

実世界におけるユーザの移動とプロフィールに基づく Web サイト間リンクの再編成

神吉 訓嘉*¹, 舟生 日出男*¹, 平嶋 宗*¹, 吉高 淳夫*¹

*¹ 広島大学大学院工学研究科

概要 近年, Web 空間上には, 飲食店, 書店など, 実世界の施設についての情報が, Web サイトによって提供されていることが多い。しかし, 多くの場合, 近接の施設の Web サイト間をつなぐリンクが提供されていない。そこで, 実世界での人の行動に着目すると, 人は意図や嗜好性, 興味関心に依りて, いくつかの施設を転々と移動することが少なくない。また, 同じような意図や嗜好性を持つ人は, 同じように移動する傾向にあると言える。こうした実態を考慮して, 本研究では, 1) 実世界における人々の施設間の滞在・移動履歴を蓄積し, 2) 人々の性質が記述されたユーザプロフィールを用いて Web サイト間のリンクを編成してユーザに提供するシステムを構築した。

A Reorganization Method of Links among Web Sites Based on Movement and Profiles of Users in the Real World

Noriyoshi Kanki*¹, Funaoi Hideo*¹, Tsukasa Hirashima*¹, and Yoshitaka Atsuo*¹

*¹ Graduate School of Engineering, Hiroshima University

Abstract: In late years, facilities in real world (ex. restaurants, bookstore, etc) provide users with their information on their Web sites. The Web sites, however, scarcely provide links among them, even if the facilities are in the neighborhood. On the other hand, people in real world usually move from one facility to another according to their intent, taste or interest. In addition, those who have similar intent, taste or interest tend to move similarly. In this study, taking such situation into consideration, we developed a system which 1) stores stay and movement history of users in real world, 2) dynamically creates links among Web sites using user profiles which represent their tastes and interests.

1. はじめに

1.1. 実世界における人々の行動と Web サイト

近年, インターネットの普及にともない, 多様な情報が Web 上に蓄積されている。その中でも, 飲食店, 書店, スーパーなど, 様々な施設に関する情報が, それぞれの施設を運営する組織の Web サイトによって提供されることが増えている。だが一方で, 地理的に近接している施設であっても, それらの Web サイト同士にはハイパーリンク(以降, リンク)がなく, Web 空間上では近接していないことも少なくない。

しかし, 実世界において, それらの施設を訪れる前に, Web サイト上の施設情報を予め調べて行動計画に活かす人が増えている。そのため, 既存のリンクだけでは, 近隣施設の Web サイト間のリンクが乏しく, 不十分である。

例えばここで, 書店 A に行くことを決め, その後, どこに行くべきかを計画する場合を考える。このとき, 書店 A の Web サイト上に, 「書店 A に滞在した後, 実世界において多くの人が次に移動していった施設の Web サイトへのリンク」があれば, そうしたリンクには人々の実際

の行動が反映されているため, 行動計画の参考になる。同様に, 他の施設のサイト上にも, そのようなリンクがあれば, 実世界における人々の振る舞いに基づいて様々な施設の Web サイトをたどっていくことができるため, 有用であると考えられる。

また, 一般に人は, 意図や嗜好性, 興味関心に基づいて複数施設を移動することが多いと言える。例えば, 女性なら, コーヒーショップでカフェオレを飲んだ後, スーパーで食材を購入し, 帰りがけに洋菓子屋に寄ってケーキを買う, 本が好きな人なら, 書店をいくつも渡り歩く, といったことが考えられる。そこで, 性別や趣味, 年齢, 嗜好などの属性に関する情報をユーザプロフィールとして記述しておけば, 1つないしは複数の属性で絞り込み, リンクを動的に再編成することで, 目的に応じた施設の Web サイトをたどることが可能となる。

こうした実態を考慮して, 本研究では, 1) 実世界における人々の施設間の滞在・移動履歴を蓄積し, 2) 人々の性質をユーザプロフィールとして記述し, 3) それらに基づいて, Web サイト間のリンクを編成してユーザに提供するシステムの構築を目指す。

