モバイルコンピューティングと
ワイヤレス通信　23−30
高度交通システム　11−30
（２００２．１１．２９）

地図を用いたWebページ検索システムのログ解析

平松 薫†

デジタルシティ京都プロトタイプにおいて「ジオリンク京都」として公開運用を行った，地図を用いた Web検索システムのログ解析結果について報告する．本システムは，運用開始の 1999年12月からの約 3年間で約 5万人に利用され，地図インタフェースを利用したユーザの各種検索行動が Webサーバ上のアクセスログに記録された．本稿では，このアクセスログに基づき，カテゴリ検索とキーワード検索の利用傾向と，地図上にアイコンで表示された Webページの選択傾向を分析するとともに，テキスト主体の検索サービスとの比較を通じて，地図を利用した Webページ検索におけるユーザの検索行動を考察する．

Log Analysis of Map-based Web Page Search System

Kaoru Hiramatsu†

This paper analyzes log data of a map-based Web page search system which is a part of the Web site of the Digital City Kyoto prototype. During 3 years open experiment (from December 1999), the system was accessed by over 50,000 users and recorded all sorts of search activities. The log data enabled us to analyze search function usages, such as category search and keyword search on the map images. In this paper, the characteristics of the map-based user interface compared with the characteristics of traditional text-based search engines are explained.

1　はじめに

現在，インターネット上に蓄積されている大量のコンテンツを利用した様々な情報サービスが展開されつつあり，都市に関する情報もインターネット経由で容易に手に入るようになりつつある．こうした中，地図を媒体とした情報提供方法が，都市に関する情報を扱う方法の１つとして注目を集めており[1, 2, 3, 4]，筆者もデジタルシティ京都プロトタイプ[5]において地図を用いた Webページ検索システムを構築し，「ジオリンク京都」としてインターネットで公開運用を行ってきた[6]．

地図を用いた Webページ検索では，Webページに記述された内容だけでなく，実際の都市における地理的な位置に基づいて情報を検索することができる．従ってその検索過程は，インターネット上の Web情報だけでなく，実際の都市空間も意識した検索過程を表すと考えられる．

そこで本稿では，Webサーバ上に記録されたログに基づきジオリンク京都の利用状況を分析し，地図インタフェースに実装した検索機能とテキスト主体の検索サービスの利用傾向の比較と，地図インタフェース上での Webページの選択傾向から，

† 日本電信電話株式会社
NTTコミュニケーション科学基礎研究所
Nippon Telegraph and Telephone Corporation
NTT Communication Science Laboratories

地図インタフェース上で都市に関連した Webページがどのように検索されていたかを明らかにする．

以下本稿*では，2章で分析対象であるジオリンク京都の概要について説明し，3章で検索機能別の利用傾向とテキスト主体の検索サービスの比較を試みる．地図イメージ上における Webページの密集とアクセスの集中に関する分析を 4章で行い，5章で全体のまとめを述べる．

2 ジオリンク京都

ジオリンク京都では，地理的な属性評価を Webページ検索に取り入れ，検索結果を地図にリンクされることにより，都市情報への入り口としての役割と実世界に即した情報検索の実現を目指している．

2.1 検索対象の特徴

ジオリンク京都では，Webページに関する所有権やプライバシといった社会的問題を解決するため，Webページの所有者から使用許諾が得られたWebページだけを検索対象としている．1999年10月に約2,600件の Webページが登録された形で公開運用を開始し，2002年10月現在，約4,800件が登録されている．

また公開運用では，京都市の広域地図イメージ1枚（図 1 参照）と東西2km南北1.5kmの詳細地図イメージ 108枚（図 2 参照）で京都市内をカバーしている．

2.2 システム構成

ジオリンク京都の地図インタフェース[6]は，サーバ上の情報検索システム GeoLinkと連携して検索を実行する．このシステムの全体構成を図 3 に示す．地図インタフェースは，Javaアプレットで構築しており，Webページの表示と同様の手順で起動することができる．また，地図インタフェースと GeoLinkは，3層クライアント・サーバ構成に

