クエリに応じたファセットの動的抽出によるWeb画像検索結果の提示

川野　悠† 大島　裕明†† 田中　克己††

† 京都大学工学部情報学科　〒606–8501 京都府京都市左京区吉田本町
†† 京都大学大学院情報学研究科社会情報学専攻　〒606–8501 京都府京都市左京区吉田本町
E-mail: †{kawano,ohshima,tanaka}@dl.kuis.kyoto-u.ac.jp

あらまし　本稿では，Web画像検索におけるクエリに応じてファセットを動的に抽出することで，Web画像検索結果をある観点から分類して表示する手法を提案する．画像にはそれぞれ特徴を表すような属性と属性値があると仮定し，複数の画像に存在するような属性をファセットとして抽出する．そこで本稿では，既存の関連語発見手法を利用してWeb検索で得られた単語から属性値となる候補を選び，それらの上位語を発見することで属性と属性値を抽出する．さらに画像特徴量を用いて，得られた属性と属性値が画像を分類するために適切かどうかを判断する．

キーワード　Web画像検索，ファセット抽出，分類

1. はじめに

近年，ブロードバンドの普及により画像や動画などのマルチメディアデータをWeb上で取り扱う機会が増えている．それに伴い，Web上に存在する大量のマルチメディアデータを検索する需要も高まっている．Yahoo! Japan(注1)やGoogle(注2)などの主要なWeb検索エンジンでは，テキストをクエリとする画像検索（TBIR）が可能である．一般に，クエリとなるキーワードを入力すると，画像検索エンジンは周辺テキストにキーワードを含むような画像に対してランキングを行い，その結果を表示する．しかし，現在のTBIRではユーザの求めている画像を即座に見つけるのは容易ではない．例えば，京都に観光に行きたいがどの季節に行くべきか悩んでいるようなユーザを想定する．そのユーザが京都の観光名所の画像を見てどの季節に行くべきかを決める場合，現在のTBIRでは「京都」「観光名所」というクエリでWeb画像検索を行い，結果の画像集合からそれぞれの季節を表すような画像をユーザ自身で判断しなければならない．しかし，このような作業は非常に面倒であり，ユーザの判断が適切である保証もない．この場合，Web画像検索の結果を季節別の画像集合に分類して提示する方が望ましい．他にも以下のような例が挙げられる．

- 人気タレントの画像を年代ごとに見る
- ジャケットを素材別に見る
- ゆるキャラの画像を都道府県別に見る

これらの例も同様に，画像を分類して提示することでユーザの目的を果たすことができると考えられる．このように，Web画像検索結果をある観点から分類することは有益であると言える．

そこで本稿では，クエリに応じてファセットを動的に抽出することで，Web画像検索結果を様々な観点から分類して表示する手法を提案する．ファセットとは物事を見る側面のことで，本稿では画像を分類するための観点として定義する．例えば，「ディズニー」というクエリによって得られる画像集合を分類するファセットとして「キャラクター」や「映画」などがある．

ファセットによって分類するメリットは多角的な検索を行うことができる点である．従来のカテゴリによる階層的な分類では，まず「男性」か「女性」を選んで，その次に「未成年」か「成人」を選んでいくようにファセットを選択する順序が決められている．しかし，これでは選択の自由が制限されるだけでなく，ファセットの数が多いと階層が深くなってしまい構造も複雑になるという欠点がある．そこで複数のファセットを提示することで，ユーザは自分の興味のあるファセットから画像を絞り込むことができる．

ファセットを用いた分類ではあらかじめ決められたファセットに従って分類することが多い．しかし，我々はファセットをクエリに応じて動的に抽出することでそのクエリ特有のファセットを発見し，より柔軟な分類を行うことを可能とした。

この提案手法が大きな効果を発揮できるケースは主に2つある．1つ目はクエリが曖昧な場合であり，2つ目はユーザがクエリに対して無知な場合である．まず前者は，クエリが曖昧だと検索結果の画像に多種多彩な画像が現れ，分類することで検索結果が見やすくなる．後者は分類後のファセットからクエリに関する情報が得られると考えられる．例えば，靴が欲しいがどのような靴があるのかわからないので靴全体の画像を見てみたいというユーザを想定した場合，従来の一元的に並べられたWeb画像検索結果から目的の靴を絞り込むことは難しい．しかし，ファセットごとに画像を表示することで「衝撃吸収に優れた靴」や「通気性の良い靴」などの機能をユーザは知ることができ，靴選びの参考になると期待される．

本研究では，画像の周辺テキストには画像の特徴を表すような語が存在することに注目した．画像には属性とその属性値が存在し，それらの単語が画像の属性値になると仮定することで，複数の画像に共通する属性が画像を分類するファセットになる

(注1)：http://www.yahoo.co.jp/
(注2)：http://www.google.co.jp/

と考えられる．しかし，画像の周辺テキストには画像の特徴とは無関係な単語も存在し，また逆に画像の特徴となりうる単語が周辺テキストに現れるとも限らない．そこで，まずクエリの特徴を表すような語を抽出した上でファセットの抽出を行う．その後各ファセットに対して画像を割り当てることで，画像のもつ特徴を表す語による分類が可能となる．最後に画像特徴量を用いて，ファセットと割り当てられた画像が分類として適切なのか判断する．

本稿の構成は以下のとおりである．2 節では関連研究について述べ，本研究と既存の研究との違いについて述べる．3 節では提案手法について詳細に述べる．4 節では実験とその評価について述べ，得られた評価についての考察を述べる．5 節では，本稿では実現できなかった発展的課題や実験を行うことで新たに発見された課題について述べる．最後に 6 節で本稿のまとめを述べる．

