# An Automatic Document Coloring and Browsing System

Tomoyuki UCHIDA[a] and Hidehiko TANAKA[b]

[a]President, Yun Factory Corporation, 7-17-10 Koyama Shinagawa-ku, Tokyo, 142 Japan
email:tomo@yun.co.jp

[b]Department of Electrical Engineering, Faculty of Engineering, The University of Tokyo, 7–3–1 Hongo Bunkyo-ku, Tokyo, 113 Japan email:tanaka@mtl.t.u-tokyo.ac.jp

## 1 INTRODUCTION

In recent years, the number of online documents, for example WWW, has increased. Nevertheless they are still expressed in traditional black-and-white. Color devices such as color CRT are available, and document expressions which make full use of this chromatic faculty is in great demand[1]. We will study effective coloring of documents to increase their legibility and understandability.

The difficulty of making the most of the good effects is due to its complicated side effects such as unpleasant feeling, distraction and individual variation[2]. Previous researches tend to consider only on gaudy primary colors and on fixed devices such as paper. But there are quiet colors and dynamic devices which can change coloration, and there is still much to investigate on document coloring.

Thus we measured by experiment the influence of color given to a reader, and studied the way to make the most of colored documents. And finally we developed an automatic Japanese document coloring and browsing system (named CERAS) was developed. From an estimation experiment, CERAS improved the speed and accuracy of reading.

## 2 COLOR EFFECT

There are various psychological effects of using colors[1]. First of all, color can express up to millions of attributes human can distinguish. Especially a phenomenon called pop-out that can discover the stimulus of a purpose in a glance from within a plural stimulus is very effective. Next, color can bring the sense of warmness, size, distance and weight. Besides that, color brings feelings such as beauty and joy, as it is now acknowledged in the advertisement field. If the effects of these colors are used effectively, it will be expected that a reader can get the outline of a document faster and understand the content of a document further and enjoy reading more.

In our experiments, coloring promotes understanding and remembering the content of a document.

The coloring rules are established taking into the advantages and side effects. But it is difficult to construct only one coloring rule that always extracts maximum effect in all styles of documents, reader's attitudes and purposes. For that reason, it is more desirable that a document browser selects best coloring rules dynamically.

## 3 AUTOMATIC COLORING SYSTEM 'CERAS'

We developed a colored text browsing system named CERAS. This system gives color to an input plain text and displays it on a computer CRT. Users can customize the coloring rules and tune the expression interactively by a GUI. Document processing is implemented on a UNIX workstation, and GUI on a Windows95 PC.

Morpheme analysis is carried out to sentences in the text, and coloring points; type of character, part of speech and keywords, are extracted. CERAS gives color expressions to extracted points considering the reader's customize information and feedback information from the GUI. The browser image is shown in Figure 1. It displays a colored text in a big window. Operation is carried out by mouse.

A "Pi menu" is used as a GUI to input user customize coloring information and to tune the color expression interactively in the system. Pi menu shown in Figure 2 is displayed on top of the text screen centering around the mouse pointer, when the left button of the mouse is clicked. The user moves the mouse pushing the button to the direction of a menu button and selects a function. The tuning expression is assigned to the first layer of the Pi menu. Customizing is assigned to the second and third layer. For example the bottom menu button of the first layer is assigned to weaken the expression. Therefore if this button is kept selected, the expression becomes more quiet and ends as a black-and-white expression. The top button is assigned to strengthen the expression and makes the expression gaudy.

Figure 2: Feedback GUI "Pi menu".

Figure 1: Appearance of the colored text browser.

CERAS acts in either General coloration mode or Specific coloration mode. In General coloration mode, documents are colored by general features, such as type of character and part of speech. Specific coloration mode is used when a user has a specific point to read, and the user can enter this mode anytime, to specify a keyword by clicking the right button of a mouse on it. In this mode, words related to the keyword are also colored by using a thesaurus. The related words are expressed by the color whose strength is associated with the strength of the relation. The user can quickly find out interesting parts within a document.

## 4 ESTIMATION OF CERAS

Keyword coloration function of CERAS was evaluated by a news article classification experiment.

A testee searches one or two designated descriptions inside a displayed document in speed-reading and classified the document into two kinds.

First, a condition sentence is presented on the CRT to a testee, and the testee chooses and clicks one word as a coloring keyword. Next, a news article of approximately 1,100 characters is presented on the CRT, and the testee classifes it and pushes either the right or the left mouse button. Two patterns of combinations of colored and black-and-white document are prepared and presented to a testee. In colored documents, keywords are colored red and colored words are displayed in deep blue.

The number of testee is 14. Time required for a testee to click the mouse button from the presentation of condition sentence is measured. It is equal to the reading time in black-and-white, because keyword is not designated in this case.