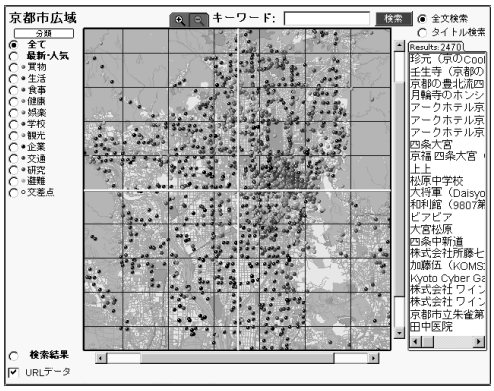

図 1: 地図インタフェースに表示した広域地図

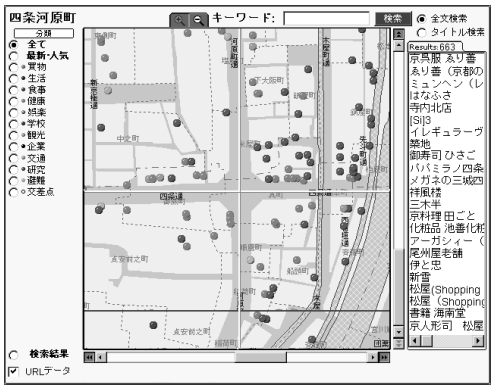

図 2: 地図インタフェースに表示した詳細地図

従っており，Webサーバ上の CGIプログラムを中継して通信を行う．

2.3 検索機能

ジオリンク京都では，地理的な属性の評価を Webページの検索に取り入れ，京都市内の Webページを対象とした近傍検索，カテゴリ検索，およびキーワード検索を実現している．各検索機能の検索結果は地図上にアイコンの分布として表示され，それぞれをマウスでクリックすると対応する Webページが表示される．

1. **近傍検索**
   マウスで地図イメージ上の任意の点をクリックすると，地図上に白い線で表示されている十字カーソルの中心がその点へと移動し，地図上に表示されている情報のタイトルが中心点

*本稿の内容は，これまでの発表[6, 7]に 2002年10月までに記録されたログの分析結果を反映させたものである．

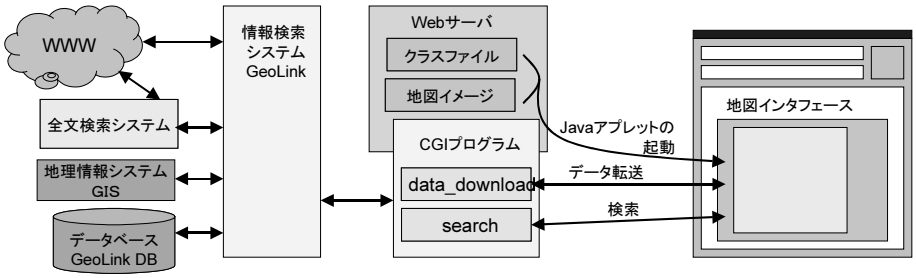

図 3: 情報検索システム GeoLinkと地図インタフェースの構成

から近い順にソートされて右側のリストボックスに表示される．

2. **カテゴリ検索**
   画面左側に表示されている分類から希望の分類を選択することが可能であり，「食事」の「中華」のように大分類と小分類を組み合わせて絞り込みを行うことができる．
3. **キーワード検索**
   画面上部のテキストボックスにキーワードを入力すると，Webページのタイトルとホームページ内部に記述されている情報，そして Webページから抽出した地理的な属性を対象とした検索が実行される．

2.4 検索行動を記録するための設定

Webサーバのアクセスログには，以下のようにユーザの行動が記録される．なお，アクセスログ中でユーザを識別するために，Webブラウザに対してクッキーを配布するように設定している．

- 地図インタフェースの起動
  → Javaクラスファイルのダウンロード
- 検索機能の利用
  → 検索処理用 CGIプログラムの呼び出し
- Webページの選択
  → リダイレクト CGIプログラムの呼び出し

Webサーバ上のアクセスログには，地図インタフェースの操作以外のアクセス履歴も一緒に記録されるため，まずパターンマッチングによるクリーニング処理を行い，地図インタフェースに関連した操作を抽出する．また，2人以上のユーザが同時

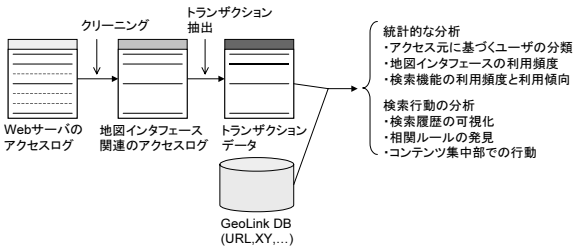

図 4: 検索行動分析手順の概要

に地図インタフェースを利用している場合に対応するため，トランザクション抽出により，地図インタフェースの起動から終了までの操作をユーザごとにまとめる．なお，地図インタフェースの設計上，明確な終了操作は記録されないので，一定時間†以上操作間隔が空いた時点でトランザクションの終了とした．最後に，Webページに対応する地理座標を付加して，地理属性を考慮した検索行動の分析を行う．以上の分析の流れを図 4 に示す．

