

## クエリ情報を用いた多言語タグ付けの効果 The effect of multilanguage tagging using the information of user query

高橋有礼<sup>†</sup>  
Arinori Takahashi

菱山玲子<sup>†</sup>  
Reiko Hishiyama

### 1. はじめに

Web上のマルチメディアコンテンツ取得の手法の1つに、コンテンツに対する記述であるタグ情報を活用することでマルチメディア情報を取得するものがある。しかし、タギングは個々の人間が恣意的に行っているため、同じ対象や性質のものであっても異なる言葉を付与されることが多く存在する。そのためタグ情報をそのまま活用する手法では、適切なタグが付与されていない場合には有効に機能せず、コンテンツ内容の記述もコンテンツに由来する特定の言語のみで行われているのが一般的である。

また情報検索全般にはユーザが目的を達成するために現在持っている知識では不十分だと感じている状態、情報要求 [1] と呼ばれる状態が存在する。この情報要求に対して Q1 から Q4 の4段階に分類した提案がされている [2]。この分類を元に画像をタグ情報を用いることによる検索を分類する。Q4である必要な情報の情報源が同定できるくらい問題が具体化された状態である「調整済みの要求」ではユーザは検索を行うだけでよい。しかし Q3である言語表現可能な「形式化された要求」であったり、Q2であるあいまいにしか言語表現することできない「意識された要求」、Q1である具体的に言語化することができない「直観的要求」の場合、そのユーザには適切なキーワードを選び出すことが困難であり、自らの意図に沿わないコンテンツを取得してしまう場合が多く存在する。

このため、ユーザの使用言語に依存した単言語による検索クエリやタグ情報を利用するだけでは、ユーザ意図に沿ったコンテンツの効果的な取得が困難と予測される。一方、ユーザにとっては一般に、検索クエリを多言語で適切に記述することは難しい。そこで、本研究では静止画像を対象として、機械翻訳サービスを利用してユーザの検索クエリを多言語化して生成する一方、不十分なタグ付け情報を自動的に適切に補うことを提案する。この提案について、ユーザ意図に沿った適切な画像取得の効果を実験から評価することを目的とする。

### 2. 関連研究

言語横断的な検索手法に関し、Partonら [3] は、検索クエリと文章の双方を翻訳することで検索精度の向上を提案している。この研究は言語横断的な多言語クエリの組み合わせによる効果が検証されていない。また Varshneyら [4] はユーザが知らない言語でも内容が記述されている情報に対しての検索をする際に、不完全な辞書や不正確な機械翻訳システム、同じ意味を表すが言語が異なる際に生じる不適切な翻訳によって適

切な結果を取得することのできない問題を解決するべく、コーパスを用いたクエリサジェストを提案している。しかしクエリから得られる地理情報などのメタ情報を考慮しておらず、検索クエリを単言語で利用している。画像検索を効果的に行う方法として、Zakariaら [5] はメタデータに依存することなく、ユーザ主体のタグ付けや地理参照メタデータによって生成された集合的な知識と組み合わせることで、コンテンツ内容記述の充実化を図った。しかし、これらにはユーザの検索意図を直に反映させた検索を考慮したものではなく、検索クエリの多言語化も検討されていない。また Leukenら [6] はユーザの要求が曖昧な場合に多様な結果を提示することで適切な画像の取得を支援するために、画像特徴量を用いた検索結果多様化手法を提案しているのだが、検索結果を得られない場合を考慮していない。

### 3. 提案

本研究では、ユーザが期待する画像の取得を実現するため、以下の方法を提案する。まず、正解画像を入手する過程に、中島ら [7] の差異増幅型フィードバックを適用し、ユーザの意図を反映した画像検索を実現する。このフィードバック手法を効果的に利用するためには、事前に検索対象となり得る画像を幅広く入手しておき、ユーザの適正画像の選択がより適切なものとなるよう導くことが必要である。すなわち、フィードバック時にユーザが適合画像として選択する画像群の候補集合を拡張することが効果的である。そこで、この候補集合の拡張を行うため、画像検索前に検索クエリの地理情報を取得し、取得した地理情報をもとに機械翻訳サービスを用いてユーザ使用言語で記述された検索クエリを多言語に翻訳すると同時に、メタデータとして多言語情報を活用しながら、効果的にこれらを組み合わせて利用する。更に内容記述が乏しいことが多い検索対象の画像群に対して、自動タグ付けを行うことで内容記述の充実を図る。今回は検索クエリの情報を元に自動タグ付けを行う。これにより、コンテンツ内容の記述を多言語で補填しながら、検索対象とするタグ候補の量を増大させることで、検索精度を向上させる。本研究の流れを図1に示す。

