

ユーザ視点によるレイアウト・色彩情報に基づく Web 検索システム

宮本洋文[†] 平川正人^{††}

概要 近年、情報化社会の構築に伴い Web の世界は飛躍的な成長を続けている。現代の Web データは個性的な特徴を持つようになってきている。しかし、従来の Web 検索のほとんどがキーワードに頼っている。個性的な特徴を持つ Web データの特徴をキーワードだけで表現することが困難であることは Web 検索における課題の一つである。本研究では、レイアウト及び色彩の情報を検索キーとして、類似する Web ページを抽出する新しい Web 検索システムを提案する。提案システムはソーシャル・ブックマーク・サービスのシステム上に構築しており、Web ページには、ユーザによりレイアウトと色彩情報が付与される。ユーザは、情報付与時と同様の方法により条件付けして関連ページの検索を行うことができる。

A Web Search System Based on Page Layout and Color Features

Hirofumi Miyamoto[†] Masahito Hirakawa^{††}

Abstract The Web continues rapid growth with the development of information-oriented society. Web pages in recent days have unique features in their contents. However, the web search is still carried out by placing textual keyword(s). Users find difficulties in choosing appropriate keywords. This paper proposes a web search system which is capable of finding similar web pages based on their layout and color. The system is built on a social bookmark system. When the user adds a certain web page to his/her bookmark, the layout and color of its blocks are specified as well by the user. It is allowed for the user to search web pages by specifying a condition of layout and/or color in the same way with page registration.

1. はじめに

近年、情報化社会の構築に伴い、Web の世界は飛躍的な成長を続けている。従来の Web の世界にあっては、ユーザは情報を享受するだけであり、Web 上のデータのほとんどがテキストデータであった。現代では、ユーザは情報を得るだけでなく、自ら発信し他者や社会と交流するようになってきている。Web 上のデータについては、コンテンツをユーザ自身でデザインし、画像や動画が用いられるようになるなど、個性的な視覚的特徴を強く持つようになってきている。

これに伴い、より個性的な特徴を持つようになった Web データに対する特徴をキーワードだけで表現し、自らが望む的確な検索結果を得ることは困難となり、現代の情報化社会における課題の一つである。ユーザが Web データにアクセスする主な窓口は検索エンジンである。検索エンジンは、その開発当初と比べると、検索エンジンに合わせたキーワード入力を補助するオプション機能、画像検索、動画検索、時間軸による検索、サイトのプレビューを添えた検索結果表示、関連検索ワードの視覚化など、様々な機能を提供するようになってきている。しかしながら、これらの検索に対してユーザは、依然としてキーワードによる条件付けしか行うことができない。

本研究では、Web ページを対象とし、ユーザの視点を表現し反映するという観点から、そのような問題にアプローチを試みる。Web ページの視覚的特徴に関して、検索の基準、ユーザの持つイメージというのはユーザごとに異なるが、このことを反映する検索、機構の提供を図る。従来も同様のアプローチは存在するが、コンピュータに処理を全面的にまかせることが多かった。我々は、コンピュータが自動で判断することの限界を認め、ユーザとの協力の下に実現することを考える。具体的には、Web ページが保持しているレイアウト及び色の情報をユーザが検索キーとして Web ページに付与することを可能にすることで、ユーザが視覚的判断の基準構築に参加する。また、これらを直感的、視覚的に記述可能にし、ユーザが多く必要の要請を獲得できる Web 検索システムを提案する。提案システムは、ソーシャル・ブックマーク・サービスのシステム上に構築しており、Web ページには、ユーザが予めレイアウトと色の情報を付与するものとする。この情報付与の方法として、表示された Web ページの上に、縦横方向の分割線を用いてレイアウトの構造を記述すると、区切られた各領域における色情報をシステム側が自動で取得する。併せて、この色情報は、ユーザに提示される。検索にあたっては、想定する Web ページのレイアウトと色についての条件付けを、情報付与時と同様の方法によりユーザ自身で作成する方式及び具体的な Web ページを

[†] 島根大学大学院 総合理工学研究科 数理・情報システム学専攻
Interdisciplinary Graduate School of Science and Engineering, Shimane University

