

# 利用者の閲覧 Web 文書から抽出される 色情報プロフィールに基づいた Web ランキング方式

森脇ゆりか<sup>†1</sup> 本多宏<sup>‡2</sup> 鷹野孝典<sup>†1</sup>

本稿では、利用者の閲覧した Web ページから抽出される色情報を色プロフィールとして蓄積しておき、この色プロフィールを利用した Web 文書のランキング方式について提案する。本研究では、実際に構築したプロトタイプを用いた検索実験により、提案方式を用いて、Web ページの内容だけでなく、Web ページに含まれる配色情報を考慮した Web ページの提供が可能であることを確認する。

## COLOR-based Ranking Method for Web documents

YURIKA MORIWAKI <sup>†1</sup> HIROSHI HONDA<sup>‡2</sup> KOSUKE TAKANO<sup>†3</sup>

In this paper, we present a color-based ranking method for web documents based on a user's color profile, which is extracted from web documents that the user has browsed. By experiments using prototype of the proposed system, we confirm that our ranking method allows a user to obtain Web documents according to the color information of the Web documents, not only their textual information.

### 1. はじめに

WWW の普及とともに、Web 検索エンジンを用いた情報獲得は、我々の生活に欠かせないものとなっている。現在利用されている Web 検索エンジンの多くは、テキストの内容に着目した Web ページのランク付けを行っており、利用者は問い合わせ単語を検索エンジンに送信することにより、検索目的に合致した Web ページの検索結果を得る事ができる。

しかしながら、Web 検索エンジンや情報推薦システム[3]を利用した Web ページの獲得において、内容に基づいた関連性のみではなく、個々の利用者にとっての読みやすさを考慮することが重要な場合がある。例えば、利用者が情報推薦システムから配信されるニュース記事を受け取るとき、好みにあった配色やレイアウトのものであれば、利用者はその記事が読みやすいと感じることができ、満足感が得られる可能性が高いと考えられる[4]。さらに、特定の色が識別できない等、色覚特性のある人々を対象とした場合、それらの人々が受容できる配色の Web ページを検索結果にて提示できることが求められる。例として、もし、ある利用者が赤色を識別するのが困難であれば、検索結果から、赤色の配色率が高い Web ページを除外し、代わりにその利用者の受容できる色を多く使用しており、かつその Web ページと同じ程度の内容を含んだものを検索結果の上位に提示することが望ましい場合がある。

本研究では、このような Web ページの獲得における配色

情報の重要性に着目し、利用者の閲覧した Web ページから抽出される色情報プロフィールに基づいた Web ランキング方式を提案する。本研究を進めるにあたり、人々は、Web ページのある特定のトピックや形式において、意識的ないしは潜在的に色の嗜好性を持っており、それに基づいた Web 文書の獲得・閲覧を行っていると仮定する。

色の嗜好性に基づいて意図的に Web ページを獲得する例としては、ある女性の利用者が、黄色やピンク色等の春らしさをイメージする色に注目して、服飾や雑貨商品が掲載されている Web サイトを閲覧している場合等が挙げられる。この場合、もし、情報推薦システムに「ファッション」というキーワードが登録されていれば、システムは黄色やピンク色等の春らしさをイメージする色の服飾雑貨商品を多く掲載している Web サイトを推薦結果として配信すべきである。

また、色情報に応じて潜在的に Web ページを選択する例としては、利用者が多彩な配色を持った広告が複数配置されているような Web ページを読みづらいつ感じているような状況等が挙げられる。このような状況において検索結果をクリックして表示された Web ページが類似の配色を持っていれば、すぐスキップしまう可能性が高くなると考えられる。このような場合、検索システムが多彩な配色を伴った Web ページの検索順位を下げることにより、その利用者は、自分にとって読みやすい配色の Web ページを獲得することができるようになると考えられる。

提案方式は、利用者の閲覧した Web ページから抽出される色情報を色プロフィールとして蓄積しておき、この色プロフィールを利用して Web ページのランキングを行うものである。Web ページのテキスト情報とともに、利用者が閲覧する Web ページの色情報を考慮することにより、利用

