

地理的関連性を用いたWeb検索技術の研究

平松 薫

概要

近年のインターネットの普及により，ネットワーク上の情報量は爆発的に増大している．こうした中，世界規模で情報を統一的に扱うインターネットにおいても，人々の生活圏である地域や都市を単位とした情報空間が新たな情報流通の形態として注目を集めており，地域ポータルサイトやデジタルシティと呼ばれる Web サイトが，各地で立ち上がり始めている．

こうした地域を単位とした情報空間には，今後，日常生活に密着した情報が大量に集積されるようになると予想される．大量の情報から必要な情報を効率よく選択するためには検索機能が重要となるが，現在インターネット上で利用可能な検索機能では不十分と考える．たとえば，キーワード検索だけで土地勘のない場所の情報を検索するのは困難であるし，ディレクトリサービスを利用する場合もある程度の事前知識が必要となる．また，地図を使った検索の場合も，特定地域に情報が密集すると，コンピュータ画面の面積や解像度の限界により，表示できる情報は限定されてしまう．

この地域情報検索に関する問題は，インターネット上の電子的な情報空間と都市の物理的な空間を融合させ，使いやすいユーザインタフェースから利用できるようにすることで解決されると予想している．たとえば，実世界に存在する様々な関連性によって情報同士を結び付けることができれば，インターネット上の電子的な情報空間を実世界に即した形に拡張することができる．そして，その拡張した情報空間を使いやすいユーザインタフェースを通して気軽に利用できるようになれば，日常生活の様々な場面で地域情報を活用できるようになると考える．

そこで本研究では，都市生活に密着した地域情報サービスを実現することを目的に，地理的関連性を用いた Web 検索技術の構築を目指す．本研究のアプローチとその成果を以下に示す．

- 地理的関連性を用いた Web 情報空間の拡張
電子化された地理情報を利用してインターネット上の Web ページ間に地理的なリンクを動的に生成し，Web 情報空間を拡張する手法を提案する．この地理的なリンクを地理的ジェネリックリンク，拡張した Web 情報空間を拡張 Web 空間とよぶ．拡張 Web 空間では，Web ページに関する条件とハイパーリンクと地理的ジェネリックリンクに関する条件を組み合わせると，Web ページの列を検索することができる．本研究

では、この拡張Web空間に対する問い合わせを SELECT-FROM-WHERE 形式で記述できる検索言語を定義するとともに、複数の検索例を用いてその記述性を確認した。

一方、拡張Web空間における検索は、半構造データの検索と捉えることができるが、その効率化を図るためには検索空間に適した手法が必要となる。そこで、拡張Web空間に基づく情報検索システムを試作し、実際の京都に関する Web情報と地理情報を利用して複数の最適化手法の比較実験を行い、拡張Web空間における問い合わせ処理に適した最適化手法を明らかにする。その結果、動的計画法を用いた事前最適化では、問い合わせに含まれるリンク数が増加すると最適な条件処理順序の作成が困難となる場合があり、その原因がリンク関係の不規則さに起因する予測の誤差の累積にあることがわかった。これに対し、最適な条件の選択とその処理を繰り返す実行時最適化では、処理状況を反映した正確な予測が可能で、手順的に予測誤差の累積も無い。実装した結果も事前最適化より良い性能を示し、最大で二倍の性能向上を観測した。

- 地域情報検索のための地図インタフェース

特定都市に関連する Webページを検索するためのユーザインタフェースとして、地図を利用した検索インタフェースを構築した。この地図インタフェースには、検索機能として、近傍検索、カテゴリ検索、およびキーワード検索を実装し、検索結果を地図上に重ね合わせて表示することによって、実世界に即した Webページ検索を可能にするとともに、クライアントにキャッシュしたデータを検索に再利用することによって、インタラクティブな検索処理とサーバの負荷の軽減を実現した。

また、構築した地図インタフェースをデジタルシティ京都の実験サイトにて公開運用し、Webサーバのアクセスログに記録されたユーザの検索履歴の分析により、実装した検索機能の利用傾向と、情報が集中した地域におけるユーザの検索行動を明らかにした。

目次

1	序論	1
1.1	研究の背景	1
1.2	研究の目的	2
1.3	論文の概要	3
2	インターネット上にみられる地域性	7
2.1	都市とインターネット	7
2.1.1	都市空間と情報検索	8
2.1.2	都市空間と情報空間の融合	10
2.1.3	コミュニケーションの場	11
2.1.4	利用動向の分析	12
2.2	デジタルシティの構築	13
2.2.1	京都における試み	14
2.3	まとめ	18
3	地理的関連性を用いた Web空間の拡張	21
3.1	都市情報とデータモデル	23
3.1.1	半構造データ	23
3.1.2	Web検索言語	24
3.1.3	地理情報システムとWWW	24
3.1.4	XMLによる情報空間の集約	25
3.2	拡張Web空間	27
3.2.1	基本的なアイデア	27
3.2.2	地理的ジェネリックリンク	28

3.2.3	検索パス式	32
3.3	検索言語	33
3.3.1	文法	33
3.3.2	意味論	34
3.4	問い合わせの記述例	36
3.4.1	地理的ジェネリックリンクを用いた検索	36
3.4.2	繰り返し表現を含んだ検索式	36
3.4.3	シナリオの検索例	37
3.4.4	3D京都との連携	39
3.5	拡張Web空間に基づく情報検索システム	41
3.5.1	情報検索システムの構成と検索方法	42
3.5.2	性能評価と最適化手法の比較	56
3.6	まとめ	66
4	地域情報検索のための地図インタフェース	67
4.1	関連事例	68
4.1.1	インターネット上の地図情報サービス	68
4.1.2	地図を利用した情報検索	69
4.2	地図を用いたWeb検索インタフェース	69
4.2.1	基本的な考え方	69
4.2.2	近傍検索	70
4.2.3	カテゴリ検索	71
4.2.4	キーワード検索	72
4.3	Javaアプレットを用いた地図インタフェースの構築	73
4.3.1	構成	73
4.3.2	検索処理の分散	74
4.3.3	性能評価	75
4.4	位置情報を用いた携帯端末用地図インタフェースの構築	78
4.4.1	Mobile InfoMap の構成	78
4.4.2	システム構成	81
4.4.3	性能評価	83
4.5	地図を用いたWeb検索行動の分析	85
4.5.1	検索行動を記録するための設定	86

4.5.2	アクセスユーザの特徴	89
4.5.3	地図インタフェースの利用頻度	91
4.5.4	検索機能の利用傾向	91
4.5.5	アクセス履歴の可視化	95
4.5.6	地理的な集中とアクセスの集中の関係	97
4.5.7	人気 Web ページに関連する相関ルール	103
4.6	まとめ	105
5	考察	107
5.1	都市空間と情報のマッピング	107
5.2	問い合わせ処理の最適化に関わる課題	109
5.3	ユーザインタフェースの機能拡張	110
5.4	コミュニケーションツールとの連携	114
6	結論	115

図一覧

2.1	WWW文書の地理的分散（参考文献[1] より）	8
2.2	3D京都	16
2.3	二条城バスツアー	18
2.4	デジタルシティ京都実験フォーラムのトップページ	19
2.5	デジタルシティ京都実験フォーラムのリンクリスト	20
3.1	拡張Web空間の位置付け	27
3.2	ハイパーリンクの例	29
3.3	地理的関連性を用いた地理的ジェネリックリンクの例	29
3.4	検索パス式による検索パスのパターン	33
3.5	検索言語の文法	34
3.6	100m以内に存在する建物Webページを検索	36
3.7	繰り返し表現を含んだ問い合わせ記述例	37
3.8	例題前半の記述例	38
3.9	例題後半の記述例	39
3.10	3D京都との連携	40
3.11	周辺情報検索の記述例	40
3.12	拡張Web空間に基づく情報検索システムの構成	43
3.13	拡張Web空間におけるインデックスの例	43
3.14	POI-DBを利用した地理的ジェネリックリンクの生成	44
3.15	図 3.13 のインデックスに対する問い合わせ	45
3.16	図 3.15 の検索処理の例	46
3.17	繰り返し表現を含む検索パス式の展開	49
3.18	動的計画法による事前最適化アルゴリズム	53
3.19	制限付き事前最適化アルゴリズム	55

3.20	性能比較のための例題 (1)	57
3.21	最適化手法と処理時間	58
3.22	制限付き事前最適化の再帰の深さと処理時間	59
3.23	制限付き事前最適化の再帰の深さと入出力量	59
3.24	性能比較のための例題 (2)	61
3.25	問い合わせに含まれるリンク数と検索処理時間	62
3.26	検索処理時間と入出力量 (リンク数4の場合)	62
3.27	性能比較のための例題 (3)	64
3.28	条件数と検索処理時間	65
4.1	近傍検索の実施例	70
4.2	飲食店カテゴリの検索結果	71
4.3	「バス停」の検索結果	72
4.4	情報検索システム GeoLinkと地図インタフェースの構成	73
4.5	拡大地図イメージに切替えた地図インタフェース	76
4.6	Mobile InfoMap の概要	79
4.7	Mobile InfoMap の検索画面	80
4.8	コミュニケーションツールの画面	81
4.9	Mobile InfoMap のシステム構成	82
4.10	測位機能付携帯端末 Naviewn と携帯電話	83
4.11	位置情報に基づいた主画面	84
4.12	検索処理時間の内訳	86
4.13	Webサーバに記録されるアクセスログの例	87
4.14	Webサーバが発行するクッキーの例	87
4.15	検索行動分析手順の概要	89
4.16	検索キーワードの傾向	93
4.17	カテゴリの選択傾向	94
4.18	各カテゴリの検索回数	94
4.19	選択された Webページ数のカテゴリ	94
4.20	カテゴリ検索 1回あたりに選択される Webページ数	94
4.21	アクセス履歴の可視化手法	96
4.22	アクセス履歴の可視化 (1)	96
4.23	アクセス履歴の可視化 (2)	96

4.24	検索対象の Web ページの分布	98
4.25	Web ページのアクセス数	99
4.26	Web ページのアクセス数とアクセス順位の関係	99
4.27	Web ページの密度 (1)	101
4.28	アクセス頻度 (1)	101
4.29	Web ページの密度 (2)	101
4.30	アクセス頻度 (2)	101
4.31	連続してアクセスされた Web ページ間の距離と周辺密度	102

表一覧

2.1	インターネット上の地図を用いた情報サービス	9
2.2	3D仮想空間の構築例	10
2.3	地域情報を扱う Web サイトと運営母体	15
3.1	属性比較の際の一致条件の緩和	31
3.2	地理的ジェネリックリンクで用いる地理的關係演算例	35
3.3	問い合わせ処理の最適化で用いる予測式	52
3.4	制限付き事前最適化の再帰の深さと条件処理順序	60
3.5	情報検索システムの入出力量の比較 (1)	60
3.6	リンク数4の場合の条件処理順序	63
3.7	情報検索システムの入出力量の比較 (2)	63
3.8	9条件が指定された場合の条件処理順序	65
4.1	地図インタフェースの動作速度の測定環境	77
4.2	検索機能の所要時間 (1)	78
4.3	検索機能の所要時間 (2)	78
4.4	図 4.11 の検索結果の詳細	84
4.5	国別アクセス件数	90
4.6	JP第二ドメイン別アクセス件数	90
4.7	県別アクセス件数	90
4.8	検索機能の利用状況	91
4.9	ユーザの利用手順を示した状態遷移確率	92
4.10	連続してアクセスされた Web ページ間の平均距離	103
4.11	人気 Web ページに関連する相関ルール	104
4.12	相関ルールに出現した Web ページ	104

第 1 章

序論

1.1 研究の背景

近年，World Wide Web (WWW)の普及に代表されるように，インターネット上の情報量は増加しており，その利用機会も急激に増加している．ビデオリサーチネットコム[2]が実施している「インターネット普及状況調査」によると，2000年 9月の日本のインターネット世帯利用率が30.5%，インターネット利用人口の推計値が 3,383万人であったのに対し，2001年 9月にはそれぞれ 44.7%，4,787万人へと増加している．この 2001年 9月の利用人口 4,787万人のうち，自宅からの利用者が 3,993万人，そのうちパソコンによる利用者が 3,306万人，携帯電話による利用者が1,826万人であった．また，インターネット利用場所の構成割合の推移をみると，2000年 9月には自宅外のみ，すなわち職場や学校でのみインターネットを利用するユーザが28.3%であったのに対し，2001年 9月には 16.6%へと減少していた．この調査から，インターネットへのアクセス手段が多様化し，ビジネスシーンだけでなく日常生活へと急激に浸透しつつあることがわかる．

しかし，同調査中の世帯内パソコンにおけるインターネット利用状況によると，検索ポータル系の Webページに週一度以上アクセスするインターネット利用者が 82.3%であるのに対し，地域・タウン情報系の Webページに同じ頻度でアクセスするインターネット利用者は 5.1% にすぎなかった．この傾向は，調査の行われた 2000年 9月から 2001年 9月にかけて，インターネットの利用者数が急激に増加し，その利用機会も多様化したにも

関わらずほとんど変化していない。従って、インターネットの利用者は、近所の学校や病院の情報など、日常生活に関わる身近な情報を得るためにも、地域・タウン情報系の Web ページではなく、世界規模で構築された大規模な検索サービスを利用しているのではないかと予想される。

こうした中、世界中の情報へアクセスできるインターネットにおいても、地域性を持った情報流通形態が注目を集めはじめており、地域ポータルサイトやデジタルシティ[3, 4, 5]の構築が各地で始まっている。都市に関する情報も増加する傾向にあり、Web ページの形でネットワーク上に大量に発信されつつある。ただ、現在発信されている都市に関する主な情報は、街中の繁華街にあるレストランの情報や駐車場の空き情報など、人口や建物と同様に都市中心部に集中する傾向にある。

1.2 研究の目的

従来から、地域限定のニュースやイベント情報、地図情報などを提供する情報サイトはインターネット上に存在している。しかし、これらのサイトが提供する情報検索手段は、都市部に集中する大量の情報から必要な情報を選び出すためには不十分と考える。例えば、ディレクトリサービスでは分類に従い情報が整理されているが、ユーザが探したい情報に関する知識をある程度持っていないと、目的の情報までスムーズにたどり着くことはできない。キーワード検索が可能な場合も、適切なキーワードを指定しないと、無関係な検索結果が大量に表示されてしまうため、その中から必要な情報を再度探し出さなくてはならない。また、従来の多くの検索サービスでは、検索結果がリスト形式で表示されるため、検索結果間の実際の都市における地理的な関係を知りたい場合には、別の手段を利用しなくてはならない。

この地域情報検索に関する問題は、インターネット上の電子的な情報空間と都市の物理的な空間を融合させ、使いやすいユーザインタフェースから利用できるようにすることで解決されると予想している。たとえば、実世界に存在する様々な関連性によって情報同士を結び付けることができれば、インターネット上の電子的な情報空間を実世界に即した形に拡張することができる。そして、その拡張した情報空間を使いやすいユーザインタ

フェースを通して気軽に利用できるようになれば，日常生活の様々な場面で地域情報を活用できるようになると考える．

さらに，こうした情報空間を都市に住む人々が共有できるようになれば，情報を媒介としたコミュニティが形成され，街への帰属意識も高まっていくと予想される．デジタルシティや地域ポータルサイトが核となり，様々なサービスが利用できるようになることが望ましいが，インターネット上の Web サイトを都市生活のための新たな社会基盤として機能させるためには，様々な要素技術の開発が必要となる．

そこで本研究では，都市生活に密着した地域情報サービスを実現することを目的に，都市に関する情報を扱う上で必要となる要素技術のうち，情報検索と情報提供に焦点を当て，以下の課題の解決を試みる．

- Web ページを中心としたインターネット上の情報空間に都市の地域性をどのように反映させるか？
- 都市の地域性を反映させた情報空間の効率的な検索方法は？
- 検索インタフェースにはどのような機能が必要か？
- 検索システムはどのように利用されるのか？

1.3 論文の概要

前節で述べた課題を解決するためには，情報を記述するためのデータモデルや検索アルゴリズムなど，システム構築に必要な理論の確立とともに，情報検索システムやユーザインタフェースの構築，そしてその実地検証が必要と考える．そこで本研究では，実際に都市情報に関する情報検索システムを構築しながら，以下のようなアプローチで検討を進めていく．

- 地理的関連性を用いた Web 情報空間の拡張
(第 3 章)
- 地域情報検索のための地図インタフェース
(第 4 章)

本論文は、本章以下、全 6 章で構成される。

第 2 章では本研究の背景として、都市とインターネットの関係とデジタルシティの構築例について述べる。本研究で提案する拡張 Web 空間と地図インタフェースは、デジタルシティ京都の情報検索システムを実現する過程で生まれたものであり、実際の Web サイトにおける公開運用を通して具体化させてきたものである。デジタルシティ京都に関わる活動は、主に 1998 年 10 月から 2001 年 9 月までの 3 年間で行われたものである。また、1999 年 9 月、および 2001 年 10 月にはデジタルシティをテーマとした国際ワークショップ[6, 7]が京都で開催されるなど、研究コミュニティも徐々に広がりを見せている。第 2 章では、このデジタルシティに関連する活動の詳細と、デジタルシティに求められる情報検索機能等についても述べる。

第 3 章では、地理的関連性を用いた Web 情報空間の拡張について述べる。インターネットの普及にともない、ネットワーク上には都市に関する情報が大量に発信されてきているが、それらを効率よく扱うためには検索機能が重要となる。しかし、現在のインターネット上で提供されているキーワード検索やディレクトリサービスだけでは、都市部に集中する実世界のオブジェクトに関する情報を十分に活用することはできないと考える。この問題を解決するために、本研究では都市における情報間の地理的な関連性に注目し、その関連性をインターネット上の情報空間に反映させる。この地理的な関連性を反映させた情報空間を本研究では、拡張 Web 空間と呼ぶ。従来のインターネット上の情報検索では、キーワード等を用いて Web ページを点で検索することができる。これに加え拡張 Web 空間では、「近くにある」「隣接している」「ハイパーリンクしている」など、Web ページ間の地理的、論理的なリンクに関する条件に基づく検索を可能にする。第 3 章では、この拡張 Web 空間における問い合わせを SELECT-FROM-WHERE 形式で記述可能な拡張 Web 空間のための検索言語を提案するとともに、複数の例を用いて検索言語の記述性を確認する。

一方、拡張 Web 空間における検索は、地理的な関連性で拡張された Web 情報空間で行われることから、半構造データの検索として捉えることができる。半構造データを対象とした検索手法は、情報検索およびデータベースシステムに関連して数多く提案されてきているが、検索システムの効率