2. 関 連 研 究

2.1 Web 画像検索結果の分類

Web 画像検索結果の画像を分類するような研究は盛んに行われている．Ambai ら [1] は VisualRank [2] を改良し，画像同士の類似度を算出した後クラスタリングを行うことで，結果を類似画像ごとに並列に提示するシステムを考案した．本研究が分類にテキストを用いるのに対して，画像類似度に基づいて分類している点が根本的に異なる．Jing ら [3] は全ての画像検索結果を実時間でセマンティックにクラスタリングする手法を提案している．この手法ではクラスタ同士の関係は考慮されておらず，ファセットを用いて分類を行う本手法とは異なる．

2.2 ファセットを用いた分類

Yee ら [4] は与えられた画像をそれぞれファセットに割り当てることで，複数のファセットから画像を検索できるシステムを考案した．このシステムはインターフェースとして Flamenco(注3) を採用している．本手法がファセットをクエリに応じて動的に抽出するのに対し，このシステムは Web 画像検索ではなくドメインが限定された画像集合内での検索であり，あらかじめ決められたファセットに分類するという点で本手法とは異なる．

2.3 特定の関係にある語の抽出

特定の関係にある語とは，ある語に対する上位語や下位語などのことであり，これらの抽出に関する研究は盛んに行われている．以後特定の関係にある語を関連語と呼ぶこととする．WordNet [5] は英語の概念辞書であり，語の簡単な定義や語同士の関係が記されている．しかし WordNet は人工的な辞書なので，全ての語を網羅することは不可能である．大規模なテキストコーパスやデータセットを利用して関連語を抽出する手法にも様々なものがある．Hearst ら [6] は “such as” のような構文パターンに着目して，上位語や下位語を抽出する手法を提案した．Ghahramani ら [7] はベイズ推定を用いて，同位語を共起テーブルのような大規模データから発見する手法を提案した．同位語とは共通の上位語をもつ，同じカテゴリに属するような語のことである．このアルゴリズムは単純かつ高速であるが，EachMovie や Grolier encyclopedia のような大量のデータセットを必要とする．Lin ら [8] は係り受け解析をした大量のテキストデータから類義語を発見する手法について提案している．大島ら [9] はある語の前後に接続する 2 種類の構文パターンを用いて関連語を抽出する手法を提案した．この手法は Web 検索結果のみを用いており，関連語の取得も高速かつ比較的精度も良いので，本手法内の関連語抽出にはこの手法を用いている．

(注3)：http://flamenco.berkeley.edu/

3. クエリに応じたファセットの動的抽出

本節では Web 画像検索のクエリに応じたファセットの動的抽出手法について述べる．まず抽出するファセットの種類について，次に手法の概要，その後に本手法を 4 つのフェーズの手法に分けて詳しく述べる．

3.1 ファセットの種類

分類を行う上で，ファセットはファセット名とファセット値を持つと考えられる．ファセット名とは画像が持つ観点の名前であり，ファセット値とは実際に画像を分類するための項目である．例えば,「川野悠」は「性別」というファセット名のファセット値「男」に分類されると同時に,「職業」というファセット名のファセット値「大学生」に分類される．

ここで，どのような種類のファセットが画像の分類に適しているか考える必要がある．例として，以下のようなファセットが挙げられる．

- 真偽を問うファセット
- 序列に関するファセット
- クラス名となるファセット

真偽を問うファセットとは例えば「リンゴである」といったもので，対応するファセット値は「真」と「偽」のみである．序列に関するファセットとは「アルファベット順」「50 音順」などで，ファセット値としては「A」から「Z」までを順番に羅列したものなどが考えられる．クラス名となるファセットとは，対応するファセット値がクラス名に対するインスタンス名となるようなものである．例えばクラスが「人」といった抽象的なものに対して,「大島裕明」や「田中克己」など具体的なものがインスタンスにあたる．多くの場合でクラスはインスタンスの上位概念になっており，またどのようなドメインに対しても適用できると考え，今回はインスタンス名に対してその上位語にあたるクラス名を取得することでファセットを抽出する手法を提案する．

3.2 本手法の概要

大まかな流れを図 1 に示す．本手法は大きく分けて，画像と付与される語の取得，ファセットの抽出，ファセットと画像の関連付け，妥当性判定の 4 つのフェーズに分けられる．

画像とそのキーワードの取得では Web 画像検索のためのクエリ q を入力とし，画像集合 $P = \{p_1, p_2, \ldots\}$ に含まれる p_i と p_i に付与される語集合 $Keywords(p_i)$ のペア $(p_i, Keywords(p_i))$ を得る．ファセットの抽出では入力に q を用いて，ファセット集合 $\mathcal{F} = \{F_1, F_2, \ldots\}$ を取得する．ただし，ファセット F_i は

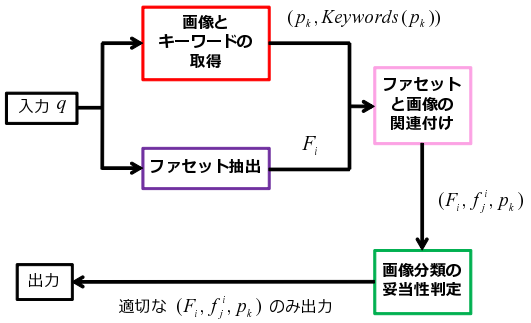

図 1　本手法の概要

$F_i = \{f_1^i, f_2^i, \ldots\}$ で構成される語集合である．本稿では f_j^i を F_i におけるファセット値と定義する．ファセットと画像の関連付けでは，ファセットの抽出によって得られた $f_j^i \in F_i$ と画像 $p_k \in P$ を関連付けて，$(f_j^i, p_k) \in F_i \times P$ を得る．妥当性判定ではファセット値 f_j^i に割り当てられた画像集合 P_j^i を用いて，(F_i, f_j^i, P_j^i) が画像の分類として適切なのか判断する．