<sup>†1</sup> 神奈川工科大学 情報学部 情報工学科  
Department of Information and Computer Sciences, Faculty of Information  
Technology, Kanagawa Institute of Technology

<sup>†2</sup> 神奈川工科大学大学院 情報工学専攻  
Graduate School of Information and Computer Sciences, Kanagawa Institute  
of Technology

者は自分が意識的または潜在的に嗜好する配色の Web ページを獲得できるようになる。

本研究では、実際に構築したプロトタイプを用いた検索実験により、提案方式を用いて、Web ページの内容だけでなく、Web ページに含まれる配色情報を考慮した Web ページの提供が可能であることを確認する。

## 2. 配色情報を考慮した Web ページ検索の利点

配色情報を考慮した Web ページ検索の利点として、下記の項目が挙げられる。

- (1) 読みやすい配色を持った Web ページの獲得
- (2) 色覚特性のある人への、認識しやすい配色を持った Web ページの提供
- (3) 検索結果一覧から読みたい Web ページを探す際の、Web ページの取捨選択の効率化
- (4) 商品情報サイト等における、好みの色の商品を掲載した Web ページの優先的な提示

(1)については、利用者が読みやすいと感じる配色の Web 文書を見続ければ、その配色情報が色情報プロフィールに蓄積されるため、結果として、その利用者にとって読みやすい配色の Web 文書を提示することができる。(3)については、例えば、検索結果一覧に表示される Web ページのスナップショット画像から、利用者は内容を熟読する前に、Web ページの配色情報に基づいた第一印象により、読むべき Web ページを選択できる可能性がある。

(4)について、図 1 に、実装したシステムを用いて、商品情報サイト等において、利用者の嗜好する色の商品を掲載した Web ページを優先的に提示している様子を示す。この例では、まず利用者は商品情報サイトにおいて青系統の色の洋服を頻繁に見ていた結果として、青系統の色が色情報プロフィールに蓄積されている。さらにその結果として、検索エンジンで「シャツ」を問い合わせとして検索した場合、緑やピンク色よりも、青系統の色のシャツを掲載した Web ページが優先的に検索されている。

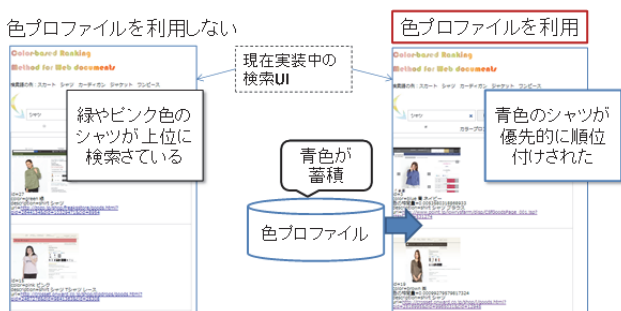


図 1 提案システムの応用例

Figure 1 Example Application of proposed system

## 3. 提案方式

### 3.1 概要

提案システムの概要を図 1 に示す。提案システムは、色情報プロフィール、色情報に基づいた検索エンジン、Web クローラ、および、色情報プロフィールと Web ページ・インデックスの各データベースより構成される。

提案方式では、まず、利用者の閲覧した Web ページのキャプチャ画像を取得し、画像から色情報を抽出することにより色プロフィールを作成する。一方、色クローラは Web ページを収集し、単語に基づいたメタデータおよび色情報に基づいたメタデータを作成しておく。

色情報プロフィールより抽出される  $n$  次元の色ベクトルと、Web ページ・インデックスより同じ形式で抽出される各 Web ページの色情報ベクトル間の類似度を算出し、さらに、利用者の問い合わせに対する Web ページの類似度を算出することで、色情報と単語情報の類似度を考慮したスコアに基づいて、Web ページのランキングを行う。

