

動画コンテンツの位置情報メタデータに関する検討

A Consideration of the Location Metadata in Video Contents

宇田川智輝*¹
Tomoki Udagawa

今井良*¹
Ryo Imai

和泉憲明*²
Noriaki Izumi

橋田浩一*²
Kôiti Hasida

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 (AIST)

In order to bridge the gap between context dependent view for users and detailed data stored as tagged information, we propose a methodology for providing the suitable view based on structured context architecture. This report enables us to adopt the criteria for the determination of the most focused content according to the user referring contents. Dynamic structure can be generated by defining the tag weight based on access logs and total times of corresponded content. Through the application to a map services on the Web, we consider an extension for general multimedia content.

1. はじめに

Google マップ API を始めとした地図 API のように、Web コンテンツと地図やカレンダー、タスクリストなどと、テキストだけでなく、静止画や動画、音楽とが、高度にリンクされ、Web コンテンツの構造は、さらなる高度化の可能性を具現化させつつある。この結果、地図を利用した画像や動画の検索や、お店の情報検索といった様々なサービスに関して、ユーザのコンテンツ利用のコンテキストに応じて、いかに、適切なビューを提供できるかが重要となる。

例えば、地図 API を利用したサービスは位置関係を把握しやすく便利であるが、近年の住所検索や緯度経度変換機能などの高度化により、コンテンツの表現対象が具体的に地図上のどの部分を指し示しているかが正確になった。一方で、正確であるが故に、小さな差も異なるとの判断根拠となり適切な抽象化が困難となる。このため、コンテンツ数が増えるに従い検索結果である地図の把握が困難となり、一覧性が損なわれる。例えば、「新宿駅周辺のラーメン屋」のコンテンツを検索したケースを考える。新宿駅周辺にはラーメン屋が多量に存在するため、地図の縮尺によっては地図が無数のマーカーで埋め尽くされ、見にくくなる。そこで、大量のマーカーを少数にまとめる必要がある。

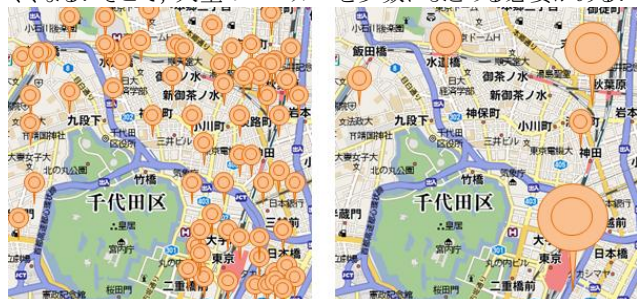


図 1. 大量のマーカーを少数にまとめる

ここで、単純に住所の行政区画構造に沿ってトップダウンでまとめた場合、表示に関する問題が生じる。具体的には、まとめたマーカーの中心とする位置を正しく決められないという問題である。

連絡先: 宇田川智輝, 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: {t_uda,yamaguti}@ae.keio.ac.jp

以上の観点から、本研究では行政表現と一般表現のギャップを埋めるために内部の詳細構造を補完するための適切なユーザのビューを提供することを目的として、コンテンツの内容に基づいた中心位置の決定法と構造化の手法を提案する。ここでは、位置情報メタデータの付加された動画コンテンツの動画の長さをコンテンツの重みとみなし、中心位置決定の基準にする。これにより、各コンテンツの重みに応じた動的な階層構造が可能になる。この構造の地図 API サービス等への応用例を述べる。最終的に、動画コンテンツだけではなくマルチメディアコンテンツ一般への拡張を検討する。

2. 場所オントロジーに基づくコンテンツの構造化

2.1 場所オントロジー

店舗名や地域名等の位置情報によるコンテンツ管理を可能とするために、あらかじめ、住所の包含関係や行政区画上の基準点、交通機関を利用した場合の移動距離に基づく地物間の概念間距離などを、場所オントロジーとして参照可能としておく。場所オントロジーは、「住所レイヤー」と「店舗名レイヤー」の2つのレイヤーとタグを持ち、それぞれ、住所と店舗名、言い換え表現に対応している。ここでは、地物(ここでは店舗名)が階層構造を持たないようにするため、ビルなどは住所の構成要素とする。

