

Web ページのローカル度検出に基づく情報フィルタリング

松本 知弥子[†] 馬 強^{††} 田中 克己^{†††}

WWW の急激な進歩と普及によって、多くの一般のユーザが多種多様な情報を受信・発信できるようになり、情報資源の量は、日々増加し続けている。ユーザが大量の情報の中から、情報を検索することは非常に困難な作業である。特に、地域密着情報のようなローカルな情報を獲得するには、従来の検索やフィルタリング手法は不十分である。本稿では、Web ページのローカル度という新しい尺度を定義し、それに基づくフィルタリング手法を提案する。ローカル度の定義方法として、1 文書内のローカルさを特徴付ける地理用語や組織名の頻度・詳細度・位置情報と、他のページとの関係からリンク構造解析や類似しているページの数を用いる。また、ローカル度の定義を評価するための予備実験の結果を示す。

Information Filtering by Detecting Localness Degree of Web Pages

CHIYAKO MATSUMOTO,[†] MA QIANG^{††} and KATSUMI TANAKA^{†††}

The vast amount of information is available on the WWW. Usually, users use the information technologies or search engines to acquire their favorite information. However, it's still not easy to acquire local information with the conventional search engines and information filtering technology. In this paper, we propose a new information filtering method based on localness degree, which is a new measure to discover local information from the WWW. The proposed notion, localness degree, is estimated by 1) the content analysis for the estimating page, such as geographical words, position and so on, and 2) the correlation between the estimating page and its linked in pages. Its similar pages in other Websites are also considered. We also show some results of our preliminary evaluation.

1. はじめに

近年の、WWW (World Wide Web) の急激な進歩と普及によって WWW サーバから提供される HTML (Hyper Text Markup Language) ページの数も 21 億ページに達するようになり (2000/7, 米 cyveillance 社調べ), 専門家のみではなく、多くの一般のユーザがインターネットに接続し、様々な情報を共有できるようになった。

ユーザの情報獲得を支援する Web 情報検索や情報

フィルタリングにおいては、従来、主にキーワードやユーザプロフィールを用いて検索・フィルタリングが行われている。しかし、Web では地域密着情報のようなローカルな情報が多数存在し、それらの情報を獲得したり、排除したりするには、従来のフィルタリング・検索手法では不十分である場合がある。

われわれは、従来のアプローチとは異なり、Web 情報の内容によるローカル度という新しい尺度に基づいたフィルタリングを提案してきた¹⁾。

地域名や組織名などの地理情報が多く含まれており、さらにその地理情報が局所であるページや、同じ環境、目標、趣味、志向を共有する一部のユーザの集まりであるコミュニティ内での話題はローカルな内容である可能性が高い。つまり、Web 上のローカルな情報として、

- 地域情報

[†] 神戸大学大学院 自然科学研究科 情報知能工学専攻
Division of Computer and Systems Engineering, Graduate School of Science and Technology, Kobe University

^{††} 神戸大学大学院 自然科学研究科 情報メディア科学専攻
Division of Information and Media Science, Graduate School of Science and Technology, Kobe University

^{†††} 京都大学大学院 情報学研究科
Graduate School of Informatics, Kyoto University

- コミュニティでの話題

を考えることができる。

いずれにしても、Web 情報のローカル度をいかにして決定するか、その概念をどのように定義するかという問題自体、個人の主観や経験によるもので、依然曖昧であり明確な基準がないのが現状である。

そこで、本研究では、Web ページを次の二つの方向から調べることによって Web ページのローカル度を定義する。

Web ページの内容解析 1 文書内に含まれている地理用語や組織名の頻度・詳細度と、その地理用語の位置情報の密度からローカル度を定義する。
他のページとの関係 他のページからどれぐらい参照されているかというリンク構造や、他のサイトで類似した内容が扱われている数からローカルさを定義する。

そして、大量な Web ページから情報を検索するときにローカル的な情報を獲得したり、排除したりできるように、ローカル度の値の範囲によって、階層性を持ったフィルタリングを行う方法を提案する。

以下、本論文の構成を示す。まず、2 章で関連研究を示す。3 章では、Web ページのローカル度を 1 ページ単位と、他のページとの関係から定義し、統合したローカル度の定義を示す。4 章では、ローカル度を実際に計算した実験の方法と結果について述べ、考察する。5 章で、ローカル度を用いたフィルタリング手法を述べる。6 章では、まとめと今後の課題について述べる。