1.2. 既存 Web リンクと実世界リンク

本研究では、「既存 Web リンク」を、「施設が提供する Web サイトが既に持っているリンク構造」、「実世界リンク」を、「人の滞在・移動データに基づいて編成された Web サイト間のリンク構造」とそれぞれ定義する。

先述したように、本研究の目的は、「実世界リンク」を編成し、ユーザに提示するシステムを開発することである。例えば、書店 A、レンタルショップ B、スーパー C、コンビニ D の Web サイトがそれぞれあり、既存 Web リンクとして、書店 A からレンタルショップ B へのリンクのみがあったとする。しかし、実世界の人々の滞在・移動の様子を眺めると、書店 A からスーパー C、レンタルショップ B からコンビニ D への移動が頻繁に見られるとする。このとき、実世界リンクとして、書店 A からスーパー C、レンタルショップ B からコンビニ D へのリンクを編成し、ユーザに提供することが、本研究の目標である。

以降、2 章では関連研究との比較を行い、3 章では滞在施設の検出方法について述べる。次に、4 章で実世界におけるユーザの行動と Web 空間との一致方法について述べた後、5 章でシステム構成について述べる。6 章で本手法の評価について述べ、7 章でまとめと今後の課題について述べる。

2. 関連研究

本章では、Web 空間上から実世界の施設関連情報を取得する研究と、実世界における人の移動履歴を利用した研究とを比較する。

Web 空間上から実世界の施設関連情報を取得する研究では、地域情報に特化した検索システムが多数開発されている。例えば、Web が持つ住所情報と地図が持つ住所情報との一致を見ることで、地図インタフェースを用いてユーザが地図上で指定する範囲内の施設情報の取得を可能にしたリ[1][2]、「検索対象」+「地理条件」の検索クエリで検索することを目指したり[3]、などである。これらの研究ではその状況にあった施設関連情報の取得を目指しており、本研究のように人の複数施設間移動に基づいた Web ブラウジングを考慮していない。

実世界における人の移動履歴を利用した研究では、GPS を用いて実世界上のどのような場所を訪れているかのデータを取得し、そのデータを基にして、ある場所から次に訪れる可能性の高い場所の予測を行うことや[6]、実世界における観光動態の取得を目指すことが行われている[7]。しかし、これらの研究では実世界の施設と Web 空間との関係が考慮されていない。

本研究では実世界におけるユーザの移動履歴と、実世界と Web 空間との関係とを考慮して、Web サイト間のリンク作成を目指す。

3. 滞在施設情報の記録方法

3.1. 施設情報

本研究では、滞在施設の位置などの情報(以降、施設情報)を以下のように定義した。

- 1) 施設 ID,
- 2) カテゴリ-ID,
- 3) 施設名,
- 4) 北西の経度・緯度(x1, y1),
- 5) 南東の経度・緯度(x2, y2)

3.2. 滞在施設情報記録システム

滞在施設情報記録システム(図 1)のハードウェアは、PDA(SONY 製 VGN-U71P)と GPS 受信ユニット(SPA 製イオ 111USB, 図 1 中央上部)から構成される。ソフトウェアは、Microsoft Visual C++ Ver.6.0 を用いて開発した。



図 1. 滞在施設情報記録システム

3.3. 滞在判定方法

本研究では、次の 2 条件を同時に満たすとき、施設に滞在していると判定することとする。

- 1) GPS の緯度・経度データが、ある施設の領域内に含まれる(施設情報の北西・南東の緯度経度と比較),
- 2) 施設領域内に、「滞在判定時間」以上、留まっている。

1)の条件では、GPS によって得られる位置座標情報に、様々な理由による誤差が含まれるために、その誤差を考慮する。[8]によれば、GPS の測位誤差は $2drms$ (The distance root mean square: 放射状測位誤差の自乗平均の正の平方根)を半径とする円内に 95%の測位点が入るとされており、

$$2drms = 2HDOP \times UERE \quad (\text{ただし, } UERE = 2.0)$$

と定義されている。

ここで、UERE は User Equivalent Ranging Error (利用者等価測距誤差)である。また、HDOP は Horizontal Dilution Of Precision (水平方向精度劣化指数)であり、GPS 電波の受信状況によって変動する。受信状況が良好な環境では HDOP は概ね 1 から 2、測位誤差は 6m 前後、一方、受信状況が悪い場合は HDOP の値は 7 から 9、測位誤差は 32m 前後と見積もられる。1)の条件については、GPS によって得られる位置座標が、施設情報で設定された施設領域内に含まれるか否かを、測位誤差 $2drms$ を加味して判断する。