化を図るためには、対象とする検索空間に適した手法を検討する必要がある。そこで本研究では、拡張Web空間に基づく情報検索システムを試作し、実際の京都に関する Web 情報を利用して複数の最適化手法の比較実験を行い、拡張Web空間における問い合わせ処理に適した最適化手法を明らかにする。比較する最適化手法は、動的計画法を用いて完全な条件処理順序を作成する事前最適化、同じく動的計画法による部分的な条件処理順序の作成とその実行を繰り返す制限付き事前最適化、最適な条件の選択とその処理を繰り返す実行時最適化と最適化なしの 4 手法とし、検索処理の実行時間と検索処理の途中で発生する予測の誤差の比較などから、問い合わせ処理の最適化におけるポイントを明らかにする。この拡張Web空間に基づく情報検索システムの試作と、その問い合わせ処理の効率化については、第 3 章の後半で説明する。

第 4 章では、地域情報検索のための地図インタフェースについて述べる。現在、インターネット上には、無料で地図を提供するサービスが複数存在しており、その多くは地図に付加価値を与えるために、様々な情報を添付して配信している。地図に情報を添付する手法は、情報の一覧性が高く直感的でわかりやすいものの、ただ情報を地図に張り付けるだけでは情報同士が重なってしまい、都市部に集中する情報を検索する上での根本的な改善とはならない。そこで本研究では、地域情報検索のための地図を利用した Web ページ検索インタフェースを構築する。この地図インタフェースには、近傍検索、カテゴリ検索、およびキーワード検索を実装し、検索結果を地図上に重ね合わせて表示することによって、実世界に即した情報利用を可能にする。また、クライアントにキャッシュしたデータを検索に用いることによって、インタラクティブな検索処理とサーバ負荷の軽減を図る。また、この構築した地図インタフェースは、デジタルシティ京都の実験サイトにて公開運用を行った。この際に Web サーバのアクセスログに記録された検索履歴の分析から、ユーザによる検索機能の利用傾向と、情報が集中した地域における検索行動を明らかにする。第 4 章では、この地図インタフェースの構成、検索機能の詳細および動作性能に関する評価と、アクセスログに基づく利用動向の分析について述べる。

第 5 章では，本研究で提案する拡張Web空間と地図インタフェースに関する考察を行う．特に，デジタルシティ京都におけるシステムの公開運用を通して得られた知見に基づき，新たに構築したシステム上でのプライバシーや著作権といった社会的問題をどのように捉えるべきかを議論し，今後のインターネット上の情報サービスにおける指針を示す．

最後に第 6 章で，本研究の結論を述べる．

第 2 章

インターネット上にみられる地域性

2.1 都市とインターネット

近年のインターネットの普及とユーザ層の広がりにより，ネットワーク上に大量の情報が発信されるようになり，普段生活する身近な街の情報やこれから訪れる場所の情報など，人々の行動に直結する情報がインターネットから入手できるようになりつつある．NTT情報流通プラットフォーム研究所で行われている「モバイルインフォサーチ実験」の「このサーチ」に関連して行われた調査[1]によると，インターネット上の全 WWW 文書中の約 28% が，住所，ランドマーク，駅などの位置関連情報を含んでいるという．また，同プロジェクトで運用している位置指向の検索システムから検索可能な WWW 文書を対応する座標にプロットすると，図 2.1 のような分布が得られる．この図から，WWW 文書の分布と日本列島の形状が良く対応していることがわかるが，一方で，地理的に均等に分散しているわけではなく，WWW 文書が都市部に集中する傾向も読み取ることができる．

このような都市部への情報の集中は，必要な情報を短時間で探し出すことを困難にする．たとえば，街角にある店舗の Web ページを検索する場合，曖昧なキーワードで検索を実行すると，検索目標とは無関係な検索結果が大量に出てきてしまう．本当に必要な情報は，その大量の検索結果の中から再度選びださなくてはならない．この他にも，インターネット上のリンク集やディレクトリサービスなど，第三者が編集したリストから情報を探す方法なども考えられるが，いずれにせよ，ある程度の事前知識をあらか

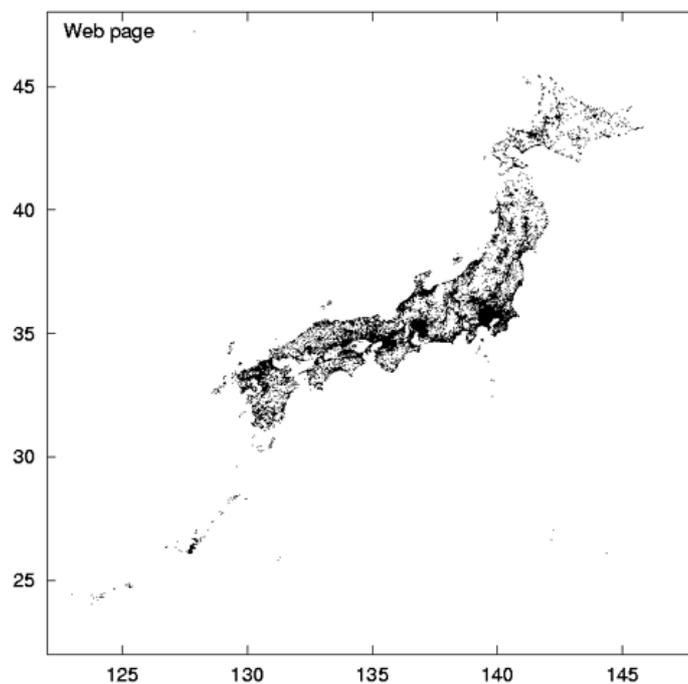


図 2.1: WWW文書の地理的分散 (参考文献[1]より)

じめ持っていないと、スムーズな情報検索ができないのが現状であろう。

2.1.1 都市空間と情報検索

事前知識が必要となる都市情報の検索を改善する手法の一つに、都市のメタファを用いたユーザインタフェースの構築が考えられる。たとえば、検索条件を指定する際に都市の電子地図や街並を再現した仮想空間を利用すると、実際の都市における地理的な場所や位置関係を条件として指定しやすくなる。また、検索結果の表示に地図や仮想空間を利用すると、都市空間のオブジェクトに検索結果をリンクさせたり、検索結果の分布を都市空間における情報の分布として捉えることが可能となる。このような地図や仮想空間を利用した情報サービスは、インターネット上の情報と実際の

表 2.1: インターネット上の地図を用いた情報サービス

名称	説明
マピオン	広告情報を登録した地図配信サービス
マップファンウェブ	携帯端末にも対応した地図配信サービス
MapQuest.com	米国の地図サービス，運転経路検索も可能

都市空間を結び付けるものであり，デジタルシティにおいて実現が期待される機能の一つと考えられる．

地図を利用した情報サービスは，これまでも数多く提案されてきており，ネットワークを利用して都市の情報を流通させるツールとして利用可能となっている．現在，インターネット経由で利用可能な地図を利用した情報サービスとしては，インターネットを新たなメディアと捉えて広告情報を登録した地図の配信を行うマピオン[8]や，地図を利用した情報配信サービスで携帯端末にも対応したマップファンウェブ[9]がある．米国においても，都市の詳細地図の検索表示や運転経路情報などの検索が可能な地図サービス MapQuest.com[10]が展開されている．これらのサービスでは，地図上に表示されたアイコンから関連情報の閲覧が可能となっているが，今後は，地図情報と WWW を融合した情報検索が，インターネットにおいて重要な役割を持つようになると考えている．

また，都市空間を仮想的な三次元空間として再現し，その空間を利用して各種サービスを行う試みも各地で行われている．地図による情報提供と比較し，より詳細に都市空間を再現することが可能であるが，仮想空間の構築コストとデータ量が大きいことから，限定的な地域を対象とした試みが多い．都市の仮想空間の構築例としては，ヘルシンキ[11, 12]，パリ[13]，上海[14]，京都[15]などがある（表 2.2）．これらの都市を仮想空間中に再現する試みの詳細については，デジタルシティに関する特集記事[5]を参照されたい．

三次元仮想空間の構築は，土木，建築，測量やバーチャルリアリティに関する研究から，新しい手法が提案されている分野である．ただこれまでには，仮想空間を構築するための手段や利用環境が整備されていなかったた

表 2.2: 3D仮想空間の構築例

名称	説明
ヘルシンキ・アリーナ 2000	都市の高速回線を利用した コンテンツサービスを指向した ヘルシンキ市の仮想空間
ル・ドゥージュエム・モンド	自分のアバターでバーチャルパリア 訪問可能
インフォポート上海	上海の地域情報ハイウェイの整備を 目的としたプロジェクトで 3D都市モデルを試作
デジタルシティ京都	3DMLを利用した三次元仮想空間で 京都・四条通や二条城を再現

め、個々の試みに採用される手法がまちまちで、データベースの相互互換性も乏しかった。しかし、今後、都市空間に関する三次元情報も地図のように標準化され、現在の地図や WWW と結び付けることが可能になれば、より豊かな情報サービスを提供するための情報空間として活用できるようになると期待される。

2.1.2 都市空間と情報空間の融合

インターネット経由でアクセスできる都市の情報は、ユーザによって作成された静的な Web ページだけとは限らない。センサやデータベースと連動し、リアルタイムで都市の状況を伝える Web ページも増加している。たとえば、天候や大気の状態など、各種センサによって検知される都市の状況や、街の各所に設置されたカメラの映像は、リアルタイムで電子化され、インターネット経由で利用できるようになってきている。これまでは、これらのセンサがそれぞれの目的ごとに設置され、異なる機関によって管理されていたため、運用方針やセキュリティ上の問題から、各種情報を統合して利用することは困難であった。今後は、こうした情報が高度道路情報システム (ITS) などで統合的に利用できるようになると予想される。デ

デジタルシティにおいても，こうした都市の状況がリアルタイムで取り込まれ，インターネット上の情報も含めて，時間的，空間的粒度の異なる大量の情報が統合的に利用できるようになることにより，様々な情報サービスが可能になると考えられる．

また，インターネットの普及にともない，既存のマスメディアもインターネットを新たな媒体として捉え，各種コンテンツの配信をはじめている．大手新聞社や通信社には，全国規模の情報を配信する Web サイトと都道府県単位で情報を再編集した Web サイトを並行して運営するところもある．また，旅行雑誌や地域情報紙などの出版社も，紙面と連動した情報を掲載した Web サイトを運営している例が多い．これらの Web サイトでは，既存の情報収集力と編集技術の高さにより，質の高いコンテンツの配信を実現している．また，多くのコンテンツが各種ポータルサイトへと配信され，コンテンツの二次利用も行われるようになってきている．

様々な発信源からの情報の混在が，地域性を持った情報空間のおもしろさであり，情報統合や情報検索における難しさとなるが，実際にインターネットを通して多くのユーザから利用され，そのフィードバックによって情報空間自体が発展していくものとする．

2.1.3 コミュニケーションの場

デジタルシティや地域ポータルサイトには，現実空間を仮想空間へと拡張した都市としての役割だけでなく，市民のための新しいコミュニケーションの場としての役割が考えられる．豊かな情報空間に人々が集い，コミュニティが形成され，コミュニケーションが行われることによって，情報空間自体も活性化されていくと期待される．

インターネットが普及する以前も，パソコン通信では電子掲示板（BBS）を利用したコミュニケーションが行われ，様々なコミュニティが形成されていたが，インターネットの普及により，通信費用と距離の関係が無くなり，人々は自由に世界中のコミュニティへとアクセスできるようになった．その一方で，インターネット上に存在するコミュニティの選択肢の多さとネットワーク経由で参加するという気軽さから，人々のコミュニティへの帰属意識は希薄になってきている．また，サービスやコストの面でのコミュ

ニティの場を提供する側の競争も激しい。こうしたコミュニティの興亡に関しては、アムステルダムデジタルシティ（DDS）が、顕著な例としてあげられるだろう。

bit の記事[16]によると DDS は、1995年、市民に公共の場を電子的に提供することを目的に設立され、地方選挙の際に市民と政治家が議論をするための場として活用された。この後、市民のためのフリーのメールサービスや Webサイト構築のためのディスクスペースの提供、革新的なユーザインタフェースの導入により、DDS内のWebサイト数を示す「家」の数は1997年には 6,500に達した。しかし 1997年以降、インターネットに登場した hotmailなどのフリーメールや geocities などのフリー Webスペースサービスとの競争が発生し、非営利企業として活動していた DDS は、資金の確保や技術革新が困難となった。その後、「家」の数や DDSへのアクセス数が減少していくなか、DDS は営利企業へと転換し、現在はインターネットへのアクセス手段の提供するプロバイダへと転進している。

この記事の中で、コミュニティの成否はコミュニケーションツールの完成度とサイトの運営力によって大きく左右されると指摘されている。またコミュニティを長期間継続させるためには、単にフリーメールやフリーの Webスペースを提供するだけでなく、継続的な技術革新と細かなサポートを提供する必要があると指摘されている。一方、企業が自社製品への反応をテストするツールとして、マーケティングに直結したコミュニティの戦略的運営などが行われるようになるなど、インターネット上のコミュニティ自体も変化しつつある。従って、地域と密着したコミュニティをインターネット上に形成し維持することは、非常に困難な問題になると予想される。

2.1.4 利用動向の分析

デジタルシティにおける情報検索システムやユーザインタフェース、そしてコミュニケーションツールに関わる技術は、実際にインターネットを通じて多くのユーザから利用され、そのフィードバックによって発展していくものとする。このためには、ユーザがシステムをどのように利用するのか、様々な角度から分析する必要がある。

インターネット上の WWWに関わるユーザの行動は、Webサーバやブ

ロキシのアクセスログとして記録され、その分析結果はジップの法則[17]や冪乗則と関連づけて論じられることが多い。Huberman ら[18]は、ユーザが Web ページに埋め込まれたハイパーリンクをたどりながら、新たな Web ページを閲覧していくものと仮定し、様々なモデルによる説明を試みている。また、WWW のグラフ構造の大規模な分析[19]も行われるなど、グローバルな視点から見た WWW の構造も明らかになりつつある。

こうしたマクロ的でかつ匿名性の高い分析に対し、コンテンツレベルの利用動向分析や Web サイトの滞在時間など、ユーザレベルの利用動向の分析も行われている。商業系の検索サービスサイトの中には、自社が提供するキーワード検索で利用された語句のランキング情報を公開する Lycos[20] や、公開中のディレクトリの利用動向に関するアンケート結果を公表する Yahoo! JAPAN[21] などがある。また、インターネットの普及状況や利用状況、およびインターネットユーザのプロフィールをレポートする専門の調査会社[2]も立ち上がりつつある。

ただこれまでのところ、都市や地域といった視点を分析に取り入れて、ユーザの行動を説明した例はほとんどない。従って、デジタルシティを含め、地域に特化した情報システムは、実際に運用し、その結果を分析することによって、その有効性など、明らかになる面が大きいと考えられる。

2.2 デジタルシティの構築

インターネットの普及にともない、ネットワーク上の情報を利用した様々なサービスが開発される中、デジタルシティや地域ポータルサイトなど、地域性をもった情報空間が注目を集めている。地域性をもった情報空間が、インターネット上においても注目されるのは、世界中の情報にアクセスできるようになった現在も、人々の行動が依然ローカルなものであるからであろう。

デジタルシティに関する動向は、デジタルシティの現状[3, 4]にて詳説され、雑誌の特集においても各国の動向[5]と技術動向[22]が紹介されている。また、デジタルシティの目的の一つには、世界規模のグローバルな情報空間を作り出しているインターネットを利用して、日常生活に役立つ豊かな情報空間を実現すること[23]が考えられる。

デジタルシティ，もしくはそれに類する名前を持つインターネット上の情報サイトは，1994年頃から，アムステルダム，ヘルシンキ，米国各都市など，世界中の各地で構築されてきている．主なデジタルシティは，インターネット上に公開され，市民に対して様々な情報サービスを提供しているが，都市により，提供しているサービス，運営形態やそのポリシーは異なる．

- アムステルダム[24]では，ネットワークは公共の場であるというポリシーのもと，ユーザに無料でスペースが提供され，独自のインタフェースによって市民の情報コロニーが形成された．1994年に非営利団体としてスタートし，2000年初頭からは営利団体として活動している．
- ヘルシンキ[11]では，市民が利用できるマルチメディア高速ネットワークを構築し，その上でヘルシンキ市の詳細な三次元モデルを利用した街の仮想空間と各種サービスの展開を目指している．構築された仮想空間は，ドライビングシミュレータなどにも応用されている．
- 米国各都市のデジタルシティ[25]は，マスコミ的な性格を持ち，統一されたインタフェースを利用して都市の最新情報を発信している．

日本国内においても，デジタルシティや地域ポータルサイトがインターネット上に構築されており，たとえば，関西地方のように広域を対象としたサイトから，商店街など狭い生活圏を単位としたサイトまで，その形は様々である．またその運営母体も，新聞，放送，出版などのマスコミ系，Webページ制作企業やプロバイダなどによる情報通信系，そして市町村などによる地方公共団体系など，非常に幅広い（表 2.3）．

2.2.1 京都における試み

こうした中，京都においても，日常生活に役立つ新たな都市の社会情報基盤の構築を目指して，デジタルシティ京都のプロトタイプの実験が始まっている[26]．デジタルシティ京都のプロトタイプ作成に関わる実験は，1998年10月，京都大学とNTTコミュニケーション科学研究所（現在のNTTコミュニケーション科学基礎研究所）の共同プロジェクトとしてオープン

表 2.3: 地域情報を扱う Webサイトと運営母体

運営母体	例
マスコミ系	Walkerplus.com, asahi.com など
情報通信系	Town@nifty, まちgoo など
地方公共団体系	コアラ(大分), 電腦山田村塾 など

ラボ形態で開始された。このプロジェクトでは、物理的な都市（フィジカルシティ）にインターネット上の情報を対応づけて再構成した公共的な情報空間を市民に提供することをデジタルシティの役割とし、市民の日常生活や地域コミュニティの支援、国際化、異文化コミュニケーションの促進を実現することを目標とした。本論文の第3章、第4章で述べる成果は、このデジタルシティ京都のプロトタイプ作成に関わる中で進められた研究に基づいている。

このオープンラボは2000年3月まで活動し、その期間中に京都のWebページを地図で検索できるようにした地図インタフェース（ジオリンク京都）、京都の四条通を三次元仮想空間で再現した3D京都、二条城バスツアー等を作成し、デジタルシティ京都実験フォーラムのWebサイトにおいて公開運用した。なお、地図インタフェースについては、第4章にてその構成や利用状況等を説明するので、本節では残りの3D京都と二条城バスツアーについて説明する。

3D京都は、3DMLを利用して構築した三次元ウォークスルーである[5, 15]（図2.2）。3DMLは、Flatland社が作成した三次元空間記述言語であり、ビットマップ上に立方体や三角柱などのブロックを配置することで、仮想空間を構築することができる。この仮想空間は、単純なモデルとテキストチャにより仮想空間が構築されるので、Webブラウザ上のプラグインで高速に表示することができる。

この3D京都の三次元ウォークスルーでは、仮想空間中のオブジェクトにリンクを設定することが可能であり、百貨店の建物からそのホームページへのリンクや、内部のパノラマ映像へのリンクを設定することができる。また、3D京都の地面のタイルには情報検索システムへのリンクが設定され

