

メタデータマッシュアップに基づくマルチメディアコンテンツの共有法

Multimedia Content Sharing Method via Metadata Mash-up

今井良*¹
Ryo Imai

森田武史*¹
Takeshi Morita

和泉憲明*²
Noriaki Izumi

山口高平*¹
Takahira Yamaguchi

*¹ 慶應義塾大学
Keio University

*² 独立行政法人 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology

Multimedia Content Sharing sites on the World Wide Web has promoted information flow, allowing users to easily gain information from others and causing search precision and recall to drop. This paper proposes a method for sharing multimedia contents using metadata mashing, and introduces a system which assists multimedia contents sharing between users. Location information structure is mashed up, and components of the structure are used as metadata. A case study shows us that our system work well with Web pages for restaurants.

1. はじめに

マルチメディアコンテンツ共有サイトの出現により、受動的なコンテンツ取得だけではなく、容易に能動的にコンテンツ提供が行えるようになった。これまでは個々で管理していたマルチメディアコンテンツは、他ユーザと共有することが可能となり、インターネット上に流通する情報量は膨大なものとなった。さらに、コンテンツに対して「タイトル」「コメント」「タグ」等のメタデータを容易に付加できるようになった。しかし、このようなマルチメディアコンテンツ共有サイトではユーザによるタグ付けではコンテンツを示すタグのみとは限らなくなり、対象物を同定するには困難な状況を作り出している。さらに、検索を行う際にはキーワードマッチしか行わない場合が多いため、欲している情報を全て検出できていない、または検出結果にゴミが多数含まれていると考えられる。

例として、ファーストフード店であるマクドナルドを対象物としたマルチメディアコンテンツを挙げる。マクドナルドには正式名称はもちろん、「マック」や「マクド」と言った略称や愛称も存在する。あるユーザが「マック」をタグとして登録したとしても、他ユーザが同じマクドナルドの店舗を対象物とした異なるマルチメディアコンテンツに対して「マクド」とタグを登録してしまえば、キーワードマッチでは同じ店と同定できず、再現率の低下につながってしまう。また、別店舗に対して同じ名称がタグとして付加されてしまうケースも考えられる。この場合、実際には機械は異なる店舗を同一のもののみならずしてしまい正答率の低下につながる。

そこで本稿では、様々なメタデータをマッシュアップすることによって可能となる分類を支援するマルチメディアコンテンツ共有システム MiNT を提案する。MiNT ではマルチメディアコンテンツの対象物を同定でき、現存する WEB アプリケーションで多々見られるタグ情報のみを用いたキーワードマッチによる分類よりも高度な分類が可能となる。さらに、これまで個々が人手で行っていた大量のマルチメディアコンテンツの分類に対して、マルチメディアコンテンツ及びマルチメディアコンテンツが対象とするもののメタデータ情報を活かすことによって、フォークソミーの様に多数の人によるマルチメディアコンテンツの分類の自動化が可能となる。さらに、既存のマルチメディアコンテンツのみならず新たに追加されるマルチメディアコンテンツの自動分類も支援

できるようになる。

2. マルチメディアコンテンツ共有システム MiNT

2.1 MiNT の概要

本節では様々なメタデータをマッシュアップすることによってマルチメディアコンテンツの対象物の同定を可能とし、ユーザ間でのマルチメディアコンテンツ共有を支援する WEB アプリケーション MiNT について述べる。本稿におけるマルチメディアコンテンツはデジタル写真を指す。MiNT は小規模 SNS の様なコミュニティ内で利用することを前提とする。

図1に MiNT のシステムフローを示す。MiNT はスポット情報とユーザが提供する情報の二種類を入力とする。スポットとは「住所を持つ特定の場所」を指し、スポット情報とは「スポットの同定に用いられるメタデータ」を指す。MiNT ではスポット情報のプライムデータ(ユーザによって入力される基本データ)としてスポットの公式 URL を用いる。公式 URL から住所と言い換えを得ることができ、さらに住所からその場所の GPS を得ることができる。ユーザが提供するプライムデータは、デジタル写真データ及びその説明をするタグである。

