

可読性向上を図る文書自動彩色システム

内田友幸 田中英彦 東京大学 大学院 工学系研究科

1 はじめに

近年、カラー CRT などカラーで動的なデバイスで文書を読むケースが増えているが、文書に対する UI は従来の紙メディアと同様の白黒で静的なものが主流である。そこで我々はこれらのデバイスを有効に利用した読みやすい文書の UI について検討を重ねている。本稿ではこれらの UI を実装した自動彩色システム (CERAS) [1] のうちインタラクティブな彩色機能について報告する。

2 自動彩色システム CERAS の概要

CERAS はプレーンなテキストファイルを入力とし、それを適宜彩色を含む表現に加工しブラウジングできる形で出力する形態をとるが、読みながら素早く表現を変化させられる GUI で、ユーザーが彩色表現を動的に有効に生かせるようになっている。システムはワークステーションにリンクしたパソコン上に実装している。

CERAS には一般的特徴を彩色する一般彩色モードと、指定した単語とそれに関連する情報を彩色する指定彩色モードがある。本稿で報告する指定彩色モードは興味の対象が絞られている時に利用し、興味のある単語をマウスの右ボタンでクリックすることでこのモードに移る。そしてクリック位置の単語をキーワードとしてシソーラスを検索し、キーワードと関連語を彩色する。この際、関連語の色は関連の強さも反映している。これを利用することでユーザーは文書の中から興味のある部位を素早く見つけ出すことができるようになる。



図 1: ブラウザ (「政界」をクリックした様子)

ブラウザ画面の様子を図 1 に示す。ここでは「政界」という単語を右クリックして「政界」が赤に、その関連語が青～紺に彩色された状態になっている。

A Study of Automatic Text Coloring System.
Tomoyuki Uchida Hidehiko Tanaka Department of Electrical Engineering, Faculty of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-Ku, Tokyo, 113, Japan
e-mail : {tomo,tanaka}@mtl.t.u-tokyo.ac.jp

3 新聞記事分類心理実験

3.1 新聞記事分類実験の概要

CERAS の持つ指定単語の彩色機能を心理実験で評価した。被験者には表示される文書中から指定した条件にあう記述を速読で探してもらい、その有無で文書を分類してもらった。この際に白黒のものも分類してもらって彩色との違いを時間と分類精度の 2 点で評価した。

実験は 3 種類行ったが、基本的には被験者にはまず分類の基準となる条件文を提示し、その次に一画面におさまる約 1100 字、全部で 54 記事の新聞記事を順次 CRT に提示する。但し、その半数は白黒のまま提示した。回答にはマウスの右か左のボタンを押してもらう。条件文は例えば「減税の財源に言及しているもの」となっている。一つ目の実験は一つの条件文につき一つの新聞記事を分類するもの。2 つ目は一つの条件文につき複数の新聞記事を分類するもの。3 つ目はマウスクリックした単語に関連の有る単語も彩色した新聞記事を分類するものとなっている。ただし、意味判断時の曖昧性を避けるため、条件の判断には意味的ではなく、条件文中の単語との一致を見てもらって判断してもらった。

また、彩色箇所を指定してもらうため、一つ目の実験で彩色文書に対する条件文の提示の場合は条件文中に有る単語を一つ左クリックしてもらった。

3 つ目の実験では指定単語で彩色される条件の他に関連単語で彩色される条件も設定し、2 条件のどちらかが合致するかどうかを判断してもらった。利用色は表 1 に示す。

表 1: 新聞記事分類実験利用色

利用色	輝度 (lx)	色度 (CIE 1931 XYZ 表色系)
白 (文字の背景)	89.8	(0.288, 0.313)
赤 (マーク単語)	15.0	(0.615, 0.330)
紺 (関連単語)	2.6	(0.152, 0.078)
黒 (漢字)	0.0	-

3.2 新聞記事分類実験結果

被験者は成年男子 12 名、女子 2 名の計 14 名。一つ目の実験で条件文の提示から読み終わってマウスクリックするまでの平均時間は 2540ms、彩色したい単語をマウスクリックするまでにかかった平均時間は 3141ms。この差の 601ms は探索するのに都合の良い単語の選定、マウスカーソルの目的単語への移動にかかった時間と推定できる。これは彩色単語を指定するインタラクティブ彩色システムの利用に必要な時間的オーバーヘッドであるといえる。

また、平均分類所要時間と正答率を表 2 に示す。

表 2: 平均分類所要時間と平均正答率

	条件数 1 白黒	# 1 色	# 2 白黒	# 2 色
分類時間 (s)	15.47	3.745	13.18	5.347
正答率	0.87	0.92	0.91	0.93

ここから彩色することで白黒の時に比べて所要時間が条件が一つの時に 75.8%、条件が二つの時は 59.4%短縮され、正答率は条件が一つの時に 5%、2つの時に 2%彩色の方が高いことがわかる。