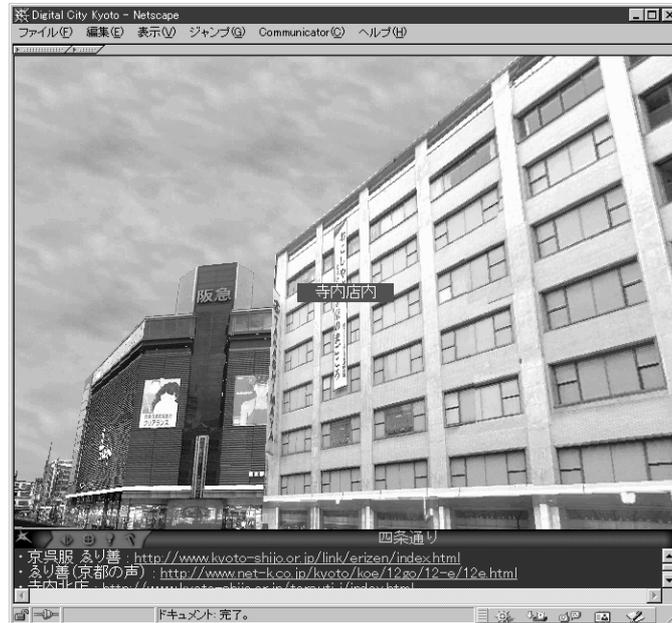


図 2.2: 3D京都

ており、タイル上をユーザが通過すると、ブラウザ下部のフレームにユーザの現在位置周辺 20mにある Webページへのリンクが表示される。この現在位置周辺の Webページ検索に関する詳細は、第 3 章で述べる。

また、3DMLで記述した三次元空間は、単純なモデルとテクスチャで再現した三次元仮想空間となるが、デジカメ等で撮影した実際の画像をテクスチャとして利用することにより、ある程度の現実感を再現することができる。3D京都の例では、四条烏丸から八坂神社下までの約 1.6km で約 400 枚の画像を市販のデジタルカメラで撮影し、画像処理ソフトで色彩や彩度の補正と変形を行い、ウォークスルー中の建物のテクスチャとして利用した。しかし、高い現実感を再現するためには高解像度の画像をテクスチャとして利用する必要があり、ウォークスルー中で高解像度の画像を利用する際には、画像ファイルの大きさや転送時間の問題を考慮しなくてはならない。3D京都の例では、ウォークスルー中のオブジェクトの配置を記述し

た 3DMLファイルが約 22KB であるのに対し、テクスチャとして利用する画像ファイルの方は、店舗の商品が判別できる程度の高解像度テクスチャだと、対象区間全体で約 16MB となっている。このテクスチャファイルの大きさは、混雑したネットワークからアクセスする場合には、転送に困難を伴う。従って、このようなコンテンツを気軽に利用できるようになるためには、ユーザ側アクセス回線の高速化が必要となる。また、複数ユーザが同時に利用するネットワーク環境では、コンテンツ配信サーバの高速化も必要となる。

一方、二条城バスツアー[27]は、エージェントによる観光案内とグループツアーを統合した実験システムである(図 2.3)。このツアーシステムは、3DMLによるウォークスルーとチャットシステム、ガイドエージェントから構成される。3DMLによるウォークスルーでは、狩野派の手による襖絵など、二条城二の丸御殿内部の画像を利用し、二条城内を仮想的に再現している。このウォークスルーにチャットシステムを組み合わせることで、仮想空間中のユーザが互いに会話できるようにしている。また、ガイドエージェントは、ユーザ間の会話の状況とユーザの現在位置がサーバ上でモニタリングしながら、あらかじめ登録されたルールとユーザの状況を照らし合わせて、建物の説明や移動先の案内を行う。なお、この二条城バスツアーのガイドエージェントには、マイクロソフトエージェントを利用している。

また、1999年 7月から 2001年 9月にかけては、デジタルシティ京都実験フォーラムが活動し、京都を実験の場として、将来のデジタルシティに必要な研究と技術開発に関する活発な議論が行われた。この期間中、京都を題材としたアニメーションを配した Webサイトのトップページ(図 2.4)の下、実験フォーラム参加機関が提供するコンテンツへのリンクリスト(図 2.5)が提供された。また実験フォーラムと並行して、NTTコミュニケーション科学基礎研究所においても、実世界指向情報統合に向けた研究プロジェクト[28]が開始されている。学術的な活動としては、1999年 9月および 2001年 10月に、デジタルシティに関する国際ワークショップ[6, 7]が京都で開催された。このワークショップでは、都市情報の検索やユーザインタフェースの構築などに関する技術的な発表から、コミュニティ形成手法等



図 2.3: 二条城バスツアー

や社会学的な実験に関する研究発表まで、幅広い議論が行われた。また、国内の学会においても、デジタルシティに関わる議論を行う場として、2001年4月に情報処理学会知的都市基盤研究グループが発足するなど、研究コミュニティも確立されつつある。

2.3 まとめ

本章では、都市とインターネットの関係と、デジタルシティや地域ポータルサイトに関する活動について説明した。

2.1.2 節で述べたように、インターネット上の情報は、個人から発信される情報だけでなく、マスコミからの情報提供や、街に設置されたセンサからの情報など、その情報源は多様化しつつあり、その量も増加している。また、地図や3Dウォークスルー等、都市のメタファを利用したユーザイ



図 2.4: デジタルシティ京都実験フォーラムのトップページ

ンタフェースがインターネット上に提供されてきてはいるが、膨大な情報をユーザが効率よく利用できるようにするためには、検索システム内部の動作を規定するモデルの構築や、インタフェースの改善が重要になると考えている。また、2.1.4 節で述べたように、実際に作成したシステムの有効性を検証するためには、実際の環境における運用とその結果の分析が重要と考えている。

デジタルシティが実際の都市において機能するためには、インフラやシステムを作るだけでなく、都市に関する大量のコンテンツと数多くの市民の参加が必要となる。また、数多くの市民の参加を募るためには、充実した良質のコンテンツが安定したシステムから安全に提供され、自由に組合せて様々な目的で利用できるようにすることも不可欠と考えている。

こうした背景に基づき、本研究では、都市に関する情報を記述するためのデータモデルやそのユーザインタフェースなど、デジタルシティのシステ

第 3 章

地理的関連性を用いた Web空間の拡張

携帯電話が高度化し、都市生活におけるインターネット利用が容易になりつつある。携帯電話が情報機器化し、対話機能を持つインタフェースエージェントが搭載される日も遠くないだろう。一方で、地理情報システム（GIS）や WWWなどを用いた都市情報の集積が進んでいる[29]。これらを統合すれば、都市生活を支援する高度な情報サービスを構築できる。以下のシナリオを考えてみよう。

「今日は土曜日で、太郎は花子と夕食の約束をしている。しかし、その前に太郎は大学で資料を完成させ、郵便局から発送したいと思っている。太郎は携帯電話の中のエージェントに問い合わせる。『大学の近くにあって、土曜日も営業している郵便局を探してくれないか？』。エージェントが探している間に、太郎は花子と連絡をとる。花子は待ちきれなくて近くまで来ると言っている。太郎はエージェントに話しかける。『レストラン情報サイトで紹介されているようなイタリア料理のおすすめ店のなかで、ワインがしっかりしていて、今探している郵便局から1km以内にある店を探してくれないか？』」

WWWには数多くの Webページが存在し、Webページ内部に記述されたハイパーリンクによって複雑に連結しあっている。また GISには、都市のオブジェクトの位置関係が格納されている。WWWと GISは独立したものであり、それぞれを対象とした検索技術は数多く提案されている。それ

ら既存の検索技術を組み合わせることができれば，WWWの Webページをそれ自身の論理的なハイパーリンクだけでなく，GISを利用して実際の都市に即した Webページ間の位置関係で関連をたどるような情報検索が可能となり，上記のシナリオの問い合わせが実現できると考えられる．

たとえば，『大学の近くにあつて土曜日も営業している郵便局』を探すには，まず大学のサイトから住所を抽出する．その大学の住所をキーにして，その近くにある郵便局を範囲を設定して GISで検索する．さらに，GISで検索した郵便局を WWWで検索し，そのWebページの中から営業時間に関する情報を抽出し，土曜に営業しているかどうかを判断する．『レストラン情報サイトで紹介されているイタリア料理のおすすめ店で，ワインがしっかりしていて，郵便局から 1km以内にある店』を探すには，レストラン情報サイト内のイタリア料理のおすすめ店のリストから各店舗の Webページを検索し，ワインに関して言及のある店舗を選択する．そして選択した店舗の Webページから住所を抽出し，その住所をキーにしてGISを検索し，郵便局からの距離が 1km以内のレストランを求めることができる．

そこで，上記のような実際の都市に即した情報検索を実現するために，WWWをGISによって拡張した拡張Web空間を提案する．拡張Web空間は，Webページ間に定義されているハイパーリンクによって構成される本来のWeb情報空間を，利用者からの問い合わせの中に出現する Webページ間の地理的関連性によって，動的に拡張可能としたものである．この地理的関連性は，Webページの持つ地理的属性を GISで評価して生成するWebページ間のリンクであり，これを地理的ジェネリックリンクと呼ぶ．また，拡張Web空間の検索では，検索空間を半構造データ[30, 31]としてとらえ，それぞれのWebページに対する条件と，Webページ間のハイパーリンクおよび地理的ジェネリックリンクに対する条件を，繰り返し表現を可能にした検索パス式によって系列化して問い合わせを表現する．

本章では，3.1 節で，WWWに関連する既存のデータモデルと都市情報との関係について説明し，3.2 節で，拡張Web空間のアイデアと，そのキーとなる地理的ジェネリックリンクと検索パス式について説明する．次に 3.3 節で，拡張Web空間における問い合わせを表現するための検索言語を定義し，3.4 節で，その検索言語を利用した問い合わせの記述例を示す．この

3.4 節では、デジタルシティ京都で公開した 3D京都との連携についても説明する。そして 3.5 節で、拡張Web空間に基づく情報検索システムの構成と適用した最適化手法の比較について説明し、3.6 節で本章のまとめを述べる。

3.1 都市情報とデータモデル

インターネット上に構築されるデジタルシティでは、WWWと様々な情報を統合することによって、都市で生活するユーザに役立つ情報を提供できると考える。特に、都市に関する情報を考えた場合、ハイパーリンクで接続された WWWの情報空間が、地理的な関連性により拡張されることにより、より有効な情報検索が可能になると考えられる。

そこで本節では、WWWの捉え方の基本となる半構造データについてと、WWWを対象とした検索言語について述べる。そして、都市に関するWebページと実際の都市空間を統合する上で必要な技術について述べ、現在進められている XMLによる情報空間の集約とその現状について分析する。

3.1.1 半構造データ

WWWは、タグとリンクにより構造化されていると考えることもできるが、データモデルとしては構造が不規則なデータの集まりであり、はっきりしたスキーマが定義できないため、半構造データとして考えた方が適切と考えられている[30, 31]。

半構造データは、データベース研究のコミュニティにおいて、1997年頃からそのデータモデルやデータ操作言語に関して、盛んに研究が行われている。田島のサーベイ論文[31]によると、半構造データに関する研究の目的の中心は、従来のデータベースの枠組みのもとで厳格にモデル化され格納されているデータと同様に、半構造データも set-orientedでかつ宣言的な問い合わせや操作を記述し、効率的に処理したいという点にあるとされる。

デジタルシティで提供される情報が、インターネット上の Webページと様々な情報を関連づけたものとする、その情報空間は半構造データとして

扱うことが適当であろう。また，その情報空間に対する問い合わせには，多様な問い合わせ表現や問い合わせ処理の最適化など，データベースにおける手法の拡張が必要になると考えられる。

3.1.2 Web検索言語

Webサイトの検索システムに対して，クライアントのユーザインタフェースから問い合わせを行う場合，何らかの形で問い合わせを表現しなくてはならない。この問い合わせの表現には，WWWを対象とした検索言語が応用できる。

WWWを対象とした検索言語は，データベース研究のコミュニティから提案されたものであり，1995年頃から始まった比較的新しい研究分野でもある。WWWを対象とした検索言語は，リンク構造とコンテンツに関する条件の指定により検索を行う第一世代と，Webページの構造に関する条件指定や検索結果の再構成が可能な第二世代に分類することができる[32]。第一世代の代表的な検索言語には，WWWを広域に展開したハイパーテキストとしてとらえたモデルを基本にした W3QL[33, 34]や，ノードやリンクに属性を追加したグラフ構造とリレーショナルモデルを利用した WebSQL[35, 36]がある。また第二世代の代表的な言語には，データ構造として hypertree を用いる WebOQL[37]や，Webサイト管理システムのために作成された StruQL[38]がある。

このうち，W3QLでは，構造の分からない Web空間から必要な情報を探し出すために，セルフループを含む検索経路を指定することができる。必要な情報を構造の分からない Web空間から検索するためには，Webページに記述された情報に対する条件や Webページ間のハイパーリンクに対する条件を組み合わせた検索経路の指定が重要となると考える。

3.1.3 地理情報システムとWWW

地理的な条件に基づき Webページを検索できるようにするためには，Webページから地理的属性を抽出して地理オブジェクトと関連づける必要がある。この Webページと地理オブジェクトの間の対応づけは，アドレスマッチングと呼ばれ，相良らが地名管理木を用いた自然言語処理による半

構造化処理を組み合わせた手法を提案している[39]。また横路らは、位置に関連した情報を含む Web ページの収集手法と、形態素解析を利用した住所を抽出する手法を構築し、「モバイルインフォサーチ実験」の中の「このサーチ」として試験サービスを行っている[1]。この 2 手法では、事前準備として、自然言語処理で利用する辞書に住所表記の追加を行うことで、その精度の向上を図っている。一方海外では、2002年に Google の主催によって行われたプログラミングコンテスト[40]で、Daniel Egnor氏による「地域内の Web ページを発見する技術」が優勝となった。この技術では、Web ページ内に記述されている地理的な情報を、米国が国勢調査のために公開している無料の地理データベースと照らし合わせることで、「自宅の近所の本屋を全て探したい」といった特定の地域の特定トピックに関連する Web ページの検索を可能にしている。以上のように、Web ページと地理オブジェクトの対応付けを行うためには、詳細な地理情報が必要になる。ただ、現在の Web ページの自由な記述に基づいてアドレスマッチングを行うと、記載の不備や表現の不一致により、Web ページと地理オブジェクトの対応が常にとれるとは限らないことを考慮しなくてはならない。

また、地理オブジェクト間の地理的な関係演算に関しては、Open GIS Consortium[41]において標準化が行われている。この標準化では、GISにおける二次元もしくは三次元の座標系を利用したオブジェクトモデルやオブジェクト間の関係演算、データベース操作のための SQL の拡張が提案されている。地理オブジェクトは、位置、形状、住所、所有者といったオブジェクトの属性に応じて構造化することが可能である。また、地理的關係演算を利用すると、距離や相対位置関係、包含、接続、交差など、オブジェクト間の地理的な関係を評価することができる。

3.1.4 XMLによる情報空間の集約

XMLは、Web技術に関する標準化団体である W3C (World Wide Web Consortium)で開発されている汎用的なマークアップ言語であり、人とアプリケーションの両方がその目的にそって理解し、処理できる形式でデータを記述できるようにすることを目的としている[42, 43]。これまでの WWW は、HTML (Hyper Text Markup Language) によって記述された静的な

Webページがほとんどであったが、今後 XMLの導入により、各種データベースに基づきアプリケーションによって自動生成される Webページも増加していくと予想される。そこで本節では、XMLによって可能となるインターネット上の情報空間の拡張のうち、リンク機能の拡張と地理情報の記述について具体的に述べてみたい。

文書間のリンク機能の拡張に関しては、XLink[44]として、W3Cから勧告が出されている。HTMLのリンクが、A (アンカー)タグにより、リンク先の URLもしくはタグのname属性を指定するだけの単純なものであるのに対し、XLinkでは HTML同様の単純リンクに加え、以下のような拡張リンクの指定が可能になる。

- リンク用のタグ以外の部分へのリンク
- 双方向リンク
- 複数個所へのリンク
- コンテンツとリンク指定の分離

このうち、コンテンツとリンク指定の分離を利用すると、Webページとは独立に、リンク情報だけを定義した文書を作成することが可能になる。たとえば、都市に関連する Webページであれば、Webページ間の地理的な関連性に基づいたリンクを Webページとは独立に定義し、蓄積できるようになる。ただし XLinkは記述法であるため、実際のリンク生成には別途システムが必要となる。

また、地理情報の記述言語に関しては、OpenGISによる GML[41]や国内の G-XML[45]など、現在 XMLに準拠した形での標準化活動が活発に進められている。これらの標準化は、GISデータベースの標準化だけでなく、将来的には Web情報と地理情報の融合につながるものと考えられる。実際、WebGISなどの普及により、地理情報の記述が WWWに広がると、Webページ中に地理情報が埋め込まれるようになり、3.1.3 節で述べたアドレスマッチング関連の処理は大幅に簡略化できるかもしれない。ただこの場合も、情報提供者側の努力によるところが大きいので、URLのように

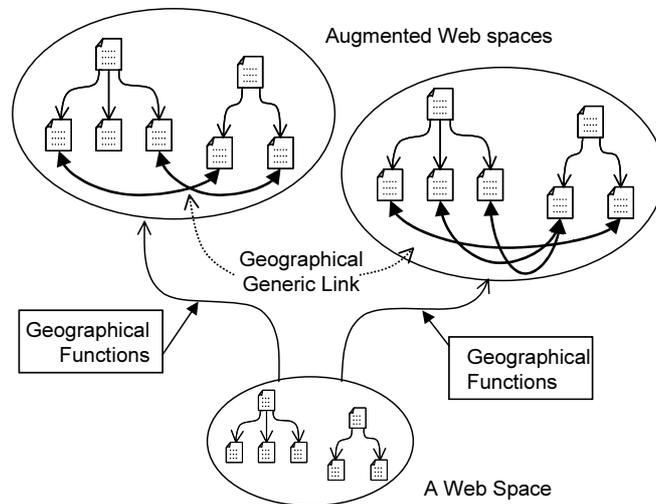


図 3.1: 拡張Web空間の位置付け

統一された表現とはならず，相互流通のための表現や語彙等の統一の課題は残ると考えている．

3.2 拡張Web空間

3.2.1 基本的なアイデア

本章で提案する拡張Web空間は，Webページ中に静的に指定されているハイパーリンクによって構成される半構造データの Web空間を，地理的ジェネリックリンクで動的に拡張したものである（図 3.1）．実際の都市でWebページ間に地理的な関係があったとしても，ハイパーリンクで構成されたWeb空間にはその関係が反映されているとは限らない．そこで，Webページ間の地理的な関係を表す地理的ジェネリックリンクを動的に生成することによって，実際の都市に即した形にWeb空間を拡張する．

拡張Web空間における情報検索は，W3QL[33, 34]および WebSQL[35, 36]に代表される WWWを半構造データとして扱う検索言語に，実際の都市に