3.3　画像とそのキーワードの取得

まず，Web 画像検索のためのクエリ q を用いて Web 画像検索を行い，画像集合 P を得る．$p_i \in P$ から $Keywords(p_i)$ を取得するためには，p_i に付随したテキストを用いる．例えば，p_i の周辺テキストに対し形態素解析を行い，名詞であると推定された語による集合を用いる方法や p_i に付与されたタグの語集合を用いる方法が考えられる．ただし，画像に付随しているテキストは Web 画像検索エンジンに依存するので，うまく使い分ける必要がある．

3.4　関連語発見手法を用いたファセットの抽出

ここではファセット集合 $\mathcal{F}$ を取得する手順を述べる．

（1） Web 画像検索のためのクエリ q を使って，ファセット値の候補となる語集合 T を得る．

（2） T の単語を節点とみなし，同位語関係を表すグラフから閉路を発見する．得られた閉路に含まれる節点集合を同位語集合 S と定義する．

（3） $t \in S$ に対し，上位語候補の集合 H_t をそれぞれ取得する．

（4） $W \subseteqq S$ である語集合 W の要素 w において，任意の w に対して H_w に含まれる共通の上位語を $Name(F_i)$ とし，W を F_i として抽出する．$Name(F_i)$ とは F_i のファセット名である．

3.4.1　ファセット値候補の抽出

(1) で述べたようにファセット値の候補となる語集合 T を取得する．T はクエリの特徴を表すような語集合であるとし，T を得る方法として Web 画像検索で得られた画像の周辺テキストから抽出する方法や Web ページ検索で得られたタイトルとスニペットから抽出する方法などが考えられる．

3.4.2　同位語集合の発見

本手法で抽出するファセット名 $Name(F_i)$ が f_j^i の上位語にあたることは先に述べた．分類を行う上で，f_j^i と f_k^i は同位語であることが望ましい．そこで，T の単語間の同位関係を調べる．我々の手法 [10] を使うと，ある語 t に対して t の同位語候補の集合 C_t が得られる．しかし，この手法を使って A の同位語として B が得られても，B の同位語として A が得られるとは限らない．このような場合 A と B が同位語なのかを判断することは難しく，同位語集合を発見するためにはこの手法だけでは不十分である．そこで T の単語を節点とみなし，各節点の同位語関係を表すグラフ $G = (V, E)$ を作成する．V と E は以下のように定義される．

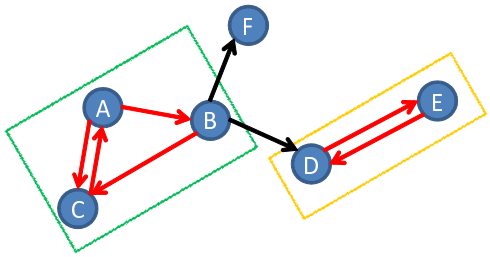

図 2　同位関係を表すグラフと得られる閉路

$$V = \{\ v_i \mid \forall i,\ v_i \text{は } t_i \in T \text{ を表す節点}\ \}$$

$$E = \{\ (t_i, t_j) \mid t_j \in C_{t_i},\ i \neq j\}$$

つまり G において節点は T 中の語であり，ある節点 t_i の同位語候補に別の節点 t_j が含まれていれば，t_i から t_j に有向枝を引くということである．次にグラフ G の各節点 v に対して，v を含む節点集合の要素数が最大の閉路を発見する．これはある節点から有向枝を辿っていき，自分自身の節点に戻ってくるならば，その過程で通過した節点は同位語である可能性が高いという仮定に基づくものである．例えば図 2 のようなグラフが得られた場合，閉路の探索を行うことで同位語集合 {A,B,C} と {D,E} が得られる．このようにして語集合 T から同位語集合をそれぞれ得る．

3.4.3　上位語の抽出

同位語集合 S ごとに，含まれる全ての語に共通する上位語を求めファセットとする．上位語の抽出には両方向構文パターンを用いた手法 [9] を用いる．ある語 t に対して t の関連語候補の集合 X_t が得られ，$x \in X_t$ はそれぞれスコア $\mathrm{Score}(X_t, x)$ を持つ．$\mathrm{Score}(X_t, x)$ の値が大きいほど x は t の関連語である可能性が高いと言える．この手法では前後に接続する構文パターンを変えることで，上位語や下位語などの様々な関連語を発見することが可能である．例えば，上位語の抽出には上位語を求めたい語の前に接続するパターンとして「といえば」，後に接続するパターンとして「である」などが考えられる．

しかし，求める上位語が全ての語の上位語候補に現れることは極めて稀である．そこで S に含まれる語は同位関係にあるという制約を用いて 2 種類の手法を提案する．

1 つ目は両方向構文パターンを用いた手法を利用する．本来両方向構文パターンを用いた手法では，前後に接続する構文パターンに現れる語のうちいずれの構文パターンにも出現する語を抽出する．ここで同位関係にあたる 2 語を入力とした場合，両方向構文パターンによる抽出条件を緩和することを考える．例えば，入力として「ミッキーマウス」と「ミニーマウス」が与えられ,「ミッキーマウス」の上位語として「キャラクター」

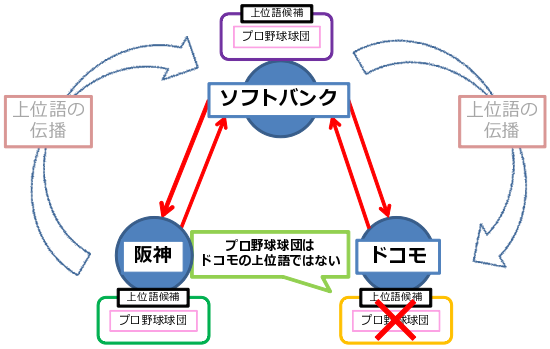

図 3　同位語集合の例

という語が得られた場合,「ミニーマウス」の上位語を求める際に「キャラクター」という語が片方の構文パターンにしか現れなかったとしても,「ミニーマウス」の上位語として「キャラクター」を抽出するようにする．このように両方向構文パターンによる手法を緩和することで，上位語候補の数が増えると考えられる．