^{††} 島根大学 総合理工学部 数理・情報システム学科
Interdisciplinary Faculty of Science and Engineering, Shimane University

例に用いる方式の2通りを提供する。現在の Web の規格を変えずに新しい検索機能の実現を図る。更には、本システムの実現により、保持する Web ページの特徴情報からユーザの視覚的感性の特徴を抽出し、その特徴情報に基づいて、ユーザ同士を繋ぎ、そこから新しい知の発見と交流を可能にすることも考えたい。

2. 関連研究

ユーザの主観的な判断基準の違いを考慮した視覚感性のモデル化手法に関する研究として、文献[1]がある。この研究では、局所性を考慮した画像解析によってユーザの視覚感性をモデル化し、ユーザの主観的な画像類似度判定基準を構築している。これに基づいて画像検索することでユーザ主観に近づくことを可能にしている。視覚感性のモデル化にあたっては、予めユーザが複数のサンプル画像を自身で分類を行う。検索キーは画像であり、モデル化された判断基準に基づき例示画像を解析し、データベース内の画像と照合する。ユーザはこの判断基準を確認することはできない。これに対し、本研究では、検索基準を描画クエリとして作成することが可能であり、ユーザの感性のモデル化を必要としない点で違いがある。しかしながら、ユーザによって注目部位や重視する特徴が異なることに着目した視覚感性の構築というユーザ視点の理念は近いものがあり共感を得る。

一方、Web ページを視覚的に検索する研究として文献[2]-[6]がある。文献[2]は、ユーザが Web ページレイアウトのイメージをスケッチすることによって、空間的な位置関係を重視し類似する Web ページを検索する。スケッチは、テキストや画像などのパーツ（オブジェクト）を配置することにより行う。なお、HTML によって作成された Web ページのみを対象としている。ユーザが Web ページのイメージを描くことで検索が可能であり、条件付けレイアウトの自由度は高い。しかし、色については用いられておらず、Web ページの解析はシステムが行っている。また、後に述べるが、対象の Web ページの形式が限定されている。

文献[3][4]は、画像検索の手法を Web ページ検索に応用している。色情報は、固定化されたレイアウト領域を画像化した Web ページに適用し、抽出される。さらに、領域内の色は決められた近似色に集約され、領域内の近似色の割合を特徴量として記述される。ユーザは、固定化されたレイアウト領域に色を指定することによって類似する Web ページを検索する。また、Web ページの解釈はシステムが行う。これらのアプローチと比べ本研究は、レイアウトと色彩の指定に自由度が高い点の特徴となる。そして、Web ページの解析にはユーザの関与を求めているという点で違いがある。

文献[5]は、画像処理によってレイアウト解析を行い、ページ同士の類似の比較を行う研究である。アイコン、画像、テキストなどを画像処理で判別して分類し、ページ

イメージの領域分割を行う。分割された領域の集合に対してグラフ理論を用いて空間的位置関係の考慮をした比較を実現している。取り扱える Web ページは、Internet Explorer6.0 に限定されている。さらに、サンプルの対象が企業サイトの TOP ページのみである。本研究では、サンプルの Web ページを提示することだけでなく、ユーザが自ら描画クエリを作成することができる。また、TOP ページだけでなくサイト全体のページを対象としている点で異なっている。

文献[6]は、通常の実験結果に Web ページ画像を添えて提示することによって、検索効率の向上を図っている。この研究においてイメージ情報とは、キーワード検索による結果の補助的役割として位置づけられる。イメージは、検索結果の Web ページにおいて、ユーザの検索要求に関連する部位の画像を抽出し提示するものである。この結果に対し、ユーザは関連するものが適切に選ばれているか評価することが可能であり、この評価によってシステムは学習的に抽出能力を修正する。テキストのみの結果表示よりもイメージを添えた結果表示の方がユーザの検索タスクを効率良くすることができるという報告がされている。Google 社の Page previews 機能は、この研究のシステムと似ている。本研究とは、検索に描画クエリを用いるかどうかの違いがある。

これらの既存アプローチと比べて本研究では、視覚の特徴情報としてレイアウト及び色彩の両方を用いつつ、検索の基準作成にユーザを参加させる。さらに、描画クエリを用いた検索を実現する。また、ユーザが解析・検索の基準に参加する中で、領域の分割を行い領域ごとの色評価を行うことで、ユーザ視点による局所性を考慮した基準構築を実現する。