おけるオブジェクト間の地理的関連性を評価するための地理的關係演算を導入して実現する。Webページに対する条件と、Webページ間のハイパーリンクおよび地理的ジェネリックリンクに対する条件は、繰り返し表現を含んだ検索パス式で系列化する。繰り返し表現を導入することで、半構造データである拡張Web空間に適応した検索経路の表現が可能となる。

情報空間の拡張としては、GISを WWWのハイパーリンクで拡張する拡張Web空間とは反対のアプローチも可能ではあるが、拡張した情報空間を実際の都市にあわせて収束させることと、インターネット上に構築するデジタルシティ京都からの利用を考慮し、本研究では WWWを GISで拡張する拡張Web空間のアプローチを採用する。

3.2.2 地理的ジェネリックリンク

拡張Web空間では Webページ間のリンクとして、ハイパーリンクに加えて、地理的ジェネリックリンクを定義する。たとえば、「デパートAの西隣にあるデパートB」、「京都市内にある小学校から 100m以内にあるバス停」といった実際の都市空間に存在するオブジェクト間の地理的関連性に基づき、Webページ間に地理的ジェネリックリンクを生成する。

ハイパーリンク（図 3.2）が、Webページ内のアンカータグに記述されたリンク先 URLに基づく Webページ間の有向リンクであるのに対し、地理的ジェネリックリンク（図 3.3）は、Webページ間に潜在する地理的關係に基づく有向リンクとなる。この地理的關係は、ハイパーリンクのアンカータグのように Webページ内に明示的に記述されているわけではない。具体的な Webページの間に潜在する地理的關係を評価し、地理的ジェネリックリンクを生成する。この生成手順を以下に示すが、もし、いずれかのステップで対応するオブジェクトが存在しない場合には、リンクの生成は失敗となる。

- 第 1 ステップ

アドレスマッチングにより Webページから地理オブジェクトへのマッピングを行う。Webページから住所や店舗名などの地理的属性を抽出し、その属性に対応するオブジェクトを GIS内から検索する。

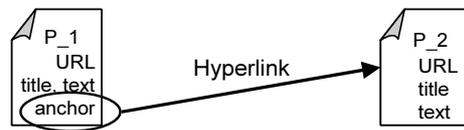


図 3.2: ハイパーリンクの例

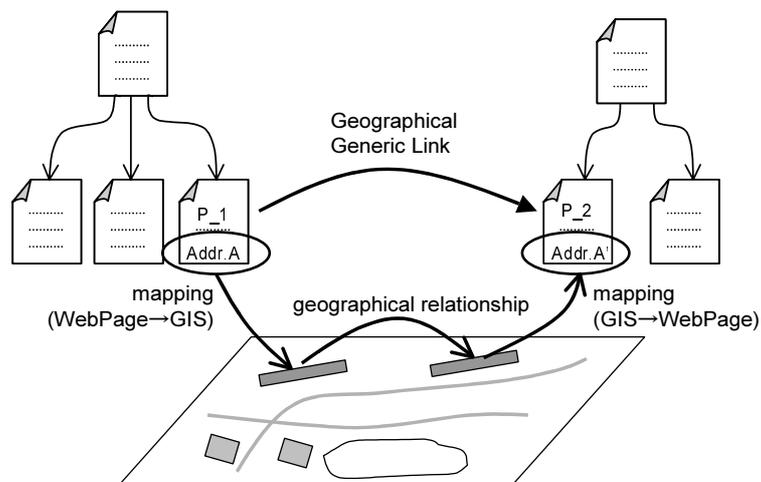


図 3.3: 地理的関連性を用いた地理的ジェネリックリンクの例

- 第 2 ステップ
第 1 ステップで検索した GIS 内のオブジェクトを起点に，条件として指定された距離関係などの地理的関係演算を実施し，指定された条件を満たすオブジェクトを GIS 内で検索する．
- 第 3 ステップ
第 2 ステップで検索した GIS 内のオブジェクトの住所，建物名，所有者といった属性に対応する Web ページを WWW で再度検索して，最終的に Web ページ間に地理的ジェネリックリンクを生成する．

上記のステップの中でリンク元となる Web ページは，単一の Web ペー

ジであっても，Webページの集合であっても構わない．もし，リンク元がWebページの集合である場合には，それぞれの要素に対して，上記の第2ステップ以降の操作を実施する．また，この生成手順から考えると，指定された地理関係にあるGIS内のオブジェクトを検索し，それに対応するWebページを検索しても良いことになるが，途中で生成されるGIS内の組み合わせ数を考えると現実的ではない．

また，地理的ジェネリックリンクでは，リンク元となるWebページに明示的なアンカーは存在しないが，その代わりに，上記の第1ステップのアドレスマッチングによりWebページからマッピングされた地理オブジェクトを仮想的なアンカーとして扱う．従って，もし1つのWebページから n 個の地理オブジェクトにマッピングされた場合には，その n 個の地理オブジェクトを仮想的なアンカーとして扱い，それぞれのオブジェクトから第2ステップの地理的關係演算を実施する．

実際のWWWおよびGISを対象に地理的ジェネリックリンクを生成する場合には，不完全な記述や表現の不一致などにより，第1ステップのWebページからGIS内のオブジェクトへのマッピング，および第3ステップのGIS内のオブジェクトからWebページへのマッピングで，比較対象とする属性がすべて一致するとは限らない．このマッピングの際の問題は，拡張Web空間に基づく情報検索システムの実装に依存するものとするが，本研究の地理的ジェネリックリンクの生成では，比較対象となる属性間の一致条件を表3.1のように順次緩めながら繰り返し検索を行い，WebページとGIS内のオブジェクトの間のマッピングを実施する．

また，第3ステップのWebページの検索では，一致条件を調整した場合も，同じ詳細度の条件で複数のWebページが検索される場合が考えられる．この際，検索されたすべてのWebページを地理的ジェネリックリンクのリンク先とすることも考えられるが，本研究では，一致した属性をその詳細度に応じて得点化し，合計得点の大小でマッピング先となるWebページを選択することにした．たとえば，表3.1の5段目の条件で比較を行い，「市町村名，区名」が一致した組み合わせと「区名，組織名」が一致した組み合わせがあった場合には，一致した属性の詳細度から後者を選択する．

この他，Webページと対応付けるGIS内のオブジェクトの種類，GIS内

表 3.1: 属性比較の際の一致条件の緩和

条件	比較属性
(厳密)	市町村名, 区名, 町名, 番地, 建物名, 組織名
↑	市町村名, 区名, 町名, 番地, 組織名
	市町村名, 区名, 町名, 組織名
↓	市町村名, 区名, 町名, 番地
(緩和)	市町村名, 区名, 組織名

のオブジェクトと Web ページの所有者の関係, Web ページの内容の正確さや詳細度などに基づくリンク先を選択, Web ページの著作権や利用許諾に基づくリンク生成の禁止などが考えられるが, これらについても, 後述する拡張 Web 空間に基づく情報検索システムに依存するものとする. また, 地理的ジェネリックリンクは, リンク元からリンク先への片方向のリンクとして定義されるが, 指定される地理条件によっては, 逆方向の地理的關係演算を利用してリンク先からリンク元を検索できる場合もある. このような場合には, 問い合わせ処理の段階の最適化で, より処理コストの少ない方向から地理的ジェネリックリンクに関する条件が処理される.

一方, 明示的に記述されたリンク以外を導入した先行研究としては, ハイパーテキストデータベースシステムを対象として, ドキュメントに存在するリンクとは独立に, クエリの組合せによるリンク (query-pair link) に変数を導入した TextLink-III[46], TextLink/Gem[47]がある. query-pair link では, 文書の任意の部分を選択してアンカーとし, それを抽象化したクエリを変数によって関連づけてリンクを生成することができる. この際, リンク元とリンク先の双方がクエリにより動的に計算される.

これに対し地理的ジェネリックリンクは, リンク元となる Web ページからリンク先が動的に計算されるが, リンク元となる Web ページから GIS 内のオブジェクトへのマッピング, GIS 内の検索, そして GIS 内のオブジェクトから Web ページへのマッピングと, 具体的なオブジェクトを追跡してリンクが生成される. この際, リンク元となる Web ページは, 地理的ジェネリックリンクに関する条件とは別に指定された条件に基づき検索される.

また，地理的ジェネリックリンクの場合には，WWWと GISの間で属性の表現が異なる場合にも地理的ジェネリックリンクの生成ができるように，属性間の一致条件を制御しながらオブジェクト間のマッピングを実施する．この属性間の一致条件の制御により，Webページと GIS内のオブジェクトの間のマッピングを制御できる点も，地理的ジェネリックリンクの実装上の特徴となる．

3.2.3 検索パス式

拡張Web空間では，Webページに対する条件と，Webページ間のハイパーリンクおよび地理的ジェネリックリンクに対する条件を，繰り返し表現を含んだ検索パス式で系列化して検索を行う．この拡張Web空間における検索パス式は，半構造データを対象とした検索言語W3QLのパターングラフ (pattern graph) を基本にしたものである．

検索パス式では，起点となる Webページから終点の Webページまでの検索経路を，Webページとリンクの繰り返しによって記述する．たとえば，図 3.2，図 3.3の検索パス式はそれぞれ，“P_1, HyperLink, P_2”，“P_1, GeographicalGenericLink, P_2” と記述する．なお，検索パス式の中では，ハイパーリンク，地理的ジェネリックリンクともに同様に記述できるものとする．

3つ以上の Webページを経由する検索経路は，図 3.4の形を基本に構成する．図 3.4 (a) は図 3.2 の例にさらにリンクと Webページを接続したものであり，その検索パス式は “P_1, Link_1, P_2, Link_2, P_3” と記述する．これに対し図 3.4 (b)は，Web空間の検索で目的の情報の正確な位置が予測できない場合に対応するため，繰り返し表現によってパス中の Webページにそれ自身へのセルフループを設定した例である．検索パス式は “P_1, Link_1, (P_2, Link_2), Link_3, P_3” と記述し，括弧でくくった “(P_2, Link_2)” の部分で繰り返しを表現する．この表現を利用すると，たとえば，内部構造の分からないレストラン情報サイトのトップページから有限回リンクをたどってワインに関する言及のあるレストランの Webページを探すといった検索表現を，起点の Webページから有限回リンクを経由して目的の Webページまで到達するという形の検索パス式で記述することが可

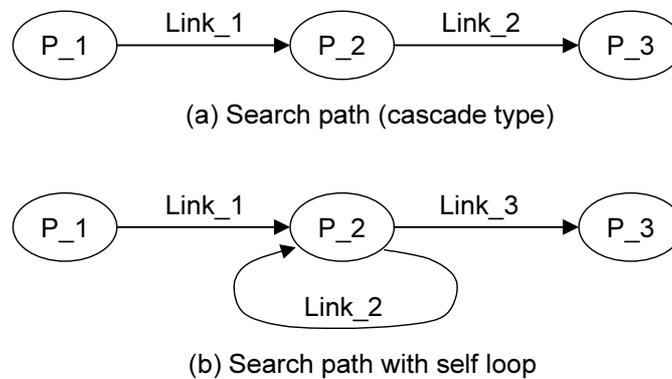


図 3.4: 検索パス式による検索パスのパターン

能となる。

3.3 検索言語

本節では、3.2 節で述べた拡張Web空間における検索を表現するための検索言語の文法を定義し、その意味論について説明する。

3.3.1 文法

本研究で定義する拡張Web空間の検索言語は、W3QL[33, 34]を基本に、Webページの属性の扱いと評価方法を変更し、Webページの持つ地理的属性を検索式中で扱えるように地理的ジェネリックリンクを導入したものである。検索式は、SELECT-FROM-WHERE形式で記述するものとし、図 3.5 のように定義する。

定義中の大文字列で始まるトークンは終端記号、小文字列は非終端記号、‘+’ は1回以上の繰り返しを意味する。*value* は、数値、文字列、文字列の正規表現、もしくは検索パス中に出現する Webページの属性を表現する。*operator* には、Perl言語で定義されている 2項演算子を利用するものとし、Webページの属性と *value* の部分一致、完全一致、もしくは大小関係の評

価するときに用いる .

```
query ::= SELECT select_clause
        FROM from_clause
        WHERE condition+
select_clause ::= object+
object ::= web_page | web_page.attribute
attribute ::= url | server | title | anchor |
            href | text | address
from_clause ::= step * web_page
step ::= web_page, link, | (web_page, link), link,
condition ::= web_page.attribute operator value |
            link AS link_type
value ::= True | False | numerical_value |
        string_value | regular_expression |
        web_page.attribute
link_type ::= HYPER_LINK | geographical_generic_link
geographical
_generic_link ::= gis_function( web_page, web_page )
operator value
gis_function ::= Equals | Disjoint | Touches | Within |
              Overlaps | Crosses | Intersects |
              Contains | Distance
```

図 3.5: 検索言語の文法

3.3.2 意味論

select_clause には , 検索結果として出力するWebページ (*web_page*) もしくは Webページの持つ属性 (*web_page.attribute*) を指定する . *from_clause*

表 3.2: 地理的ジェネリックリンクで用いる地理的關係演算例

関数名	意味	戻り値
Equals	同値	True/False
Disjoint	共通部無し	True/False
Touches	接触	True/False
Within	内部に存在	True/False
Overlaps	重複部有り	True/False
Crosses	交差	True/False
Intersects	共通部あり	True/False
Contains	包含關係	True/False
Distance	距離	連続値

には，検索の起点の Web ページから終点の Web ページまでの経路を，3.2.3 節で述べた検索パス式で記述する．また，括弧を用いて *web_page* と *link* の繰り返しを表現し，Web 空間の半構構性に対応した検索経路を実現する．

where_clause には，検索パス式上に出現する Web ページもしくはその属性に対する条件，および Web ページ間のリンクの種類とリンクに対する条件を記述する．記述された条件は，*from_clause* に記述した検索パス式に従って評価する．もし，同じ Web ページもしくは同じリンクに対して複数の条件が指定されている場合には，それらを条件の積集合として処理する．

リンクに関する条件は，まず AS によってリンクがハイパーリンクであるか，地理的関連性によって生成する地理的ジェネリックリンクであるかを指定する．ハイパーリンクの場合は，リンク元の Web ページの内部に記述されている情報に従ってリンク先の Web ページを求める．また，地理的ジェネリックリンクの場合には，3.2.2 節で説明した手順に従いリンク先の Web ページを検索し，リンクを生成する．地理的関連性に関する条件は，地理的関係演算の種類により，数値で評価する場合と，真偽値で評価する場合に分けられる（表 3.2）．

3.4 問い合わせの記述例

本節では、拡張Web空間で可能な問い合わせの例を4つ示し、それぞれの検索言語による記述例について説明する。

3.4.1 地理的ジェネリックリンクを用いた検索

起点となるWebページに対応する建物から100m以内に存在する建物のWebページを、地理的ジェネリックリンクを用いて検索する問い合わせの記述例を図3.6に示す。

```
1 SELECT target.url
2 FROM   start,distance_link,target
3 WHERE  start.url eq 'http://www.xx.yy/'
4        distance_link AS
          Distance(start,target) < 100m
```

図 3.6: 100m以内に存在する建物Webページを検索

1行目で目的のWebページのURL(target.url)を検索結果として出力するように記述し、2行目で起点のWebページ(start)、地理的ジェネリックリンク(distance_link)、目的のWebページ(target)からなる検索パス式を指定している。3行目で起点となるWebページのURLを指定、4行目でdistance_linkが距離条件による地理的ジェネリックリンクであることを記述し、距離が100m以内という条件でリンクを生成する。

この検索例では、targetに対する条件が設定されていないので、地理的ジェネリックリンクだけが条件として評価され、100m以内に存在するGISオブジェクトに関連したWebページがすべて列挙されることになる。

3.4.2 繰り返し表現を含んだ検索式

検索パス式中に、図3.4(b)で示した繰り返し表現を含んだ問い合わせの記述例を図3.7に示す。

```

1 SELECT tel.url
2 FROM   start,link,(middle,link),link,tel
3 WHERE  link AS HYPER_LINK
4       start.url eq 'http://www.xx.yy/'
5       middle.host eq 'www.xx.yy'
6       tel.text =~ /telephone/

```

図 3.7: 繰り返し表現を含んだ問い合わせ記述例

1行目で、検索結果として目的の Web ページの URL (tel.url) を出力するように記述、2行目には“(middle,link),” と繰り返し表現を含んだ検索パス式を指定する。検索パス式上のリンク link は Web ページ内に記述されたハイパーリンクであり (3行目)、URL で指定された起点の Web ページ (start, 4行目) から、同じホストに存在する Web ページへと接続したリンクをたどり (5行目)、“telephone” という文字列を含む目的の Web ページ (tel) を検索する。

繰り返し表現を導入することで、情報サイト www.xx.yy の内部構造が正確に分からない場合でも、“telephone” というキーワードを含む Web ページをすべて検索することができる。

3.4.3 シナリオの検索例

本章のはじめに示したシナリオは、GIS を利用して大学の近くの郵便局を検索し、検索した郵便局に対応する Web ページを調べて営業時間の条件に合う郵便局を検索することと、検索した郵便局から 1km 以内にある、レストラン情報サイトで紹介されているイタリア料理のおすすめ店のうち、ワインに関する言及のある店を検索することに分けることができる。

前半の「大学の近くにあって土曜日にも営業している郵便局」の検索では、まず起点を大学に設定し、GIS で大学の近くにある郵便局を検索し、検索した郵便局の Web ページを WWW で検索して、その中で営業時間に関する条件が合うものを結果として出力する。この検索では、「近く」という

条件を距離 500m未満と設定し，以下のように検索式を記述する．

```
1 SELECT schedule.url
2 FROM  university,distance_link,
        (postoffice,link),link,schedule
3 WHERE link AS HYPER_LINK
4     university.url eq 'http://www.zz.edu/'
5     distance_link AS
        Distance(university,postoffice) < 500m
6     postoffice.title =~ /postoffice/
7     schedule.text =~ /open on saturday/
```

図 3.8: 例題前半の記述例

1行目で検索結果として出力する `schedule.url` を指定，2行目でその検索パスを指定する．検索パス式では，まず，大学の Web ページ (`university`) を起点に，500m以内という距離条件に基づく地理的ジェネリックリンク (`distance_link`, 5行目) と Web ページのタイトルにキーワードである郵便局 “`postoffice`” が含まれるという条件 (6行目) で郵便局の Web ページを検索する．そして検索パス式後半 “`(postoffice,link),link,schedule`” により，郵便局サイトの内部構造にかかわらず，ハイパーリンクをたどってサイト全体の検索を行い，7行目の条件，営業時間に関するキーワード “`open on saturday`” を含む Web ページを発見して，その Web ページの URL (`schedule.url`) を結果として出力する．

後半のイタリア料理店の検索では，レストラン情報サイトのイタリア料理のおすすめ店リストを起点に，先に検索した郵便局との距離が 1km以内で，イタリア料理店のメニューの中でワインに関して言及のある店を検索結果として出力する．