2. 関連研究

モバイルインフォサーチ²⁾では、ある特定の場所に関する情報をネットワーク上から収集・選択・加工し、ユーザの状況を考慮し、ユーザに適切な情報を提供する「位置指向の情報統合」を目的としている。そして、ネットワーク上のバーチャルワールドと、リアルワールドの双方向の情報を効率良くやりとりし、現在地に関する情報や、世の中の流行、お勧めをユーザに提供する実験³⁾を行っている。本研究では、1 文書の内容と、他の関連ページとの比較から Web ページのローカル度という新しい概念を定義し、それを用いたフィルタリング手法を提案しているところが異なる。

Google⁴⁾ は、リンク構造解析を用いた、検索エンジンである。Google のアルゴリズム⁵⁾ は、ページ A からページ B へのリンクをページ A によるページ B への指示投票とみなし、この投票数、つまりリンク数によりそのページの重要性を判断し、また、リンクを貼ったページについても再帰的に分析し、「重要度」の高いページによって貼られたリンクはより高く評価されて、リンクを貼られたページを「重要なもの」にしていくというものである。Google はこの重要度により、ページのランキングをしている。本研究は、リンク構造を用いてローカル度を定義している。また、リンク構造だけでなく、対象とするページの地理用語の割合も考慮して、情報のフィルタリングを行おうとする点が異なる。

Clever Project⁶⁾ は、リンク構造解析を使った、検索新エンジンプロジェクトである。Clever は最良の Authority と Hub を決定するためにリンク構造解析を繰り返す。時には、あるページが Authority と Hub の両方で、トップランクになることもある。Clever のアルゴリズムでは、最もページ間リンクが密集している場所を探し出し、コミュニティを発見することもできる。Google との主な違いは 2 つである。1 つは Google はあらかじめ各ページにリンクによるランク付けをしているが、それはどんな検索質問に対しても同じである。一方 Clever は各検索キーワードに対し異なるページの集合を集め、それに含まれるページをランク付けする（つまり、検索キーワードによって各ページのランクが変わる）。これに伴い、Google の方が速い検索が可能である。もう 1 つの違いは、Google の基本原理が「誰が、重要なページによって参照されているか」という、リンクの順方向のみを考慮したもののなのに対し、Clever は、「誰がその重要なページを参照しているか」というリンクの逆方向も同時に考慮に入れている点である。本研究では、リンク構造の他に地理用語の割合とページ内の地理用語の緯度・経度の散らばり具合を用いて、ローカル度を定義し、ローカル度によって、情報のフィルタリングを行おうとする点が異なる。

Stanford 大の Buyukkokten らの研究⁷⁾では、Web 上のページが、実際の地理上にどのように分布しているかサーバの IP アドレスや、ページの内容の地理情

報、郵便番号などのデータベースから求め、地図上に視覚的にプロットするシステムを作り、Web のページがどの地域に密着しているか分かるようにしている。本研究では、ローカル度を定義するときに、地理用語の割合とリンク構造を用いて、リンクされているページがローカルかどうかをも考慮して、ローカル的な情報を獲得したり、排除するためのローカル度を定義している点が異なる。

3. ローカル度

Web ページを 1 文書単位と他のページとの関係の二つの方向から調べることによってローカル度を定義する。

3.1 Web ページの内容解析

ローカル度を、1 ページ単位内に出現する、ローカルさを特徴付ける地理用語の頻度・詳細度と、その地理用語の位置情報によって定義する。

3.1.1 地理用語の頻度

一般的に地域への依存度の高い Web 資源・ニュースは、国・県・市・町村名などの地理情報を表す地理用語を多く含み、その文章が地域に依存した情報であるということを特徴付けている。この特徴を生かして、本研究では、記事に含まれている地理 (geographical) 用語・組織名の割合が大きい場合に、ローカル度が高く、割合が小さい場合にローカル度が低いと定義する。