3. システム概要

3.1 Web データへのアプローチ

本研究における処理手順を図1に示す。これについては、ソーシャル・ブックマーク・サービスのシステム形態を参考にしている。大別すると、ユーザが Web ページに対してレイアウトと色の情報を付与する登録システム、レイアウトと色の情報に基づき類似する Web ページの検索を行う検索システム、URL アドレスとユーザによるレイアウトと色の情報を保持しているシステム内 Web ページ領域からなる。

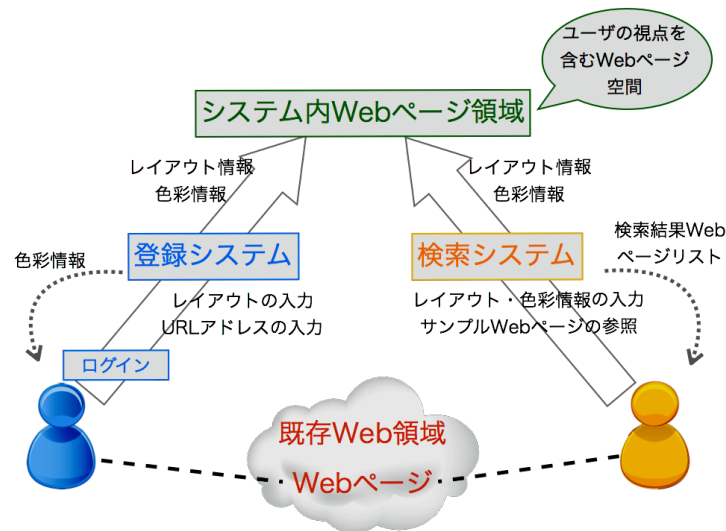


図 1 処理手順

3.2 システム構成

次に、提案するシステムのシステム構成を図 2 に示す。提案システムは、ユーザインタフェースを実現するアプリケーション部、データ通信とデータベースを実現するサーバ部からなる。アプリケーション部は、サーバ部に対して稼働の有無の確認とレイアウト・色情報などのデータ通信を行う機能、既存 Web ページ群に対してアドレスの有無と Web データを取得する機能を提供する。サーバ部は、アプリケーション部に対してデータ通信を行うとともに、データベースを操作・管理する。

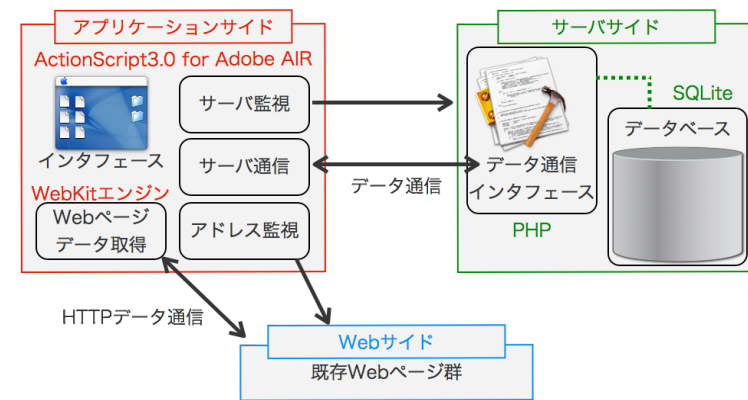


図 2 システム構成

4. 登録システム

3.1 節で述べた登録システムについて説明する。登録システムのユーザインタフェースを図 3 に示す。登録システムは、ベース情報入力部とレイアウト・色情報入力部からなる。ベース情報入力部では、通常のソーシャル・ブックマーク・サービスと同様に、登録する Web ページにタイトル、URL アドレス、タグ、コメントを記述する。ベース情報入力部で URL アドレスが入力されると、Web ページが参照可能であるか調べ、レイアウト・色情報入力部に Web ページ画像が表示される。レイアウト・色情報入力部では、表示された Web ページ画像上に縦線と横線を用いてユーザ自身が抱くレイアウトの構造を描く。縦線と横線は Web ページ画像上部にあるアイコンを選択し、Web ページ画像上にマウスで描画する。また、戻るボタンとクリアを用いることで入力を取り消すことができる。縦線と横線によって構造が描かれている様子を図 4 に示す。ユーザが入力したレイアウト情報に基づき、分割された領域内の色情報はカラーボタンを押すことによって確認することができる。色情報は、領域内の色から特徴的な 6 種を取得している。色情報を表示している様子を図 5 に示す。