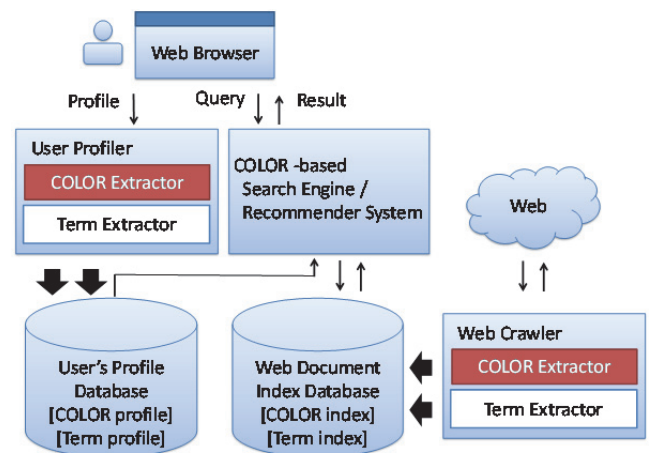


図 2 提案システムの構成

Figure 2 The architecture of proposed system

### 3.2 色情報プロフィールの生成

利用者がブラウザを通して閲覧した Web ページを継続的にキャプチャする事により、Web ページの色情報  $c_w$  を抽出し、ユーザ色情報プロフィール上に蓄積する。この際、閲覧した Web 文書には、あらかじめ定められたトピック  $p_i$  が割り振られる。検索処理における相関量計算のために、色情報プロフィールより、 $n$  色 ( $c_1, c_2, \dots, c_n$ ) の特徴ベクトル  $c_u$  として抽出する。この処理において、Web 文書に含まれる各色は、 $c_i$  の近似色に割り当てられる。

$$c_u = [c_{u1}, c_{u2}, \dots, c_{un}] \quad (1)$$

ここで、 $c_{ui}$  は該当する近似色  $c_i$  のピクセル数である。また、

色ベクトルは、トピックを指定することにより、そのトピックのみを対象とした色ベクトルとして抽出することも可能である。

$$\mathbf{c}_{u,p} = [c_{u1,p}, c_{u2,p}, \dots, c_{un,p}] \quad (2)$$

現時点での実装において、利用者の閲覧した Web ページは画像としてキャプチャされ、Web ページの色情報  $c_u$  はキャプチャ画像から抽出される。このため、色情報はテキストのフォント色やページの背景色だけではなく、閲覧した Web ページ中に含まれる画像からも抽出される。

### 3.3 Web クローラによる色情報メタデータの生成

検索対象となる各 Web 文書の色情報メタデータは、Web クローラ中の色情報抽出モジュール(color extractor)により抽出・収集される。ここで、Web 文書  $d_x$  の色情報メタデータは、色情報プロファイルと同様の手順・形式で生成され、色情報ベクトル  $\mathbf{c}_u$  と同様の色定義を持った色情報ベクトル  $\mathbf{c}_x$  として抽出される。

$$\mathbf{c}_x = [c_{x1}, c_{x2}, \dots, c_{xm}] \quad (3)$$

ここで、 $c_{xi}$  は  $d_x$  で使用されている色の近似色  $c_i$  のピクセル数である。

### 3.4 単語メタデータ抽出

Web 文書  $d_x$  の単語メタデータは、単語集合  $T_x$  として抽出される。 $T_x$  は  $d_x$  に含まれる  $m$  個の単語群より構成される。

$$T_x = \{t_{x1}, t_{x2}, t_{x3}, \dots, t_{xm}\} \quad (4)$$

ここで、 $t_{xi}$  ( $i = 1, 2, \dots, m$ ) は  $d_x$  中に出現する単語である。全ての Web 文書中に出現する総単語集合  $T$  は、 $T_x$  を用いて次のように表される。

$$T = \bigcup_{i=1}^x T_i \quad (5)$$

総単語集合  $T$  に含まれる各単語を特徴単語とした Web 文書空間  $S$  を生成し、Web 文書  $d_x$  のベクトル化を行う。

$$\mathbf{d}_x = [u_{x1}, u_{x2}, \dots, u_{x|T|}] \quad (6)$$

ここで、 $u_{xi}$  は Web 文書  $d_x$  における  $t_i$  の出現頻度である。

### 3.5 Web 文書のランキング

Web 文書のランキング処理において、利用者の問い合わせは、色情報プロファイルから抽出される色ベクトル  $\mathbf{c}_u$  (または  $\mathbf{c}_{u,p}$ ) と、利用者が入力した単語群  $q$  に基づく問い合わせ単語ベクトル  $\mathbf{q}$  から構成される。