このアイデアは、例えば世界を旅するツーリストの観光情報の検索を支援するために有益である。特に同じ場所に何度も観光に行くような人に、価値ある情報を提供することができる。初めての観光客は、その観光地に存在する一般的な観光資源に興味がある [8]。一方、何度もその地を訪れているリピーターは、その観光スポットに関連しつつも、より一般的ではない観光資源にまつわる情報や訪問体験などに関心を持つようになり、更に詳しい情報を知りたがる [9]。つまり、図2に示すように、ヨーロッパへ何度も観光に訪れている

<sup>†</sup>早稲田大学創造理工学研究科経営システム工学専攻

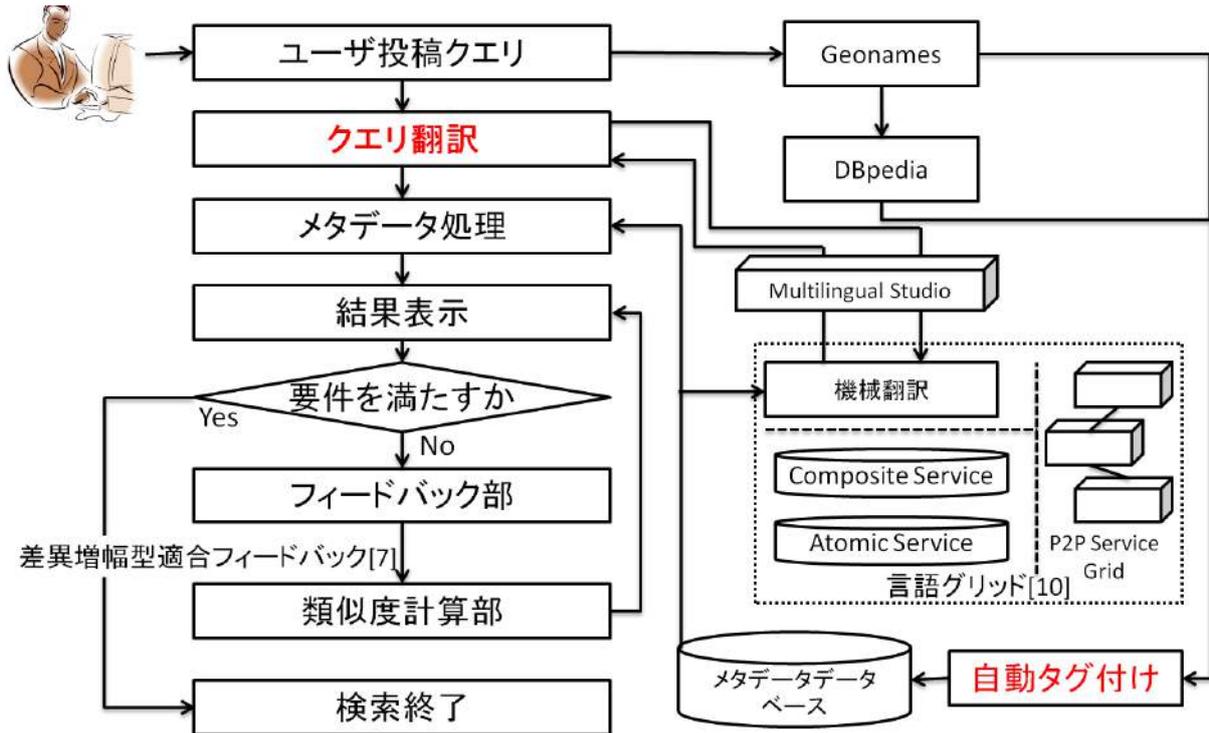


図1: 提案フロー

リピーターは、ロンドンにある観光スポットのより詳細な情報を求めるようになる。このように、同じ検索対象に対しても、検索者により異なる画像検索ニーズがあると考えられる。例えば、期間限定でライトアップされる色が違う画像であったり、主要な位置以外から撮られた画像である。こうした普段とは違う状態を知ること、観光客は次に観光に行く際に見るものを想像しやすくなる。