ただし「姫路」を多く含んでいる記事は「日本」を多く含んだ記事よりもローカルに依存していると考えられる。すなわち、地理名詞のレベルを考慮してローカル度を計算することが必要であると思われる。そこで、場所の範囲によって i 番目に出てくる地理用語: $geoword_i$ には重み: $weight(geoword_i)$ をつける。重みは、地域国名 < 組織名 < 地域一般名 の 3 段階でつけている。地域一般名の中でも県名とその他の重みは変えている。組織名は、会社名・学校名・団体名などである。

組織名を考慮するのは「市役所」などの組織は、地域に依存して存在するものなので、組織名を多く含んだページもローカルさに関係があると考えられるからである。

以上より、記事: art_x のローカル度: $Lcl_g(art_x)$ を、

記事: art_x に書かれている名詞の総数: $totalword(art_x)$ に対する、重み: $weight(geoword_i)$ をつけた地理用語: $geoword_i$ の割合から、計算する。ただし、名詞の総数: $totalword(art_x)$ から、どの文章にも普遍的に出現し、文章を特徴付ける語にはならない「わたし」、「彼」などの stop word の数は省く。

定義式は式 (1) の通りである。

$$Lcl_g(art_x) = \frac{\sum_{i=1}^n weight(geoword_i)}{totalword(art_x)} \quad (1)$$

3.1.2 地理用語の位置情報

各記事に出現する地理名詞に対して、緯度・経度を考慮し、地理名詞の散らばり具合を調べる。

一般的に、局所的な地域の話題の記事の場合、文章に出てくる地理名詞も、局所的な地域のものである。反対に、広い範囲に渡る話題の記事の場合、文章に出てくる地理名詞も、広い範囲に渡って現れることが多い。

そこで、ページ内の地理名詞を、緯度・経度データに従って地図上にプロットした時に、全ての点を含む最小の範囲の面積を調べる。この時、MBR (Minimum Bounding Rectangle)^{8),9)} という考え方を使う。

MBR 直交する x 軸、 y 軸上にプロットされた点の最も大きい x 座標、 y 座標と、最も小さい x 座標、 y 座標を選び、

- (最も小さい x 座標, 最も大きい y 座標)
- (最も小さい x 座標, 最も小さい y 座標)
- (最も大きい x 座標, 最も小さい y 座標)
- (最も大きい x 座標, 最も大きい y 座標)

の 4 点を直線で結んでできる、プロットされた点すべてを内側に含む長方形。

本稿の場合、 x 軸の方向を緯線の方向、 y 軸の方向を経線の方向として MBR を作る。

緯度・経度のデータとしては、全国 3252 個の都道府県庁、市区役所、町村役場の所在地データを用いる。それよりも細かい地名の場合は、1 つ上のレベルに上げて、その場所の地名を用いる。

例えば、兵庫県姫路市大津区 … の場合、兵庫県姫路市の市役所所在地のデータを用いる。

記事の中の地理名詞を緯度・経度データとマッチする場合、全国には同じ名前の地名がたくさんあるという問題がある。しかし、ほとんどの場合において最初

郵便住所の町名までのデータ

表 1 日本の都道府県市町村役場の所在地の緯度・経度の最大最小
Table 1 Maximum(Minimum) Latitude and Longitude of Japan

緯度最大	45 °24	北海道稚内市
緯度最小	24 °49	沖縄県伊良部町
経度最大	145 °11	北海道羅臼町
経度最小	123 °00	沖縄県与那国町

に地名が出てくる時は， 県 市や， 県 町という風になっているので，絞り込むことができる．

また，既出の地理名詞の場合には突然，文書の中に 市と出てくる場合があるので，この場合は，既出の地理名詞かどうかを調べて判断する．

MBR の緯度・経度データを用いた計算式は，文書内に出てくる地理用語の緯度，経度の最大を $lat_{max}, long_{max}$ とし，最小を $lat_{min}, long_{min}$ とすると，

$$(lat_{max} - lat_{min})(long_{max} - long_{min}) \quad (2)$$

である．ここで出てきた記事： art_x の MBR の面積： MBR_{art_x} の数値は，日本全国を含む MBR_{japan} を式 (2) で計算した数値に対してどれぐらいの範囲を占めるかによって正規化を行う．そして，面積が小さいほどローカル度が高く，面積が大きいほどローカル度が低いので，逆数をとってローカル度を定義する．日本の都道府県市町村役場の所在地の緯度・経度の最大最小は表 1 に示す．定義式は，式 (3) である．

$$Lcl_i(art_x) = \frac{MBR_{japan}}{MBR_{art_x}} \quad (3)$$

3.1.3 地理用語の密度

3.1.1, 3.1.2 から，ページ内に出現する地理用語を地図上にプロットしたときに，それらを全て内側に含む最小の長方形：MBR を使って求めた面積の中に，どれぐらいの地理名詞がプロットされているかという，密度を求めることができる．その密度が大きいほど，ローカル度が高く，密度が小さいほど，ローカル度が低いと考えられる．