2)の条件では、「滞在判定時間」を閾値としているのは、単なる通過ではなく、「滞在して、何らかの行動を行

つたと考えられる施設」のみを検出するためである。本研究では「滞在判定時間」を、施設領域の周囲を徒歩で通過するのに要する平均時間とし、これを超えて、その施設領域内に利用者が存在する場合に、「滞在了」と判定する。

滞在判定時間 TS_i は、施設情報の北西・南東の緯度・経度を基に、施設 F_i の矩形領域の縦幅を H_i 、横幅を W_i 、GPS の測位誤差を $2d_{rms}$ 、平均的な歩行速度を v として、次式で求める。

$$TS_i = (H_i + W_i + 2 \times 2d_{rms}) / v$$

なお、GPS は、建物に入ると電波を受信できなくなるため、その期間、その建物内の施設に、ユーザが「滞在了」と見なすことができる。そのため本研究では、「一つの建物 = 一つの施設」を前提としている。

3.4. ユーザプロフィール

ユーザの特性や、興味・関心、嗜好性を記述するために、14 種類の属性とそれらの属性値を設定した(表 1)。予め各ユーザにユーザプロフィールを記述してもらうことで、実世界リンクを構成する際に、特定の属性に基づいて、リンクを絞り込むことができる。例えば、「性別」で「女」を、かつ、「趣味」で「買物:服」を選択すれば、服を買うことが好きな女性の移動・滞在履歴に基づいた実世界リンクが構成されることになる。

表 1. ユーザプロフィール

属性	属性値
性別	男, 女
職業	大学院生, 大学生
健康状態	体力がない, 普通, 体力がある
⋮	⋮
活動時間	朝方, 夜型
⋮	⋮
嗜好	広島風, 関西風, … 料亭, ファミリーレストラン, … ラーメン, うどん, そば, … ファーストフード(全 23 種類)
趣味	アニメ, カラオケ, 音楽, 買物: 本, 買物: 服, … スポーツすること: テニス, … スポーツ観戦: 野球(全 31 種類)

3.5. 予備実験

被験者 16 名(男 12 女 4)に、滞在施設情報記録システムを用いて、施設への滞在・移動に関する情報(以下、自動取得データ)を自動取得してもらうとともに、滞在・移動についての正解情報(以下、正解データ)を手動で記録してもらった。

また、被験者 16 名の内、15 名(男 11 女 4)には、ユーザプロフィールを記述してもらった。

実験のべ期間は 106 日であり、滞在のべ回数は 368 回、1 日の平均滞在回数は 3.5 回となった。

自動取得データの Recall, Precision を正解データに

基づいて算出したところ、

$$\text{Recall} = 253 / 368 = 0.67$$

$$\text{Precision} = 221 / 250 = 0.85$$

となった。

この結果を検証するために、施設をカテゴリーで分けて滞在数を比較する(表 2)とともに、正解データと自動取得データそれぞれを用いて、地図上に「実世界リンク」を表示して比較を行った。

表 2 に示すように、自動取得データに含まれる滞在数は、正解データのそれよりも少ないが、164/209 = 78.5%を網羅していることが分かる。また、図 2, 図 3 の丸で囲った部分のように、自動取得データでは作成できていないリンクが見られるが、それは一部にとどまっている。さらに、表 2 の滞在数を χ^2 検定した結果、有意差は見られなかった($\chi^2(3) = 0.20, ns$)。このことから、滞在数が特定のカテゴリーに偏っていることはなく、全体的な傾向には差がないことが分かる。これらの結果から、実験システムによって、滞在施設データをほぼ正確に自動で取得できていると言える。

表 2. 滞在数の比較

	正解滞在・移動データ (209 件)	自動取得滞在・移動データ (164 件)
食事	16.7% (35/209)	17.7% (29/164)
買物	25.4% (53/209)	25.0% (41/164)
遊び・スポーツ	5.3% (11/209)	6.1% (10/164)
その他	52.6% (110/209)	51.2% (84/164)