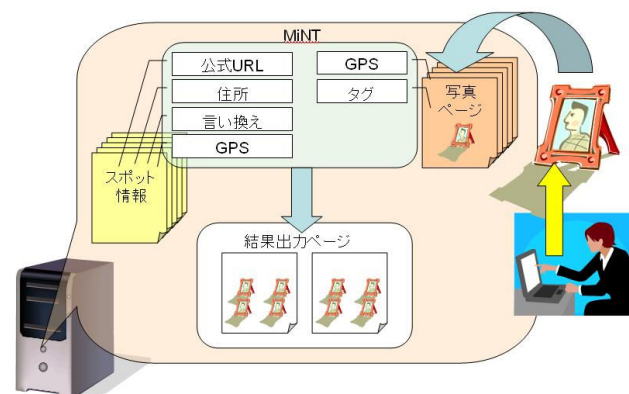


図1: MiNT のシステムフロー

MiNT のユーザにはスポット情報を提供する管理ユーザとデジタル写真データを提供する一般ユーザの二種類がある。本研究はコミュニティ内の利用を想定しているので、ある一人のユーザが両役割を持つことも可能である。

既存の WEB 上のサービスではタグのキーワードマッチによる分類が主流だが、MiNT はスポット名による分類と地域名による分類を可能とする。MiNT は公式 URL から様々なスポット情報を

連絡先: 今井良, 山口高平, 慶應義塾大学理工学部
〒223-8522 神奈川県横浜市港北区日吉 3-14-1
Tel: 045-566-1614, E-mail: {r_imai,yamaguti}@ae.keio.ac.jp

得るための支援も行う。よって、本来ならばコストのかかってしまうスポット情報の入力、MiNT を用いることによってコスト低減される。

管理ユーザ及び一般ユーザによるプライムデータを個々で利用した場合、精度の高い分類を行うことはできない。しかし、MiNT にプライムデータを入力し、それにより得られるスポットの言い換え、住所、GPS やデジタル写真データの GPS といったメタデータを蓄積することにより、精度の高い分類が可能となる。

2.2 スポット情報に基づく位置情報オントロジー構築

MiNT のスポット名及び地域名による分類を可能とするには、住所が包含関係を持つ必要があり、スポット名の表記揺れに対応できる必要がある。スポット情報に基づいた位置情報オントロジーにより上記を解決する。

位置情報オントロジーは図 2 の様な仕組みになっている。上から「住所レイヤー」「店舗名レイヤー」「エリアレイヤー」の三つのレイヤーがある。住所レイヤーはスポットの住所を用いて作成する。店舗名レイヤーはスポットに対して登録時に付けられたスポット名を代表する名称である。エリアレイヤーはスポット名の略称や通称と言った言い換えや表記揺れをまとめたものである。

住所は「面積の大きい順に記述される」特性があるため、例えば「東京都千代田区外神田イマイビル」という住所があるとすると、「東京都>千代田区>外神田>イマイビル」という包含関係を得ることができ、そのままオントロジーの階層概念に適応させることが可能である。