この検索は，図 3.9 のように記述する．

1行目で，検索結果として `restaurant` を指定し，2行目でその検索パスを指定する．レストラン情報サイトのトップページ `info_top` を4行目で指定し，検索パス式内の繰り返し表現と 5行目の条件により，ホスト “`www.info.com`”

```

1 SELECT restaurant
2 FROM   info_top,link,(info,link),link,
        restaurant,distance_link,postoffice
3 WHERE  link AS HYPER_LINK
4   info_top.url eq 'http://www.info.com/'
5   info.host eq 'www.info.com'
6   restaurant.text =~ /italian/
7   restaurant.text =~ /wine/
8   postoffice eq 'http://www.post.com/'
9   distance_link AS
        Distance(restaurant,postoffice) < 1km

```

図 3.9: 例題後半の記述例

にある Web ページを 1 回以上経由して、キーワード “italian” を含むイタリア料理店でワイン “wine” に関する記述のある Web ページ (restaurant) を検索する。そして、前に検索した郵便局サイトとの間で距離を条件とした地理的ジェネリックリンクを生成し、条件に合致するレストランサイトの URL を検索結果として出力する。

3.4.4 3D 京都との連携

Flatland 社が作成した 3DML プラグインを利用して、京都市四条通の約 1.6km を 3D ウォークスルー (図 3.10) で再現した。3DML は、HTML とほぼ同様の形式を持った簡易型三次元空間記述言語であり、格子上のマップにオブジェクトを配置していくことで、仮想的な三次元空間を作成することができる。この 3D ウォークスルーは、Web ブラウザ上で利用することが可能であり、ユーザは仮想空間中をマウスやカーソルキーを利用して自由に移動することができる。

この 3D ウォークスルーの路面には、情報検索システムでクエリを実行するための CGI プログラムへのリンクが埋め込まれており、ユーザの移動

3.5 拡張Web空間に基づく情報検索システム

前節で説明した検索言語で記述された拡張Web空間に対する問い合わせでは、*select_clause* に指定された項目が検索結果として出力されるが、その際、*from_clause* に指定された検索パスと *where_clause* に指定された条件に従い、ハイパーリンクもしくは地理的ジェネリックリンクで接続された Web ページの列 (Web ページ列) を検索する。従って、この Web ページ列の検索が、拡張Web空間に基づく情報検索システムの目的となる。

指定された条件を満たす Web ページ列を検索する方法としては、インターネット上の Web サーバから逐次 Web ページを取り込み、指定された条件を評価しながら目的の Web ページ列を検索する方法と、あらかじめ作成した Web ページの属性に関する各種インデックスを利用し、指定された条件を満たす Web ページ列を検索する方法が考えられる。

前者は、Web crawling と呼ばれる方法であり、crawler と呼ばれる Web ロボットが、インターネット上の Web ページを効率的に収集するために用いられる。この crawler には、大量の Web ページを効率よく収集することを目的とした WebFountain プロジェクトの crawler[48] や、局所的に集中した特定の話題に関する Web ページを効率良く収集する Focused Crawler[49]、アンカータグ内の文字列を見て位置関連情報を含む Web ページを高確率で収集する位置関連情報の選択的収集[1]などがある。また Aggarwal らは、特定の話題に関する Web ページを収集する crawler の評価基準として、収穫率 (harvest rate, (条件を満たすページ数)/(収集したページ数)) を導入している[50]。Web crawling で Web ページ列を検索する場合は、この収穫率の高い方法が効率的と考えられるが、処理時間で考えると、逐次 Web ページを取り込み、その中に記述された情報に従って次に収集する Web ページを決定するため、インターネット経由で Web ページにアクセスする時間がその大半を占めることになる。このアクセス時間は、収集する Web ページ数に比例して増加するので、複雑な問い合わせをインタラクティブに処理する必要がある場合には、この方法は適さない。

これに対し後者の各種インデックスを利用する方法は、あらかじめ作成したインデックスを利用して問い合わせ処理を行う。このため、検索される情報の最新性は失われるものの、crawler を利用した方法よりも高速に

処理を行うことができる。この方法は、情報検索とデータベースシステムに関する技術の応用であり、オブジェクト指向データベース（OODB）、XMLデータ、半構造データを対象としたシステムで多く用いられている方法でもある。たとえば、XMLデータベースと半構造データを対象としたデータベース管理システム Lore[51, 52] では、オブジェクトやラベルつきエッジ、属性値などに関するインデックスと、データベースの構造を要約したパスインデックスを用意し、検索言語 Lorel[53]で記述された問い合わせの処理を実現している。OODBにおいても、複数のインデックスを利用した問い合わせ処理[54]が実現されている。

両者ともそれぞれの利点があり、どちらの方法でも、拡張Web空間に対する問い合わせで指定された条件を満たす Webページ列を検索することができるが、本研究では、問い合わせ処理の速度を優先し、後者の方法により情報検索システムを構築する。

3.5.1 情報検索システムの構成と検索方法

各種インデックスを利用する情報検索システムは、入出力モジュール、インデックス検索モジュール、地理的ジェネリックリンク生成モジュール、検索制御モジュール、最適化モジュールで構成する（図 3.12）。

入出力モジュール

入出力モジュールは、入力された問い合わせの字句解析を行い、指定された条件と検索パスを取り出す。また、後述するモジュールが検索した Webページ列から問い合わせで指定された項目を取りだし、整形出力する。

インデックス検索モジュール

インデックス検索モジュールは、Webページのタイトルに関するインデックス、Webページに含まれるキーワードのインデックス、ハイパーリンクのリンク元とリンク先に関するインデックスなど、あらかじめ収集した Webページから作成したインデックスを利用して、指定された条件を満たす Webページを検索する。このインデックス中では、Webページの URLは IDで管理され、インデックス検索の結果もその IDで返される。たとえば、図

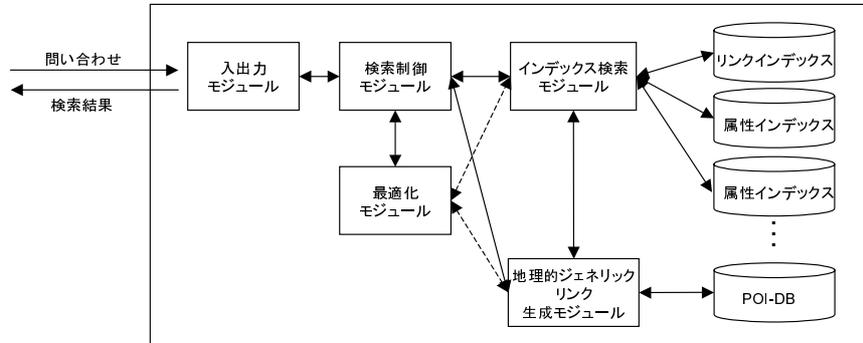


図 3.12: 拡張Web空間に基づく情報検索システムの構成

URLリスト	リンク先インデックス	リンク元インデックス	キーワードインデックス
1 http://www.aaa.com/	1 2,3,4	1 2,3	apple 1,2
2 http://www.bbb.com/	2 1	2 1	orange 3,4,5
3 http://www.ccc.com/	3 1	3 1,4	lemon 4,5
4 http://www.ddd.com/	4 3,5	4 1,5	
5 http://www.eee.com/	5 4	5 4	

図 3.13: 拡張Web空間におけるインデックスの例

3.13 のようなインデックスの場合，キーワード “apple” を含む Web ページの ID は 1 と 2 で，その ID に基づき URL リストを参照すると実際の URL を得ることができる。

なお，後述の性能評価では，インデックスに含まれている Web ページとその属性を問い合わせの条件として指定するものとする。また，Web ページからの動的な属性抽出やインターネット上の Web ページへの逐次アクセスを含むような問い合わせ処理は，本研究の情報検索システムでは扱わないものとする。

地理的ジェネリックリンク生成モジュール

地理的ジェネリックリンク生成モジュールは，図 3.14 に示す方法でリンクを生成する。Web ページの URL とその地理座標を関連付けた表 (POI-DB) を利用した Web ページの URL から地理座標へのマッピング，問い合わせ

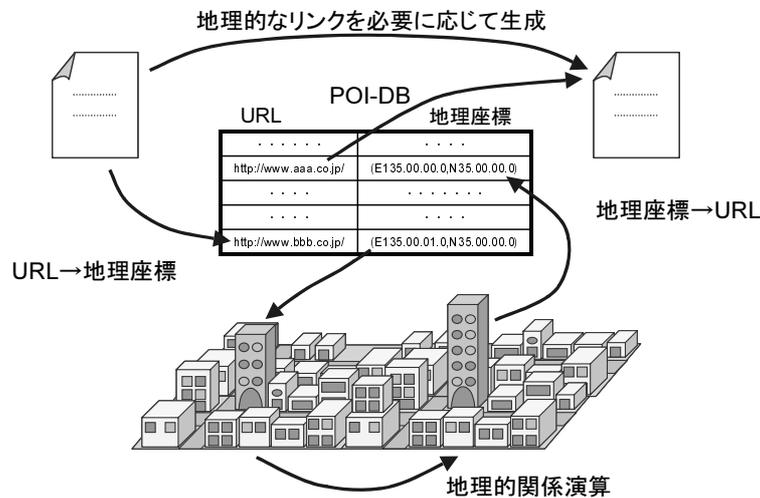


図 3.14: POI-DBを利用した地理的ジェネリックリンクの生成

で指定された条件に基づく地理座標間の地理的關係演算，そして POI-DB を利用した地理座標から Web ページの URL へのマッピングにより，地理的ジェネリックリンクを生成する．なお本研究では，表 3.2 に示した地理的關係演算のうち，距離条件によるリンク生成に限定して議論を進める．

地理的ジェネリックリンクを生成する際に利用する POI-DB は，Web ページの URL とその Web ページから抽出された地理属性に対応する地理オブジェクトの代表点の座標を関連付けたものである．3.2.2 節で説明した手法に従い，条件を緩和させながらオブジェクト間のマッピングを行うアドレスマッチングを利用して，他の属性に関するインデックスと同様にあらかじめ生成しておく．従って，POI-DB に登録されている Web ページの URL とその地理座標は，1対1対応であるとは限らない．また，インターネット上への公開を考慮し，位置の精度と利用許諾などの観点から，POI-DB に登録する情報の選別を行っている．

```

1 SELECT P2.url
2 FROM   P1,link,P2
3 WHERE  P1.url eq 'http://www.aaa.com/'
4        link AS HYPER_LINK
5        P2.text =~ /orange/
6        P2.text =~ /lemon/

```

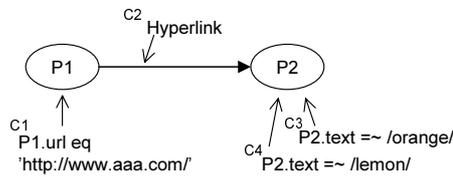
図 3.15: 図 3.13 のインデックスに対する問い合わせ

検索制御モジュール

検索制御モジュールは、入出力モジュールが問い合わせから取り出した検索パスと条件に従い Web ページ列を検索する。この際、指定された条件を満たす検索結果を中間結果に登録し、複数の条件による検索結果を統合しながら検索処理を進める。なお、問い合わせに指定された条件の処理順序は、条件間の依存関係とその処理コストを考慮して、最適化モジュールが決定する。この最適化モジュールによる条件処理順序の決定方法は、次節で説明する。

拡張 Web 空間に対する問い合わせで指定される条件は、Web ページの URL やタイトル、Web ページに含まれるキーワードなどに関する属性条件と、Web ページ間のハイパーリンクおよび地理的ジェネリックリンクに関するリンク条件に分けられる。そこで、それぞれの条件に基づく検索処理と検索処理中に行われる中間結果の操作を、図 3.15 に示す問い合わせ例を使って説明する。なお、この問い合わせ例は、図 3.13 で例示したインデックスを対象としたものである。

まず、図 3.15 の検索言語で記述された問い合わせから抽出した検索パスと条件 C1 ~ C4 を図 3.16 (a) に示す。この問い合わせで指定された条件のうち、C1, C3, C4 が属性条件、C2 がリンク条件である。また、指定された条件が C1, C2, C3, C4 の順で処理された場合の中間結果の状態を図 3.16 (b)、C1, C3, C4, C2 の順で処理された場合の中間結果の状態を図 3.16 (c) に示す。指定された条件に基づく検索結果は、中間結果に Web ページ単位で登



(a) 検索パスと条件

ノード	URL(ID)	リンク先	リンク元
P1	1		

(b-1) C1の検索結果の登録

ノード	URL(ID)	リンク先	リンク元
P1	1	2,3,4	
P2	2		1
P2	3		1
P2	4		1

(b-2) C2の検索結果の登録

ノード	URL(ID)	リンク先	リンク元
P1	1	2,3,4	
P2	2		1
P2	3		1
P2	4		1

(b-3) C3による絞込み

ノード	URL(ID)	リンク先	リンク元
P1	1	2,4	
P2	3		1
P2	4		1

(b-4) C4による絞込み

(b) C1,C2,C3,C4 の順序による検索処理中の中間結果

ノード	URL(ID)	リンク先	リンク元
P1	1		

(c-1) C1の検索結果の登録

ノード	URL(ID)	リンク先	リンク元
P1	1		
P2	3		
P2	4		
P2	5		

(c-2) C3の検索結果の登録

ノード	URL(ID)	リンク先	リンク元
P1	1		
P2	3		
P2	4		
P2	5		

(c-3) C4による絞込み

ノード	URL(ID)	リンク先	リンク元
P1	1	4	
P2	4		1
P2	5		

(b-4) C2による絞込み

(c) C1,C3,C4,C2 の順序による検索処理中の中間結果

図 3.16: 図 3.15 の検索処理の例

録される。この表に登録される情報は、検索結果の Web ページの検索パス上の位置を示すノード名、検索された Web ページの ID、リンク先 Web ページの ID、そしてリンク元 Web ページの ID である。また、図 3.16 (b),(c) の中間結果の中で、太字で記した項目が登録された情報、横線で消されている項目が削除された情報を表す。

1. 属性条件に基づく検索

属性条件の対象となっている属性に関するインデックスを検索し、指定された条件を満たす Web ページを検索する。この検索結果は、問い合わせ処理の中間結果に統合するが、処理対象となった検索パス上のノードに対応する中間結果の登録状態により異なる処理を行う。

- (a) 対象ノードに対応する中間結果が未登録の場合
検索結果の Web ページを中間結果として登録する。図 3.16 の例では、(b-1), (c-1), (c-2) がこれにあたる。たとえば (c-2) の場合は、条件 C3 によって検索された Web ページ 3, 4, 5 が中間結果に登録される。なお、この時点では、検索された Web ページのリンク先及びリンク元は不明なので、リンク情報は登録されない。
- (b) 対象ノードに対応する中間結果が既に登録されている場合
検索結果 Web ページと登録済みの中間結果を比較し、一致しなかった中間結果を登録から削除する。図 3.16 の例では、(b-3), (b-4), (c-3) がこれにあたる。たとえば (b-3) の場合は、条件 C3 によって検索された Web ページ 3, 4, 5 と、条件 C3 の対象となっているノード P2 の中間結果として登録されている Web ページ 2, 3, 4 を比較し、一致しなかった Web ページ 2 を中間結果から削除する。

2. リンク条件に基づく検索

ハイパーリンクに関する条件の場合は、リンク先に関するインデックスを検索して、リンク先 Web ページを求める。また、地理的ジェネリックリンクに関する条件の場合には、地理的ジェネリックリンク生成モジュールを利用して、リンク先 Web ページを求める。検索されたリンク先 Web ページは、検索パス上のリンク先ノードに関する中間結果に統合するが、属性条件の場合と同様に、中間結果の登録状況に応じて異なる処理を行う。

- (a) リンク先ノードに対応する中間結果が未登録の場合
検索結果の Web ページをリンク先とリンク元情報と一緒に、中間結果に登録する。図 3.16 の例では、(b-2) がこれにあたり、Web ページ P1 のハイパーリンク先である Web ページ 2, 3, 4 がそのリンク情報とともに中間結果に登録される。
- (b) リンク先ノードに対応する中間結果が既に登録されている場合
検索結果の Web ページと登録済みの中間結果を比較し、一致し

た中間結果にはリンク情報を追加し、一致しなかった中間結果は登録から削除する。図 3.16 の例では、(c-4) がこれにあたり、Webページ P2 の中間結果として登録されている Webページ 4, 5 のうち、Webページ 1 からリンクされていない Webページ 5 が中間結果から削除される。この際、Webページ 1, 4 には、検索されたリンク情報がそれぞれ追加される。

なお、本研究で実装した情報検索システムでは、ハッシュを利用して中間結果と検索結果の統合処理を行う。

もし上記の統合処理により、中間結果に登録されている Webページのリンク先がゼロであることが判明した場合は、その時点でその Webページを中間結果から削除する。たとえば、図 3.16 (b-4) で、ノード P2 に対応する中間結果が全て削除された場合には、その削除された Webページのリンク元の Webページ 1 も中間結果から削除する。この処理は、登録されている中間結果全体に対して再帰的に行う。

なお、リンク元に関するインデックスによりリンク元 Webページが検索可能な場合、もしくは逆方向に地理的ジェネリックリンクを生成できる場合には、後述する最適化モジュールが、それらを考慮に入れて条件処理順序を決定する。もし、逆方向に処理を行う場合には、上記の処理のリンク先とリンク元を入れ替えて処理を行う。

また、図 3.17 左のように問い合わせの検索パス式に繰り返し表現が含まれている場合は、図 3.17 右に示すように繰り返し部分を展開して複数の検索パスを作成する。そして、展開した検索パスそれぞれについて検索処理を行い、最終的にノード P1 に対応する Webページから 1 回以上ノード P2 に対応する Webページを経由してノード P3 に対応する Webページへとリンクで接続した Webページ列を検索結果とする。この繰り返し部分の展開は、XMLデータベースシステム Lore[55] で行われている正規表現を含む検索パスの展開を応用したものである。検索言語の仕様では、条件を満たすリンク先 Webページがある限り展開を続けることになるが、本研究の情報検索システムでは、検索処理の停止性を考慮し、システムの初期設定として最大展開数があらかじめ設定されるものとする。また、同一 Web

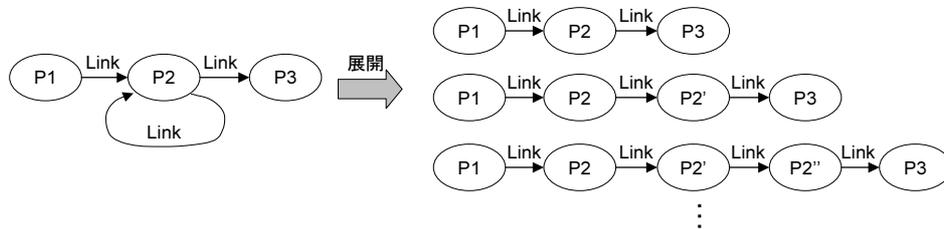


図 3.17: 繰り返し表現を含む検索パス式の展開

ページを繰り返し部分の中間結果として複数回登録しないように、登録済み Web ページをリスト形式で管理する。この他の処理については、リンク条件と属性条件を組み合わせた通常の間い合わせ処理と同様である。なお、この展開後の検索パス式に基づく検索処理における条件処理順序は、繰り返し部分のノードに対して指定される条件とその検索結果として登録される中間結果に基づき、後述の最適化モジュールが決定する。