図 2. 正解データ



図 3. 自動取得データ

4. 施設と Web 空間との一致

予備実験で取得した正解データを基に、ユーザが滞在した施設の内、どの程度の数が Web サイトを持っているのかを調査した。検索エンジン Google に対して、それぞれの施設名と住所を基にした検索要求を与え、その結果の上位 30 件までを対象として、当該施設の Web ページの有無を判断した。このとき、施設が持つ Web ページをサイト単位で扱った。

4.1. 施設が持つ Web サイトの定義

[9]に基づき、Web サイトを、「一般的に各個人、組織によって提供されていると認識される情報の単位」、つまり「特定の個人あるいは組織が作成・管理するひとま

とまりのページ群」と定義した。具体的には、Web 集合を、トップページを根として、そこから URL の根のアドレスが変化するまでリンクで辿れる Web ページ群を1つのサイト単位と考える。このように Web ページを扱うのは、サイト単位でまとめた方が、ユーザの Web ブラウジングが簡便になると考えられるからである。

4.2. Web サイトのトップページの定義

施設の Web サイトのトップページを、以下の条件で定義した。

- 1) Web ページ名にその施設名、もしくは、その施設を運営するグループに関する名称が含まれている。
- 2) URL のファイル名部分の内、本体部分が index, top, welcome など、拡張子部分が html, htm, jsp, php などである。
- 3) トップへ戻るなどのバックリンクを持つ。ただし、単一の Web ページのみで構成されている場合はバックリンクを持たないので、そのページをトップページとする。

4.3. Web サイトを持つ施設数

調査の結果、予備実験でユーザが滞在した施設数は 104、その内、Web サイトを持つ施設数は 57 であり、 $57/104 = 54.8\%$ の施設が、Web サイトを有していることが分かった。低い割合にとどまっているが、その理由は、小規模の施設では Web サイトを提供していないことが多いためであると考えられる。

5. システム構成

システム構成を図 4 に示す。ユーザが、起点となる Web サイトを選択する(図 4 ①)と、システムは、その Web サイトのトップページをユーザに返す(図 4 ②)。そのページからの実世界リンクを見るには、ユーザプロフィールの属性値を選ぶ(図 4 ③)。すると、システムは、その選択された属性値と一致するユーザの滞在・移動データを抽出して、そのデータに基づく実世界リンクを構成し(図 4 ④)、ユーザに提示する(図 4 ⑤)。ユーザは、提示された実世界リンクの中から、任意の Web サ

イトを選択し、閲覧する。この作業を繰り返して Web ブラウジングを行う。

なお、ユーザが本システムを携帯して、実世界において、施設間を移動し、滞在することで、ユーザの滞在・移動データが、データベースに蓄えられる。

6. 評価実験

6.1. 既存 Web リンクと実世界リンクとの比較

6.1.1. 目的

提案システムによって提供される実世界リンクと、既存 Web リンクとを比較し、実世界リンクには、既存 Web リンクには見られない新しいリンクが生成されていることを検証する。

6.1.2. 方法

評価で取り扱う施設が持つ Web サイトは、予備実験で取得した施設であり、かつその施設が Web サイトを持っている場合を対象とする。それらの Web サイト群について、既存 Web リンクと、予備実験で得られた正解データを基にした実世界リンクとを比較する。方法としては、1) Web サイトのリンク数による数値比較と、2) そのリンク構造を地図上に表示しての視覚による比較を行う。

6.1.3. 考察

既存 Web リンク構造と実世界リンク構造とを数値比較した結果を表 3 に示す。

まず、実世界リンクの数は、既存 Web リンクのそのものの 2.8 倍となっており、既存のリンクにはない、新しいリンクが構成されていることが分かった。

次に、同一サイト内のリンク数を比較すると、既存 Web リンクでは、全リンク数のほぼ半数となっているのに対して、実世界リンクでは、ごく少数にとどまっている。このことから、実世界リンクでは、既存のリンクにはない、異なる組織の Web サイト間のリンクが数多く構成されていると言える。この点について確認するために、地図上にリンク構造を表示したところ、実世界リンクは図 5、既存 Web リンクは図 6 のようになった。これらの図の丸で