The time for choose a keyword and point it by moving the mouse is 601ms. This means that the overhead time for the utilization of interactive coloration function of CERAS is 601ms.

Table 1: Average classification time

| | mono(1target) | color(1) | mono(2) | color(2) |
|---|---|---|---|---|
| Match | 12.86 | 3.936 | 10.37 | 4.048 |
| Unmatch | 16.47 | 3.675 | 15.98 | 6.647 |
| Total(sec) | 15.47 | 3.745 | 13.18 | 5.347 |

Next, the average classification time is shown on Table 1. The required time is reduced to 24.2% by coloration in one target condition and to 40.6% in one of two targets condition. And it costs less time to classify by coloration in all cases.

The correct answer rate is raised by coloration in 3 out of 4 cases. In total, it rises 5% in one condition, 2% in two conditions.

There is a linear relationship between the location of a search goal inside an article and classification time. Therefore we can estimate the reading speed according to this relationship. It can be said that there is a correlation between distance from the top of an article to the search goal and classification time in the case of a black-and-white news article, because a testee is apt to search the target from the top to the end.

The required time to read one content word is estimated 27.71ms and 31.22ms in one condition and two conditions in the case of a black-and-white document. The time to check around a mark is 923.3ms for a keyword, and 123.6ms for a related word.

We consider that a user searches an unknown word which satisfies a semantic target in the related words.

There are n documents on the same condition. The number of keyword marks, related word marks, content words from top to target are a, b and t.

According to the fact that the time required for semantic search costs 24.4% more than lexical search[3], semantic search costs $123.6 \cdot 1.244 = 153.8$(ms) for each related word marks. The overhead of coloring is $\frac{601}{n}$(ms). The average time required for searching in the coloration documents is $\frac{601}{n} + 923.3a + 153.8b + 3073$ (ms).

In black-and-white document, we consider to 10% priming effect additional to the above. It costs $\frac{462.3 \cdot 1.244 \cdot 0.9}{14.81} = 34.95$(ms) for each content words. Therefore, it takes $34.95t + 2311$(ms) to search the target.

For that reason, in the case under the following condition, CERAS is advantageous.

$$t > \frac{17.2}{n} + 26.42a + 4.401b + 21.80$$

$$t > 0, a \geq 0, b \geq 0, n > 0$$

We tried to apply this condition to news articles using this experiment as an example. The parameters are as following: $a = 0.4500, b = 18.05, t = 355.4, n = 20$

When $b$ is free, the condition is $b < 72.8$. This means CERAS can use 72.8 related word marks from the top to the target in average. It means 3.03 marks per line, too. In addition that the limit value is 4 times as much as that in the present system.

In a news article speed-reading case, it is presumed that related word coloration is effective for speed when the number of coloration marks are less than 3 per line on an average.

# 5 CONCLUSION

The number of online documents and those who read documents on color devices is increasing very much. We estimated the effect of coloring a document by experiment, and developed CERAS which generates color expression automatically and responds quickly to the user's customize and tuning request through a GUI.

In General coloration mode, CERAS presents colored expression which improves comprehension. And an user can customize its color and tune interactively while reading. In Specific coloration mode, CERAS has many advantages in fast-reading. Adding a special feature makes it easy to search a keyword, which has generally been carried out in document searching area. But, there are no systems which can change keywords and expressions interactively nor give expressions to words related to the keywords. People frequently search their interesting point when they read relatively less important documents such as news paper articles. For that reason, raising efficiency of this search is generally effective. To raise this efficiency, it is effective that a system adds a special feature to the interesting point, and raises the search speed. CERAS accomplishes this in interactive keyword specification and coloration to keywords and related words. Coloration is suitable to speedup a search utilizing its pop-out function. Moreover, interactively specification of keywords interactively is considered effective when the object of interest is changed while reading. And there is the effect of preventing the oversight of an interest point in relative word coloration. An user can find an interest point which does not include the keyword but its synonym. From an estimation experiment, CERAS is effective for fast-reading of news articles, both in time and in accuracy in Specific coloration mode. And there is enough margin till a defect appears.

Still, there is room of precise estimation and examination, but this system has important significance as the tool that estimates the possibility of color expression of a document. From now on, we will modify the system and establish a presentation method which will increase the effectiveness of colored documents.

# References


[1] William Winn,"Color in Document Design," IEEE Trans. Professional Commun., vol .34, no.3, pp.180-185, 1991.

[2] R. John Brockmann,"The Unbearable Distraction of Color," IEEE Trans. Professio nal Commun., vol.34, no.3, pp.153-159, 1991.

[3] Ken Goryo,"What's reading," Tokyo Univ. Press, 1987.