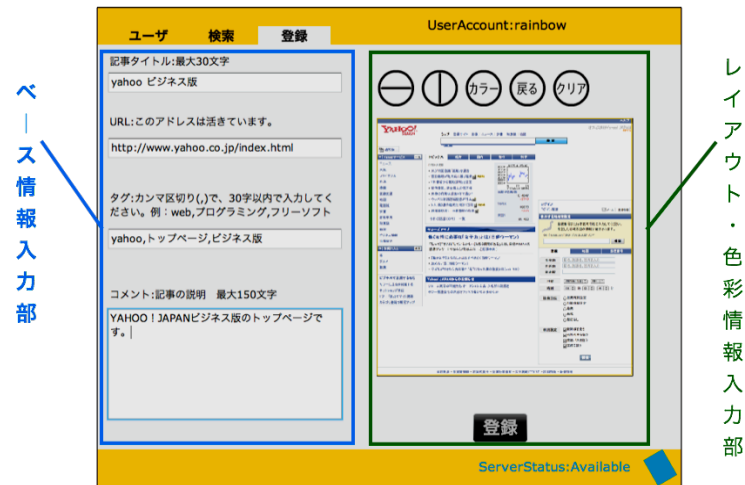


図 3 登録システムユーザインタフェース

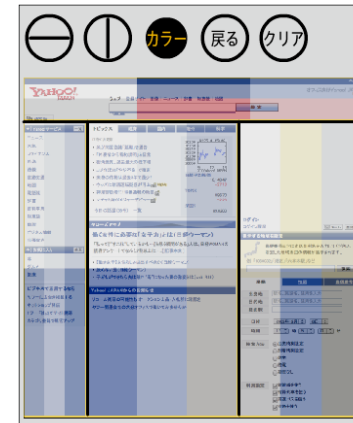


図 5 色彩情報の表示

レイアウト情報の入力完了し、ユーザが登録ボタンを押すとシステムはサーバへユーザ名、タイトル、URL アドレス、タグ、コメント、レイアウト情報、領域と色情報を送信する。レイアウト情報と色情報の詳細は、第 6 章で説明する。

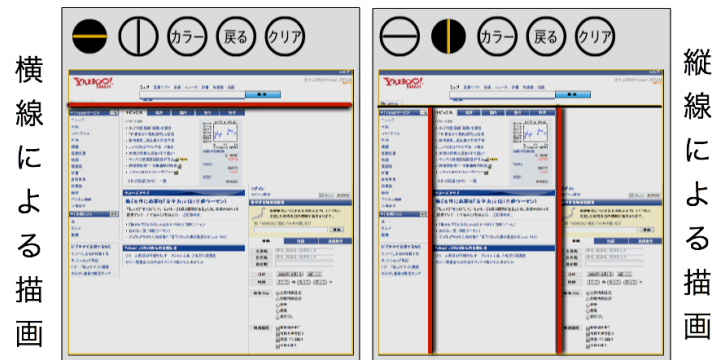


図 4 縦線と横線による構造の描画

5. 検索システム

3.1 節で述べた検索システムについて説明する。検索システムのユーザインタフェースを図 6 に示す。検索システムは、検索結果リスト部とレイアウト・色情報入力部からなる。検索結果リスト部で検索サンプルを選択する方法とレイアウト・色情報入力部でユーザ自ら描画クエリを生成する方法の 2 通りで検索することができる。検索結果リスト部では、レイアウト・色情報入力部で入力された描画クエリに基づき検索を行った結果がリスト表示される。また、初期状態では、データベース内の最新登録順にリスト表示される。リスト内の情報は、タイトルと登録ユーザ、タグ、登録時間のみの簡素な表示にしている。検索結果リスト部でリストオブジェクトを選択すると、詳細な登録情報が表示されるとともに、登録されているレイアウト・色情報がレイアウト・色情報入力部にセットされる。自ら描画クエリを生成するには、検索ボタン上部にあるサンプル生成を選択し、レイアウト・色情報入力部に、登録システムの場合と同様に縦線と横線を用いてユーザ自身が抱くレイアウトの構造を描く。