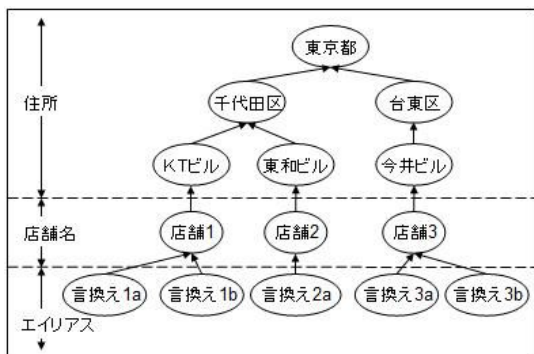


図 2: 位置情報オントロジーの例

さらに、位置情報オントロジーには範囲の特定できない地域名にも対応させる必要がある。位置情報オントロジーに「秋葉原」といった特殊な地域名を追加したいとする。「秋葉原」は住所に現れる単語ではあるが、秋葉原駅周辺を漠然と指す概念でもある。そのため、地域名分類においてこのような概念としての地域名を利用したいユーザもいると思われる。位置情報オントロジーに組み込むためには、包含関係が必要となるが、このような地域名は人によって範囲の定義が違うので難しい。そこで、新たに特殊な地域名を追加する場合は追加する地域名の一つ下位の地域名と同等とすることにする。例えば、「東京都>秋葉原>千代田区」「東京都>秋葉原>台東区」という包含関係を与えたいとする。この場合、「東京都>秋葉原=秋葉原」「東京都>秋葉原=台東区」と考える。つまり、元々ある位置情報オントロジーの枝の形を変えずに、秋葉原を千代田区と台東区と同等の扱いとするのである。これにより、分類クエリに「秋葉原」がきた場合には「千代田区+台東区」で分類されたものと同じ結果を得ることができるようになり、漏れがなくなる。

2.3 位置情報オントロジー構築支援

位置情報オントロジーの構築を手で行うと、多大な人的コストがかかってしまう。MiNT は位置情報オントロジー構築支援のために、公式 URL よりスポットの言い換えと住所を半自動的に抽出する。

スポットの公式 URL はスポットの言い換えと住所を提供してくれる上質なソースとなる。公式 URL に複数の住所が記述されている場合は、スポット情報の登録者が選択する。公式 URL から住所を抽出できなかった際には、手入力する。エリア作成のためには、インターネット上の公式 URL にリンクをしてある HTML 文書における、公式 URL のアンカーテキストと alt 属性を抽出する。抽出結果を手でクレンジングすることにより、エリアの作成が可能となる。ここで抽出した文字列は多数の人によるスポットの表現の仕方であり、多様性があるため、分類時の再現率向上につながると考えられる。

2.4 一般ユーザによるメタデータ付与

MiNT において一般ユーザはマルチメディアコンテンツのアップロードを行う際に、タグと GPS 種類のメタデータを付与する。一般ユーザは各自のコンピュータ内にあるマルチメディアコンテンツをアップロードし、タグ付けを行う。タグは複数登録することも可能だが、最低条件としてスポットの名称を一つ記入する必要がある。最近のデジタルカメラや携帯電話のカメラを用いればマルチメディアコンテンツに GPS を付加することができる。しかし、GPS 情報にはその性質から誤差が生じる可能性があり、さらに GPS 情報を付加できない環境にあるユーザも存在すると考えられる。MiNT はマップ上で選択した場所の GPS 情報をマルチメディアコンテンツに付加できるため、誤差が生じていた場合の修正や元々 GPS 情報が付加されていないマルチメディアコンテンツにも対応できる。

2.5 メタデータマッシュアップ

分類時にはスポット情報のメタデータと一般ユーザの入力したマルチメディアコンテンツのメタデータをマッシュアップする。分類を行いたい場合、ユーザは MiNT 上で分類クエリを入力する。分類クエリは位置情報オントロジーを参照するときベースとなる単語である。地域名分類では、分類クエリに地域名がきたときに、位置情報オントロジーにおいてそのノード以下のものを全てをマルチメディアコンテンツに付加されているタグの検索対象とし、その地域に存在すると思われるマルチメディアコンテンツを全て見つけてくる。図 3 は地域名分類を表した図である。

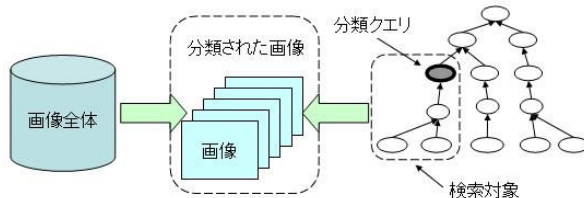


図 3: 地域名分類の様子

スポット名分類は、分類クエリにスポット名がきたときに、そのスポットを対象とすると思われるタグを持つマルチメディアコンテンツを一般ユーザより提供されたコンテンツ全体から分類する。エリアに基づくタグの表記揺れに対応した分類である。どちらの分類の際にも分類結果の一部として GPS 情報を用いた位置情報の地図表示も行う。