3 検索機能別の利用状況

ジオリンク京都の地図インタフェースの運用は，インターネット上に公開されたデジタルシティ京都プロトタイプにおいて行った．この Webサイトの構築については，1999年 5月に公式な報道発表を行い，それ以降，継続的に新聞，雑誌，書籍において紹介されてきた．また，Yahoo! Japan や，goo，MSNなどのポータルサイトから検索可能であることも確認している．

†本分析では 2時間とした．

表 1: 検索機能の利用状況 *

	利用者数	総利用回数	頻度
地図インタフェース	2,524	8,110	
検索機能の利用者	1,339	5,446	4.07
利用なし	1,185	2,664	2.25
近傍検索	411	1,102	2.68
カテゴリ検索	1,128	6,563	5.82
キーワード検索	594	2,230	3.75

* 対象期間：2000/2/4 ～ 2000/3/23

3.1 アクセスユーザとリピータ

Webサーバのアクセスログに記録されたクッキーに基づき，地図インタフェースの利用者数を集計したところ，約 3年間‡で延べ 50,770人であった．このうち，2000年2月4日からの 50日間の利用者数と利用回数の集計を表 1 に示す．この期間中，地図インタフェースに実装した検索機能を一度以上利用したユーザは 1,339人で，延べ 5,446回の利用があった．一方，検索機能を 1 度も利用しなかったユーザは 1,185人で，延べ 2,664回の利用であった．また，地図インタフェースの利用回数を利用ユーザ数で割った利用頻度で比較すると，検索機能を利用したユーザが 4.07 であったのに対し，利用しなかったユーザは 2.25 であった．

この差は，実際に地図インタフェースで検索機能を利用したユーザと利用しなかったユーザの差を示しており，実際に検索機能を利用したユーザの方が，地図インタフェースを繰り返し使っていたことが分かる．

また，地図インタフェースに実装した検索機能の中では，カテゴリ検索が最も利用されていた．

3.2 カテゴリ検索の利用傾向

地図インタフェースのカテゴリ検索で利用されたカテゴリの割合を図 5 に示す．この割合をみると，食事，買物，交通，観光などのカテゴリの利用率が高く，緊急避難，研究，健康，企業などのカテゴリの利用率が低かったことがわかる．

‡集計期間は 1999/12/10 ～ 2001/10/26

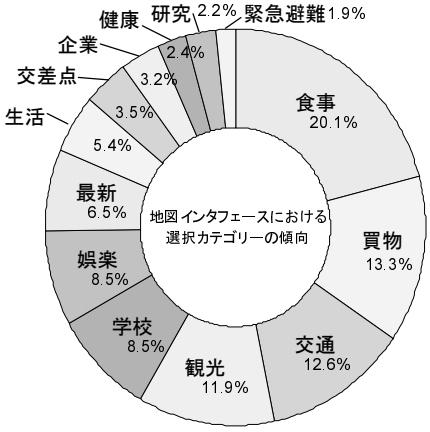

図 5: カテゴリの選択傾向

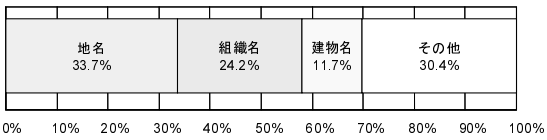

図 6: 検索キーワードの傾向

3.3 キーワード検索の利用傾向

地図インタフェースのキーワード検索で利用されたキーワードの割合を図 6 に示す．この割合をみると，住所や通り名，駅名など地名による検索が 33.7%，学校名や企業名など組織名による検索が 24.2%，ビルなど建物名による検索が 11.7%と，都市に実在するオブジェクトの名称をキーワードとして利用した検索が全体の 69.6%を占めた．これに対し，「ホテル」「ラーメン」といった一般名詞やその一部分など，その他の表現を利用した検索は 30.4%であった．

3.4 テキスト主体の検索サービスとの比較

地図インタフェースのカテゴリ検索の利用傾向は，ディレクトリの利用に関する Yahoo! JAPAN のアンケート結果[8]と類似しており，現時点でのインターネットの利用者層やその利用目的の影響が大きかったと思われる．また，実験を行った場所の影響により，実際に京都を訪問した時に役立つような情報が含まれるカテゴリの利用が顕著で