本方式では Web 文書のランキングのために、まず、色情報  $c_u$  と  $c_x$  についての類似スコア  $S_c$  をコサイン尺度[1]を用いて下記のように算出する。

$$Sc(c_u, c_x) = \frac{\mathbf{c}_u \cdot \mathbf{c}_x}{\|\mathbf{c}_u\| \|\mathbf{c}_x\|} \quad (7)$$

同様に、利用者の問い合わせ単語群  $q$  と Web 文書  $d_x$  についての類似スコア  $S_t$  を下記のように算出する。

$$St(q, d_x) = \frac{\mathbf{q} \cdot \mathbf{d}_x}{\|\mathbf{q}\| \|\mathbf{d}_x\|} \quad (7)$$

色情報に関する類似スコア  $S_c$  と単語情報に関する類似スコア  $S_t$  の双方を用いて、各 Web 文書  $d_x$  のランキングスコア  $S$  を算出する。

$$S = \alpha \cdot S_c + (1 - \alpha) \cdot S_t \quad (8)$$

ここで、 $\alpha$  ( $0 \leq \alpha \leq 1$ ) は  $S_c$  と  $S_t$  をバランスさせるためのパラメータである。

Web 文書は、スコア  $S$  を用いてランキングされ、利用者は色情報プロファイルに基づいた Web 文書を獲得することができる。

## 4. 実験

### 4.1 実験目的

本実験では、実際に構築したプロトタイプを用いた検索実験により、提案方式を用いて、Web ページの内容だけでなく、Web ページに含まれる配色情報を考慮した Web ページの提供が可能であることを確認する。

### 4.2 実験環境

実現したシステムにおいて、利用者が閲覧した Web ページから色情報を抽出するための色情報プロファイラを Google Chrome の拡張機能を用いて実装した。色情報プロファイラは、バックグラウンドで動作し、利用者が閲覧した Web ページを継続的に画像としてキャプチャし、色情報を抽出する。抽出された色情報は、色情報プロファイルに蓄積される。また、色情報メタデータ機能および単語メタデータ抽出機能を備えた Web クローラを Java 言語を用いて

構築した。色情報プロフィール、および Web ページ・インデックスを構築するデータベースとして PostgreSQL を利用した

検索実験データとして、3つの異なる配色を持った5つのweb文書を $d_1$ - $d_5$ を用意し、全体として、15個(=3x5)のweb文書を作成した。表1と図3に、作成したWeb文書の単語メタデータと配色セットの例を示す。

表 1 Web 文書の単語メタデータの例  
 Table 1 Example of term-based metadata

Doc ID	トピック	単語メタデータ
$d_{1a}$	ファッション	coordination, fresh, girl, ecology, knit, dungaree, beige, navy
$d_{1b}$		
$d_{1c}$		
$d_{2a}$	料理	cook, recipe, ingredient, butter, pasta, penne, gnocchi, rice, bean
$d_{2b}$		
$d_{2c}$		
$d_{3a}$	IT ビジネス	solution, agile, software, development, network, service
$d_{3b}$		
$d_{3c}$		
$d_{4a}$	コンピ ュー ター	research, publication, lecture, member, database, mobile
$d_{4b}$		
$d_{4c}$		
$d_{5a}$	音楽	lesson, service, support, audio, chorus, instrument, mobile, smartphone
$d_{5b}$		
$d_{5c}$		