イメージを図 1 に示す．図 1 の上では，MBR の内側に 4 個の点が含まれており，下では，8 個の点が含まれている．上の例の方がローカル度が高く，下の例の方がローカル度が低いとする．そして，各地理用語には重みがついている．

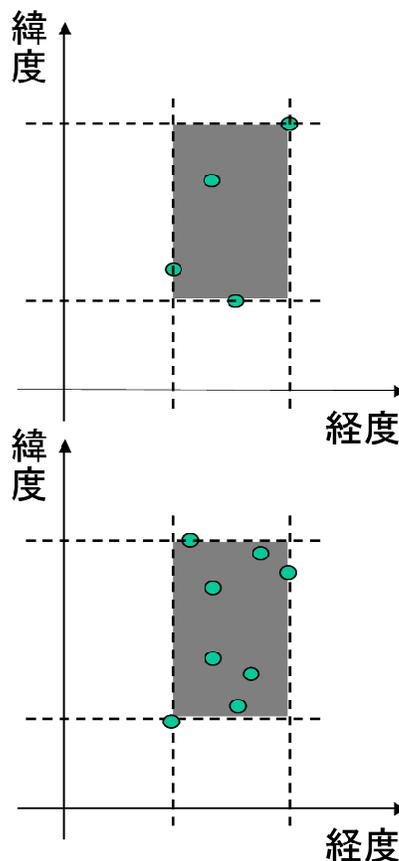


図 1 密度の例

Fig. 1 Example for Density Based Localness

これらの考え方から，1 文書内の単位での地理用語の密度によるローカル度を定義する．重みつき地理用語の記事： art_x 内の個数の合計は， $\sum_{i=1}^n \text{weight}(geoword_i)_{art_x}$ で，地理用語の位置情報から MBR を求めて，ローカル度を定義したものが， $Lcl_d(art_x)$ である．

$$Lcl_d(art_x) = \sum_{i=1}^n \text{weight}(geoword_i)_{art_x} \times Lcl_i(art_x) \quad (4)$$

3.2 関連ページとの比較

ローカル度を，参照されているページがローカルか，そうでないか，また，類似記事はどれぐらいの数があるかという他のページとの関係から定義する．

3.2.1 リンク構造解析

10 億¹⁰⁾ を超えるハイパーリンクが，WWW の膨大な情報をつないでいることから，リンク構造を解析

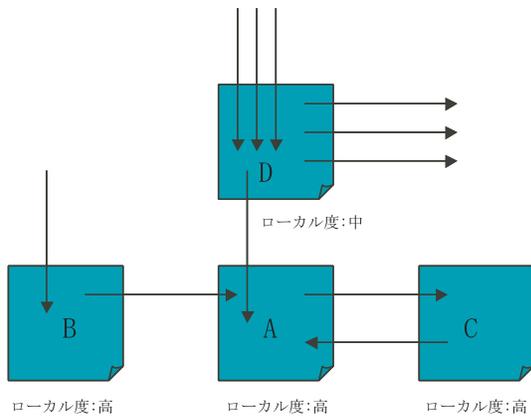


図2 ページのリンク構造とローカル
Fig. 2 link structur and localness of pages

する^{5),6)} ことにより、各ページを評価することができる。このことは、ページのローカル度を定義する時にも当てはまると考えられる。

図2で、矢印はリンクを表し、ページAのように、ローカル度の高いページ:B,Cから参照されているページは、ローカル度が高いと考えられる。また、ローカル度の低いページから参照されているページはその反対にローカル度が低いと考えられる。従って、リンク構造に基づいてローカル度を定義することが考えられる。

art_x が、ページ p_1, p_2, \dots, p_n の n 個のページからリンクを貼られているとき、リンク構造によるローカル度 $Lcl_1(art_x)$ を、

$$Lcl_1(art_x) = \frac{1}{n} \sum_{i=1}^n (Lcl_1(p_i)) \quad (5)$$

と定義する。

このようなリンク構造は、参照ページへのリンクの他にも同じ環境、目標、趣味、志向を共有するコミュニティ内部で、局所的にリンクを張り合うことによっても生まれるので、ローカルなコミュニティを発見することにもつながる¹⁰⁾