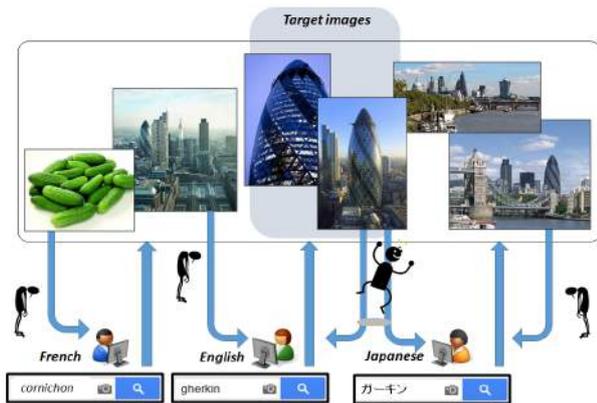


図2: Unusual Part

提案システムは、画像検索部、クエリ多言語化部、地理情報取得部、自動タグ付け部の4つの部分により構成されている。以下では、図1の各部分の機能について記述する。

#### 提案システムフロー

画像検索部は初めにユーザが投稿したクエリを受け取る。そしてそのクエリをメタデータ処理部に渡すことで地理情報を取得し、それをデータベースに入れる。次にクエリ翻訳部に取得した地理情報とともにユーザの使用言語で記述されたクエリを渡すことで、多言語化されたクエリを入手する。その後新たなクエリでメタデータ検索を行い、その検索結果をユーザに提示する。この時表示する画像数は、一般的な画面に収まるような量である30枚とした。表示結果が要件を満たさない場合ユーザは提示された画像の中から任意の画像を選択しフィードバックをかける。そして選択された画像とデータベース内の画像に対して類似度計算を行い、その結果をユーザに再度提示する。また検索対象に付与されているタグは予め翻訳し付与しておく。

#### 画像検索部

画像検索部では初めにユーザが投稿したクエリを受け取る。そしてそのクエリを地理情報取得部に渡すことで地理情報を取得し、それをデータベースに入れる。次にクエリ多言語化部に取得した地理情報とともにユーザの使用言語で記述されたクエリを渡すことで、多言語化されたクエリを入手する。その後新たなクエリでメタデータ検索を行い、その検索結果をユーザに提示する。この時表示する画像数は、一般的なPCディスプレイに収まるような量である30枚とした。表示結果が要件を満たさない場合ユーザは提示された画像の中から任意の画像を選択しフィードバックをかける。そして選択された画像とデータベース内の画像に対して類似度計算を行い、その結果をユーザに再度提示する。

### クエリ多言語化部

クエリ多言語化部は、ユーザ使用言語で記述されたクエリを、地理情報取得部によって取得された検索対象の現地言語を元に言語グリッド [10] を通すことで、現地言語や英語に翻訳し検索クエリに追加する。これにより、新たに多言語化された検索クエリが生成され、それを画像検索部に返す。言語グリッド [10] とは言語サービスを提供する Web サービスプラットフォームである。ユーザは言語グリッド上の多様な言語サービスを組み合わせて利用することができる。例えば、ユーザのクエリが「エッフェル塔+赤」であった場合言語グリッドを介することで図 4 に示すように現地言語であるフランス語の「tour+eiffel+rouge」に翻訳され、最終的に出力されるクエリとしては「エッフェル塔+赤+tour+eiffel+rouge」となる。

### 地理情報取得部

地理情報取得部は、検索クエリをもとに既存の Web サービス連携により地理情報を取得する。具体的な処理は以下のとおりである。まず、Geonames を用いて、対象を検索する。Geonames とは、全世界の地理データを収集検索可能としている Web サービスである。これにより対象を、地名・建造物名とそれ以外の 2 種類に分類する。また、DBpedia により、Geonames で地名・建造物名かどうかを判別したものと同一のものを取得するため、Geonames が独自に割り振っている GeonameID を取得する。その後、地名・建造物名に対しては、DBpedia から関連する地名を取得し、これをメタデータとして追加する。DBpedia とは、Wikipedia から情報を抽出して Linked Open Data として公開している Web サービスである。具体例をあげると、図 3 のように投入されたものがフランス語、英語、日本語で記述されたものであっても 1 つ目の Web サービスである Geonames を通すことで、地名・建造物名であれば 1 つの geocode に統一される。その後 2 つ目の Web サービスである DBpedia を介し、nativeNameLang 等を参照することで現地言語を取得することができる。