2 つ目は先ほどの同位語関係を表したグラフを利用する．S に含まれる語 t_i と t_j に対し，$(t_i, t_j) \in E$ かつ $(t_j, t_i) \in E$ であれば，t_i と t_j は相互関係にあると定義し，お互いの上位語候補を共有することとする．手順は以下のとおりである．

- 各同位語集合に含まれる語 $t \in S$ に対し，両方向構文パターンを用いて上位語候補の集合 H_t をそれぞれ取得する．
- $t_i, t_j \in S$ が相互関係にあれば，H_{t_i} と H_{t_j} は互いを共有する．つまり以下の式が成立する．

$$H_{t_i}^{new} = H_{t_j}^{new} \equiv H_{t_i} \cup H_{t_j}$$

ただし，上位語の共有はその他の語には伝播させないものとする．例えば，図 3 のような同位語集合を想定する.「阪神」の上位語として「プロ野球球団」が得られたとすると，上位語の共有により「ソフトバンク」の上位語として「プロ野球球団」が加えられる.「ソフトバンク」と「ドコモ」が相互関係にあるからといって,「阪神」の上位語であった「プロ野球球団」を直接の相互関係でない「ドコモ」にも加えてしまうことは「ドコモ」の上位語に「プロ野球球団」が存在するということになり，不都合が生じる．上位語の共有に制限を設けることはこのような事態を避けるためである．

これら 2 種類の手法はいずれも，本来得られた上位語候補には現れなかった上位語候補を取得するための手法であり，これらの手法を用いることで共通の上位語をより多く取得できると考えられる．

次に任意の $t \in S$ に対して共通の上位語が存在すれば，同位語集合 S を F_i，その上位語をファセット名 $Name(F_i)$ として抽出する．

$$\bigcap_{t \in S} H_t^{new} \neq \phi \Longrightarrow F_i = S,\ Name(F_i) = h_{max}$$

ただし，h_{max} は $h \in \bigcap_{t \in S} H_t^{new}$ に対し，$\mathrm{Score}(H_t^{new}, h)$ の平均値が最大の h である．

しかし，S に共通の上位語が存在するとは限らない．図 3 は S に「阪神」と「ソフトバンク」,「ドコモ」と「ソフトバンク」という 2 つの同位語集合を含み，共通の上位語は存在しない．そこで，S に複数の同位語集合が含まれる場合を考える．n 個の上位語 $h_1, h_2, \ldots h_n$ を選択し，それらの同位語集合 W_i の和によって S が網羅されているか判断する．以下の条件が成立する場合，n 個の上位語 $h_1, h_2, \ldots h_n$ を要素とする集合 Hyp_n を得る．

$$\forall w \in W_i,\ h_i \in \bigcap_i H_w^{new} \tag{1}$$

$$\bigcup_i^n W_i = S \tag{2}$$

ただし，n は (1) と (2) の式を満たす Hyp_n が少なくとも一つ存在する最小の自然数である．S に共通の上位語が存在するのは，$n = 1$ で (1) と (2) の式を満たす Hyp_1 が得られる特別な場合である．

上位語 h_i の選び方によっては (1) と (2) の式を満たす h_i の組み合わせは複数考えられる．本来ならば，任意の $t \in S$ から生成される全ての上位語候補 $\bigcup_{t \in S} H_t^{new}$ の中から (1) と (2) の式を満たす n 個の上位語を選ぶことで Hyp_n を取得し，Hyp_n が複数存在する場合はその中から最も適切な Hyp_n を選択する必要がある．しかし，n が大きくなるにつれ，$\bigcup_{t \in S} H_t^{new}$ の中から n 個の上位語を選ぶ組み合わせは膨大なものとなり，その計算量はとても実用的であるとは言えない．そこで，$\bigcup_{t \in S} H_t^{new}$ に含まれる上位語をある規則に従って並び替えたものを SHL とし，SHL の先頭に近い上位語を優先的に選択していき，(1) と (2) の式を満たす Hyp_n を取得する．$\bigcup_{t \in S} H_t^{new}$ の要素を並べ替えるための規則は以下のとおりである．

- $h_i, h_j \in \bigcup_{t \in S} H_t^{new}$ に対して，$n(W_i) \neq n(W_j)$ であれば，$n(W_i)$ と $n(W_j)$ のうち大きい方の上位語を先頭寄りに並び替える．$n(W_i)$ は h_i の同位語集合 W_i の要素数である．
- $n(W_i) = n(W_j)$ であれば，$average(h_i)$ と $average(h_j)$ のうち大きい方の上位語を先頭寄りに並び替える．ただし，$average(h_i)$ は (1) を満たす w に対する $\mathrm{Score}(H_w^{new}, h_i)$ の平均値である．

つまり，$\bigcup_{t \in S} H_t^{new}$ に含まれる上位語 h を $n(W)$ に対して降順でソート，$n(W)$ が等しい h に対しては $average(h)$ に対して降順でソートをして得られたものが SHL である．SHL は $n(W)$ の大きいものから順に並んでいるので，SHL の先頭に近い上位語を優先的に選択していくことで，できるだけ少ない数の上位語で Hyp_n が得られると考えられる．

このようにして得られた Hyp_n の要素 h を $Name(F_i)$，その同位語集合 W を F_i としてそれぞれ抽出する．以上の工程を全ての同位語集合について行い，最終的にファセット集合 $\mathcal{F}$ が得られる．