続いて記事内の探索目標の位置と分類時間の関係についての結果を調査した。白黒の場合は目標までの自立語数、彩色の場合は目標までの彩色マーク数を横軸に、分類時間を縦軸にして関係を調べたものを表 3 に示す。

表 3: 目標位置と分類時間の相関と回帰直線

	条件数 1 白黒	# 1 色	# 2 白黒	# 2 色
相関係数	0.603	0.789	0.803	0.915
傾き	27.71	923.3	31.22	123.6
y 切片	4032	1815	2311	3073

この結果を見るとそれぞれ直線的な傾向が有ることがわかる。また、回帰直線の傾きは要素をチェックするのに必要な時間に相当すると言える。

3.3 新聞記事分類実験考察

以上のような結果から、単語表現を手がかりに読むポイントを捜すような速読時に彩色は時間短縮と精度向上に有効なケースがあるといえる。

続いて以上の結果を利用して、より現実の利用形態に近い意味的に条件を満たす未知の単語も含めて探索する場合を検討する。

関連語中から意味的にキーとなる単語を含んだ条件の探索にかかる時間は、意味的探索が単なる探索にくらべて 24.4%増加する事 [2] を使えば彩色関連語 1 語あたり 153.8ms かかると考えられる。また、順次単語を追っていく白黒文書はプライミング効果が 10%あるとすれば 1 自立語あたり 34.95ms かかるといえる。

同じ条件で n 文書閲覧し、目標となる彩色マークを含むそれ以前にあるキーワード彩色マークが平均して a 個、関連語彩色マークが平均して b 個有り、目標は自立語を数えて t 単語目に有るとすると、彩色した場合の平均所要時間は $\frac{601}{n} + 923.3a + 153.8b + 3073(\text{ms})$ 。白黒文書の場合は $34.95t + 2311(\text{ms})$

よって、意味的探索の場合

$$t > \frac{17.2}{n} + 26.4a + 4.40b + 21.8$$

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

という条件を満たす場合彩色文書は有利であると考えられる。この条件を、一つの例として今回のような新聞記事を対象にした時に当てはめると $a=0.4500$, $b=18.05$, $t=355.4$ であるから、 t と a に上記の数字を

代入して考えると、 $b < 72.8$ であるので、関連語彩色マークはこのような場合平均 72.8 個、平均行数が 24 行なので 1 行当たりでは 3.03 個まで増やしても彩色の効果を得られると言える。この上限値は現在システムの約 4 倍の値に相当する。

4 結論

ブレンテキストをインタラクティブに自動彩色し、表現を自由に変えられるブラウザで提示する一貫したシステム (CERAS) を作成した。CERAS を利用した評価実験の結果から CERAS の指定彩色モードの彩色機能は多くのケースにおいて速読の速度向上と精度向上に有効であることが示された。

文書検索を行った場合、キーワードに探索しやすい特徴を付加することは一般に行われてきた。しかし、指定彩色モード時のように関連単語にまで特徴を付加したり、さらにインタラクティブにキーワードを変更して表現を変更できるシステムは無かった。

一般に精読の対象としていない不要な情報を多く含む文書を読み進む場合、興味のあるところを頻繁に拾い読みしながら読み進むことになる。そのため、この拾い読みの効率を上げることは有効なことであるといえる。この効率を上げるには興味のあるところに特徴を付加し探索速度を上げることが有効であると考えられるが、CERAS はこれをインタラクティブな指定をした単語とその関連語に対する彩色で実現している。実験結果からこの彩色は新聞記事を対象にすると 1 行に 3 個程度まで付加しても彩色の効果を得られる事が分かったが、これは現在の 4 倍となる量である。新聞記事の速読の用途では現システムは多くの場合速く読める事が期待され、彩色が邪魔になるレベルまでに十分なマージンが取れていると言える。

今後の課題としては、現在は動的に指定できるキーワードは 1 種類だけであるが、これを複数にすることでより興味の対象に近い関連語を絞りこむことができると考えられるため、将来的に検討していきたい。

また、現在の検討では、心理実験の規模が必ずしも十分ではなく、より精密な結論を引き出すことができていない。有効な彩色ルール、使いやすい GUI など基礎的な点を心理実験を行って確立していきたい。

謝辞

形態素解析には京都大学長尾研究室、奈良先端大松本研究室によって開発された JUMAN を、また、新聞記事の素材には CD-毎日新聞 94 年版を利用させてもらいました。この場を借りて感謝の念を表します。

参考文献

- [1] 内田 友幸, 田中 英彦: “自動文書彩色システムによる可読性向上の評価” 情処第 53 回全国大会, 2Q-3, Vol.4, pp.143-144
- [2] 御領 謙, “読むということ,” 東京大学出版会, 1987.