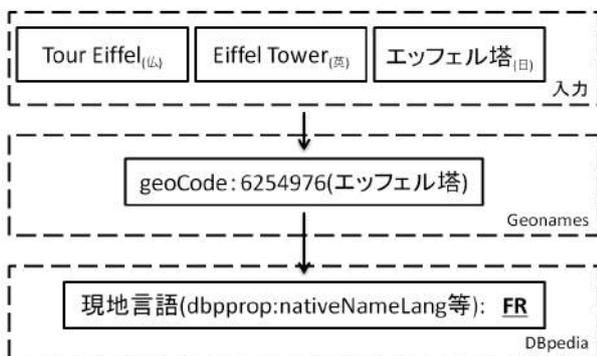


図 3: 言語情報取得フロー

### 自動タグ付け部

自動タグ付け部では、クエリ多言語化部と同様のプロセスで、既に画像に付与されている画像投稿者の使用言語で記述されているタグを検索クエリの情報をも

とに言語グリッド [10] を通すことで、多言語化し付与しなおしている。例えば、既に付与されているタグの「rouge」であった場合言語グリッドを介することで図 4 に示すようにクエリ言語である日本語の「赤」に翻訳され、最終的に付与されるタグとしては「rouge, 赤」となる。

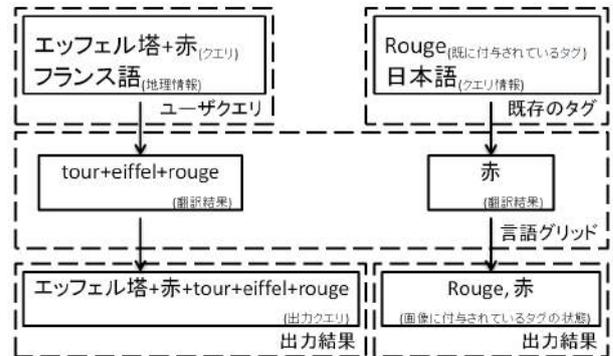


図 4: 多言語化フロー

### フィードバック部

フィードバック部はユーザの検索意図を反映した検索を実現する機能であり、中島ら [7] により提案された差異増幅型フィードバックを適用する。この方法は、ユーザが選択したコンテンツと選択しなかったコンテンツの差異を利用し、ユーザの意図をより強く反映させるものである。ここで用いる関数  $f_{DA}$  は正事例  $S_{pi}$ 、正事例の周辺画像  $neighbor(S_p)$ 、正事例の周辺画像数  $N_{S_{pi}}$  で定義される。 $\alpha$  は差異増幅の増幅係数であり、 $\epsilon$  は差異を増幅させるかの閾値である。 $S_{pi}$  に基づく次の検索質問  $Q_{i+1}$  を以下の式 (1) と定義する。

$$Q_{i+1} = (1-\beta) \times Q_i + \beta \times f_{DA}(S_{pi}, neighbor(S_p)) \quad (1)$$

ただし

$$f_{DA}(S_{pi}, neighbor(S_{pi})) = \begin{cases} S_{pi} + \alpha \times \frac{\sigma_{\epsilon neighbor(S_{pi})}(S_{pi}-S)}{N_{S_{pi}}} & (|\frac{\sigma_{\epsilon neighbor(S_{pi})}(S_{pi}-S)}{N_{S_{pi}}}| \geq \epsilon) \\ S_{pi} & (|\frac{\sigma_{\epsilon neighbor(S_{pi})}(S_{pi}-S)}{N_{S_{pi}}}| < \epsilon) \end{cases}$$

### 類似度計算部

類似度計算は、ユーザが選択した画像と検索結果の画像との類似度を 64 色に減色した状態で計算する。減色は画像の各ピクセルに対して、RGB の各成分を 4 等分し中央値に置き換えることで行う。そして減色後に 64 次元のカラーヒストグラムを作成する。カラーヒストグラム生成後は類似度を計算する。類似度計算には

HistogramIntersection[11]を用いる。この式(2)では $H_1$ が画像クエリのヒストグラムであり、 $H_2$ は比較対象のヒストグラムである。この二つのヒストグラムの同次元の要素 $i$ を比べ、値が小さい方を足し合わせていくことで2つの画像の類似度を求める。ただし画像の大きさにヒストグラムの要素数は依存するため、正規化を行っている。この類似度が高い順に画像を整頓し、その結果をユーザに対して示す。

$$\frac{\sum_{i=0}^{63} \min(H_1[i], H_2[i])}{\sum_{i=0}^{63} H_1[i]} \quad (2)$$