GPS 情報には二種類ある。それぞれ、スポット情報に登録さ

れた住所よりわかる GPS 情報(以下正式 GPS)と、マルチメディアコンテンツの Exif に記述されている GPS 情報(以下画像 GPS)である。正式 GPS と画像 GPS を用いることによって、実際に写真が撮られた場所とそのスポットの位置関係が視覚化できる。GPS 情報は分類結果に直接は影響しない。しかし、分類結果に表示することによって分類の信頼度をユーザ自身が判断できるようになる。位置情報を視覚化しユーザ自身がそれを分類の信頼性に加味することによって、正答率が向上すると考えられる。

なお、スポット名分類が行われた際は、該当するスポット情報に登録されている住所の位置もマップ上に表示し、その周りに円を描く。これにより、分類クエリにスポット名がきた時に、そのスポットと分類されたマルチメディアコンテンツが実際に撮影された場所が明確になり、あまりにも正式 GPS とかけ離れた場所に画像 GPS が存在していればそのコンテンツは分類しているスポットとは別物との判断をユーザが下すことも可能となる。

3. MiNT の実装

MiNT は WEB アプリケーション開発プラットフォームである YOSEE(Yarn Of Semantically Enhanced Entities) [和泉 07]を用いて実装する。本章では YOSEE 及び位置情報オントロジーの構築支援機能について説明する。

3.1 YOSEE

YOSEE は、インターネット上で流通可能な様々な情報コンテンツを作成、蓄積、公開するための CMS(コンテンツマネジメントシステム)のアーキテクチャとその実装である。YOSEE では特に次の 3 つの機能の実現を目指している。

「貯める」機能: WIKI, ブログ, スケジュール, デジカメの画像や動画, WEB ページのクリッピング情報, ファイルなど電子的データを蓄積する。

「つくる」機能: 蓄積したコンテンツを組み合わせて、新たなコンテンツを作成することを支援する。

「公開する」機能: 蓄積したコンテンツを見つけやすくするための検索の仕組みを提供する。

YOSEE では以上の機能を実現するために、フォークソノミーを応用した共同文書管理のしくみである WIKI 機能をベースとしている。WIKI を拡張することにより、YOSEE では文書のみならず、画像、動画のほか PC 上で扱える多くのファイルを複数のユーザで蓄積管理することができる。

YOSEE は、コンテンツを蓄積するだけでなく、各コンテンツには属性(タグ)をつける機能により、意味によるコンテンツ同士の関連付けができる。この属性による関連付けを使用することで、全文検索方式とは違う知的なコンテンツ検索を実現することが可能である。

YOSEE の「貯める」機能を用いてスポット情報及びマルチメディアコンテンツ自体の管理をし、「つくる」機能を用いて分類されたマルチメディアコンテンツをまとめる。さらに、「公開する」機能で分類結果を表示させる。

図 5 は YOSEE をベースに作成した WEB アプリケーション MiNT のインターフェース表示例である。左から順に、スポット情報ページ、マルチメディアコンテンツページ、分類結果ページである。

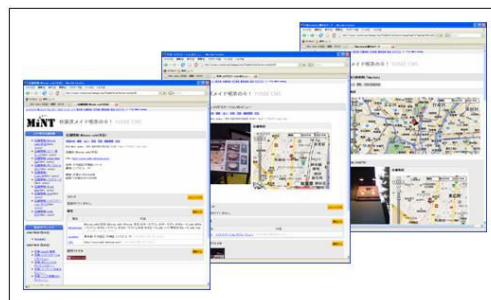


図 5: MiNT のインターフェース例

3.2 位置情報オントロジー構築支援機能の実装

位置情報オントロジー構築支援機能として、公式 URL から住所及びエイリアスを抽出するプログラムを実装した。

住所抽出プログラムでは、はじめに、公式 URL からリンクされている WEB ページを獲得するために、Jericho HTML Parser[Jericho]を用いて公式 URL の WEB ページ内の link 要素を抽出する。次に、公式 URL の WEB ページ及び公式 URL からリンクされている Web ページ内の住所を正規表現により抽出する。