3.2.2 他のニュースサイトとの比較

ニュース記事が多くのサイトで扱われている場合、その記事の話題は、広く扱われていると判断し、ローカル度を下げる。反対に、少ないサイトでしか扱っていない場合、その記事の話題は、広く扱われていないと判断し、ローカル度を上げる。

ページAとページB間の類似度はVSM(Vector

Space Model)を用いて与えられる。

$$sim(A, B) = \frac{\overline{v(A)} \cdot \overline{v(B)}}{|\overline{v(A)}| |\overline{v(B)}|} \quad (6)$$

ここで、 $\overline{v(A)}$ 、 $\overline{v(B)}$ は、それぞれページA,Bのキーワードベクトルである。

3.3 統合ローカル度

最終的に、式(4)、(5)を統合して、

$$Lcl(art_x) = Lcl_d(art_x) \times Lcl_1(art_x) \quad (7)$$

をローカル度の定義式とすることが考えられる。

ただし、Webページの内容解析と、他のページとの関係の統合は、この組み合わせだけでなく、3.2.2で述べた、他のニュースサイトで類似記事が扱われている数を組み合わせるなど、他の定義も考えられる。

4. 予備実験

ローカル度の定義を評価する予備実験を行った。

4.1 実験

実験には、<http://www.asahi.com/>の記事を使用した。各記事のHTML文書ソースには、記事の始まる前後に、`<!-- Start of kiji -->`、`<!-- End of kiji -->`の記述があるので、その内側に書かれた記事だけを形態素解析の対象にし、それ以外のサイト内へのリンクや、広告の記述は外した。

予備実験1では、<http://www.asahi.com/>の記事の5日分約200ページのローカル度をページ内の地理用語の頻度から、式(1)により計算し、筆者の判断で正解記事を選び、適合率と再現率を求めた。システムの正解は、ローカル度0.1以上とした。その結果、再現率は0.481で、適合率は0.844という結果になった。

予備実験2では、同じ実験材料で、ローカル度を式(4)を用いて地理用語の頻度だけでなく、地理用語の位置情報や、密度を考慮して計算し、適合率と再現率を求めた。その結果、再現率は0.532で、適合率は0.933という結果になった。

図3、図4に、実験の結果を示す。実験材料全記事のうち、実際にローカルな記事が左の円、システムがローカルだと判断した記事が右の円である。それぞれの内訳の記事数を示している。

予備実験1と予備実験2について、適合率と再現率を表2にまとめて示す。

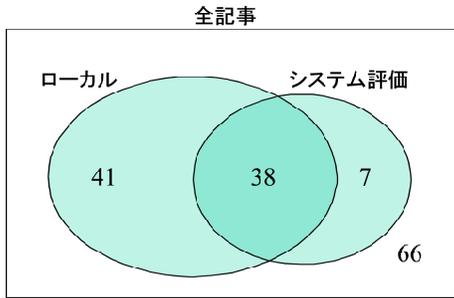


図 3 定義式 (1) による実験結果

Fig. 3 Results of Preliminary Evaluation: Case of Function (1)

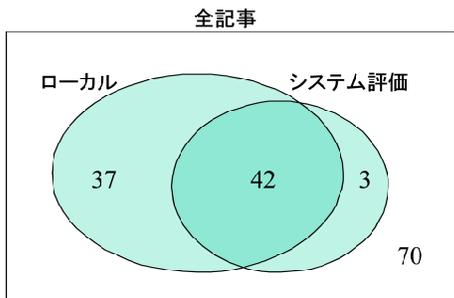


図 4 定義式 (4) による実験結果

Fig. 4 Results of Preliminary Evaluation: Case of Function (2)

表 2 適合率, 再現率

Table 2 Results of Preliminary Evaluation: Recall and Precision Ratio

	適合率	再現率
予備実験 1	0.84	0.48
予備実験 2	0.93	0.53

4.2 考察 1

予備実験 1 の結果, 図 5 のように各地の地理情報を散りばめたページの場合, 実際にはローカルな記事ではないが, 地理用語の割合が大きくなり, ローカル度が高くなって判定が失敗している. また, 局所的な地域の話題でローカルな記事の場合でも, 図 6 のように長い文章の場合, 地理用語の割合が小さいため, ローカル度が小さくなり判定が失敗している.