場所オントロジーの構築は人的に行うにはコストが大きすぎるため、スポットの公式URLより住所や言い換え表現の半自動抽

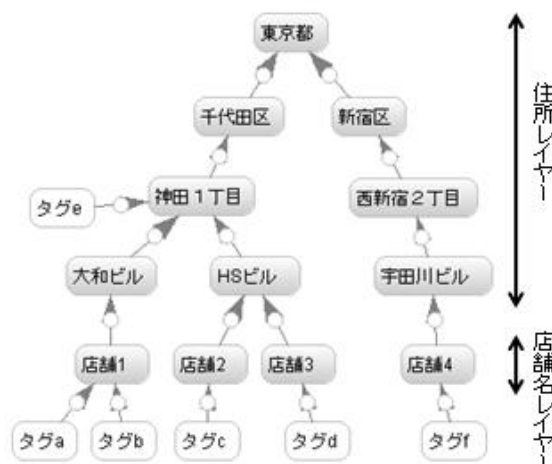


図 2. 場所オントロジー

出を行う。エイリアス作成のためには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。抽出結果を人手でクレンジングすることにより、エイリアスタグ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方法であるため表現の多様性があり、分類時のキーワードの網羅性の向上につながる。

2.2 コンテンツへのメタデータ付加

本システムにおいて、一般ユーザは動画コンテンツのアップロードを行う際に、タグと位置情報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動画コンテンツをアップロードし、タグ付与を行う。タグは複数登録することが可能とし、位置情報の登録は、携帯電話やデジタルカメラを利用したGPSによる入力、他、地図を利用した入力補助を行う。

入力された位置情報を基に、コンテンツの場所オントロジーへの関連付けを行う。

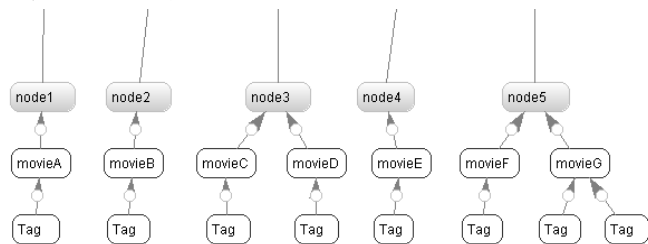


図3：動画コンテンツの場所オントロジーへの関連付け

3. 動的な階層構造の構築及び利用

3.1 検索クエリの入力

システムの利用者はコンテンツを検索する際に、システム側に検索クエリを送信する。検索クエリを受け取ると、システム側は場所オントロジー、コンテンツのタグと照らし合わせ、一致する動画コンテンツ及び、一致する場所オントロジーのノード以下に関連付けられているコンテンツを返す。

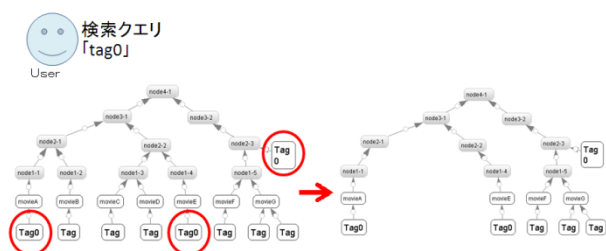


図4：クエリを入力後、場所オントロジーに沿って検索

3.2 重心の計算

本節では、コンテンツを少数のマーカーにまとめた時の中心位置を決定する方法を述べる。この中心位置のことを以後「重心」と呼ぶ。ここでは、動画コンテンツの再生時間を「スコア」として重心の決定に用いる。スコアの定義は次のようになる。

動画コンテンツのスコア=再生時間の長さ
ノードのスコア=関連付けられている動画のスコアの総和

あるノードに関連付けられているコンテンツ群のうち最もスコアの大きなコンテンツの位置を、コンテンツ群の重心とみなし上位ノードに位置情報として持たせる処理を行う。アルゴリズムは以下のようになる。

$N_1, N_2 \dots$: 場所オントロジーのノード
 $s_1, s_2 \dots: N_1, N_2 \dots$ のスコア
 N_p : $N_1, N_2 \dots$ の親ノード
 s_p : N_p のスコア

Step1: 場所オントロジーの末端のあるノードを N_1 とする。
Step2: N_1 の親ノードを N_p とする。 N_p の他の子ノードを $N_2, N_3 \dots$ とする。
Step3:

$\text{Max}\{s_1, s_2, s_3 \dots, s_n\}$
となるノード番号を求め、その番号を X とする。

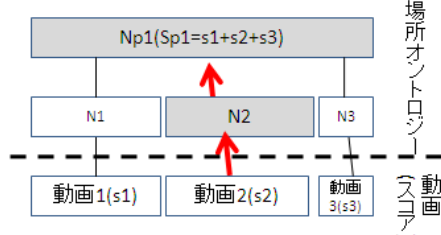


図5：スコアの最も大きいノードを抽出

Step4: $s_p = (s_1 + s_2 + s_3 \dots + s_n)$ とする。 N_p の位置情報として、 N_X の位置を持たせる。
Step5: 同じ階層の他のノードについて、まだ s_p の決まっていない N_p を持つノードがある場合、そのノードを N_1 として Step2 に戻る。ない場合、 $N_1 = N_p$ として Step2 に戻る。
Step6: N_1 が最上部のノードになったら終了。

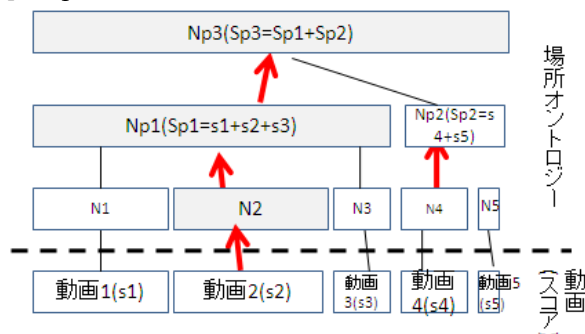


図6：ボトムアップに最上位ノードまで繰り返す

3.3 ビューの作成

3.2 で決定された各ノードの重心位置及びスコアは、検索クエリに対応するコンテンツの内容を反映した情報といえる。この2つの情報を基にして多量のマーカーを少数にまとめ、地図上に縮尺に応じて見やすくプロットした形でユーザに結果を返す。

ユーザに返す地図の縮尺は、3.1 で返された場所オントロジーの最上位のノードを参照して決定する。場所オントロジーの最上位ノードを縮尺に取ることで、結果として返すコンテンツを全て網羅できる。以後、地図の縮尺の基準となったノードを「縮尺ノード」と呼ぶ。

次に、地図にプロットするマーカーのサイズを決定する。以下にアルゴリズムを記す。

$N_1, N_2 \dots$: 縮尺ノードの1段階下のノード
 $s_1, s_2 \dots: N_1, N_2 \dots$ のスコア
 $size_1, size_2 \dots: N_1, N_2 \dots$ のマーカーのサイズ

Step1: $s_1, s_2 \dots$ の平均を求め、求めた平均を \bar{s} とする。

Step2: 初期値として, $n=1$ とする.
 Step3: s_n/s を求める. この値を N_n のマーカーサイズの指標 $size_n$ とする.
 Step4: $n=n+1$ とし, Step3 に戻る. 全ての N についての $size$ を求めることが出来たら Step5 へ.
 Step5: 基準のマーカーサイズを決め, 各マーカーのサイズを, 基準サイズ $\times size_n$ の面積となるように決定する.

3.4 マーカーの分解と合成

コンテンツがまとめられるマーカーの数は, 多すぎても, 少なすぎても見にくくなってしまふ. そこで, 閾値に基づいてマーカーの数を調整する処理を行う. 以下に, 2種類のケース別にアルゴリズムを表記する.

t : 閾値

P_0 : 現状マーカー数

P_a : 合成・分解後のマーカー数

i). マーカー数が閾値よりも少ない場合,

$P_0 \leq t$ である場合, スコアの最も大きなマーカーを木構造に沿って1段階下の構造に分解することを考える.

Step1: 縮尺ノードの1段階下のノードの中の, スコアの最も大きなものを N とする. N の1段階下のノード数を n とする

Step2: N を一段階下の構造に分解した時のマーカー数 P_a を求める. $P_a = P_0 + n - 1$ である.

Step3: 判定式: $\min\{|t - P_0|, |t - P_a|\}$ の計算を行い, 絶対値の小さくなる方の選択肢を採用する. 前者の絶対値が小さかった場合は分解をせず, 後者の方が絶対値が小さかった場合は N を分解する選択を採る. 分解をしなかった場合, 終了.

Step4: $P_a = t$ である場合, 終了. $P_0 < t$ である場合, Step1 に戻る. $P_a > t$ である場合, アルゴリズム ii) へ進む.

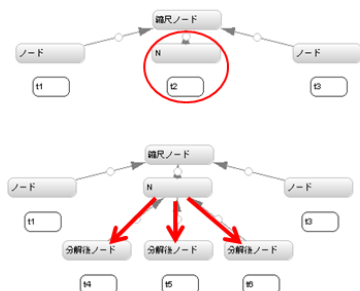


図7: マーカーの分解

ii). マーカー数が閾値よりも多い場合

$P_0 > t$ である場合, スコアの最も小さなマーカーを, そこから最も距離の近いマーカーに合成する.