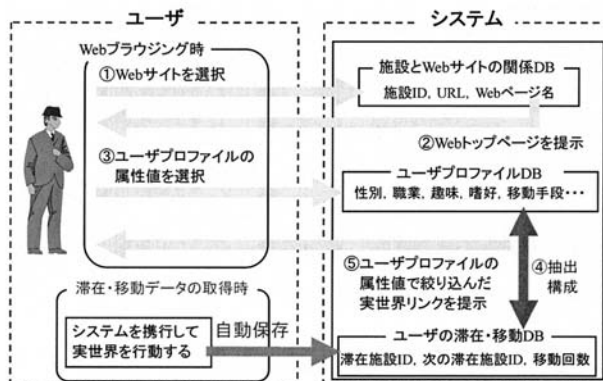


図 4. システム構成図

囲っている部分を見比べると明らかなように、実世界リンクには、既存 Web リンクにない Web サイト間のリンクが張られていることが分かった。

表 3. 既存 Web リンクと実世界リンクの数

	Web サイト間	同一サイト内	計
既存 Web リンク	25	26	51
実世界リンク	143	2	145



図 5. 実世界リンク



図 6. 既存 Web リンク

6.2. ユーザプロフィールに基づく実世界リンクの変化

6.2.1. 目的

実世界リンクについて、ユーザプロフィールの属性に応じて、滞在・移動履歴をグループ化し、それぞれのグループ毎にリンクを構成した際に、それらのリンク構造には、グループ毎の特徴が反映されているか否かを評価する。これにより、ユーザプロフィールによってリンクを構成・提示することで、ユーザの意図や嗜好性により適合した Web ブラウジングが可能となるかを考えることができる。

6.2.2. 方法

プロフィールを取得した被験者 15 名(男 11 女 4)の正解データの内、Web サイトを持つ施設のみを対象とし、ユーザプロフィールの属性値の中から「性別: 男」、「性別: 女」、「趣味: 買い物(本)」、「趣味: カラオケ」で絞り込み、計 4 つの実世界リンクを作成した。それらを地図上に表示するとともに、数値で示して比較を行った。

6.2.3. 考察

15 名全員のデータを基にした実世界リンクを図 7、表 4 に示す。これに対して、ユーザプロフィールの属性値、「性別: 男」、「性別: 女」、「趣味: 買い物(本)」、「趣味: カラオケ」で絞り込んだ実世界リンクをそれぞれ、図 8~11、表 5~8 に示す。

これらを比較すると、まず、「全員」の実世界リンクと属性値で絞り込んだそれとでは、リンク構造が異なっていることが読み取れる。さらに、「性別: 女」は、「買物(スーパー・デパート)」へのリンクが強くなる傾向が見られた。また、「趣味: 買物(本)」では「本」へのリンクが、

「趣味: カラオケ」では「カラオケ」へのリンクが強くなる傾向が見られた。これらの結果から、ユーザプロフィールの属性値を用いた絞り込みにより、実世界リンクの構造が動的に変化する可能性が示されたとと言える。



図 7. 全員のデータを基にした実世界リンク

表 4. 全員のデータを基にした実世界リンク構造

大分類	小分類	割合 (実数値)
食事		16.7% (35/209)
買物		25.4% (53/209)
	スーパー・デパート	12.4% (26/209)
	本	7.2% (15/209)
遊び・スポーツ		5.3% (11/209)
	カラオケ	2.4% (5/209)
その他		52.6% (110/209)



図 8. 「性別: 男」



図 9. 「性別: 女」



図 10. 「趣味: 買物(本)」

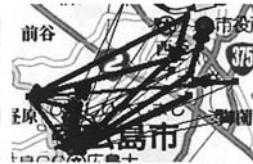


図 11. 「趣味: カラオケ」

表 5. 「性別: 男」の実世界リンク構造

大分類	小分類	割合 (実数値)
食事		16.8% (33/196)
買物		24.5% (48/196)
	スーパー・デパート	11.7% (23/196)
	本	7.7% (15/196)
遊び・スポーツ		5.6% (11/196)
	カラオケ	2.6% (5/196)
その他		53.1% (104/196)