予備実験 2 では, 上記のような問題が解消している. これらの結果から, ローカル度の定義についてある程度の精度が得られ, ローカル度の判定に用いることができると言える.

しかし, 実験に用いたデータ数がまだ少ないことや, 局所的な地域の話題でも, 大事件の場合には, ローカ



図 5 $Lcl_g(x)$ が 0.1 以上で,

かつ, 実際はローカルでない記事 不正解例

Fig. 5 not local but $Lcl_g(x)$ is 0.1 and more: the example of inaccurate

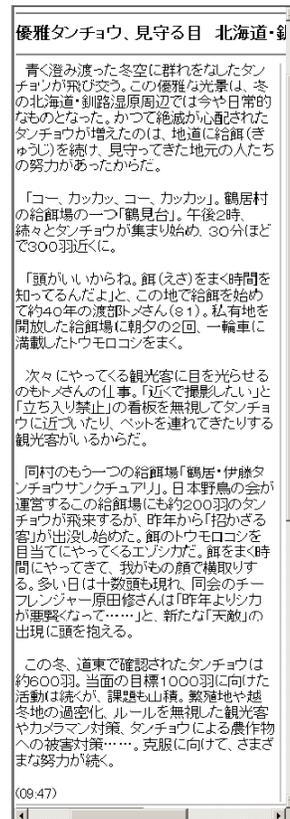


図 6 $Lcl_g(x)$ が 0.1 以下で,

かつ, 実際はローカルである記事 不正解例

Fig. 6 local but $Lcl_g(x)$ is 0.1 and less: the example of inaccurate



図 7 フィルタリングのモデル図
Fig. 7 Filtering Model Based on Locality

ルでない場合があり、判定が間違ふ場合があることなどから、今後データ数を増やし、リンク構造解析を用いた、式(5)も考慮してローカル度を判定していく必要があると思われる。

4.3 考 察 2

ローカル度の定義をさらに精度の高いものにするために、次のような改善案が考えられる。

行政単位 MBR の考え方を用いるとき、式(3)で算出した数値が同じでも、都市のスケール・分布によってローカルさは異なることが多い。そこで、その地名が実際にカバーする範囲を入れて MBR を計算することが考えられる。

例えば(大阪市,堺市)と(大阪市,神戸市)の場合には、大阪市と堺市はともに大阪府に所属し、大阪市と神戸市を含む最小の行政単位は、近畿地方で、大阪府 < 近畿地方なので(大阪市,堺市)のローカル度を高くする。

段落構造 文章に地理名詞が出てくる場合、最初の段落に、どの地域が関係しているか示されることが多い。そのため、長い記事の場合には、式(1)の値が相対的に小さくなってしまふ。

そこで、段落構造を考慮して、地理用語の割合を計算することも考えられる。

日常語の割合 地名、組織名以外にも日常生活や、ある領域に関係する語の割合も考慮すると、さらに精度がよくなると考えられる。

例えば、名産品、名所旧跡、レジャー施設、文化施設、宿泊施設、お食事処、交通施設、行政施設、イベントなどの日常語である。

5. フィルタリング機構

前章までで定義したローカル度を用いて、Web ページをフィルタリングする方式を提案する。

検索エンジンなどで、検索窓にキーワードを入力して、情報を得たい場合、ローカルに依存した情報を排除したいことがある。その場合、ローカル度の高いページを検索結果から外す。

反対に、ローカルな情報を得たい場合は、ローカル度の高いページを検索結果の上位にランキングするようにする。

ただし、フィルタリングを行う時、各ページには、計算したローカル度情報があるので、その値の範囲によって、階層性を持ったフィルタリングを行う。

フィルタリングモデルを図7に示す。ユーザが検索質問をすると、WWW サーバから検索エンジンが情報を検索し、その結果からローカル度を用いて、ローカルな情報を獲得したり、排除したりできるようにフィルタリングを行う。