Step1: 縮尺ノードの1段階下のノードの中の, スコアの最も小さなものを N , N から最も距離の近いノードを N_0 とする.

Step2: N を N_0 に合成する. N_0 のスコアに N のスコアを加え, N のマーカーを消す. $P_0 = P_0 - 1$ となる.

Step3: $P_0 = t$ となった場合, 終了. $P_0 > t$ である場合, Step1 に戻る.

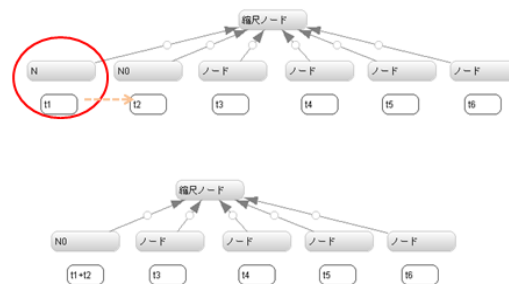


図8: マーカーの合成

4. コンテンツへのタグ付与

4.1 動画共有サイトを用いたタグ付与

2節では, 動画コンテンツに適切なタグがついた状態を前提としてきた. タグには, ジャンルや内容に関するもの, コンテンツの言い換え表現や俗称, 登場人物など様々な情報が含まれているが, 本節では動画コンテンツへのタグの付与方法について紹介する.

動画コンテンツへのタグ付与の方法として, 前述のようにユーザが任意にタグを付与出来る他, YouTubeなどの既存の動画サイトを利用したタグ付与支援を行う. 場所オントロジーを持つ店舗名や会社名などの「スポット名」をキーワードとして, YouTube等の既存の動画サイト上の動画のタグを自動抽出しタグ付与に利用する.

4.2 タグの汎化

タグは, 各個人が自由な表現でタグ名をつけることができるため, 表現の多様性を持つ性質がある. しかしその反面, 同一物への呼称の表記ゆれの性質も併せ持つ. 例えば, 「秋葉原」の同義表現は「アキバ」「秋葉」等が考えられるが, これらの単語が同じものを指していることを表す必要がある. ここで, ある程度市民権を得ている俗称を正式名称とみなす処理を行う.

あるノードに関連付けられたコンテンツのうち, 同一のタグが付けられたコンテンツが半数以上ある場合, そのタグを親ノードに対して付加する. アルゴリズムは以下のようになる.

$t_0, t_1, t_2 \dots$: タグ

n : 場所オントロジーのノード

N : コンテンツの個数

$M_1, M_2, M_3 \dots$: 動画コンテンツ

Step1: 初期値として, $m=0$ とする.

Step2: t_m のついている動画の親のノード n よりも下に関連付けられているタグの総数を求め T とする.

Step3: n に関連付けられているコンテンツの個数 N を計算し,
 $(N/2) \leq T$

が成り立つかを判定, 成り立つ場合 t_m を n のタグとして付加する.

Step4: $t_m = t_{m+1}$ として, Step1 へ戻る. タグがなくなったら Step4 へ.

Step5: n を同じ階層にある他のノードに変更, Step1 へ戻る. 同じ階層のノードがなくなったら Step6 へ.

Step6: 階層を1つ上げ, Step1 へ戻る. 最上階層まで到達したら終了.

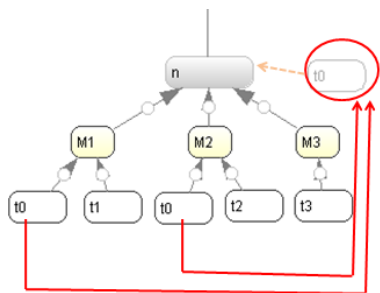


図 9 : タグの汎化

5. 実装及び実験方法ならびに実験結果の概要

5.1 実装システム

本研究での提案手法をマルチメディアコンテンツ共有システム MiNT[今井 07]として実装した。以下に、MiNT の主な 4 つのモジュールの既存機能及び追加機能について記述する。