XML データベースや OODB を扱うシステムで、検索パスに推移閉包や Kleene 閉包、正規表現が含まれる場合には、条件を満たすオブジェクトの検索に加え、オブジェクト間の先祖-子孫関係など、検索されたオブジェクト間の接続関係の検証も、検索処理を高速化するためのポイントとなる。この接続関係の検証方法としては、オブジェクトに付与した番号の比較により接続関係を確認できるようにした Dietz の番号付け[56] および XISS の番号付け[57]がある。また S 式の形式でオブジェクト間のリンク関係を記述し、そのインデックスにより接続関係の検証を可能にする path dictionary index[58]もあるが、情報空間が木構造であることを前提としているため、拡張 Web 空間のようにループを含む情報空間には適さない。

また、演繹データベースのボトムアップ推論で用いられるマジックセット法[59]では、再帰的ルールが効率的に処理されるように、ルールの変形と付加的な述語とルールの作成を行う。これにより、推論時にゴールに関係する部分だけが処理されるようになるが、上述した処理方法では、検索パスが繰り返し表現を含む場合も含めて、リンク条件に基づくリンク先 Web

ページの検索と属性条件に基づく Web ページの検索により検索結果が限定される。従って、マジックセット法の付加ルールで削除されるようなゴールに全く関係しない Web ページは、インデックス検索の段階で排除されることになる。

最適化モジュール

最適化モジュールは、中間結果の登録状況と、検索処理コストと検索結果数の予測に基づき、問い合わせに指定された条件の処理順序を決定する。Lore の問い合わせ最適化[51, 52]では、コスト予測に基づき最適な条件処理順序とパスをたどる戦略をそれぞれ決定し、問い合わせ処理を進める。これに対し、本研究の情報検索システムの場合は、リンク条件の検索結果を中間結果へ登録する作業が Web ページ間のリンク関係の確認となるので、問い合わせに指定された属性条件およびリンク条件の処理順序を一括して決定する必要がある。

検索処理コストと検索結果数の予測は、情報検索システムの入出力量の予測に基づき計算する。この予測式を表 3.3 に示す。予測式中の $|P_n|$ は、検索パス上のノード P_n に対応する中間結果として登録されている Web ページ数を表す。 $|est_result|$ は、指定された問い合わせ条件に基づいたインデックス検索の結果として予測される Web ページ数を表す。 $attr_index_size$, $link_index_size$, $poedb_size$ は、それぞれ属性に関するインデックスの大きさ、ハイパーリンクに関するインデックス、POI-DB の大きさを表す。

まず $|est_result|$ であるが、対象条件が属性条件の場合は、インデックスを先読みして予測する。ハイパーリンクに関する条件の場合は、中間結果に登録されている Web ページの URL が判明している場合には、インデックスを先読みして予測を行い、判明していない場合には、Web ページが含むハイパーリンクの平均とリンク元の候補として登録されている Web ページ数の積を予測とする。地理的ジェネリックリンクに関する条件については、予測時にリンク元の候補の URL が具体的に判明している場合には、POI-DB を走査して実際に地理属性をもつ Web ページ数を求め、POI-DB 内で指定された距離条件を満たす組み合わせ数[60]との積により検索結果を予測する。また、URL が判明していない場合には、リンク元の Web ペー

ジ数と統計情報から求めた Webページの地理属性の含有確率と距離条件を満たす組み合わせ数の積を予測とする。

検索処理コストの予測は、インデックス検索に必要な処理コストの予測と中間結果への検索結果の登録に必要な処理コストの予測により求める。インデックス検索の処理コストは、指定された条件を満たす検索結果を得るために必要なデータ読み込み量から予測する。インデックス検索モジュールは、ソートされたインデックスを二分探索するので、インデックスの大きさを N とすると $\lceil \log_2(N) \rceil$ 回のキー値の読み込みと比較処理が必要となる。なお、属性条件の場合はインデックス検索が 1回で済むが、リンク条件の場合は中間結果に登録されているリンク元 Webページ数だけインデックス検索を繰り返す必要がある。また、地理的ジェネリックリンクの場合は、POI-DBの検索と地理的關係演算が行われるので、それぞれの処理コストの予測を足し合わせる。地理的關係演算については、起点となるオブジェクト数 $|P_n|$ と検索結果のオブジェクト数 $|est_result|$ の和を入出力量とする。

中間結果への検索結果の登録に必要な処理コストは、属性条件の対象となったノードもしくはリンク条件のリンク先ノードの中間結果として登録状況に応じて予測する。対象ノードの中間結果が未登録の場合は、検索結果数の予測 $|est_result|$ が登録処理の予測コストになり、この検索結果数が中間結果にも登録される。対象ノードの中間結果が登録されている場合は、登録されている中間結果から検索結果に含まれない Webページを削除することになるので、中間結果として残る Webページ数の最大 $\min(|P_n|, |est_result|)$ を検索結果数の予測とし、この際の比較処理に必要なデータ読み込み量 $|est_result| + |P_n|$ と削除される Webページ数の予測 $|P_n| - \min(|P_n|, |est_result|)$ の和を処理コストの予測とする。なお、Webページの属性同士を比較する条件の場合はインデックス検索は不要であるが、中間結果に関する処理コストの予測は属性条件の場合に準じる。

この検索処理コストの予測に基づき条件処理順序を決定する場合は、動的計画法を用いた方法[61]が有効である。しかし、半構造データである拡張Web空間では、Webページに記述されている内容や Webページ間のリンク関係の不規則さにより、検索処理コストの予測に誤差が発生し、動的

表 3.3: 問い合わせ処理の最適化で用いる予測式

問い合わせ条件	中間結果	処理コストの予測式
P _n .attribute operator Constant	未登録 (P _n)	$1 * \lceil \log_2(attr_index_size) \rceil + est_result $
	登録済 (P _n)	$1 * \lceil \log_2(attr_index_size) \rceil + P_n + est_result + P_n - \min(P_n , est_result)$
link AS HYPERLINK (P _m , link, P _n)	未登録 (P _n)	$ P_m * \lceil \log_2(link_index_size) \rceil + est_result $
	登録済 (P _n)	$ P_m * \lceil \log_2(link_index_size) \rceil + P_n + est_result + P_n - \min(P_n , est_result)$
link AS Distance(P _m , P _n) operator Constant	未登録 (P _n)	$(P_m + est_result) * \lceil \log_2(poidb_size) \rceil + P_m + est_result + est_result $
	登録済 (P _n)	$(P_m + est_result) * \lceil \log_2(poidb_size) \rceil + 2 * (P_n + est_result) + P_n - \min(P_n , est_result)$
P _m .attribute operator P _n .attribute (m ≠ n)	登録済 (P _n , P _{n+1})	$ P_m + P_n + P_n - \min(P_m , P_n)$

計画法では最適な条件処理順序が決定できない可能性が考えられる。このような状況の対処法としては、適応型問い合わせ処理[62]が考えられる。この適応型問い合わせ処理としては、条件処理順序を決定する際に参照した予測値と実際の値の間の誤差が一定基準を越えた段階で処理順序を再検討する方法[63]や、有効な予測に基づく部分的な処理順序の決定とその処理をインターリーブする手法を取り入れた Tukwila[64]などがある。

そこで本章では、拡張Web空間における問い合わせ処理に適した最適化手法を検討するため、複数の最適化手法を情報検索システムに実装し、実際の問い合わせ例を用いてそれぞれの処理性能を比較する。比較する最適化手法は、動的計画法による事前最適化、制限付き事前最適化、実行時最適化の3つである。それぞれの処理手順の詳細を以下に示し、処理性能と比較結果については次節で説明する。

create_plan_by_DP(C, R)

入力：未処理条件の集合 C, 中間結果 R

出力：条件処理順序 query_plan, 予測処理コスト cost, 予測される中間結果 result

```
1: C' = feasible_condition( C )
2: if C' =  $\phi$  { return( {}, 0, R ) }
3: foreach c  $\in$  C' {
4:   r = R
5:   cost = cost_estimation( r, c )
6:   result = result_estimation( r, c )
7:   result_register( r, result )
8:   ( plan_c, cost_c, result_c ) = create_plan_by_DP( C \ c, r )
9:   if ( cost + cost_c ) < best_plan_cost {
10:    best_plan = { c, plan_c }
11:    best_plan_cost = cost + cost_c
12:    best_plan_result = result_c
13:   }
14: }
15: return ( best_plan, best_plan_cost, best_plan_result )
```

図 3.18: 動的計画法による事前最適化アルゴリズム

- 動的計画法による事前最適化

問い合わせに指定されている条件の予測処理コストの和が最小となるように、図 3.18 のアルゴリズムを再帰的に呼び出して条件処理順序を作成する。この作成した条件処理順序に従い、検索制御モジュールが実際の検索処理を行う。このアルゴリズムの入力は、問い合わせに指定された条件の集合と空の中間結果、出力は条件処理順序とその予測処理コストと予測中間結果である。もし、問い合わせに指定されたリンク条件が逆方向にも処理可能で、リンク先からリンク元を検索できる場合には、その条件も入力の場合に追加する。

まず、1行目で未処理条件の集合から処理可能な条件を取りだし、2行目で処理可能な条件の有無を確認する。もし、処理可能な条件の集合が空集合の場合は、空の条件処理順序とコスト0と渡された中間

結果をそのまま返す．ここで未処理条件の集合が空集合の場合は，再帰呼び出し前の段階で処理可能な全ての条件に関する予測が終了していることを意味する．

3行目から，処理可能な条件に対する処理を始める．4行目で中間結果のコピーを作成し，5行目の関数 `cost_estimation` で検索制御モジュールが対象条件を処理する際に必要となる入出力量を予測し，6行目の関数 `result_estimation` で対象条件によるインデックス検索で得られる検索結果数を予測する．これらの予測は中間結果と対象条件に基づき行われ，Webページを単位とした予測値が返される．また，実際の検索処理では Webページの IDを中間結果に登録するが，最適化処理の最中は検索結果として予測される Webページ数を登録する．

7行目の `result_register` で，対象条件の予測結果 `result_c` を中間結果のコピー `r` に統合する．この際，対象条件の対象ノードに関する中間結果が未登録の場合は検索結果として予測される Webページ数を登録，既登録の場合は予測される Webページ数と既登録の Webページ数の最小値を再登録する．

次に，8行目で最初に与えられた未処理条件から現在の対象条件を除いた条件集合と，対象条件までの処理による予測中間結果を引数として，本アルゴリズムを再帰呼び出しする．これにより，対象条件の処理以降の最適な条件処理順序 `plan_c` とその予測コスト `cost_c`，そして対象条件から処理を進めた場合の予測結果 `result_c` がそれぞれ得られる．

9~12行目で，対象条件の予測処理コストとその以降の予測処理コストの和が最小となる条件処理順序とその予測処理コストと予測中間結果をそれぞれ `best_plan`, `best_plan_cost`, `best_plan_result` に保存し，16行目でそれらを最適な条件処理順序として返す．この際 `best_plan` には，最適となった条件とそれ以降の最適条件処理順序を連結した条件処理順序が保存される．

- 制限付き事前最適化

制限付き事前最適化による検索処理では，最適化モジュールによる

create_plan_by_IDP(C, R, K)

入力：未処理条件の集合 C, 中間結果 R, 再帰の深さ K

出力：条件処理順序 query_plan, 予測処理コスト cost, 予測される中間結果 result

```
1: C' = feasible_condition( C )
2: if ( C' =  $\phi$  ) or ( K = 0 ) { return( {}, 0, R ) }
3: foreach c  $\in$  C' {
4:   r = R
5:   cost = cost_estimation( r, c )
6:   result = result_estimation( r, c )
7:   result_register( r, result )
8:   ( plan_c, cost_c, result_c ) = create_plan_by_IDP( C \ c, r, K - 1 )
9:   if ( cost + cost_c ) < best_plan_cost {
10:    best_plan = { c, plan_c }
11:    best_plan_cost = cost + cost_c
12:    best_plan_result = result_c
13:   }
14: }
15: return ( best_plan, best_plan_cost, best_plan_result )
```

図 3.19: 制限付き事前最適化アルゴリズム

部分的な条件処理順序の作成と、それに従った検索処理を繰り返して、実際の検索処理を進めていく。この部分的な条件処理順序を作成するためのアルゴリズムを図 3.19 に示す。このアルゴリズムでは、図 3.18 の事前最適化のアルゴリズムに再帰の深さ K を設定することにより、部分的な条件処理順序を作成する。従って、もし、再帰の深さに問い合わせに指定された条件数が設定された場合には、事前最適化の場合と同等の検索処理を行うことになる。

この制限付き事前最適化では、2回目以降の条件処理順序の作成に実際の中間結果が反映されるので、より正確な予測に基づき条件処理順序が作成されることになる。その反面、完全な条件処理順序が作成されないため、問い合わせに指定された条件全てが最適な順序で処理されとは限らない。

- 実行時最適化

実行時最適化では，最も予測処理コストの小さい条件の選択とその条件に基づく検索処理を繰り返して，検索処理を進める．従って，この方法は，図 3.19 の制限付き事前最適化アルゴリズムで再帰の深さを 1 にした場合と同等になる．

この方法では，事前最適化のように予測結果に基づく予測が行われないので，予測誤差の累積は発生しない．たとえば図 3.19 の 5 行目の `cost_estimation` では，実際の検索結果が登録された中間結果に基づき処理可能な条件の処理コストが予測され，その処理コストが最小の条件が次に処理されることになる．従って，処理される条件の最適性は局所的なものになるため，制限付き事前最適化と同様に，問い合わせに指定された条件全てが最適な順序で処理されるとは限らない．

3.5.2 性能評価と最適化手法の比較

本節では，実際のインターネット上の Web 情報と地理情報から作成したインデックスと 3 つの問い合わせ例を用いて，情報検索システムの性能評価を行う．この性能評価は，脚注¹に示した環境で行い，検索処理と最適化処理にかかる実時間の測定²と，条件処理順序を決定する際に最適化モジュールが参照した予測値と，決定された条件処理順序に従い検索制御モジュールが検索処理を行った際の実際の入出力量を比較する．なお，以下に示す各処理時間は，評価環境における 10 回の試行の平均である．また，比較結果中の最適化モジュールの予測値および検索制御モジュールの入出力量の実測値は，中間結果に登録される Web ページを単位としている．

本節の性能評価で比較する最適化手法は，事前最適化，制限付き事前最適化，実行時最適化，そして最適化なしの 4 手法とする．このうち，最適化なしの場合には，問い合わせに指定された検索パスの先頭から順に，指定されている条件を処理していく．

¹評価環境：DELL Precision530, Xeon2.0GHz x 2, 1GB RAM, 20GB HDD, Redhat7.1, PostgreSQL7.1.3, Perl5.6.0

²Perlのプロファイラ (Devel::DProfモジュール) で測定．

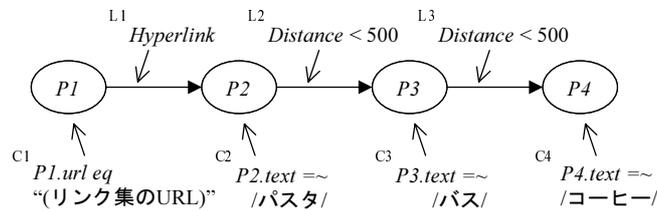


図 3.20: 性能比較のための例題 (1)

また、性能評価で用いる POI-DBには、実際にデジタルシティ京都からリンクしている 4,974件の URLとその地理座標を登録した。この POI-DBの管理には、PostgreSQL[65]を利用している。そして、POI-DBに登録されている URLと Yahoo! JAPAN のディレクトリの「地方情報 → 日本の地方 → 京都」に登録されている URLをシード URLとし、深さ 3の設定で Webロボットを利用して再帰収集した Webページ 94,007ページからキーワードインデックスやリンク先インデックスなどの各種インデックスを作成した。

例題 (1) による性能比較

はじめに、図 3.20 に示す例題 (1) を用いて最適化手法の性能比較を行う。この例題では、属性条件 $C1 \sim C4$ 、およびリンク条件 $L1 \sim L3$ を満たす Webページ列 ($P1, P2, P3, P4$) を検索する。まず、この例題 (1) における検索処理時間と最適化処理時間を図 3.21 に示す。なお、この図 3.21 の結果を得た際の制限付き事前最適化の再帰の深さは、後述する実験で得た最適値の 2である。各最適化手法の検索処理時間を比較すると、制限付き事前最適化が 0.276秒、実行時最適化が 0.278秒とほぼ同等であったのに対し、事前最適化の処理時間は 0.439秒とそれらに及ばなかった。最適化処理時間に関しても、実行時最適化と制限付き事前最適化がほぼ同等であったのに対し、事前最適化の処理時間はそれらを大幅に上回っていた。

本研究で実装した動的計画法を用いた事前最適化には、特に枝刈り処理等が含まれていないため、条件数が増加すると動的計画法により比較する

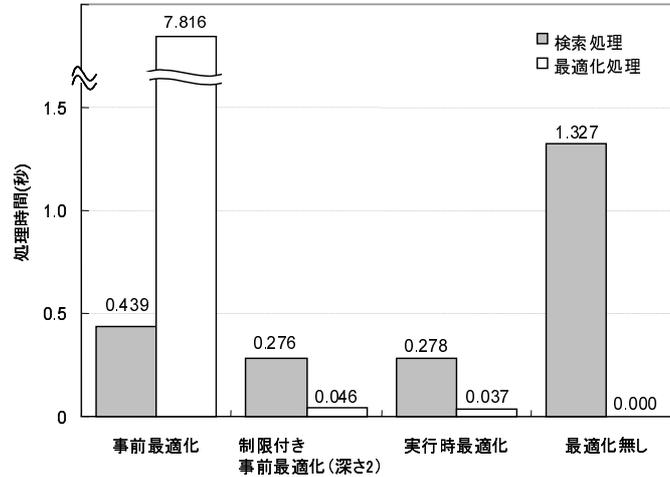


図 3.21: 最適化手法と処理時間

条件処理順序が指数的に増加する．従って，最適化処理時間の増大は自明と考えられるが，例題（1）の場合には最適な条件処理順序を作成することもできなかった．

この原因を調べるために，制限付き事前最適化で設定する再帰の深さを変化させて再実験を行った．この実験結果を図 3.22 に示す．例題（1）には条件が 7つ含まれているので，再帰の深さを 7にすると事前最適化と同等となり，1にすると実行時最適化と同等になる．その途中区間の最適化処理時間をみると，再帰の深さが 4を越えた段階で検索処理時間を上回り，その後急激に増大した．また検索処理時間をみると，深さ 2のとき最短の 0.276秒であったが，再帰を深くしたからといって必ずしも検索処理時間は改善されなかった．