#### 4. 実験

評価実験は、自動タグ付けを行わない検索結果比較(実験1)、検索クエリの地理情報を用いた自動タグ付けを行った後の検索結果比較(実験2)により行った。

##### 4.1. 各実験の設定

各実験は検索対象が3つ存在する。1つ目の対象は「見上げた角度、赤色、エッフェル塔」を満たすものとし、2つ目の対象は「ロゴを読むことができる、赤色、アリアンツアリーナ」を満たすもの、3つ目の対象は「見上げた角度、緑色、エンパイアステートビル」を満たすものとした。対象の画像はFlickrから検索対象に関するタグがつけられた画像を取得した。検索は適合率が0の時、またはフィードバック回数が5回を超えた場合とし、終了時の結果はそれまでのピーク値とした。フィードバックをかける画像の選択基準は、検索結果最上位の画像とした。但し、同一画像は他の画像を選択することができる限り避け、その場合は次の順位の画像を選択した。

検索対象の選定基準は対象画像が多く、色が変化するものとした。例えば、シドニーオペラハウスでは、色の種類は豊富であるが、画像数が少ないため不適切である。また、検索クエリに「ライトアップ」や「イルミネーション」等の単語を追加しても、「色」が示す画像は「ライトアップ」の示す画像を含んでいるため検索精度に影響を及ぼさない。そのため、検索クエリは「場所の名前+色」とした。また検索クエリの言語は、現地言語と日本語、そして現地言語+日本語の3種類とした。

#### 5. 結果と考察

##### 5.1. 結果

各実験の、開始時と終了時の適合率を表1、表2、表3、表4、表5、表6に示し、検索結果の例として検索結果を図5、図6、図7、図8、図9、図10、図11、図12、図13、図14に示す。開始時結果とは図1の提案フローにおいてメタデータ検索後の初回の結果であり、終了時結果とは検索成功/失敗条件を満たした際の結果である。また、各検索結果として得られた画像を「異なる対象」、「ランドマークは満たしている」、「クエリは満たしている」、「正画像」の4つに分類し数え上げた。その結果を図15、図16、図17に示す。

表 1: 実験 1-1 結果

		仏	日	日+仏
適合率	開始時	0.30	0.00	0.33
	終了時	0.43	0.00	0.60
再現率	開始時	0.24	0.00	0.27
	終了時	0.35	0.00	0.49
F 値	開始時	0.27	0.00	0.30
	終了時	0.39	0.00	0.54

表 2: 実験 1-2 結果

		独	日	日+独
適合率	開始時	0.43	0.00	0.50
	終了時	0.60	0.00	0.63
再現率	開始時	0.16	0.00	0.19
	終了時	0.23	0.00	0.24
F 値	開始時	0.24	0.00	0.27
	終了時	0.33	0.00	0.35

表 3: 実験 1-3 結果

		英	日	日+英
適合率	開始時	0.200	0.00	0.267
	終了時	-	-	-
再現率	開始時	0.069	0.00	0.092
	終了時	-	-	-
F 値	開始時	0.002	0.00	0.003
	終了時	-	-	-

表 4: 実験 2-1 結果

		仏	日	日+仏
適合率	開始時	0.37	0.37	0.37
	終了時	0.57	0.57	0.53
再現率	開始時	0.30	0.30	0.30
	終了時	0.46	0.46	0.43
F 値	開始時	0.33	0.33	0.33
	終了時	0.51	0.51	0.47

表 5: 実験 2-2 結果

		独	日	日+独
適合率	開始時	0.40	0.00	0.37
	終了時	0.80	0.00	0.73
再現率	開始時	0.15	0.00	0.14
	終了時	0.33	0.00	0.28
F 値	開始時	0.22	0.00	0.14
	終了時	0.44	0.00	0.40

表 6: 実験 2-3 結果

		英	日	日+英
適合率	開始時	0.167	0.200	0.267
	終了時	-	-	-
再現率	開始時	0.057	0.069	0.092
	終了時	-	-	-
F 値	開始時	0.002	0.0002	0.003
	終了時	-	-	-