i) 場所オントロジー構築モジュール

URL を入力し、HP から住所や言い換え表現を抽出し場所オントロジーを構築するモジュール。

ii) コンテンツアップロードモジュール

タグと共に、コンテンツをアップロードするためのモジュール。

4.2 節にて述べた「タグの汎化」処理機能を実装した。

iii) 検索モジュール

検索クエリを入力し、場所オントロジーに沿って結果を返すモジュール。

コンテンツ単位ではなく、場所オントロジーの一部を結果として返す。結果は直接ユーザに返すのではなく、重心計算・マーカー処理モジュールを通してユーザに返される。

iv) 重心計算・マーカー処理モジュール

場所オントロジーとスコアに従って 3.2 節にて述べた「重心の計算」を行うモジュール。

3.4 節の「マーカーの分解と合成」の機能も持ち、最終的な結果をユーザに返す。

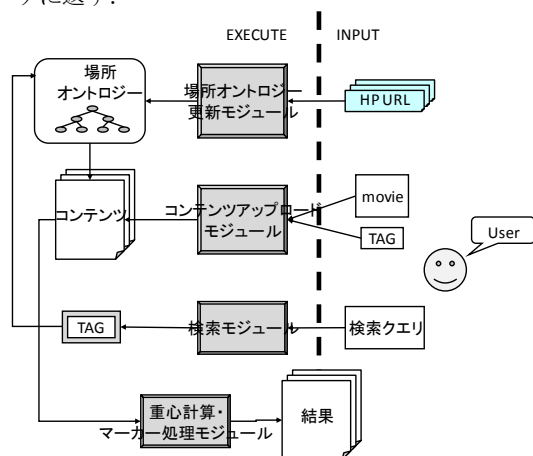


図 10 : MiNT のシステムフロー

5.2 実験

動画共有サイト YouTube からの収集及び自分で撮影することで大量に動画コンテンツを用意し、位置情報とタグを付加した。また、それぞれのスポットの公式 URL から場所オントロジーを構築した。検索クエリは「秋葉原」とし、重心がどこに決まり、マーカーがどのように出力されるかを検証した。「秋葉原」をクエリに選んだ理由は、行政表現上の秋葉原の位置と秋葉原が指す

一般的な位置との間にはギャップがあり、本提案の効果が明確に出ると考えられるからである。なお、3.4 節マーカーの分解と合成のアルゴリズムにおける閾値は、 $t=6$ とした。

5.3 結果

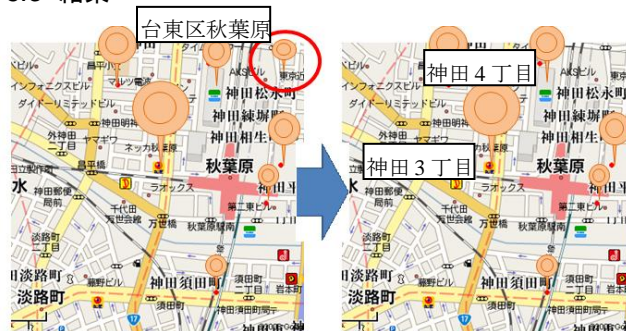


図 11 : マーカーのまとめり方の様子

行政区画表現上での秋葉原は「台東区秋葉原」であるが、いわゆる「アキバ」と呼ばれる店舗が集中している地域は「神田 3丁目」「神田 4丁目」付近である。神田 3丁目、4丁目にうまくマーカーを立てることができており、コンテンツ内容を反映した重心位置を決定できたと言える。

また、左の地図の「台東区秋葉原」のマーカーは、3.4 ii 節のマーカー合成のアルゴリズムによって、左隣りの「神田 4丁目」のマーカーに合成されている。行政区画をまたいで台東区秋葉原と千代田区外神田のノードが1つの「秋葉原」のマーカーとしてまとめられており、実際の感覚にかなり近いまとめ方ができた。

6. おわりに

地図 API サービスの大量マーカーによる一覽性低下の問題の解決手段にはマーカーを少数にまとめる必要があり、そのためにはコンテンツ内容に則したマーカーの中心位置を決めることが重要である。本研究では、動画コンテンツの長さや内容を重みとして捉えて分類基準とすることでコンテンツ内容を反映した重心を決める手法を提案した。これを利用することで、地図上の大量のコンテンツを少数にまとめることが可能になる。この方法は特に、観光地やドラマといった、行政区画表現とコンテンツの内容の指す位置との間にギャップが生まれやすいジャンルにて特に効果を発揮する。

例えば、観光地というジャンルにおける「旭川」の指す位置は市役所ではなく、旭川動物園であるといったような例が考えられる。

再生時間をページ数やファイル数に置き換えることにより、動画コンテンツからホームページや画像といったマルチメディアコンテンツ一般に適用範囲を拡大していくなどが考えられるが、これらは、今後の課題である。

参考文献

[今井 07]今井良, 森田武史, 和泉憲明, 山口高平: 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, 第 21 回人工知能学会全国大会, 2B3-1.