また、ユーザの望む場所に関するページを獲得するために、検索のオプションとして、場所を指定することなどが考えられる。

6. おわりに

Web 情報検索や情報フィルタリングにおいては、従来、主にキーワードやユーザプロフィールを用いて検索・フィルタリングが行われている。本研究では、これらのアプローチとは異なり、Web 情報がどの程度地域やコミュニティに特化したローカルな情報なのかという視点が、情報検索や、情報フィルタリングの時

に、ローカルな情報を獲得したり、ローカルな情報を排除したりすることを可能にするためにローカル度という新しい尺度を定義した。

Web 情報を取得する上で、ユーザがどの程度のローカルさを持つ情報に興味があるかは、ユーザによって大きく異なる。そこで、本研究では、Web 情報のローカル度をいかにして決定するか、その概念をどのように定義するかという基準の定義を提案してきた。

本論文で提案した、ローカル度の定義をまとめると、以下のとおりである。

- 各ページのローカルさを特徴付ける地理用語の頻度・詳細度とその地理用語の位置情報からローカル度を求める。
- 他のページからどれくらい参照されているかというリンク構造や、他のサイトで類似した内容が扱われている数からローカルさを定義する。

予備実験で、<http://www.asahi.com/> の記事のローカル度を地理用語の割合から計算すると、再現率は 0.48 で、適合率は 0.84 という結果になり、地理用語の密度を考慮して計算すると、再現率は 0.53 で、適合率は 0.93 という結果になった。一定程度の精度は得られ、情報検索、フィルタリングの基準とすることができると考えられる。

今後は、ローカル度の定義をさらに見直し、情報を受信するユーザが望む程度のローカルさの情報を配信する方法などを検討していく予定である。

謝辞 本研究の一部は、文部省科学研究費基盤 (C) 「分散型ハイパーメディアからの構造発見とアクセス管理」(課題番号 12680416) 及び、日本学術振興会未来開拓学術研究推進事業における研究プロジェクト「マルチメディア/コンテンツの高次処理の研究」(プロジェクト番号 JSPS-RFTF97P00501) によっております。ここに記して謝意を表すものとします。

参 考 文 献

- 1) 松本知弥子, 馬強, 田中克己: 地理情報を用いたニュースのフィルタリング機構, 情報処理学会第 62 回全国大会, Vol. 3, pp. 505-506 (2001).
- 2) 三浦信幸, 高橋克巳, 横路誠司, 島健一: 位置志向の情報統合 ~ モバイルインフォサーチ 2 実験 ~, 情報処理学会第 57 回全国大会, Vol. 3, pp. 637-638 (1998).
- 3) MIS2 (モバイルインフォサーチ 2 実験): <http://www.kokonon.net/>.

- 4) Google: <http://www.google.com/>.
- 5) Brin, S. and Page, L.: The anatomy of a large-scale hypertextual Web search engine, *Computer Networks and ISDN Systems*, Vol. 30, No. 1-7, pp. 107-117 (1998).
- 6) Kleinberg, J. M.: Authoritative sources in a hyperlinked environment, *Journal of the ACM*, Vol. 46, No. 5, pp. 604-632 (1999).
- 7) Buyukkokten, O., Cho, J., Garcia-Molina, H., Gravano, L. and Shivakumar, N.: Exploiting Geographical Location Information of Web Pages, *WebDB (Informal Proceedings)*, pp. 91-96 (1999).
- 8) Guttman, A.: R-trees: A dynamic index structure for spatial searching, *Proc. ACM SIGMOD Conference on Management of Data*, Vol. 14, No. 2, pp. 47-57 (1984).
- 9) Zaniolo, C., Ceri, S., Faloutsos, C., Snodgrass, R. T., Subrahmanian, V. S. and Zicari, R.: *Advanced Database Systems*, The Morgan Kaufmann (1997).
- 10) Chakrabarti, S., Dom, B., Kumar, S. R., Raghavan, P., Rajagopalan, S. and Tomkins, A.: ハイパーリンクを賢く使う (梶谷浩一訳), 日経サイエンス (1999).