エイリアス抽出プログラムは、Google SOAP Search API[Google]を用いて実装した。はじめに、公式 URL にリンクしている Web ページを Google SOAP Search API を用いて獲得する。次に、Jericho HTML Parser を用いて、獲得した Web ページの中から公式 URL を link 要素の href 属性とするアンカーテキストを抽出する。さらに、link 要素内に img 要素がある場合にはその alt 属性の値を抽出する。

4. 実験

公式 URL より得られるエイリアスと住所の精度は分類に用いるための位置情報オントロジーの精度に直接つながる。よって、いかに公式 URL からエイリアスと住所を正確に抽出し、スポット情報として登録するかが重要となる。ここでは、実際に MiNT を用いてスポット情報登録を行い、いくつかのエイリアス用の言い換えが抽出できるのか及びどれぐらい正確に住所を抽出できるかの実験を行った。実験の際に用いたスポットは、「秋葉原及び中華街に存在する飲食店」からインターネット上にホームページを持つものをランダムに選んだ計 86 個のスポットである。

4.1 エイリアス抽出結果

表 1 は実験対象としたスポットのエイリアスの抽出結果を示すものである。表 1 の A 正答率は、抽出した言い換えのうち実際にスポットと関連があると判断できるものの比率である。表 1 の B 正答率は抽出した言い換えのうち実際にスポットの言い換えとして成り立つものの比率である。なお、抽出数の数値は重複を全て排除した後のものである。

表 1: 実験対象としたスポットのエイリアス抽出結果

| | 秋葉原 | 中華街 | 全体 |
|-------------|--------|--------|--------|
| 登録したスポット数 | 34 | 52 | 86 |
| 合計抽出数 | 190 | 169 | 359 |
| 1スポットの最高抽出数 | 27 | 23 | 27 |
| 1スポットの最低抽出数 | 0 | 0 | 0 |
| 1スポットの平均抽出数 | 3.25 | 5.59 | 4.17 |
| A正答率 | 72.11% | 79.88% | 75.77% |
| B正答率 | 52.11% | 28.99% | 41.23% |

表 1 より、登録した各スポットにおいて平均 4 個強の言い換えが抽出された。フォークソノミーのように様々な人の言い換えが得られるため、「通な人」が利用する言い換えも抽出することができていることがわかった。

4.2 住所抽出結果

表 2 は住所抽出の際に得た結果を示すものである。テキスト率とは、公式 URL の HTML において、住所が記述されている確立である。画像・flash・pdf 率は、トップページまたはトップページのリンクページに住所は表記されているが HTML にテキスト記述されていない確率である。リンク内率とは、テキスト記述の住所がトップページからリンクを二つ以上辿らなければいけないページにある確率である。C 正答率は対象とした全ての公式 URL を見たときに住所を正しく抽出できた率を指す。D 正答率は全ての公式 URL においてテキスト表記されているページの中の住所を正しく抽出できた確率を指す。なお、登録したスポットはエイリアス抽出の実験で用いたものと同じなのでスポット数はここでは省略してある。

表 2: 実験対象としたスポットの住所抽出結果

| | 秋葉原 | 中華街 | 全体 |
|----------------|--------|--------|--------|
| テキスト率 | 79.41% | 75.00% | 76.74% |
| 画像・flash・pdf 率 | 11.76% | 13.46% | 12.79% |
| リンク内率 | 2.94% | 13.46% | 9.30% |
| C 正答率 | 70.59% | 57.69% | 62.79% |
| D 正答率 | 88.89% | 76.92% | 76.74% |

作成した住所を抽出プログラムでは、HTML 内に記述されているテキストを見て住所を検出するため、住所の表記が画像化されていたり、flash の中になったり、ページ自体が pdf 化されていると住所を抽出することはできない。本システムでは Jericho HTML Parser より抽出されたページ及び公式 URL のみを住所の抽出対象としているため、リンク内率で表されているような公式 URL からリンクを二つ以上辿った場所に住所が記述されていれば抽出することができない。