3.5　ファセットと画像の関連付け

前のフェーズで得られたファセット F_i と Web 画像検索によって得られた画像 $p_k \in P$ の関連付けを行う手順を説明する．f_j^i と p_k の割り当てには p_k に付与される語集合 $Keywords(p_k)$ を用いる．$Keywords(p_k)$ には p_k の特徴を表す語が含まれて

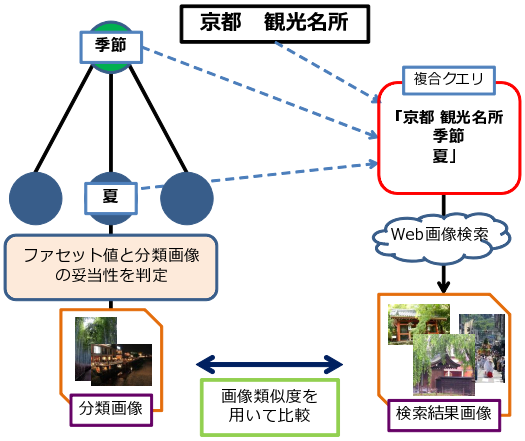

図 4　画像分類の妥当性判定手法の例

図 5　システムのスクリーンショット

いると考えられるからである．そこで，$\mathrm{Rel}(f_j^i, Keywords(p_k))$ を導入する．$\mathrm{Rel}(f_j^i, Keywords(p_k))$ は f_j^i が p_k の特徴を表す語となるような度合いを示す．$\mathrm{Rel}(f_j^i, Keywords(p_k))$ がある閾値 σ 以上の値をとれば，ファセット値 f_j^i に p_k を割り当てて (f_j^i, p_k) を得る．

3.6　画像類似度を用いた画像分類の妥当性判定

これまでの過程でファセット名 $Name(F_i)$ と F_i のファセット値 f_j^i，f_j^i に割り当てられた画像集合 P_j^i を得ることができた．しかし，F_i が画像の分類として適切なのかまでは考慮されていない．例えば,「京都 観光名所」というクエリに対して，ファセット名として「季節」，ファセット値のひとつとして「夏」が得られたとする．ここで,「夏」に分類された画像が夏を表すには程遠い画像であることも起こりうる．これは夏を表す画像に付随したテキストに「夏」という語が含まれるとは限らないからである.「夏」に分類された画像集合に夏を表す画像が含まれないことは，画像の分類として適切であるとは言い難い．そこで我々は画像類似度を用いて，$(Name(F_i), f_j^i, P_j^i)$ の妥当性を統合的に判断する手法を提案する．本手法では，画像類似度を測る特徴量として Color Coherence Vector [11] を用いる．Color Coherence Vector(CCV) は空間上の色の疎密を表現でき，本手法にも十分適用できると考えられる．手順は以下のとおりである．

（1）　入力クエリ q に対する F_i のファセット値 f_j^i ごとに，q と $Name(F_i)$，f_j^i を空白で接続した複合クエリを用いて Web 画像検索を行う．図 4 の例では，クエリは「京都 観光名所 季節 夏」となる．

（2）　Web 画像検索結果の上位 k 件の画像集合を R とし，f_j^i に割り当てられた画像 $p \in P_j^i$ と R 内の画像 r の画像類似度を以下のように定義する．

$$\mathrm{Sim}(p, r) = \frac{\mathrm{CCV}(p) \cdot \mathrm{CCV}(r)}{|\mathrm{CCV}(p)||\mathrm{CCV}(r)|}$$

ここで，$\mathrm{CCV}(i)$ は画像 i の Color Coherence Vector である．全ての r に対して $\mathrm{Sim}(p, r)$ を計算し，その平均値がある閾値 θ 以上となれば p はこの分類に適切であると判断する．

（3）　P_j^i 中の δ 以上の画像が適切であるとされた場合，P_j^i は F_i，f_j^i への割り当てとして妥当であると判断し，f_j^i をファセット値として残す．妥当でないと判断された場合は，F_i から f_j^i を取り除く．

（4）　1. と 3. を全ての f_j^i について行い，$F_i = \phi$ となれば F_i を $\mathcal{F}$ から取り除く．

（5）　1. から 4. までを全ての F_i に対して行う．

以上が提案手法の詳細である．本手法には閾値 σ，θ，δ が含まれているので，精度良く抽出するためにはそれらを適切に設定する必要がある．

4.　実　　験

本手法では，Web 画像検索結果の画像を分類するのにふさわしいと考えられるファセットを抽出した後，ファセットに割り当てられた画像がそのファセットによって表されるような特徴を持つのかを判断するため，分類の妥当性を判定している．そこで，ファセットの抽出と妥当性判定の有用性を確かめるためシステムのプロトタイプを C# を用いて実装し，2 種類の実験を行った．実装の際に必要となる Web 検索エンジンからのデータの取得や自然言語処理には，C# ライブラリの SlothLib [12] を使用した．

4.1　システムの実行例

図 5 に今回実装したプロトタイプの実行例を示す．ユーザが検索したい画像に対するクエリを入力することで，システムがクエリに応じて動的に抽出したファセットを提示する．さらにその中から興味のあるファセットを選択するとそのファセットに関する項目が表示され，ユーザは選択した観点によって分類された画像をみることができる．この例では「神戸 夜景」というクエリに対して,「場所」というファセットから分類した際の「モザイク」の画像を表示している．

4.2　実 験 設 定

今回の実験では，画像とそのテキスト情報の収集に Yahoo! Japan によって提供されている API(注4)を使い，検索結果の中で周辺テキストがついた上位 100 枚の画像とその周辺テキストを取得した．得られた周辺テキストから，形態素解析器