図 5: 実験 1 エッフェル塔+赤+Eiffel+Tour+Rouge 開始時結果

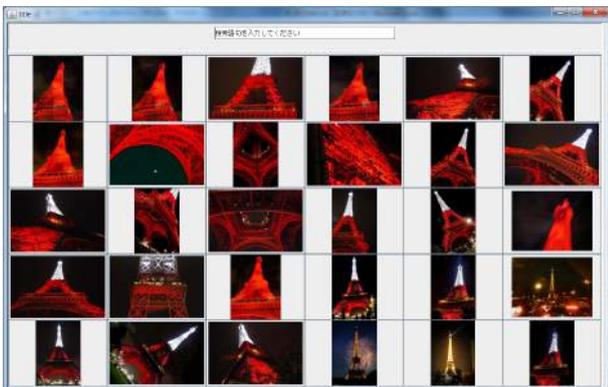


図 6: 実験 1 エッフェル塔+赤+Eiffel+Tour+Rouge 終了時結果

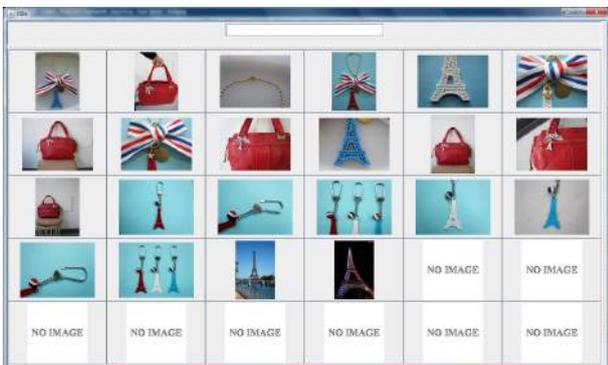


図 7: 実験 1 エッフェル塔+赤 開始時結果

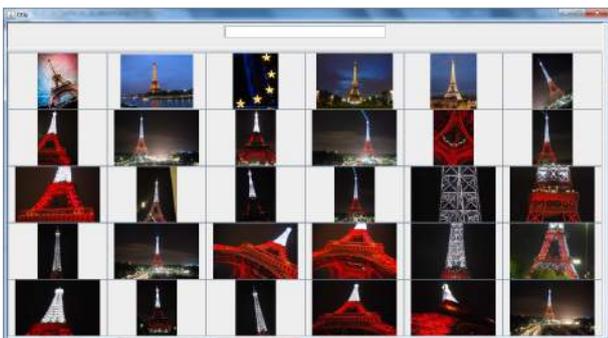


図 8: 実験 2 エッフェル塔+赤+Eiffel+Tour+Rouge 開始時結果



図 9: 実験 2 エッフェル塔+赤+Eiffel+Tour+Rouge 終了時結果

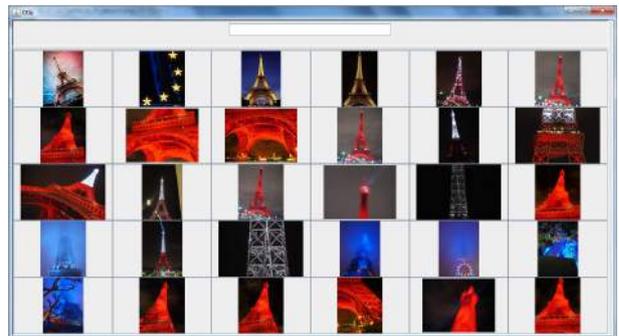


図 10: 実験 2 エッフェル塔+赤 開始時結果



図 11: 実験 1 Allianz+Arena+Rot 開始時結果

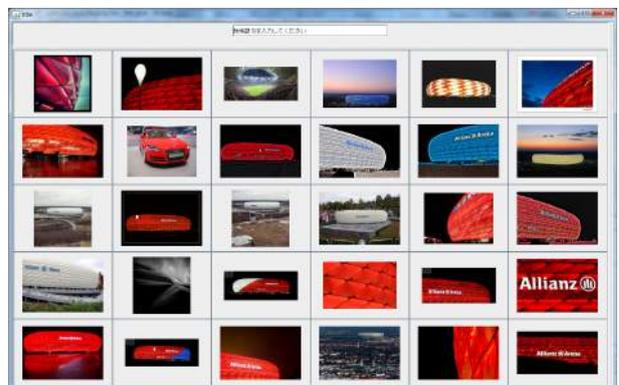


図 12: 実験 2 Allianz+Arena+Rot 開始時結果

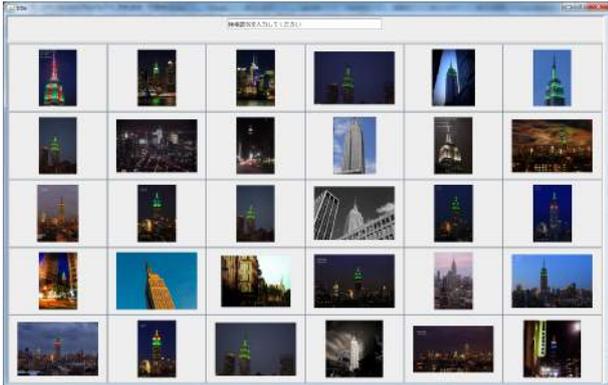


図 13: 実験1 エンパイアステートビル+緑+EmpireStateBuilding+Green 開始時結果