次に，再帰の深さと条件処理順序の関係を示した表 3.4 をみると，途中の順序には差があるものの，いずれの場合も地理的ジェネリックリンクに関するリンク条件 $L3$ が最後に処理されていたことがわかる．この $L3$ を処理する際，再帰の深さが 1,2,3,6の場合は，その直前までの処理結果を反映した予測に従い $L3$ が順方向に処理されたのに対し，深さが 4,5,7の場合は，

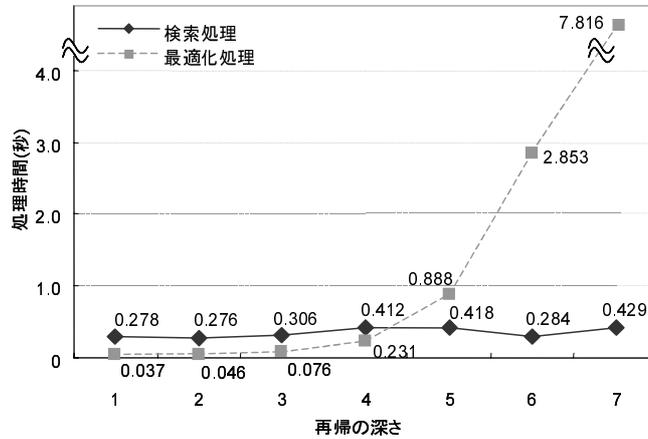


図 3.22: 制限付き事前最適化の再帰の深さと処理時間

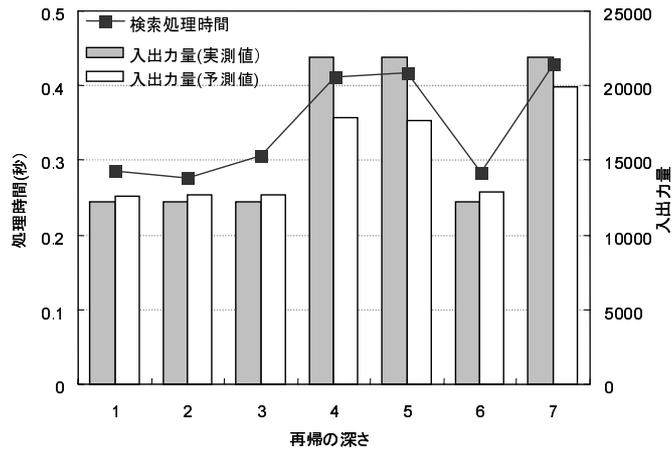


図 3.23: 制限付き事前最適化の再帰の深さと入出力量

$L3$ 以外の条件を含む条件処理順序に基づき, $L3$ が逆方向に処理されていたことがわかる. また, 評価システムの入出力量を示した図 3.23 をみると, 前者のグループの方が検索処理時間が短く, 入出力量の実測値と予測値の間の誤差も小さかったことがわかる. このうち, 検索処理時間が最短

表 3.4: 制限付き事前最適化の再帰の深さと条件処理順序

深さ	条件処理順序
1	C1 L1 C2 C4 C3 L2 L3
2	C1 L1 C2 C4 C3 L2 L3
3	C1 L1 C2 C3 L2 C4 L3
4	C1 L1 C2 C4 C3 L2 L3 ⁻
5	C1 L1 C2 C3 C4 L2 L3 ⁻
6	C1 L1 C3 L2 C2 C4 L3
7	C1 L1 C3 L2 C2 C4 L3 ⁻

1 Lⁿ⁻ はリンクを逆方向で処理したことを示す。
 2 条件間の“|”は検索プランの作成を示す。

表 3.5: 情報検索システムの入出力量の比較 (1)

処理 順序	深さ2		深さ7	
	条件	実測値 (誤差)	条件	実測値 (誤差)
1	C1	2 (0)	C1	2 (0)
2	L1	43 (6)	L1	43 (6)
3	C2	144 (11)	C3	1460 (43)
4	C4	161 (10)	L2	1754 (1464)
5	C3	1460 (43)	C2	140 (9)
6	L2	1754 (1291)	C4	161 (10)
7	L3	8627 (872)	L3 ⁻	17657 (2830)

の深さ 2の場合と最長の深さ 7の場合の条件処理順序と入出力量を示した表 3.5 をみると、両者とも条件 L2 の処理コストの予測で同程度の誤差が発生しているが、深さ2の場合は、条件 L2 の処理結果が反映された予測に基づき条件 L3 の処理が行われたのに対し、深さ7の場合はその誤差を累積したまま条件 L3 の処理方法が決定されたことがわかる。従って例題 (1) の場合、条件処理順序を作成する際の予測の誤差の累積が、事前最適化で最適な条件処理順序を作成できなかった原因と考えられる。

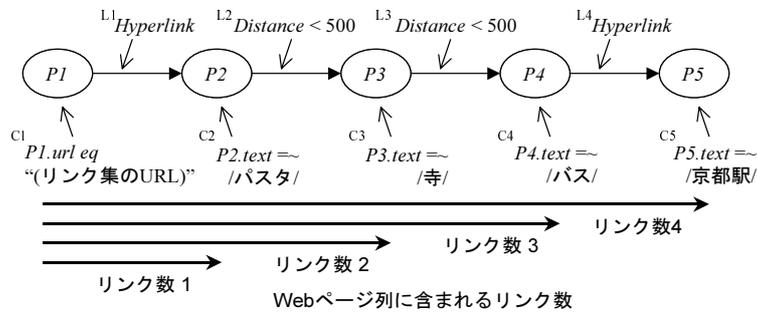


図 3.24: 性能比較のための例題 (2)

例題 (2) による性能比較 – リンク数と最適化の関係

次に、問い合わせに含まれるリンク数と最適化処理の関係を調べるために、図 3.24 に示す 4つのリンクが含まれる例題 (2) から、リンク数が 1 から 4までの問い合わせを抽出して、合計 4パターンの検索実験を行う。

まず、各パターンの検索処理時間の比較を図 3.25 に示す。この図 3.25 をみると、含まれるリンク数が 2以下の場合、最適化手法間に大きな差はみられないが、含まれるリンク数が 3以上になると、事前最適化の検索処理時間が実行時最適化の場合の 2倍以上となる場合もあった。

前節と同様に、事前最適化で最適な条件処理順序を作成できなかった理由を調べるため、含まれるリンク数が 4の問い合わせについて、制限付き事前最適化に設定する再帰の深さを変化させて再実験を行った。まず、図 3.26 に設定した再帰深さごとの検索処理時間と、評価システムの入出力量の実測値と予測値を示す。このうち、検索処理時間をみると、深さ1の場合が最短で、深さ2,4と続き、深さ3,9の場合は最短の場合の 2倍以上の処理時間がかかっていたことがわかる。入出力量の実測値と予測値の間の誤差についても、処理時間のかかった深さ3,5,6,9の場合の誤差は、それ以外の場合よりも大きかった。

また、制限付き事前最適化により最適化モジュールが作成した条件処理順序を設定した再帰深さごとに示した表 3.6 をみると、深さ2の場合は最

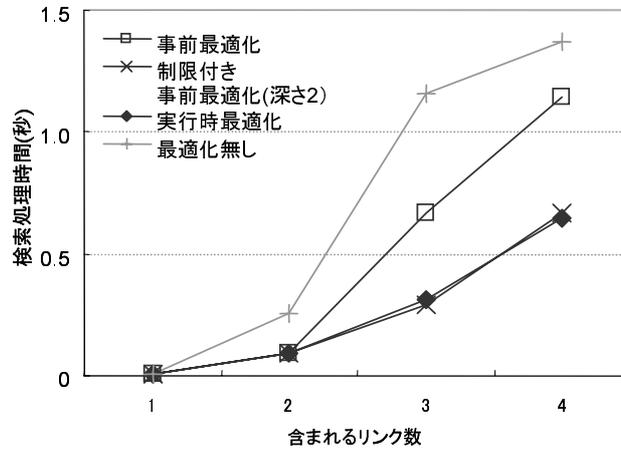


図 3.25: 問い合わせに含まれるリンク数と検索処理時間

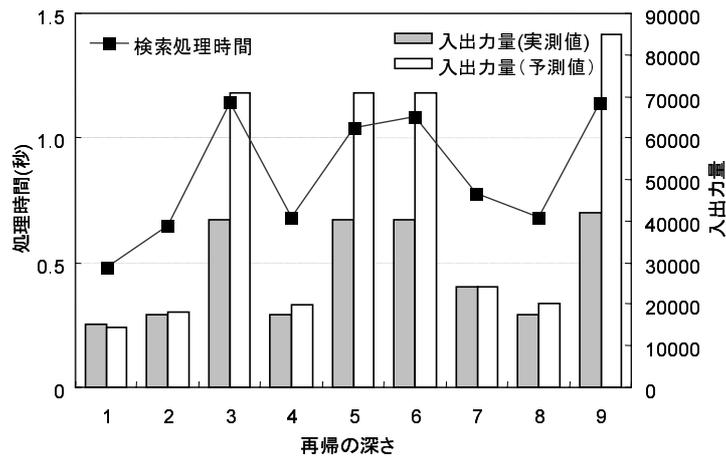


図 3.26: 検索処理時間と入出力量 (リンク数4の場合)

後にリンク条件 $L3$ だけを含む条件処理順序を作成しているのに対し、深さ3の場合は3つのリンク条件を含む条件処理順序を作成していた。この3条件を処理する際の検索制御モジュールの入出力量を示した表 3.7 をみると、深さ3の場合の条件 $L3$ の予測値の誤差が大きく、その累積が条件処理

表 3.6: リンク数4の場合の条件処理順序

深さ	条件処理順序
1	$C1 L1 C2 C3 C5 C4 L2 L3 L4$
2	$C1 L1 C2 C3 C4 C5 L2 L3^- L4^-$
3	$C1 L1 C2 C3 C4 C5 L4^- L3 L2$
4	$C1 L1 C2 C3 L2 C4 C5 L4^- L3^-$
5	$C1 L1 C2 C3 C5 C4 L4^- L3 L2$
6	$C1 L1 C2 C3 C4 C5 L4^- L3 L2$
7	$C1 L1 C3 L2 C2 C4 C5 L3 L4^-$
8	$C1 L1 C3 L2 C2 C4 C5 L4^- L3^-$
9	$C1 L1 C3 C5 C4 L3 L2 C2 L4^-$

表 3.7: 情報検索システムの入出力量の比較 (2)

処理 順序	深さ2		深さ3	
	条件	実測値 (誤差)	条件	実測値 (誤差)
7	$L2$	1727 (426)	$L4^-$	10923 (1396)
8	$L3^-$	11194 (1125)	$L3$	23998 (28861)
9	$L4^-$	1200 (800)	$L2$	1972 (181)

順序の質の悪化につながったと予想される。従って、例題(2)の場合も例題(1)と同様に、予測誤差の累積が事前最適化で最適な条件処理順序を決定できなかった原因と考えられ、その中でも特に、リンク条件に関わる予測誤差の累積が大きな原因になったと考えられる。

例題(3)による性能比較 – 条件数と最適化の関係

最後に、一つのWebページに対して複数の属性条件が指定される場合の問い合わせ処理の最適化に関する実験結果を示す。この実験では、図 3.27 に示した例題(3)を用い、Webページ $P2$ および $P3$ に指定する属性条件の数をそれぞれ1から3まで変化させた9パターンと、 $P2$ で検索を終了させ

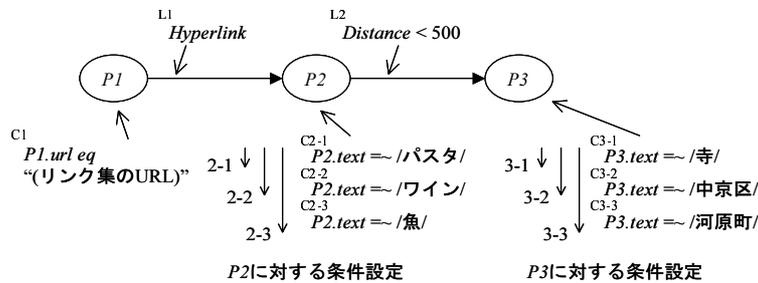


図 3.27: 性能比較のための例題 (3)

る場合の3パターンの合計12パターンを用意し、各最適化手法の比較を行った。なお、制限付き事前最適化を用いた場合の結果については、再帰深さ2が場合の結果を表示している。

まず、各最適化手法ごとの検索処理時間を示した図 3.28 をみると、問い合わせの各パターンで各最適化手法の検索処理時間に差がほとんどなかったことがわかる。例題(2)の実験結果においても、リンク数が1,2の場合には、各最適化手法で検索処理時間に差がほとんどなかったことから、問い合わせに含まれるリンク条件が少ない場合には、属性条件の数に関わらず、最適化手法による検索処理時間の差はほとんど出ないと考えられる。特に、 $P2$ に対して3条件が指定された4パターンでは、最適化なしの場合も含めて検索処理時間に差がほとんどなかった。従って、本節の例題(3)の場合には、条件処理順序の差異よりも、属性条件による絞り込み効果の方が、検索処理時間に大きな影響を与えたと考えられる。

また、 $P2, P3$ それぞれ3条件が指定された合計9条件の問い合わせの場合の条件処理順序を示した表 3.8 をみると、最適化手法により属性条件の処理順序の差はあるが、リンク条件の処理順序については一致がみられる。前節までの実験結果で、リンク条件に関する予測誤差の累積が性能悪化の原因となったことから、このリンク条件の処理順序の一致が、各最適化手法による検索処理時間に大きな差が出なかった原因の一つと考えられる。

以上、例題(1)～(3)を用いた性能比較から、拡張Web空間における

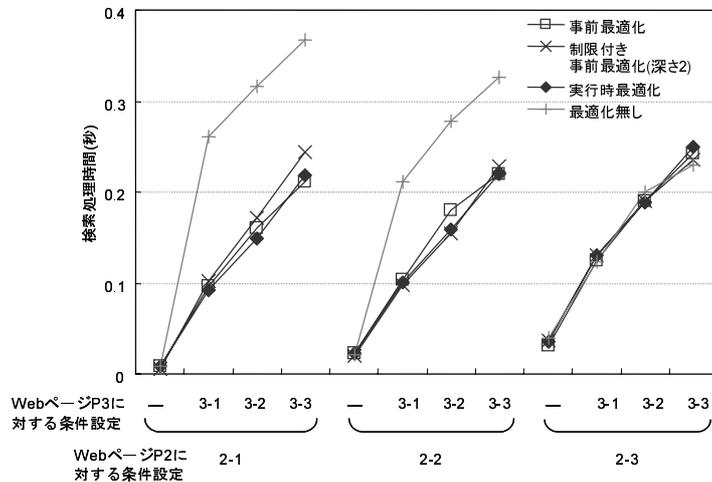


図 3.28: 条件数と検索処理時間

表 3.8: 9条件が指定された場合の条件処理順序

最適化手法	条件処理順序
事前最適化	C1 L1 C3-1 L2 C2-1 C2-2 C2-3 C3-2 C3-3
制限付き 事前最適化*	C1 L1 C2-1 C2-3 L2 C2-2 C3-1 C3-3 C3-2
実行時最適化	C1 L1 C2-1 C2-3 L2 C2-2 C3-1 C3-3 C3-2
最適化なし	C1 L1 C2-1 C2-2 C2-3 L2 C3-1 C3-2 C3-3

* 深さ2

問い合わせ処理を各種インデックスを利用した情報検索システムで行う場合には、リンク条件の処理順序を最適に決定することが検索処理の効率化において重要といえる。また、動的計画法を利用した事前最適化であらか

じめ条件処理順序を作成するよりも，再帰の深さを制限した事前最適化や実行時最適化により，条件処理順序の作成とその実行を繰り返す方が，リンク条件に関する予測誤差の累積が回避され，その結果，処理性能の悪化を回避できることがわかった．

3.6 まとめ

本章では，Web情報空間を GISを用いて動的に拡張する拡張Web空間を提案し，その拡張Web空間における問い合わせを記述するための検索言語を定義した．また，拡張Web空間に基づく情報検索システムを構築して複数の最適化手法の比較を行い，あらかじめ条件処理順序を作成する事前最適化よりも，検索処理中に部分的な条件処理順序の作成とその処理を繰り返す制限付き事前最適化や実行時最適化の方が，半構造化を有する拡張Web空間の問い合わせ処理で良い性能を示すことを観測した．

また本章では，拡張Web空間における問い合わせの記述例は示したが，その機能を十分に生かすためのユーザインタフェースの開発は今後の課題である．情報検索システムについては，対話型インタフェースなど，インタラクティブな問い合わせ処理への対応や，実際の運用を考慮した負荷分散，インターネットや GISを利用した効率的なインデックスの構築が今後の課題と考えている．

第 4 章

地域情報検索のための地図インタフェース

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，携帯電話やカーナビなどからも利用できるようになるなど，都市におけるインターネット利用は容易にかつ便利なものになりつつある．

こうしたインターネット上の情報システムのうち，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いる [8, 9, 66, 67]．地図を利用する利点は，都市に関連した電子的な情報が実際の都市における地理的な属性に基づいて参照できるようになる点にある．特に，インターネット上の Web ページと地図がリンクすることによって，より多くの多様な情報が地理的な属性に基づいて参照できるようになると予想される．しかし，地図上で扱う Web ページ数の増加や地理的な集中に対処するためには，地図を利用したインタフェースと検索機能の統合が必要不可欠となる．

そこで，都市に関する大量の Web ページを効率的かつインタラクティブに検索できるようにするため，地図を利用した多角的な Web ページの検索機能を組み入れた地図インタフェースを提案する．本地図インタフェースでは，検索機能として近傍検索，カテゴリ検索，およびキーワード検索を取り入れ，検索結果を地図上に重ね合わせて表示することによって，実世界に即した情報利用を可能にする．また，動作中にクライアント側へ転送したデータをキャッシュとして利用することによって，クライアント側の地図インタフェースとサーバ上の情報検索システム GeoLink の間で検索処

理の分散を図り，地図インタフェースにおけるインタラクティブな検索を可能にする．GeoLinkは，筆者らが地図インタフェースとあわせて開発した特定地域の Web ページを対象とした情報検索システムであり，インターネット上に存在する都市に関連した Web ページをそれぞれの地理的属性に基づいて検索することができる．

以下本章では，この地図インタフェースの Java アプレットによる実装と，測位機能を備えた携帯情報端末への適用例について説明する．次に，地域ポータルサイトとして公開実験を行っているデジタルシティ京都[23]で行った地図インタフェースの性能評価と利用状況の分析について報告する．そして最後に，地図インタフェースの今後の課題について述べる．

4.1 関連事例

4.1.1 インターネット上の地図情報サービス

デジタルシティのように都市に関する情報を集めたインターネット上のサービスでは，地図を利用した情報提供が重要になると考えられる．現在，インターネット上で公開されている地図を利用した情報サービスの代表的なものとしては，インターネットを新たなメディアと捉えて広告情報を登録した地図の配信を行うマピオン[8]や，地図を利用した情報サービスで携帯端末にも対応したマップファンウェブ[9]がある．