(注4)：http://developer.yahoo.co.jp/webapi/search/

表 1　実験に使用したクエリ一覧

抽象的なクエリ	「仏教」「正月」「ディズニー」「入学式」「宇宙」
具体的なクエリ	「花束」「神戸 夜景」「家電」「和食」「日食」

MeCab(注5)を用いて名詞または名詞句と推定される語を抽出し，それぞれの画像に対するキーワード集合を得た．上位語の抽出に関しては，事前実験により精度のよい上位語候補が比較的多く得られた両方向構文パターンによる抽出条件を緩和した手法を用いた．今回ファセットと画像の関連付けは以下の場合に行った．あるファセット F_i におけるファセット値 f_j^i と画像 p_k は，ある $keyword \in Keywords(p_k)$ に対して，f_j^i が表す概念が $keyword$ の表す概念を含んでいれば f_j^i に p_k を関連付ける．例えば「家電」というファセット F_i のファセット値 f_j^i に「テレビ」，ある画像 p_k のキーワード $keyword$ に「液晶テレビ」が含まれていた場合,「テレビ」という概念は「液晶テレビ」という概念を含んでいると考えられ，f_j^i に p_k を関連付ける．

4.3　ファセット抽出の精度を測る実験

ファセット抽出の精度を測るにあたって，画像が割り当てられたファセットを対象として 抽出されたファセット数とその中に含まれる正解ファセットの数を人手で調べた．あるファセット F_i を正解であると判断するのは以下の基準を両方満たす場合である．

（1） ファセット名 $Name(F_i)$ が，入力クエリ q によって指定されるドメインが持つと考えられる観点となっている

（2） $Name(F_i)$ と F_i に含まれるあるファセット値 f_j^i が上位語と下位語の関係となっている

今後 (1) と (2) の基準を満たすファセット名 $Name(F_i)$ とそのファセット値 f_j^i のペアを正解ペアと呼ぶ．

また同時に，ファセット値候補の抽出に画像に付随するキーワード集合を用いる手法と，入力クエリを用いた Web ページ検索で得られたタイトルとスニペットを用いる手法では，ファセットの抽出の精度にどの程度の影響を与えるのかについても調査した．画像に付随するキーワード集合を用いる手法では，先ほど述べた画像の収集で取得したキーワード集合に含まれる語から出現頻度の高い上位 100 語をファセット値の候補とした．Web ページ検索で得られたタイトルとスニペットを用いる手法では，検索で得られた上位 100 件のページのタイトルとスニペットから名詞または名詞句と推定される語を抽出し，DF 値の高い上位 100 語をファセット値の候補として用いた．今回，概念や行事などそれ自身を画像として表現しにくいクエリを抽象的なクエリ，逆にそれ自身を画像として表現しやすいクエリを具体的なクエリと定義し，それぞれ 5 つずつ，計 10 個のクエリを用いて実験を行った．表 1 に使用したクエリの一覧を示す．

表 2 は全てのクエリに対して，平均取得ファセット数，平均正解ファセット数，平均適合率などを表したものである．この表を見て分かるとおり，ファセット値候補の抽出に Web ページ検索を利用した手法の方が正解ファセットの数，適合率共によい結果となっている．また Web ページ検索を利用した手法で

(注5)：http://mecab.sourceforge.net/

表 2　全てのクエリに対する結果

	Web ページ検索による抽出	画像キーワード集合からの抽出
平均取得ファセット数	9.3	7.2
平均正解ファセット数	2.6	1.3
最大正解ファセット数	5.0	3.0
最小正解ファセット数	1.0	0.0
平均適合率	0.28	0.18

表 3　抽象的なクエリに対する結果

	Web ページ検索による抽出	画像キーワード集合からの抽出
平均取得ファセット数	9.8	8.6
平均正解ファセット数	2.6	1.2
平均適合率	0.27	0.14

表 4　具体的なクエリに対する結果

	Web ページ検索による抽出	画像キーワード集合からの抽出
平均取得ファセット数	8.8	5.8
平均正解ファセット数	2.6	1.4
平均適合率	0.30	0.24

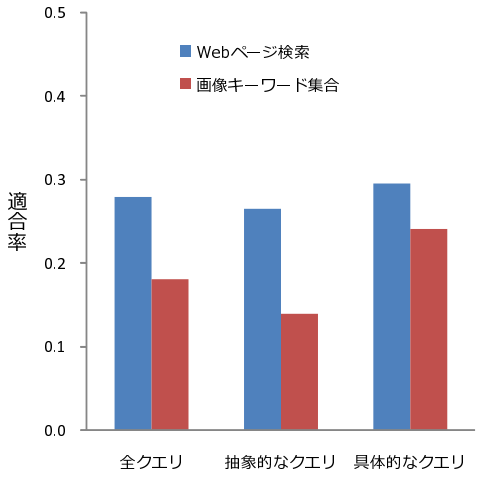

図 6　クエリの種類別の適合率の比較

は全てのクエリに対して最低 1 つのファセットが抽出されたが，画像に付随したキーワード集合を利用した手法ではファセットが抽出できないクエリも存在した．

これは Web 画像検索で得られた画像に付随した語の出現頻度が大きいからといって，その語が検索クエリに関連のある語であるとは限らないからである．また今回画像の周辺テキストからキーワード集合を得るために，形態素解析を行い名詞または名詞句と推定される語の抽出を行ったが，名詞句の抽出がうまく行われていない場合が見受けられた．例えば，画像の周辺テキストにある人物の肩書と氏名が続けて書かれている場合，肩書と氏名が分割されないまま名詞句として抽出されてしまっていた．このように現時点での特徴語の抽出手法には問題点がいくつか存在するので，改良の余地がある．