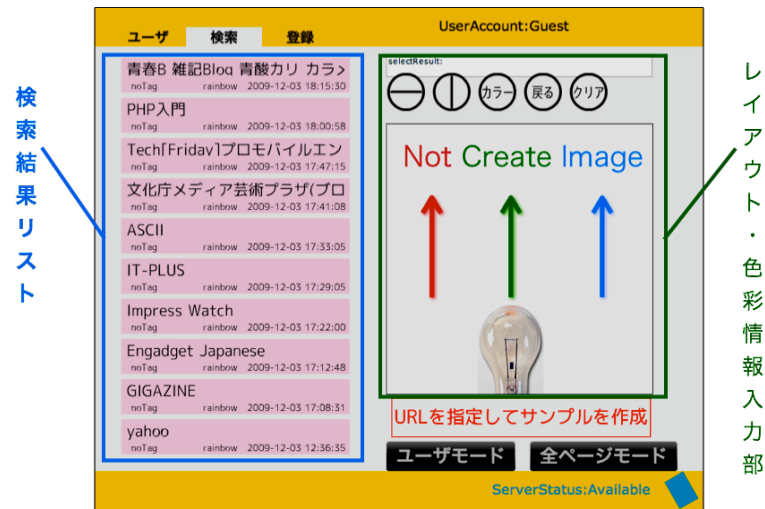


図 6 検索システムのユーザインタフェース

情報の入力が終わると検索ボタンによって検索を実行することができる。検索ボタンは、ユーザページモードと全ページモードの2つがある。ユーザページモードで検索すると、ユーザが登録したページ郡に対して検索が実行され、全ページモードで検索すると、システムに登録されている全ページ郡に対して検索が実行される。

本システムの検索システムは、現在構築中であり、レイアウト情報や色情報などの検索基準の組み合わせを指定できる機能を追加予定である。

6. 特徴情報

6.1 レイアウト・色情報の特徴量

本研究では、レイアウトと色彩の情報を特徴量として用いている。レイアウト情報は、ユーザが入力した分割線それぞれについて縦・横の別、線の始点座標、線の長さの情報を保持する。また、線によって区切られた領域において、領域の中心点座標、領域の幅と高さの情報を保持する。色情報は、それぞれの分割領域ごとの特徴的な色を6種のいずれかに集約して記述する。この処理にあたっては分割領域内の画素に対して $L^*a^*b^*$ 色空間での KMeans 法による非階層型クラスターリングを用いた[7]。 $L^*a^*b^*$ 色空間は、知覚的な色彩の距離・間隔に類似するよう設計されている色空間として広く利用されており、ユークリッド距離で色差を求めることが可能である。クラ

スタリングを行うにあたって実用的な処理時間を実現するため、クラスは6、ノード数(画素数)は4000に縮尺すると設定した。

これらの特徴情報は、登録システムを経てデータベースに格納される。データベースでは、分割線の情報、領域の情報、色の情報を個別に利用することができる設計にしている。また、本システムは Web ページ画像を生成するが、この画像の蓄積は行わない。

6.2 特徴情報の評価空間

類似ページの検索は、特徴情報が持つ評価空間において特徴量を比較することで行う。特徴情報の評価空間のイメージを図7に示す。色情報に関しては、 $L^*a^*b^*$ 色空間のユークリッド距離による色差感覚の評価としてNBS単位(米国標準局)による評価基準を利用する。この色差感覚の評価基準を表1に示す。

