the results obtained in this experiment cannot be generalized to other tasks.

4.2 Difference between Sighted and Blind Subjects

Two-way ANOVA results in subsection 3.4 shows that subject factor, sighted or blind, affects the task completion time. Figure 5 shows that the differences between sighted and blind subjects differ from task to task.

For tasks 2 and 3, task completion times for the structured site averaged about 56 ± 28 and 36 ± 15 seconds for sighted subjects and about 106 ± 33 and 69 ± 30 seconds for blind subjects. Thus, blind subjects tend to take about twice as long to complete a task for both structured and unstructured sites.

The impact of lack of structuring on blind subjects seems to be more severe for task 4. While blind subjects needed around twice as long (vs. sighted subjects) to complete when sites were structured, they needed around three times as long when sites were not structured. Further research is needed to study the interaction between type of users (sighted vs. blind) and type of task on task completion times.

However, in general the benefits of marked up heading elements to task completion time differ between sighted and blind users.

4.3 Effect of Structure Markup for Blind Subjects

Two-way ANOVA results in subsection 3.4 showed a significant interaction between type of subject (sighted or blind) and structure (structured or not).

Figure 5 showed that, except for task 1, the average ratio of task completion time with blind subjects over that of sighted subjects was about 2.

Thus, it can be said that structured sites, marked up with appropriate heading elements, decrease the gap between sighted users.

4.4 Effect of Heading Elements in Terms of Usability and Accessibility

From the discussions above, it can be said that

- 1) Heading elements improves usability because task completion time was reduced considerably as a result of using them. This finding was supported by questionnaire data that showed higher user satisfaction with the structured site.
- 2) Heading elements also improve accessibility since the benefits of structuring are sometimes stronger for blind users.

5. ACKNOWLEDGMENTS

The author gives special thanks to Yukari Shimura for designing and preparing Web sites and carrying out experiments. The author also thanks to Mae Kawabata and Risa Matuda for writing down recorded video tapes and Masako Muraoka for assisting experiments.

This work was supported by the Ministry of Education, Culture, Sports, Science and Technology, Grant-in-Aid for Scientific Research on Priority Areas, #16091210, 2004.

6. REFERENCES

- [1] Caldwell, B., Chisholm, W., Slatin, J., and Vanderheiden, G. *Web Content Accessibility Guidelines 2.0 W3C Working Draft* (27 April, 2006)
Retrieved February 26, 2007 from <http://www.w3.org/TR/2006/WD-WCAG20-20060427/>
- [2] Caldwell, B., Chisholm, W., Slatin, J., and Vanderheiden, G. *Understanding WCAG 2.0*
Retrieved February 26, 2007 from <http://www.w3.org/TR/UNDERSTANDING-WCAG20/Overview.html>
- [3] Caldwell, B., Chisholm, W., Slatin, J., and Vanderheiden, G. *Techniques for WCAG 2.0*
Retrieved February 26, 2007 from <http://www.w3.org/TR/2006/WD-WCAG20-TECHS-20060427/Overview.html>
- [4] Chisholm, W., Vanderheiden, G., and Jacobs, I. *Web Content Accessibility Guidelines 1.0* (May, 1999)
Retrieved February 26, 2007 from <http://www.w3.org/TR/WCAG10/>
- [5] Henry, S.L. *Essential Components of Web Accessibility*
Retrieved February 26, 2007 from <http://www.w3.org/WAI/intro/components.php>
- [6] Japanese Standard Association. *JIS X 8341-3:2004 Guideline for older persons and persons with disabilities - information and communications equipment, software and services- Part3: Web Content*
- [7] Shimura, Y. *The Effects of Structured Web in terms of Usability and Accessibility* (in Japanese), graduation thesis, Tokyo Woman's Christian University, Tokyo, JAPAN, 2006.
- [8] Watanabe, T. *Capability Survey of Japanese User Agents for person with visual disabilities* (in Japanese).
Retrieved February 26, 2007 from <http://www.comm.twcu.ac.jp/~nabe/data/UAResearch2005/>
- [9] Watanabe, T. and Umegaki, M.: Capability Survey of Japanese User Agents and Its Impact on Web Accessibility, In *Proceedings of International Cross-Disciplinary Workshop on Web Accessibility*, (May 2006, Scotland)
- [10] Watanabe, T. and Umegaki, M.: Capability Survey of User Agents with the UAAG 1.0 Test suite and Its Impact on Web Accessibility, to be appeared in *Special Issue in Universal Access in the Information Society*, Springer.