表 3 と表 4 はそれぞれ抽象的なクエリと具体的なクエリに対

表 5 「日食」におけるファセットの抽出例

ファセット名	現象	世界	離島
ファセット値	皆既日食	日本	奄美大島
	部分日食	中国	屋久島
	ダイヤモンドリング	アフリカ	

する，平均取得ファセット数と平均正解ファセット数，平均適合率を表したものである．また，図 6 は全てのクエリ，抽象的なクエリ，具体的なクエリに対する平均適合率を比較したものである．図 6 によるといずれの抽出方法においても，抽象的なクエリより具体的なクエリの方が高い適合率を誇っている．これは抽象的なクエリの場合，そのクエリに関連する語の範囲が幅広くクエリに関連する語と関連しない語における特徴量の差異が小さくなってしまい，うまく特徴語を抽出できなかったためだと考えられる．

特に，ファセット値候補の抽出手法に画像に付随するキーワード集合を用いた方は適合率にして 1 割以上の差がある．ここで注目すべきことは，表 3 と表 4 を比べると適合率の差を生みだしているのは正解ファセット数の差ではなく，取得ファセット数の差であるという点である．これは先ほど述べた，クエリに関連する語を表す特徴量がスパースになっていることを物語っており，特徴語ではない語同士で小さな同位語集合を形成した結果，クエリが持ちうる観点とは異なる上位語を抽出しているものだと考えられる．このことはファセット値候補の抽出手法に Web ページ検索を利用した手法にも言えるが，Web ページは一般的にある主題について書かれているので，そこから抽出されたタイトルやスニペットには同位語が含まれやすいのに対し，一枚の画像には様々な観点から得られるような語は含まれるが，同位語が同時に含まれることは少ない．よって，Web ページ検索を利用した手法は抽象的なクエリに対しても比較的精度を維持した状態で抽出できたのではないかと思われる．

表 5 に実際に「日食」というクエリで抽出したファセット名と対応するファセット値を示す．この場合，日食の画像を皆既日食や部分日食などの現象ごとや，日本や中国などの地域ごとに見ることができる．ここで「離島」というファセット名に注目すると，対応するファセット値である「奄美大島」や「屋久島」は「日本」の中の「離島」である「奄美大島」や「屋久島」であり，二つのファセット「世界」と「離島」は粒度の異なるファセットとみなすことができる．ユーザに対して粒度の異なるファセットを提示することは重要であるが，他のファセットと並列に提示することは望ましくない．もし粒度の異なるファセットが抽出されたら，例えばまず粒度の大きいファセットを提示し，ユーザがそのファセットを選択すれば粒度の小さいファセットを提示するなどの工夫が必要である．

4.4 画像分類の妥当性判定能力を測る実験

今回はファセット値候補の抽出に Web ページ検索を用いた場合のファセットの抽出で得られた 68 の正解ペアに対して，正解ペアに割り当てられた画像集合の中に含まれる適合画像を人手で判断し，適合画像を含む割合が δ 以上の場合の適合率，再現率，F 値を θ を変えてそれぞれ調査した．θ は二つの画像が類似しているかどうかを判断するための閾値で，δ はファセット値に割り当てられた画像集合の中に適合画像がどのくらいの割合で含まれていれば，そのファセット値が分類として適切かどうかを判断するための閾値である．妥当性判定手法の中で用いるクエリとファセット名，ファセットに対応するファセット値を組み合わせた複合クエリを用いた Web 画像検索によって得られる画像の数は 5 枚とした．妥当性判定を行わない場合をベースライン手法とし，θ が 0.0，0.5，0.6，0.7，0.8，0,9 の場合について実験を行い，その結果を表 6 に示す．ちなみに，δ は 0.8 に固定し，θ が 0.5 以下の場合は θ を 0.5 にした場合とほぼ同じ結果だったので，対象外とした．

表 6 妥当性判定手法の実験結果

閾値 θ	適合率	再現率	F 値
0.0(ベースライン)	0.26	1.00	0.42
0.5	0.28	0.94	**0.44**
0.6	0.26	0.83	0.40
0.7	0.30	0.78	**0.43**
0.8	0.43	0.50	**0.46**
0.9	1.00	0.22	0.36

この表から閾値 θ を大きくするにつれ，適合率は向上する一方で再現率は下降することがわかる．これは情報検索において適合率と再現率がトレードオフにあることを示しており，当然の結果と言える．F 値に注目すると，θ が 0.5，0.7，0.8 の場合にベースラインである閾値 $\theta = 0.0$，つまり妥当性判定を行わない場合を上回っていることがわかる．これより，閾値 θ を適切に設定することで妥当性判定手法は有効であると言える．

次に妥当性判定がうまく行えなかった例を挙げる．例えば，「ディズニー」というクエリでファセットの抽出を行った場合，「キャラクター」というファセット名に対して,「ミッキー」や「スティッチ」といったファセット値はどのような閾値に対してもうまく判定が行えていたが,「作品」というファセット名に対するファセット値「アニメ」は閾値によって判定結果はバラバラであった．これは「ディズニー」の「作品」の「アニメ」に関連のある画像は幅広く，典型的な画像が存在しないため「ディズニー 作品 アニメ」というクエリで Web 画像検索を行っても検索結果画像の集合に比較対象となる画像に類似したものが含まれなかったためだと思われる．

また別の例として,「神戸 夜景」というクエリに対してファセット名「場所」とファセット値「六甲山」という正解ペアが得られた際，その正解ペアに割り当てられた画像は全て典型的な夜景の画像だったにも関わらず，不適切な画像集合と判定されてしまった．これは妥当性判定を行うための検索クエリを作る際に,「神戸 夜景 場所 六甲山」という 4 語からなるクエリとなってしまい，検索結果画像が得られなかったためである．ちなみに,「神戸 夜景 六甲山」というクエリによって得られた画像を利用して妥当性判定を行うとうまく判定されていた．このように画像分類の妥当性判定を行う際に，適合画像として Web 画像検索で得られた結果画像集合を用いることで，妥当性判定の精度が Web 画像検索の精度に依存してしまう危険性がある