あったと考えられる．

一方，キーワード検索は，ニュースなどに関連した事柄が多く検索されるテキスト主体のキーワード検索のランキング(参考 [9])とは大きく異なっている．地図インタフェースのキーワード検索では，地名や組織名，建物名といった都市に実在するオブジェクトの具体的な名称が利用されることから，ユーザは表示された地図の影響を受け，キーワードを入力していたと考えられる．

4 Webページの密集とアクセスの集中

次に，検索結果として表示される Webページごとの利用頻度を分析してみる§．まず，地図インタフェースから検索可能な Webページの分布を図 7 に示す．'+' が Webページの位置を表しており，縦軸が南北を，横軸が東西を示す．また各軸の座標値は，十進形式の緯度経度で表示している．

図 7 中央やや右，(135.77, 35.00) 付近にみえる Webページの集中は，京都市内の四条河原町交差点周辺に対応するものである．この交差点周辺は，数多くの店舗が密集した京都市内の繁華街であるが，地図インタフェース上においても Webページの密集がみられた．この交差点周辺の Webページの密度は $1km^2$ あたり 599件に達している．

一方，Webサーバに記録されたアクセスログを集計し，Webページそれぞれに対するアクセス数を高さ方向の軸に表示したものを図 8 に示す．この図からは，二種類のアクセスの集中を読み取ることができる．

一つは，図 8 中央左 (135.73, 35.03) 付近にみられる人気 Webページへのアクセスの集中である．同様の集中は京都市内に数件存在するが，それぞれの周辺の Webページのアクセス数から突出していることから，地理的な原因によるものではないと考えられる．

もう一つは，図 8 中央右の四条河原町交差点周辺の Webページ密集部へのアクセスの集中である．

§分析対象期間は 1999/12/1 ～ 2001/4/26

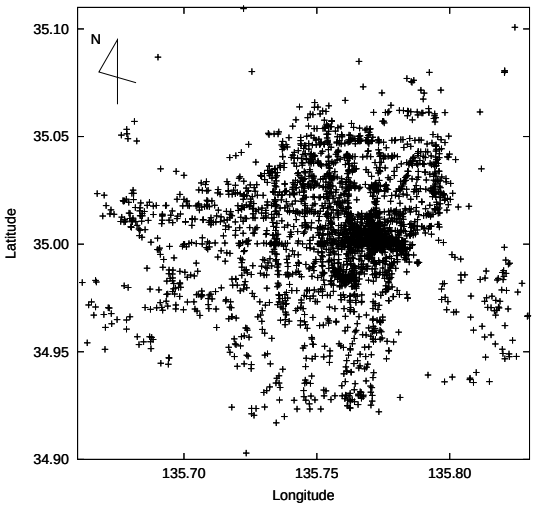

図 7: 検索対象の Webページの分布

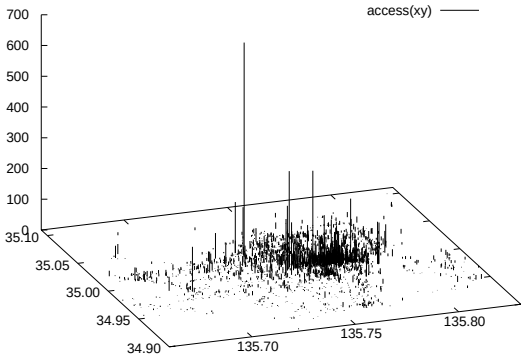

図 8: Webページのアクセス数

この範囲では特定の Webページへのアクセスの集中がみられないことから，特定範囲に Webページが密集することにより，アクセスの集中が発生したと考えられる．

そこで，この Webページの密集とアクセスの集中の関係を分析するために，図 7 の対象範囲を格子状に区切り，各区画ごとの Webページの密度とアクセス頻度の関係を考えてみる．まず，緯度経度で 0.005度四方の区画に分割し，各区画の Webページの密度を示した図 9 をみると，四条河原町交差点周辺付近に，Webページの密集を示すピークが一つあり，周辺に行くのに従い密度が漸次減少していることがわかる．これに対し，各区画に属す