5. 検討

5.1 実験結果の検討

MiNT 作成した WEB アプリケーションでは、エイリアスの B 正答率は平均で 41.23%であった。しかし、実際に抽出された文字列全てをエイリアスとして利用するわけではなく、一度ユーザーにフィードバックしユーザーに選んでもらう形を取るため、実際にはより精度の高いものがエイリアスとして活用されることになる。よって、ここでは正答率よりも、1スポットあたりの平均抽出数のほうがポイントとなる。エイリアスの抽出数にはかなりのバラツキがあり、最高 27 個、最低 0 個、平均 4.17 個であった。これはスポットの有名度やホームページの作成日時、さらに相互リンクの有無等が影響していると考えられる。エイリアスの最小抽出数は 0 個であったが、これはまだ店舗ができたばかりか、もしくは相互リンク等を一切行っていないホームページ等であり、必然的なものと思われる。エイリアスの抽出数が少なすぎる場合、多数の人による表記揺れに対応できない。そのため、抽出数が極端に低いスポットに関しては、人手でエイリアスの追加登録を行ったほうがよい。また、MiNT では英字の半角や大文字小文字等の表記の違いを考慮していないため、英字の表記が分類

クエリと異なると正しく分類されない。よって、英字の表記揺れにも対応できる仕組みを組み込む必要があると考えられる。誤って抽出された語彙に関しては、「ホームへ」「トップページへ」等が多数含まれていたがこれらはスポット情報登録の際にユーザーのクレンジングが入れば問題ない。

表 2 の D 正答率を見ると、住所が HTML 内にテキストで記述されている場合はかなりの高確率でその住所を正しく抽出できていることがわかる。作成したプログラムでは、住所を抽出する際にトップページ及びトップページから直接リンクが記述されているページの中の住所を抽出し、リンクをさらに辿ったページは見えない。これは、ページを辿りすぎるとまったく関係のないページからも住所を抽出してしまうからである。さらに、WEB サイトによってはトップページと内部ページが違うドメインのものもある。よって、ドメインがトップページと同じページのみを見ようとするとならばトップページのみしか見なくなってしまう可能性がある。これらのことを考慮し、現状のプログラムにたどり着いた。

本稿ではスポットの住所と言い換えを得るために公式 URL の入力が必要であるが、将来的には検索エンジンにスポット名を入力するだけで住所と言い換えを得られるようなシステムを考えている。これには検索エンジンを用いてスポット名から公式 URL が得られる仕組みが必要となる。

5.2 正式 GPS についての検討

スポットの正式 GPS は住所より得たものであり一点を示す一次元的なものだが、実際のスポットは二次元的に面積の広がりを持つ。よって、正式 GPS のみではスポットの GPS を表しきることはできない。だが、同じスポットを対象物とするマルチメディアコンテンツが多数アップロードされこれらの画像 GPS を総じて見ることで、そのスポットの大きさが見えてくるようになる。

6. おわりに

公式 URL と GPS 情報を用いることによって得られた位置情報の構造をマッシュアップした様々なメタデータを用いることによって、位置情報オントロジーを自動生成できる WEB アプリケーションを作成した。これにより、ユーザーがより容易にマルチメディアコンテンツの共有ができることが可能であると示した。

今後は、マルチメディアコンテンツのみならず動画コンテンツにも対応させていくつもりである。時間概念を持つ動画コンテンツに対応させることにより、様々な粒度のイベントに対しても同定を行うことが可能となると考えられる。

参考文献

- [和泉 07] 軽量オントロジーに基づくコンテンツ駆動型プラットフォーム, 和泉 憲明, 橋田 浩一, 人工知能学会 セマンティック Web とオントロジー研究会 第 15 回 SIG-SWO-A603-09, 2007
- [Jericho] Jericho HTML Parser, <http://jerichohtml.sourceforge.net/doc/index.html>
- [Google] Google SOAP Search API, <http://code.google.com/apis/soapsearch/>