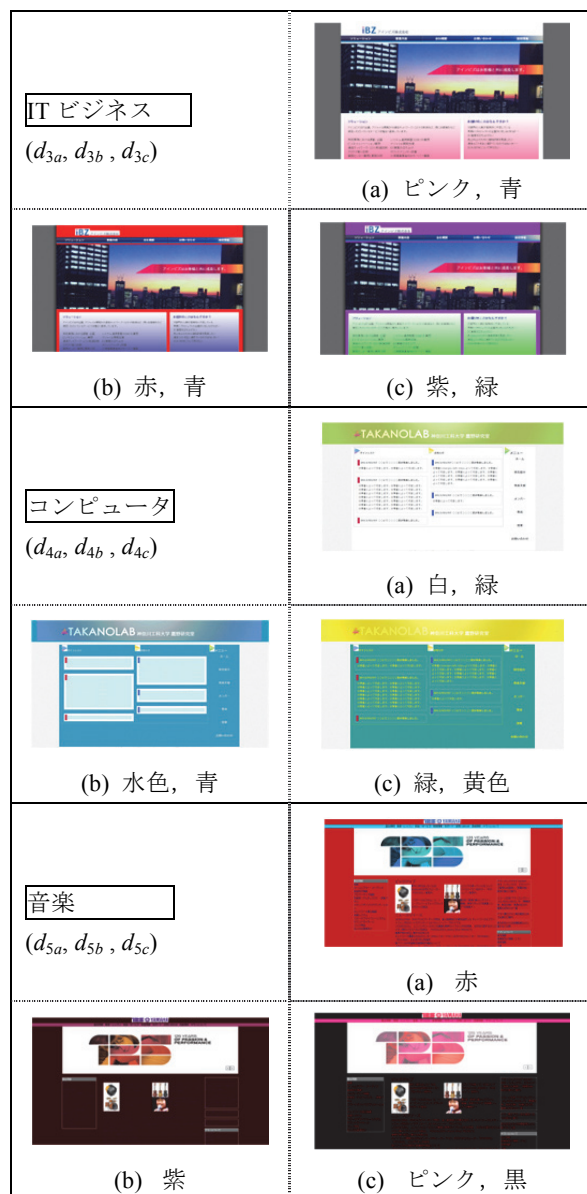
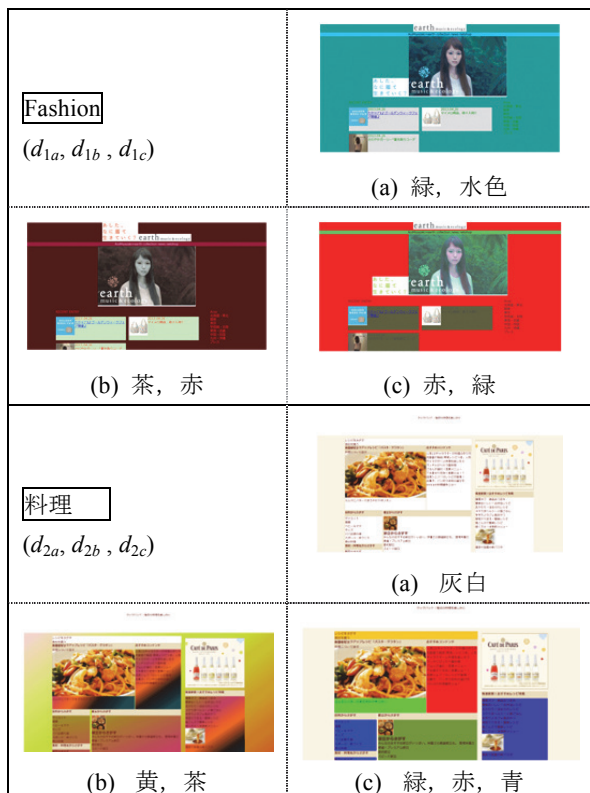


図 3 検索データの配色例

Figure 3 Example of colors used in retrieval data

### 4.3 実験方法

実験では、下記の問い合わせ単語、および2つの色情報プロフィールを用いて、提案方式を用いたランキング性能についての評価を行う。

問い合わせ単語: mobile, system

色情報プロフィール1: (主な色) 赤, カーキ

色情報プロフィール2: (主な色) 緑, 白

また、3章で示した、Web文書のスコアの算出式(8)のパラメータ $\alpha$ の値として、 $\alpha=0.0, 0.3, 0.5, 1.0$ を用いる。ここで、 $\alpha$ が大きい値を取るほど、色情報を重視してWeb文書のランキングスコアを算出していることを示している。例