ので，より頑健な手法を考える必要がある．

また絵画の画像を検索する場合などで，あるファセット値に割り当てられた画像集合が必ずしも色の疎密だけで特徴づけられるとは限らない．例えば「絵画」というクエリが与えられ，「印象派」というファセット名に対して「ゴッホ」というファセット値が得られた場合，割り当てられている「ゴッホ」の絵の特徴を表すものが色の疎密であるとは限らず，力強いタッチであったり，他の印象派画家が描かないようなものを描く点であったりと色の疎密以外の特徴量で判断できる場合もある．このように画像類似度に用いる画像特徴量を CCV に固定せず，ファセット値に割り当てられた画像集合の特徴をより顕著に表す画像特徴量を用いた画像類似度で妥当性を判定するなど，手法を改良しなければならない．

5. 今後の課題

本稿では実現できなかった課題や実験を行うことで新たに発見された課題について述べる．

5.1 クエリとファセット間の関係性の考慮

本手法ではまず入力クエリの特徴語を抽出し，得られた特徴語の上位語を求めることでファセットの抽出を行っている．これは入力クエリから直接ファセットを求めることが困難であるため，入力クエリの特徴語を介してファセットを抽出しているという考え方もできる．よって本手法では入力クエリとファセット間の関係について直接的には考慮されていない．今回の実験の中でも適切な同位語集合が抽出できていたにも関わらず，上位語がうまく抽出されていないことが多々あった．これはできるだけ多くの同位語を含むような上位語を抽出しているため，漠然とした意味の上位語が抽出されてしまいファセットとして適切とは言えないものが抽出されてしまったと考えられる．そこで入力クエリとファセット間の関係を考慮した手法を加えることで，より精度の良いファセット抽出を行うことができると思われる．

5.2 システムの実時間処理

本手法中の重要な処理として，両方向構文パターンによる関連語発見と CCV による画像類似度の計算がある．両方向構文パターンによる関連語発見手法を用いてある語の関連語を発見するのに Web に数回アクセスするため，本手法のように数十件単位で同位語や上位語を抽出するには相当な時間を要する．また CCV による画像類似度の計算も同様に時間のかかる処理である．よって，本手法で実時間でファセットを抽出し，ファセットごとに画像を提示することは極めて困難である．現在のシステムでは到底実用的であるとは言えないので，Web アクセスを最低限に抑えつつ，できるだけコストのかからない手法で分類された画像を提示するシステムの実現を考えている．

6. ま と め

本稿では，Web 画像検索におけるクエリに応じてファセットを動的に抽出することで，Web 画像検索結果をある観点から分類して表示する手法を提案した．まずファセットの抽出の下準備としてファセット値の候補となる語集合の抽出を行い，それらを用いてファセットの抽出を行った．その後ファセットと画像の関連付けを行い，画像特徴量を用いてそれらが画像の分類として適切であるのかを判定した．2 通りのファセット値候補抽出手法でファセットの抽出の精度を評価し，ファセット値候補抽出手法として Web ページ検索を利用した場合，クエリあたり平均 2 個から 3 個のファセットを抽出することができた．また画像分類の妥当性判定手法を行う場合と行わない場合で，有意な差があるのか評価し，パラメータを適切に調節した場合妥当性判定手法が有効に働くことを示した．また，今回実験を行う中で新たな知見や課題が浮き彫りとなった．今後はシステムの実時間処理化を目標として，今回得られた課題にも取り組んでいく予定である．

謝辞

本研究の一部は，京都大学 GCOE プログラム「知識循環社会のための情報学教育研究拠点」，および，文部科学省科学研究費補助金（課題番号：18049041，21700105）によるものです．ここに記して謝意を表します．

文 献

[1] M. Ambai and Y. Yoshida: “Multiclass VisualRank: image ranking method in clustered subsets based on visual features”, Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, pp. 732–733 (2009).

[2] Y. Jing and S. Baluja: “VisualRank: Applying PageRank to large-scale image search”, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1877–1890 (2008).

[3] F. Jing, C. Wang, Y. Yao, K. Deng, L. Zhang and W. Ma: “IGroup: web image search results clustering”, Proceedings of the 14th annual ACM international conference on Multimedia, pp. 377–384 (2006).

[4] K. Yee, K. Swearingen, K. Li and M. Hearst: “Faceted metadata for image search and browsing”, Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 401–408 (2003).

[5] G. Miller: “WordNet: a lexical database for English”, Communications of the ACM, **38**, 11, pp. 39–41 (1995).

[6] M. Hearst: “Automatic acquisition of hyponyms from large text corpora”, Proceedings of the 14th International Conference on Computational linguistics, pp. 539–545 (1992).

[7] Z. Ghahramani and K. Heller: “Bayesian sets”, Proceedings of the 19th Annual Conference on Neural Information Processing Systems, pp. 435–442 (2005).

[8] D. Lin: “Automatic retrieval and clustering of similar words”, Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics, pp. 768–774 (1998).

[9] 大島裕明, 田中克己：“両方向構文パターンを用いた Web 検索エンジンからの高速関連語発見手法”, 情報処理学会研究報告, **88**, pp. 37–42 (2008).

[10] 大島裕明, 小山聡, 田中克己：“Web 検索エンジンのインデックスを用いた同位語とそのコンテキストの発見”, 情報処理学会論文誌 (トランザクション) データベース, **47**, pp. 98–112 (2006).

[11] G. Pass, R. Zabih and J. Miller: “Comparing images using color coherence vectors”, Proceedings of the 4th ACM international conference on Multimedia, pp. 65–73 (1997).

[12] 大島裕明, 中村聡史, 田中克己：“SlothLib: Web サーチ研究のためのプログラミングライブラリ”, DBSJ Letters, **6**, 1, pp. 113–116 (2007).