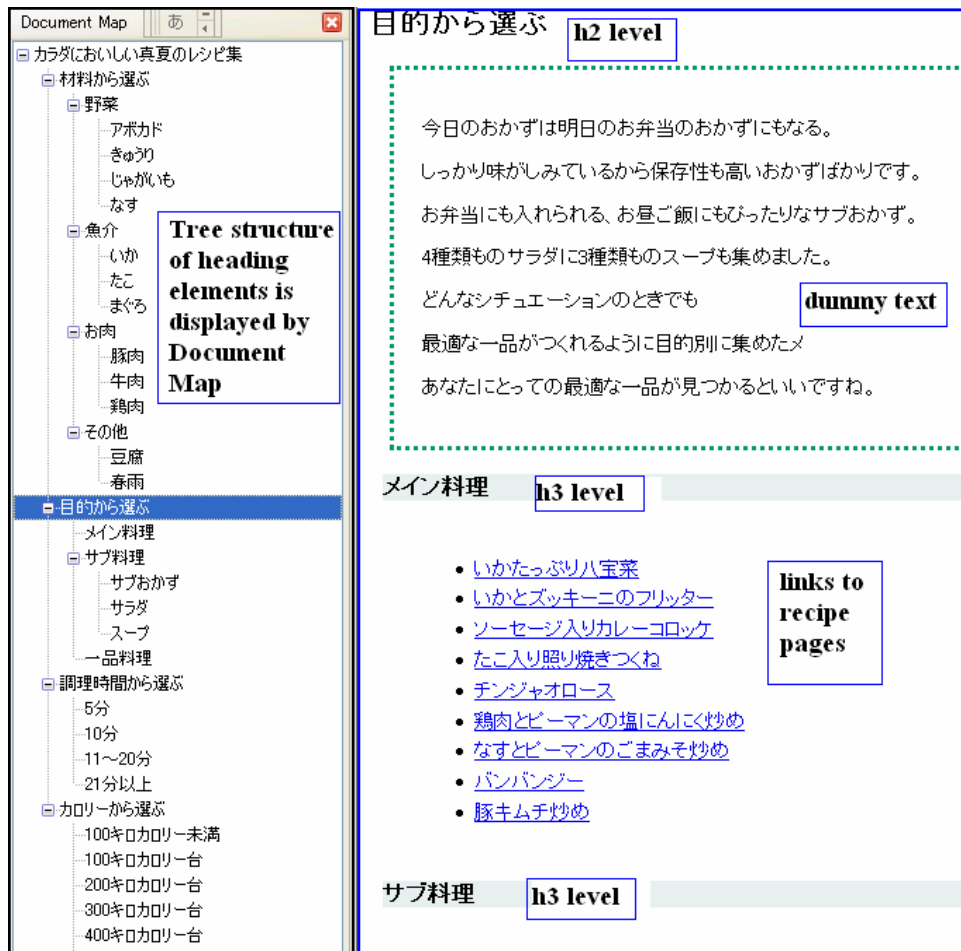


Figure 1a. Document Map shows tree-structure of heading elements in the sidebar of Firefox

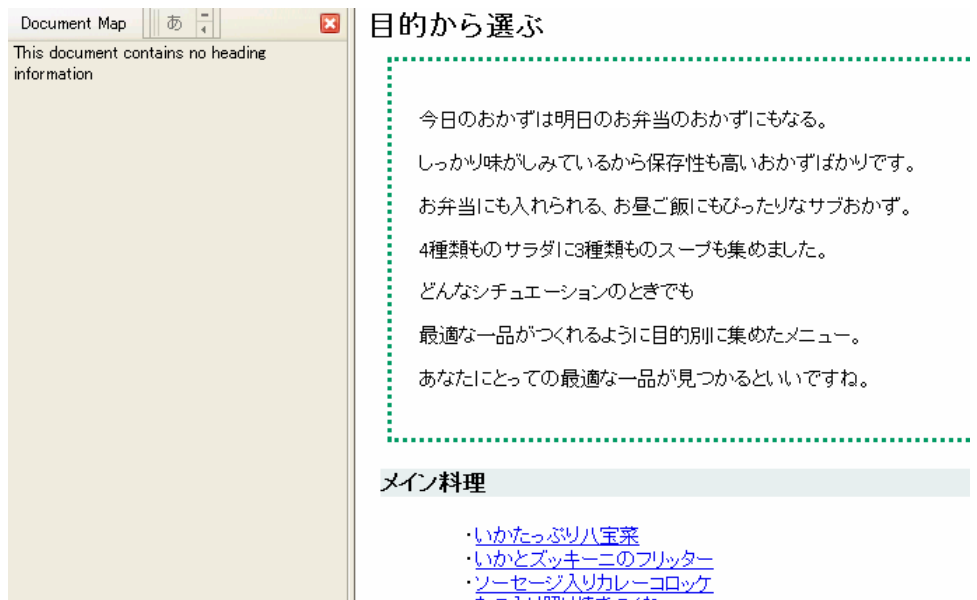


Figure 1b. Document Map shows nothing if content is not marked up with heading elements