例えば、 $\alpha=0.0$  の場合は、テキスト情報のみを用いてランキングスコアを算出し、 $\alpha=0.5$  の場合は、色情報とテキスト情報の重みを等しくして、ランキングスコアを算出する。

#### 4.4 実験結果

表 2 と 3 は、問い合わせ単語「mobile, system」について、それぞれ色情報プロフィール 1 と 2 を用いた結果を示している。

表 2 検索結果 (色情報プロフィール 1)

Table 2 Search result (User color profile 1)

Rank	$\alpha$			
	0.0	0.3	0.5	1.0
1	$d_{4a}$ 0.182	$d_{5c}$ 0.278	$d_{1c}$ 0.433	$d_{1c}$ 0.867
2	$d_{4b}$ 0.182	$d_{1c}$ 0.260	$d_{5c}$ 0.353	$d_{5c}$ 0.540
3	$d_{4c}$ 0.182	$d_{3b}$ 0.239	$d_{3b}$ 0.280	$d_{2b}$ 0.445
4	$d_{3a}$ 0.176	$d_{4a}$ 0.217	$d_{4a}$ 0.240	$d_{3b}$ 0.385
5	$d_{3b}$ 0.176	$d_{4b}$ 0.193	$d_{2b}$ 0.222	$d_{2a}$ 0.301
6	$d_{3c}$ 0.176	$d_{3a}$ 0.174	$d_{4b}$ 0.200	$d_{4a}$ 0.298
7	$d_{5a}$ 0.166	$d_{4c}$ 0.166	$d_{3a}$ 0.173	$d_{2c}$ 0.246
8	$d_{5b}$ 0.166	$d_{5a}$ 0.155	$d_{4c}$ 0.155	$d_{4b}$ 0.219
9	$d_{5c}$ 0.166	$d_{5b}$ 0.147	$d_{2a}$ 0.150	$d_{3a}$ 0.169
10	$d_{1a}$ 0.000	$d_{3b}$ 0.140	$d_{5a}$ 0.148	$d_{5a}$ 0.130
11	$d_{1b}$ 0.000	$d_{2b}$ 0.133	$d_{5b}$ 0.134	$d_{4c}$ 0.129
12	$d_{1c}$ 0.000	$d_{2a}$ 0.090	$d_{2c}$ 0.123	$d_{5b}$ 0.101
13	$d_{2a}$ 0.000	$d_{2c}$ 0.074	$d_{3c}$ 0.117	$d_{1a}$ 0.075
14	$d_{2b}$ 0.000	$d_{1a}$ 0.022	$d_{1a}$ 0.037	$d_{3c}$ 0.057
15	$d_{2c}$ 0.000	$d_{1b}$ 0.010	$d_{1b}$ 0.017	$d_{1b}$ 0.035

表 3 検索結果 (色情報プロフィール 2)