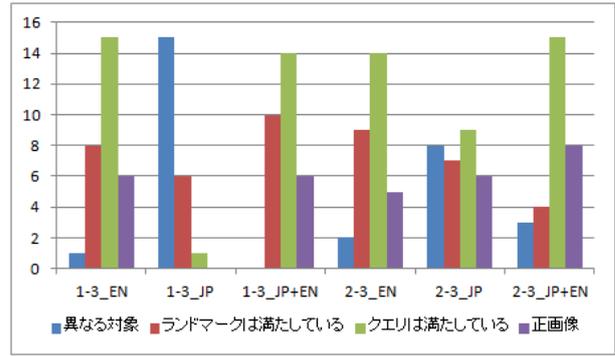


図 17: エンパイアステートビル検索結果

## 5.2. 考察

### 5.2.1. 成功事例

表1, 表3, 表4, 表6からは実験1と比較して, 実験2は開始時適合率が高いことがわかる. 実験2の日本語での検索時の適合率が実験1よりも向上した理由には, 正解画像に付与されているタグが現地語のみであったり, 英語のみである画像が存在したため, これらを翻訳をして追加することで適合画像が取得できるようになったためと考えられる.

また図15, 17にて実験1と実験2を比較すると, 正画像数はそれほど増加していないがユーザの入力したクエリを満たしている画像数が増加している一方で, ランドマークに限り満たしている画像数やまったく異なる対象の画像数は減少していることがわかる. そして図16では現地言語による検索時多言語クエリによる検索時では, 正画像こそ減少しているが, クエリは満たしている画像数はそれほど差がない状態でランドマークは満たしている画像数が増加し, まったく異なる対象の画像数が減少していることが分かる. また図17ではタグを多言語化した後に多言語クエリを用いることで, クエリを満たしている画像数と適合画像数が増加している. このことから, 画像に付与されているタグをクエリから得られる情報を用いて多言語化することで, ユーザが言語化することのできた要求をより満たすことができるようになると思われる.