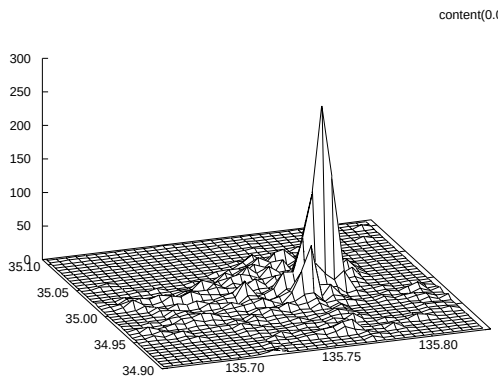

図 9: Webページの密度 (1)

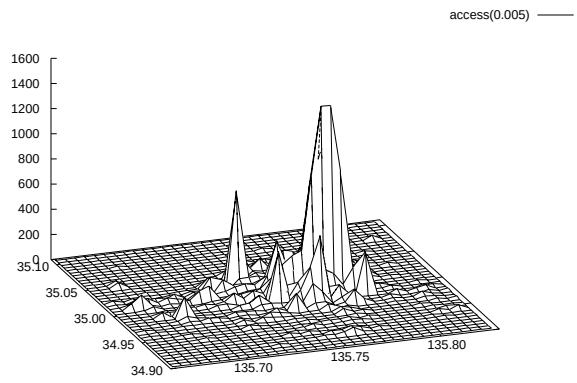

図 10: アクセス頻度 (1)

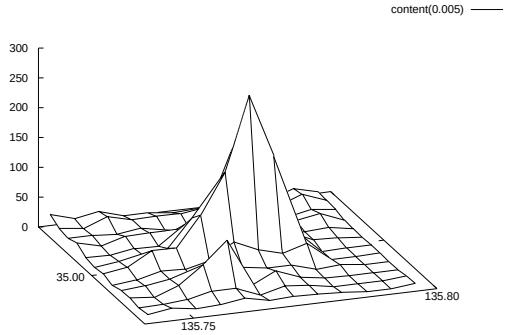

図 11: Webページの密度 (2)

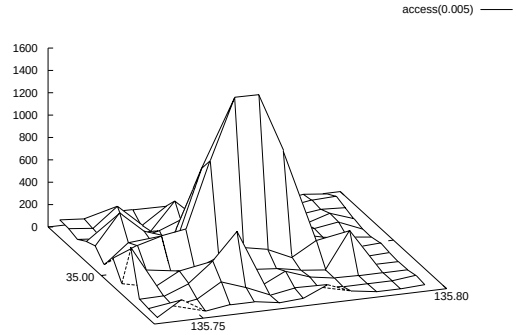

図 12: アクセス頻度 (2)

る Webページへのアクセス頻度を示した図 10 では，これまでの指摘のように二つのアクセス集中がそれぞれピークとして現れている．まず左側のピークは，人気 Webページへのアクセスの集中がピークとして現れたものである．一方，右側のピークは，四条河原町交差点周辺に属する Webページへアクセスが集中したことにより現れたピークである．この二つのピークの形状を比較すると，左側の人気 Webページによるピークが鋭く尖っているのに対し，右側の四条河原町交差点周辺のピークは頂点は比較的曖昧で太いものとなっている．また，図 9 と図 10 で右側のピークを比較すると，Webページの密度のピークよりアクセス頻度のピークは鈍化し，中腹部が広がっていることがわかる．

四条河原町交差点周辺のピークの差異は，四条河原町交差点周辺部を拡大した図 11 と図 12 を比較するとよりはっきりする．アクセス頻度を表した図 12 におけるピークの鈍化と中腹部の広がりは，Webページが密集するエリアへのアクセスが，その周辺部にも波及したためと考えられる．しかし，図 12 のアクセス頻度のピークが，図 11 の Webページの密度のピークより曖昧になっていることから，密度の増加によるアクセスの増加には限界があると予想される．一方，人気 Webページによるアクセス頻度のピークの鋭さは，周辺の Webページのアクセス頻度が低かったことを示しており，Webページ密集部にみられたような周辺部へのアクセスの波及効果が，人気 Webページにはなかったと考えられる．