Table 3 Search result

Rank	$\alpha$			
	0.0	0.3	0.5	0.7
1	$d_{4a}$ 0.182	$d_{1a}$ 0.292	$d_{1a}$ 0.486	$d_{1a}$ 0.973
2	$d_{4b}$ 0.182	$d_{4c}$ 0.279	$d_{4c}$ 0.344	$d_{4c}$ 0.507
3	$d_{4c}$ 0.182	$d_{4b}$ 0.232	$d_{4b}$ 0.265	$d_{4b}$ 0.349
4	$d_{3a}$ 0.176	$d_{4a}$ 0.185	$d_{4a}$ 0.188	$d_{2a}$ 0.194
5	$d_{3b}$ 0.176	$d_{3a}$ 0.158	$d_{3a}$ 0.145	$d_{4a}$ 0.193
6	$d_{3c}$ 0.176	$d_{5c}$ 0.147	$d_{5c}$ 0.134	$d_{2b}$ 0.177
7	$d_{5a}$ 0.166	$d_{5a}$ 0.141	$d_{5a}$ 0.124	$d_{2c}$ 0.150
8	$d_{5b}$ 0.166	$d_{3b}$ 0.137	$d_{5b}$ 0.116	$d_{3a}$ 0.114
9	$d_{5c}$ 0.166	$d_{5b}$ 0.136	$d_{3b}$ 0.111	$d_{5c}$ 0.102
10	$d_{1a}$ 0.000	$d_{3c}$ 0.134	$d_{3c}$ 0.107	$d_{5a}$ 0.082
11	$d_{1b}$ 0.000	$d_{2a}$ 0.058	$d_{2a}$ 0.097	$d_{5b}$ 0.065
12	$d_{1c}$ 0.000	$d_{2b}$ 0.053	$d_{2b}$ 0.088	$d_{3b}$ 0.045
13	$d_{2a}$ 0.000	$d_{2c}$ 0.045	$d_{2c}$ 0.075	$d_{3c}$ 0.037
14	$d_{2b}$ 0.000	$d_{1c}$ 0.004	$d_{1c}$ 0.007	$d_{1c}$ 0.015
15	$d_{2c}$ 0.000	$d_{1b}$ 0.004	$d_{1b}$ 0.006	$d_{1b}$ 0.013

表 2 と 3 の結果から、 $\alpha$  が大きい値を取るほど、色情報プロフィールに蓄積された色情報の主要色に似た配色の Web 文書が上位に検索されていることがわかる。例えば、表 2 において、色情報プロフィール 1 (主要色: 赤, カーキ) を用いた場合は、 $d_{5c}$  (主要色: 赤, 水色) が 9 位 ( $\alpha=0.0$ ) から 1 位 ( $\alpha=0.3$ ) に検索順位を上げている。また、 $d_{5c}$  の単語メタデータから、 $d_{5c}$  はスマートフォン等のモバイルデバイスを対象とした音楽配信サービスに関連したトピックであることがわかる。

以上の結果は、提案方式を用いることにより、利用者は Web 文書の内容に加えて、Web ページに含まれる配色情報を考慮した Web ページの獲得が可能であることを示している。

#### 5. おわりに

本研究では、利用者の閲覧した Web 文書から抽出される色情報プロフィールに基づいた Web ランキング方式を示

した。実験結果から、提案方式を用いることにより、利用者は Web 文書の内容に加えて、Web ページに含まれる配色情報を考慮した Web ページの獲得が可能であることを確認できた。

本研究では、人々は、Web ページのある特定のトピックや形式において、意識的ないしは潜在的に色の嗜好性を持っており、それに基づいた Web 文書の獲得・閲覧を行っているかと仮定している。しかしながら、日常的に閲覧している新聞サイトの記事や、こまめにチェックしているブログ記事等、配色に関する嗜好性とは無関係であるが、閲覧頻度の多さから色情報がプロフィールに蓄積されてしまう Web ページが多数存在する。これらの Web ページの色情報がプロフィールに蓄積すると、利用者にとって関心の高い色と判断される配色を持った Web ページの提示が困難となる。このような問題に対処するには、個々の利用者にとって、色の嗜好性が反映されるトピックを抽出し、かつ習慣的に閲覧しているような特定の Web サイトのページを除外する機能等が有効であると考えられる。

今後の課題として、複数の利用者から得られる色情報プロフィールを用いた検索実験により、提案方式の有用性を検証するとともに、利用者の配色に関する嗜好性と閲覧する Web ページの配色の関連性を評価するための実証実験を行っていくことが挙げられる。

## 参考文献

- 1) Baeza-Yates, R. and Ribeiro-Neto, B. Modern Information Retrieval, Addison Wesley, 1999.
- 2) Deerwester, S., Dumais, S., Furnas, G.W., Landauer, T.K. and Harshman, R., Indexing by latent semantic analysis, Journal of the American Society for Information Science, Vol. 41, No. 6, pp.391-407, 1990.
- 3) Jannach, D., Zanker, M., Felfernig, A., Friedrich, G, Recommender Systems: An Introduction, Cambridge University Press, 2010.
- 4) Hearst, M. A., Search User Interfaces, Cambridge University Press, 2009.