

印象語による Web 検索

小林達矢[†] 千種康民[†] 服部泰造[‡]

[†]東京工科大学 [‡]東京国際大学

1. はじめに

検索エンジンの研究は盛んに行われており、テキスト文書の検索システムはかなり実用的になっている。しかし、画像検索については、テキストとの連動においてまだ研究の余地がある。

本研究では、印象語に注目し、特定の印象語の有するカラーパターン・データベースを用いて、そのデータベースを介して、キーワードとしての印象語と Web サイトにて利用されているカラーパターンの相関を調べるシステムを考案し、その有効性について報告をおこなう。

処理方法としては、Web ページを 1 枚の画像として扱い、使用されている色とその画素数による式を定義し、それによって印象語を決定する。



図 1 検索エンジン

2. 研究内容

2.1 セーフカラーへの減色

セーフカラーとはコンピュータの違いによらずどんな環境でもこれらの色が必ず表示できると言われている色である。セーフカラーでは RGB の値をそれぞれを 6 段階(0, 51, 102, 153, 204, 255)に分けて 216 色が定義されている。元の Web ページのフルカラー画像をこのセーフカラーに減色する。

2.2 画像中の色情報の解析と背景色の決定

減色された画像をピクセルごとに使用色とその使用ピクセル数を集計する。背景色は、図 2 のように画像 4 隅の 100×100 ピクセルを別に集計し、その部分で多く使われており、且つ画像全体でも多く使われている色を背景色とする。



図 2 背景色の決定方法

2.3 リファレンスについて

文献[1]では 160 語の印象語についての配色が収録されている。1 語につき 10~16 色が示されており、これらの色を使うとその印象を持たせることができる、という意味となっている。[1]の各印象語において使用されるカラーパターンをリファレンスとする。

2.4 印象語の決定

ここでは各印象語と比較したい Web ページの類似度の計算式を示す。この式は、比較したい Web ページに対してリファレンス中の全印象語 160 種類とその差を計算し、その値が小さいほどその印象語に近くなるように定式化を行う。Web ページで使用されている色を集計した結果、最も使われている色と背景色を含めた N 色($i=1, 2, \dots, N$)を R, G, B の値を持ったベクトルを u_i とし、リファレンスの k 番目に対する色集合、最大 16 色($j=1, 2, \dots, 15, 16$)の R, G, B の値を持ったベクトルを v_{jk} とする。集計した Web ページ中の各色に対する画素数を S_i 、リファレンス中の色の優先度による係数を P_{jk} とすると、k 番目のリファレンスに対する評価値は以下のように定義する。

$$E_k = \sum_{i=1}^N \min_j (|u_i - v_{jk}| \cdot S_i \cdot P_{jk}) \quad (1)$$

Search engine by Impression-Word

Tatsuya KOBAYASHI[†], Yasutami CHIGUSA[†] and
Taizou HATTORI[‡]

[†]Tokyo University of Technology

[‡]Tokyo International University

ただし、 j の数値に関しては同じ値を用いないようにする。したがって、最終的には

$$E = \min_k (E_k) \quad (2)$$

を求めることにより Web ページに最も近いリファレンス探し、その Web ページの印象とする。

3. 解析結果例

$N=5$ として図3～図7の Web ページの解析を行うと、表1～表5のような結果になる。

結果の表は合致すると思われる印象語上位5語を掲載している。数値が小さいほどその印象語が合致するということになる。

図3は「機械的な」「正式な」「インテリ風」、図4は「激しい」「衝撃的な」「充実した」、図5は「粋な」「インテリ風」「日本的な」、図6は「冷静な」「機械的な」「儀礼的な」、図7は「インテリ風」「機械的な」「正式な」などの印象を与える Web ページであるという結果になる。見た目の印象どおりの印象語が選ばれている場合もあれば、そうでない場合もあることから、色の画素数による影響 S_i やリファレンス中の色の優先度による係数 P_{jk} の調整など、まだ研究の余地があると言える。

4. システムの評価

このシステムでの印象語抽出について評価を行う。 $N=5$ として60のサンプル Web ページを解析し、それぞれの結果の上位10以内の印象語がその Web ページの印象に合致しているかを手作業で集計した。結果、約73%はその Web ページに合致した印象が取得できた。

5. まとめ

現段階では解析結果の約73%がその Web ページに合致する印象語を抽出してくれ、本システムの有効性を確認した。色の画素数による影響 S_i やリファレンス中の色の優先度による係数 P_{jk} の調整により、さらに精度は上昇すると思われる。

そして、このような処理をさまざまな Web ページに対して行い、出力された数値を利用して検索エンジンを実現、また Web だけでなくパンフレットや広告、ラベルなどのデザインなどに対しても同様の処理で印象評価を行うなど、さまざまなことに応用できるのではないかと考えている。

参考文献

[1]セーフカラーによる配色ガイド Web カラーコーディネート 南雲治嘉著

表 1 解析結果 1

機械的な	0
正式な	9. 807692
インテリ風	14. 28216
モダン	16. 06084
ネイビー	30. 42946

図 3 Web ページ 1

表 2 解析結果 2

激しい	35. 31937
衝撃的な	46. 19776
充実した	60. 51619
豪華な	61. 36137
重厚な	61. 52675

図 4 Web ページ 2

表 3 解析結果 3

粋な	41. 8176
インテリ風	44. 0351
日本的な	48. 24543
モダン	48. 32443
気高い	50. 35985

図 5 Web ページ 3

表 4 解析結果 4

冷静な	13. 35646
機械的な	13. 35646
儀礼的な	13. 35646
男性的な	13. 60847
フォーマル	13. 73808

図 6 Web ページ 4

表 5 解析結果 5

インテリ風	23. 76478
機械的な	30. 76023
正式な	31. 95128
シャープな	37. 14096
格調のある	45. 39881

図 7 Web ページ 5