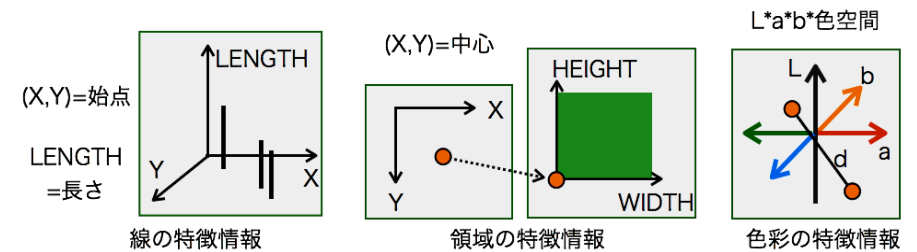


図 7 特徴情報の評価空間イメージ

表 1 色差感覚評価基準

dE (色差値)	色差感覚	
0~0.5	Trace	かすかに感じる
0.5~1.5	Slight	わずかに感じる
1.5~3.0	Noticeable	かなり感じる
3.0~6.0	Appreaceable	目立って感じる
6.0~12	Much	大きい
12以上	Very Much	非常に大きい

7. 考察

本システムにおける課題と展望について述べる。本システムは、2つの課題を持っている。そのひとつとして、本システムは登録時にユーザ指定によるレイアウトと色情報を格納する。この情報は、その後に対象 Web ページが大きく変化されたとしても自動的に追従するようには設計されていない。故に、実際にレイアウトと色彩による検索を行って出てきた結果が異なる可能性がある。しかし、登録したユーザ本人の記憶の面では、Web ページの変化前の記憶を保持していることがほとんどであるため、記憶のイメージから検索するには適していると言えよう。但し、Web ページ作成者がこのシステムを利用し、変更と同時にレイアウトと色の情報を変更することで対応することはできる。2つ目の課題は、動画や動きのある広告などは、取得することができないことである。WebKit エンジンでのレンダリングにより、リアルタイム且つ Web ページの記述形式に左右されなくなったものの、動画像の領域は、写し出し、色情報を取得することができない。

次に、本研究の展望として、領域に対する操作の拡大が考えられる。文献[2]の研究にもあるように、テキストやイメージなどオブジェクトの情報を付与することで、より詳細な構造を記録・指定することが可能になり、文献[1][5]のように位置関係を考慮することで検索精度の向上に繋がるのではないかと考えられる。また、ユーザが登録によって保持している Web ページの特徴情報からユーザの視覚的感性の特徴を構築することが考えられる。このユーザ自身の特徴情報により、他のユーザ同士を繋ぎ、そこから新しい知の発見と交流を実現していきたいと考えている。

8. おわりに

本研究では、ユーザが Web ページに対して思い描くレイアウト及び色の情報に基づいた Web 検索システムについて提案した。今後は、ユーザの登録システムの利用を通して、自身の検索要求に対してフィードバックが得られているかどうかを評価する必要がある。また、検索基準に関して、レイアウト情報のみを用いた場合、色情報のみを用いた場合、両方組み合わせ合わせた場合、また評価式の調整によるユーザの体感の変化を盛り込みたいと考えている。

参考文献

1) 多田昌裕, 加藤俊一: 階層的分類を用いた視覚感性のモデル化と類似画像検索への応用, 情報処理学会論文誌, Vol.44, No. SIG 8, pp. 37-45 (2003).

- 2) 橋本泰成, 五十嵐健夫: レイアウトによる WEB ページ検索, 情報処理学会シンポジウム論文集, Vol.2004, No.5, pp. 113-120 (2004).
- 3) 石井幹直, 細川宜秀, 高橋直久: 色とその配置位置に基づいた視覚的印象による Web ページ検索手法の実現方式, 電子情報通信学会 データ工学ワークショップ, 4B-i6 (2005).
- 4) 能美礼, 大野成義, 石川博: 色情報を利用した Web ページ検索手法について, 情報処理学会研究報告. データベース・システム研究会報告, Vol.2008, No.88, pp. 313-318 (2008).
- 5) Takama, Y. and Mitsuhashi, N.: Visual Similarity Comparison for Web Page Retrieval, IEEE/WIC/ACM International Conference on Web Intelligence, pp. 301-304 (2005).
- 6) Xiao-Bing Xue, Zhi-Hua Zhou, Zhongfei(Mark)Zhang: Improving Web Search Using Image Snippets, ACM Transactions on Internet Technology, Vol.8, No.4 pp. 1-28 (2008).
- 7) 白仁田和彦, 林健一郎, 大坪昭文: 色彩に基づいたデザイン画像の印象語付与システム, 日本機械学会論文集, Vol.70, No.689, pp. 192-199 (2004).