表 6.「性別: 女」の実世界リンク構造

大分類	小分類	割合 (実数値)
食事		15.4% (2/13)
買物		38.5% (5/13)
	スーパー・デパート	23.1% (3/13)
	本	0.0% (0/13)
遊び・スポーツ		0.0% (0/13)
	カラオケ	0.0% (0/13)
その他		46.2% (6/13)

表 7.「趣味: 買物(本)」の実世界リンク構造

大分類	小分類	割合 (実数値)
食事		16.1% (26/161)
買物		24.8% (40/161)
	スーパー・デパート	11.8% (19/161)
	本	9.3% (15/161)
遊び・スポーツ		1.9% (3/161)
	カラオケ	0.0% (0/161)
その他		57.1% (92/161)

表 8.「趣味: カラオケ」の実世界リンク構造

大分類	小分類	割合 (実数値)
食事		10.1% (9/89)
買物		15.7% (14/89)
	スーパー・デパート	6.7% (6/89)
	本	2.2% (2/89)
遊び・スポーツ		9.0% (8/89)
	カラオケ	5.6% (5/89)
その他		65.2% (58/89)

7. まとめと今後の課題

本論文では、実世界における人の移動に基づいて Web サイト間リンクの再編成する「実世界リンク」を提案した。また、ユーザの嗜好性や興味関心を記述したユーザプロフィールと組み合わせることで、目的にあわせて実世界リンクを動的に編成する方法についても説明した。

評価実験の結果、人の移動に基づく実世界リンクには、既存 Web リンクには存在しなかった、web サイト間のリンクが多く含まれることが示された。また、ユーザプロフィールに基づいて実世界リンクを分類することで、人の趣味・嗜好にあったリンク構造を提示できる可能性についても示された。しかし、今回の評価実験では、被験者数が必ずしも十分ではないため、可能性を示すにとどまっている。

今後は、被験者数を増やして、実世界リンクの既存 Web リンクとの差を検証する予定である。また、ユーザプロフィールを用いた絞り込みによって動的に編成された実世界リンクについて、統計的に検証することも考えている。

参考文献

- [1] 横路誠司, 高橋克己, 三浦信幸, 島健一: “位置指向の情報の収集、構造化および検索手法”, 情報処理学会論文誌, Vol. 41, No. 7, pp. 1987-1998 (2000)
- [2] 平松薫, 石田亨: “地域情報サービスのための拡張 Web 空間”, 情報処理学会論文誌, データベース, Vol. 41, No. SIG6 (TOD7), pp. 81-90(2000)
- [3] 大井峻, 遠山元道: “地理条件を伴う Web ページ検索におけるランク手法の提案”, 情報処理学会研究報告. DBS, Vol. 2005, No. 68, pp. 547-551 (2005)
- [4] 服部峻, 湯本高行, 田中克己: “モバイル環境における実空間相応度に基づく情報ランキング手法”, 情報処理学会研究報告. DBS, Vol. 2005, No. 68, pp. 553-560 (2005)
- [5] 深澤祐介, 長沼武史, 藤井邦浩, 倉掛正治: “タスクモデルを利用したユーザの行動予測に基づくサービス提示システム”, 情報処理学会研究報告. ICS, Vol. 2006, No. 2, pp. 35-42 (2006)
- [6] 大田智数, 波多野賢治, 吉川正俊, 植村俊亮: “ウェアラブル環境における行動履歴を用いた情報提示”, 電子情報処理学会技術研究報告. DE, Vol. 101, No. 110, pp. 1-7 (2001)
- [7] 長尾光悦, 川村秀憲, 山本雅人, 大内東: “GPS ログマイニングに基づく観光動態情報の獲得”, 観光情報学会誌, Vol. 1, No. 1, pp. 38-46(2005)
- [8] 安田明生: “GPS 技術の展望”, 電子情報通信学会論文誌, B, 通信, Vol. J84-B, No. 12, pp. 2082-2091 (2001)
- [9] 池田哲夫, 森憲一, 竹野浩, 佐藤哲司: “Web ページ集合からのサイト再構成の一手法”, 情報処理学会論文誌, Vol. 44, No. 3, pp. 751-759(2003)