この Webページの密度とアクセス波及効果の関係は，連続してアクセスされたWebページ間の距離からも読み取ることができる．この Webページ

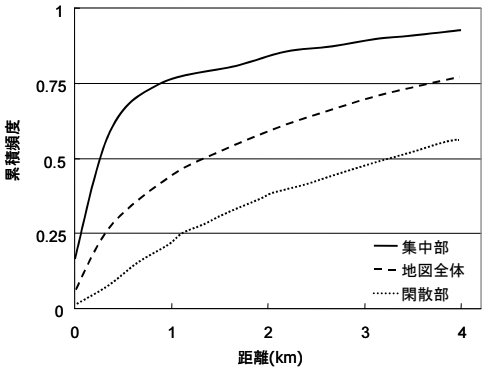

図 13: 連続アクセスの Webページ間の距離

間の距離は，2.4 節で抽出したトランザクションの中で，ユーザが連続してアクセスした Webページの地理的な座標に基づき計算した．

連続してアクセスされた Webページの起点が，四条河原町交差点周辺部の Webページ密集部に属する場合と，山間部など Webページが疎らな散在部に属する場合で比較すると，地図全体の平均値が $2.380km$ であるのに対し，密集部の平均値が $1.035km$，散在部の平均値が $4.154km$ であり，それぞれの間に有意な差がみられた¶．この Webページ間の距離とその累積頻度の関係を示した図 13 をみると，密集部へアクセスしたユーザは約 3/4 の確率でその次の Webページを $1km$ 以内から選択しており，最初に選択した Webページの近傍に留まり次の Webページを選択する傾向が強いことがわかる．また，地図上の Webページをランダムに選択する連続アクセスのシミュレーションを行い，密集部もしくは散在部に属する Webページを起点とする場合のシミュレーションの結果‖とアクセスログにおける平均値を比較（表 2）したところ，密集部，散在部，地図全体それぞれで有意な差がみられた**．

以上の比較から，地図イメージ上の Webページ

¶t検定（分散が等しくないと仮定した2標本による検定）で $P < .01$ （両側）

‖それぞれの場合について 10,000回の試行を実施．

**t検定（分散が等しくないと仮定した2標本による検定）で $P < .01$ （両側）

表 2: 連続アクセスの Webページ間の距離

	密集部	地図全体	散在部
アクセスログ	1.035	2.380	4.154
シミュレーション	3.369	4.978	7.109

※ 単位は km

の密集部では，他のエリアより近くの Webページをユーザが選択していたといえる．また，多くのユーザのアクセスが密集し，そこに留まるという意味では，地図イメージ上における情報の密集が街として捉えられ，インターネット経由でアクセスしたユーザの行動に影響を与えていたともいえるだろう．

5 おわりに

本稿では，ジオリンク京都のアクセスログを分析し，地図を利用した Webページ検索におけるユーザの検索におけるユーザの検索行動を考察した．検索機能に関する各種統計情報や Webページ密集部におけるユーザ選択行動は，地図イメージで表示された都市空間を介して，ユーザがインターネット上の情報をどのように捉えているかを理解するためのヒントになると考えられる．

特に，アクセス頻度と検索機能の利用傾向の関係や，検索結果のアクセス状況に関する分析は，情報検索システムを改善するためだけでなく，都市に関連する Webページの発信戦略を決定する際にも有用と考えられることから，本稿の分析が，システム開発者，情報の利用者や発信者へのフィードバックとなれば幸いである．

謝辞

ジオリンク京都に対して，コンテンツを提供してくださった皆様に感謝します．そして，デジタルシティ京都の構築に関して深くご指導頂いた京都大学石田亨教授，デジタルシティ京都実験フォーラムに参加された皆様，NTTオープンラボに協力してくださった方々に感謝します．

参考文献

[1] マピオン: http://www.mapion.co.jp/.

[2] マップファンウェブ: http://www.mapfan.com/.

[3] 高木悟: 分散型電子地図プラットホーム JaMaPS, 画像ラボ, Vol. 9, No. 12, pp. 27–32 (1998).

[4] 星野隆, 長谷川靖, 池田哲夫, 星隆司: ディレクトリを用いた情報提供に関する一手法〜地図情報への適用〜, 情報処理学会第120回データベースシステム研究会報告, pp. 137–144 (2000).

[5] Ishida, T.: Digital city Kyoto, *Communication of the ACM*, Vol. 45, No. 7, pp. 76–81 (2002).

[6] 平松薫, 小林堅治, Benjamin, B., 石田亨, 赤埴淳一: デジタルシティにおける情報検索のための地図インタフェース, 情報処理学会論文誌, Vol. 41, No. 12, pp. 3314–3322 (2000).

[7] Hiramatsu, K.: Log Analysis of Map-based Web Page Search on Digital City Kyoto, *Digital Cities II*, Lecture Note in Computer Science, Vol. 2362, pp. 233–245 (2002).

[8] Yahoo! JAPAN: 第6回ウェッブ・ユーザー・アンケート (1999). http://www.yahoo.co.jp/docs/result/result6/.

[9] Lycos: Lycos 50. http://50.lycos.com/.