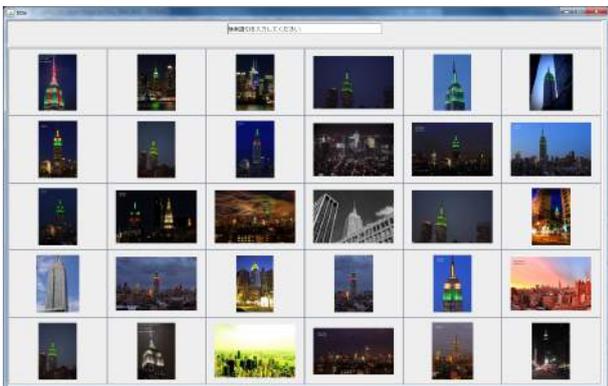


図 14: 実験2 エンパイアステートビル+緑+EmpireStateBuilding+Green 開始時結果

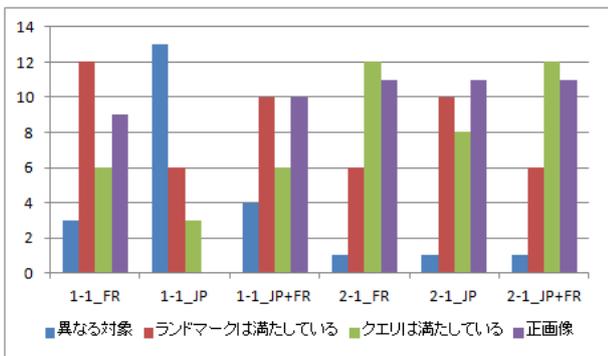


図 15: エッフェル塔検索結果

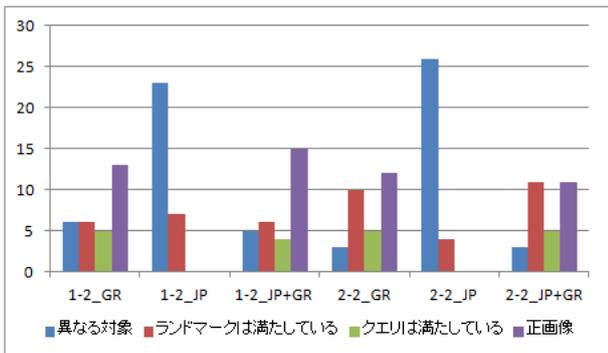


図 16: アリアンツアレーナ検索結果

### 5.2.2. 課題の残る事例

表2, 表5では逆に適合率が減少している. この実験の検索対象のアリアンツアレーナは, タグを多言語化する以前に既に半分近くが適合画像であるのと同時に, 他の検索対象と異なり日本語でつけられたタグを有する画像に正画像が1枚も存在しない. そのため, タグを多言語化したことにより負画像の量が増加したが正画像の量は増加しなかったため, 検索結果として提示される正画像が減少したと考えられる. ゆえに, 複数語により構成されているクエリは修飾語の部分に関してクエリの重要度を変化させるなどの検討を行う必要があると考えられる.

また実験1-3, 2-3にて終了時しなかった原因はフィードバックをかけても適合率が上昇しなかったため, 開

始時の適合率がピーク値となったためである。今回の実験において類似画像検出には色ヒストグラムを用いているのだが、この実験の検索対象であるエンパイアステートビルは緑色に発光している部分が非常に少ない。そのため、色ヒストグラムによる類似画像検索時に緑色以外の部分が強調されてしまったと考えられる。そのため、フィードバックをかけた意図とは異なる類似度が求まり、適合率が上昇しなかったのだと考えられる。ゆえに発光部の小さい検索対象に対しても効果のある画像特徴量を用いることや、メタデータ処理の結果を強く反映させるような類似画像検索を行う必要があると考えられる。

## 6. おわりに

本研究では、自動タグ付けを行うことで、ユーザからみて記述が難しい多言語クエリを活用し、画像に含まれる情報及びメタデータの効果的利用から、より適切な画像の入手を実現した。本研究により、画像データベース内の画像により多くの適切なタグが付与されていることが検索精度の向上に寄与することがわかった。特に多言語クエリによる検索に対し、多言語自動タグ付けを行う際には、検索対象の地理情報を用いて現地語を特定しタグ付けを行うことがユーザによるフィードバックを支援し、適合画像を容易に取得できることがわかった。

今後の課題は、更に多様な画像による評価、実際に被験者実験によりユーザによる評価を実施することである。タグ付けを促進するアイデアの導入や、より高精度なタグ付けを自動で行うための方法の検討、検索クエリ投入時に誤って付与されているタグを検出し自動修正する機能の導入、クエリの中の修飾語の重要度の変更、類似画像検索時にメタデータ処理の結果を考慮、等を行うことで、検索精度が更に高まるだろう。

## 謝辞

本研究は、JSPS 科研費 (S)(24220002, 2012-2016) の助成を受けたものです。

## 参考文献

- [1] 徳永健伸：情報検索と言語処理. 東京大学出版会 (1999).
- [2] Taylor, S.R.: Question-negotiation and information seeking in libraries, *College & research libraries*, Vol. 29, No. 3, pp. 178-194(1968).
- [3] Parton, K, et al.: Simultaneous multilingual search for translingual information retrieval, *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, pp.719-728(2008).
- [4] Varshney, S. and Bajpai, J.: Improving performance of english-hindi cross language information retrieval using transliteration of query terms, *arXiv preprint arXiv*, 1401.3510(2014).
- [5] Zakaria, L. Q., Hall, H. and Lewis, P.: Modelling image semantic descriptions from web 2.0 documents using a hybrid approach, *Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services*, pp.306-312(2009).

- [6] van Leuken, H. R., et al: Visual diversification of image search results. *Proceedings of the 18th international conference on World wide web*, pp.341-350(2009).
- [7] Nakajima, S. and Zettsu, K.: ImageAspect finder/difference-amplifier: focusing on peripheral information for image search and browsing, *Web Technologies Research and Development-APWeb 2005*, pp. 1070-1074(2005).
- [8] Li, X. R., et al. "A systematic comparison of first-time and repeat visitors via a two-phase online survey, *Tourism Management 29.2*, Vol.29, No.2, pp.278-293(2008).
- [9] Andrew, S.: Using major events to promote peripheral urban areas: Deptford and the 2007 Tour de France, *International Perspectives of Festivals and Events: Paradigms of analysis*, pp.3-19(2009).
- [10] Murakami, Y., et al.: Language Service Management with the Language Grid, *7th International Conference on Language Resources and Evaluation (LREC 2010)*, pp. 3526-3531(2010).
- [11] Swain, M. J., and Ballard, D. H.: Color indexing, *International Journal of Computer Vision* Vol.7, No.1, pp. 11-32(1991).