これらの情報サービスでは，イメージマップ機能を利用することによって，地図上のアイコンから詳細情報を参照することができる．イメージマップとは，イメージ中の任意の領域から関連情報へのリンクを設定できるようにハイパーリンクを拡張したリンク機能であり，検索内容の複雑さに応じて，ユーザの入力座標をサーバで処理するサーバサイドイメージマップと，クライアントで処理するクライアントサイドイメージマップ[68]がある．

インターネット上の地図を利用した情報提供サービスを従来のスタンドアロン型の GIS と比較すると，Web ブラウザ上で実現できるインタフェースの機能の制限と，ネットワークの帯域幅によるデータ転送速度の遅さのため，限られた検索機能をバッチ処理的にしか実行できないのが現状であ

る。ただ、従来の GISは、スタンドアローンもしくはイントラネットで運用するものがほとんどであったが、今後は、分散型電子地図プラットフォームJaMaPS[66]や MM情報マップディレクトリ[67]のように、ネットワーク上に分散した情報を地図と関連づけて統合的に利用できるシステムが、インターネットにおいて重要な役割を持つようになると考えている。

4.1.2 地図を利用した情報検索

電子地図を利用した情報検索は GISの基本機能であるが、その主な検索スタイルには、空間的な情報を文字情報に置き換えて行うキーワード検索、キーワードに相当する図形を入力に利用するアイコン検索、地図の拡大・縮小・スクロールを行いながら地理的な条件を指定する地図独特のビジュアルな検索がある[69]。

情報検索の側面から考えた場合、地図アプリケーションは検索結果の一覧性において優れているものの、地図上で可能な操作が表示された情報の直接選択が中心となるため、Webページ間のリンクをたどって情報を探す拡張Web空間の検索のような、複雑な条件指定が困難な側面もある。

4.2 地図を用いたWeb検索インタフェース

4.2.1 基本的な考え方

本研究では、都市に関する大量の Webページを効率的に検索できるようにするため、Webページの検索に地図を利用した多角的な検索機能を組み入れた地図インタフェースを提案する。本地図インタフェースでは、地理的な属性の評価を Webページの検索に取り入れることで、地図上に表示する 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する。検索結果はクライアントサイドイメージマップのように、地図上にアイコンとして表示され、それぞれをマウスでクリックすることによって検索された Webページを表示することができる。地理的な属性評価を取り入れた Webページの検索と、地図を利用した検索結果へのリンク機能により、地図インタフェースを都市情報へのハブとして利用することが可能となり、その結果実世界に即した情報利用が実現できる。

4.2.2 近傍検索

地図インタフェースの近傍検索では、地図上の任意の点の近くにある Web ページを検索することができる。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。また、リストボックス中の検索結果をダブルクリックすると、その Web ページが表示される。図 4.1 は、表示している地図イメージの中心付近を指定して近傍検索を実行した結果である。

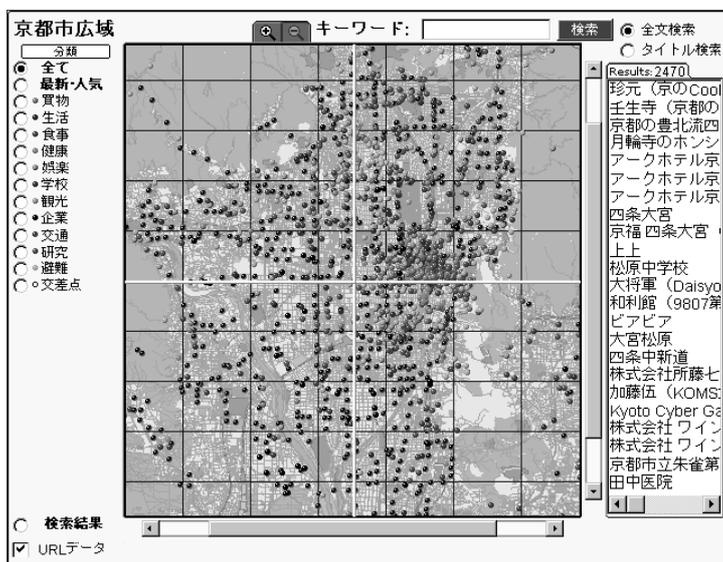


図 4.1: 近傍検索の実施例

近傍検索を使うと、ユーザは Web ページ間の距離的な関連性を利用して情報検索を行うことができる。たとえば、建物や交差点といった地図上に存在するオブジェクトを中心にその周辺の Web ページを検索したり、検索結果として見つかった Web ページから再度近傍検索を行って検索した情報の関連情報を検索したりすることができる。

4.2.3 カテゴリ検索

カテゴリ検索では、画面左側に表示されている分類から希望の分類を選択して検索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る絞り込みを行うことができる。検索結果は地図上にアイコンとして表示され、結果全体を検索した分類に属する Web ページの分布として捉えることもできる。

図 4.2は、「食事」カテゴリを選択した検索例である。この検索結果から「食事」に関する情報の分布傾向が分かり、京都市内の飲食店が市の中心部に集中していることが予想できる。また、地図中に点で表示されているアイコン1つ1つが検索結果を示しており、アイコンをクリックするとその検索結果に関連した Web ページが表示される。

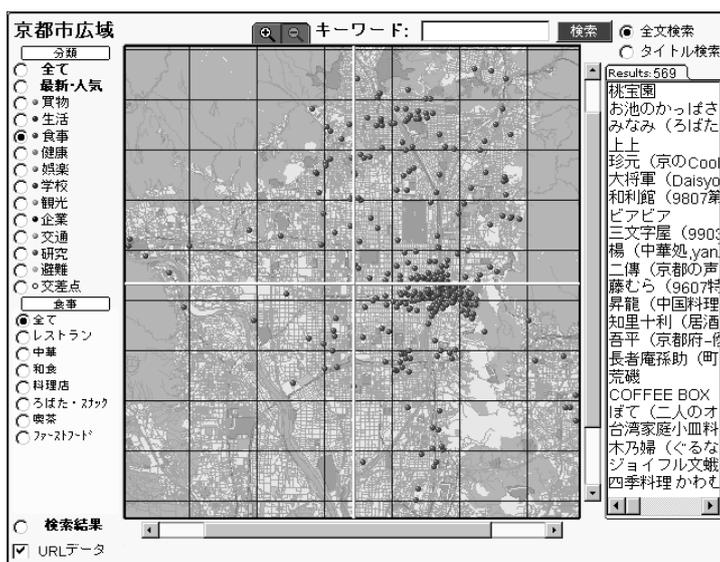


図 4.2: 飲食店カテゴリの検索結果

4.2.4 キーワード検索

キーワード検索は、画面上部のテキストボックスから入力されたキーワードに基づき、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それぞれの地理的な属性を対象に検索を実行する。Webページの内容だけでなく地理的な属性も検索対象にすることにより、住所や建物名といった地図と関連の強い地理的な表現を利用したキーワード検索も可能になる。キーワード検索の結果もカテゴリ検索の結果と同様に、地図イメージ上にそれぞれアイコンとして表示される。

図 4.3は、「バス停」をキーワードとして行ったキーワード検索の結果である。この結果の分布を見ると、京都市内のバス停が主要な通り沿いに配置されていることが分かる。

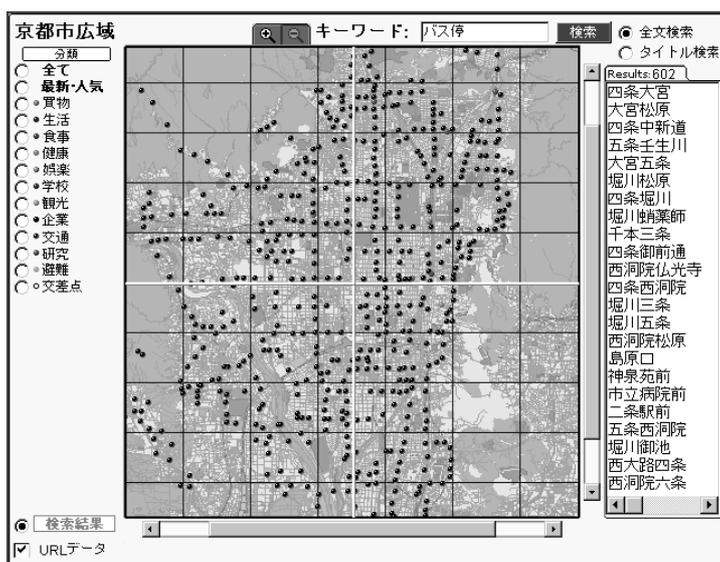


図 4.3: 「バス停」の検索結果

4.3 Javaアプレットを用いた地図インタフェースの構築

4.3.1 構成

本研究で実装した地図インタフェースは、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。地図インタフェースとサーバ上の GeoLinkなどを含めたシステムの全体構成を図 4.4 に示す。地図インタフェースは、Javaアプレットで構築することによって、Webページの表示と同様の手順で起動できるようにしている。地図インタフェースと GeoLinkは、3層クライアント・サーバ構成に従い、Webサーバ上の CGIプログラムを中継して通信を行う。

なお、地図インタフェースを Javaアプレットで実装するため、Javaアプレットのセキュリティ上の制約から、Webサーバとデータを中継するための CGIプログラムを同じホスト上で動作させる必要がある。

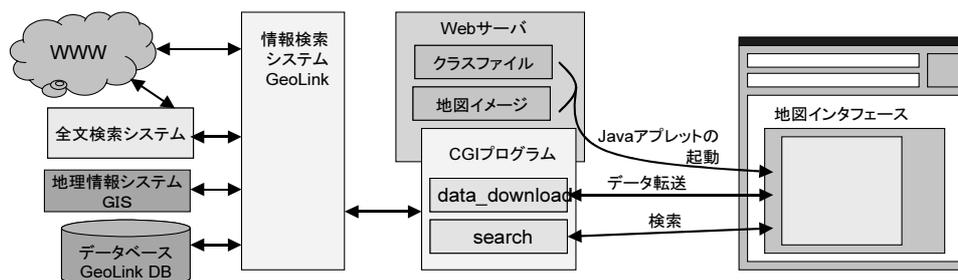


図 4.4: 情報検索システム GeoLinkと地図インタフェースの構成

地図インタフェース

地図インタフェースは、起動手順の簡略化と、クライアント側で能動的な検索処理や通信を行うことを可能にするために、Javaアプレットで構築している。また、地図インタフェースはユーザからの検索要求を受け付け、その検索結果を地図イメージ上にアイコンとして重ね合わせて表示する（図 4.1 参照）。この際、地図インタフェースは Webサーバより所定範囲のラ

スタ形式の地図イメージをダウンロードし、地図表示に利用する。このラスタ形式の地図イメージは、国土地理院発行の数値地図2500（空間データ基盤）から生成した。また、地図インタフェース上の地図の表示領域よりも大きな地図イメージを利用し、Javaアプレットのオフスクリーン描画機能とあわせて利用することによって、拡大・縮小やスクロールなど、ユーザの行う操作に対する反応性の向上や、地図イメージのダウンロード回数の削減を実現している。

情報検索システム GeoLink

情報検索システム GeoLinkは、Webページを対象とした全文検索システムとGIS、WebページのURLとタイトル、カテゴリ、地理座標などの属性を表形式で蓄積したデータベース GeoLink DBで構成する。なお、GeoLink本体はPerlで実装し、全文検索システムにはWWW全文検索システム Namazu[70] を利用し、GeoLink DBの管理にはPostgreSQL[65] を利用している。

GeoLinkは、CGIプログラム経由で受け付けた地図インタフェースからの問い合わせに応じて、GeoLink DBを対象とした検索か、全文検索システムを利用したキーワード検索のいずれかを実行する。そして、検索結果のWebページのURL、タイトル、カテゴリ、地理座標などの属性を表形式にまとめ、地図インタフェースへ返す。

4.3.2 検索処理の分散

地図インタフェースは、利用する検索機能に応じて、以下のようにクライアント側とサーバ側とで処理を振り分け、負荷分散と処理の高速化を図る。このうちクライアント側で行う検索は、動作中クライアント側に転送したデータをキャッシュとして利用することで、検索処理の高速化を実現する。

1. 近傍検索

地図イメージに重ね合わせて表示するアイコンの座標を活用して、表示している情報を対象にクライアント側で実行する。クライアン

ト側に転送した情報を利用して処理を行うため、高速に検索を実行することができる。地図インタフェースに実装した近傍検索では、アイコンの座標を利用するため精度は若干落ちるが、地図上に表示されたリンクを検索する上ではほとんど問題ない。

2. カテゴリ検索

起動後のデータ転送状況に応じて、キャッシュしている情報に関してはクライアント内で検索を実行、キャッシュしていない情報に関しては CGIプログラム経由で GeoLinkへアクセスして検索を実行する。

3. キーワード検索

クライアント側だけで行うタイトル検索と、GeoLinkと連携して実施する全文検索の2段階の検索を並行して実行する。第1段階のタイトル検索では、アイコン情報としてキャッシュしているデータを活用し、クライアント側の地図インタフェース内で検索を実行する。第2段階の検索は、CGIプログラム経由で GeoLinkに問い合わせ、後述のGeoLink DBに登録されている Webページを対象とした全文検索と地理的な属性に関する検索を実行する。第1段階の検索が、キーワードをタイトルに含む関連性の強い結果を即応的に出力するのに対し、第2段階では、検索対象となる Webページの属性を網羅的に扱うことができる。

4.3.3 性能評価

本節では、デジタルシティ京都において公開実験を行った際の地図インタフェースと GeoLinkの構成について説明し、実際に公開しているデータを利用して地図インタフェースと GeoLinkの動作性能を評価する。

評価環境

デジタルシティ京都では、京都市内に関連する Webページを検索するサービスをインターネット上に一般公開しており、その検索システムおよびインタフェースに、本研究の GeoLinkおよび地図インタフェースを利用して

いる。デジタルシティ京都では、Webページに関する所有権やプライバシーといった社会的問題を解決するため、検索対象となるホームページの所有者それぞれに対して使用確認を行い、使用許諾が得られたホームページを4.3.1節で述べたGeoLink DBに蓄積して、情報検索システムを公開運用している。なお、性能評価時にGeoLink DBに蓄積されているWebページは、約2,600件であった。

公開用の地図インタフェースでは、京都市の広域地図イメージ1枚(図4.1参照)と東西2km南北1.5kmの拡大地図イメージ108枚(図4.5参照)で京都市内をカバーする。なお、図4.5の中で表示されている地図イメージは、拡大地図イメージの一部、500m x 500mの範囲である。

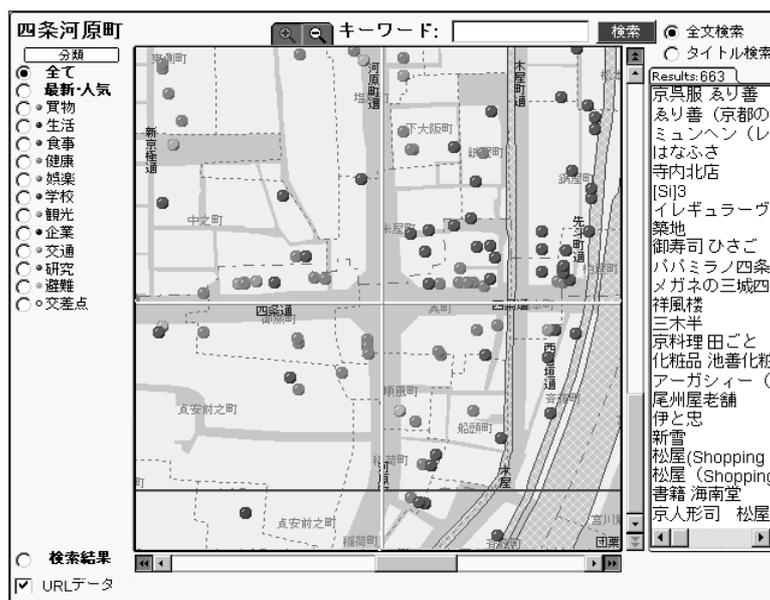


図 4.5: 拡大地図イメージに切替えた地図インタフェース

動作性能

地図インタフェースの動作速度の測定は、表 4.1 に示したデジタルシティ京都の Webサーバを設置しているイントラネット内において実施した。

表 4.1: 地図インタフェースの動作速度の測定環境

測定環境	
サーバ	SunEnterprise250, UltraSPARC300MHz×2, 512MB, Solaris2.6, Webサーバ (apache1.3.9)
クライアント	DELL Precision210, PentiumII450MHz×2, 256MB, WinNT4.0, Webブラウザ (InternetExplorer5.0)
ネットワーク	100BASE-TX

まず地図インタフェースの起動時間を計測したところ、平均 6.01秒であった。この起動時間には、地図インタフェースを表示するためのHTMLファイルと其中で指定されている Javaアプレット (約105KB) のダウンロードと、Java仮想マシンの初期化と地図インタフェースの初期化、そして初期画面として設定されている「京都市広域、分類：全て」を表示するための地図イメージ (約58KB) と、その範囲内に存在する Webページの座標、タイトルおよびカテゴリ情報 (約161KB、2453件) のダウンロード時間が含まれる。

次に、地図インタフェースに実装した検索機能のうち、近傍検索とカテゴリ検索の所要時間の測定結果を表 4.2 に示す。Webサーバのアクセスログに記録された利用者の検索履歴から検索結果がゼロでないものをランダムに 100件選び、地図インタフェースの時間関数を利用して各検索機能の所要時間を測定した。検索の平均所要時間は、近傍検索が 45.6ms、カテゴリ検索が 121.8msであった。また、比較対象として作成したサーバ側ですべての検索処理の実行するキャッシュなしの地図インタフェースでは、近傍検索の所要時間が 9461.3ms、カテゴリ表示の所要時間が 7145.6msであった。この比較から、キャッシュの利用によって検索の所要時間が大幅に短縮できていることが分かる。また後述の公開実験における検索履歴を

表 4.2: 検索機能の所要時間 (1)

検索機能	平均所要時間 (ms)	
	キャッシュあり	キャッシュなし
近傍検索	45.6	9461.3
カテゴリ検索	121.8	7145.6

表 4.3: 検索機能の所要時間 (2)

検索機能	平均所要時間 (ms)	
	タイトルデータ検索	全文検索
キーワード検索	87.9	909.2

見ると、近傍検索およびカテゴリ検索の 97.6% がクライアント側にキャッシュされた情報を利用して実行されており、実際の利用状況からもキャッシュが有効に機能することが確認された。

最後に、表 4.3 にキーワード検索の平均所要時間を示す。キーワード検索についても同様に、Webサーバのアクセスログに記録された利用者の検索履歴から検索結果がゼロでないものをランダムに 100件選び、地図インタフェースの時間関数を利用して所要時間を求めたところ、第 1 段階のタイトルデータを対象とした検索の終了までが 87.9ms、第 2 段階の全文検索の終了までが 909.2msであった。第 1 段階のタイトル検索は、検索対象は狭いものの、他の検索機能とほぼ同等の即応性を実現している。

4.4 位置情報を用いた携帯端末用地図インタフェースの構築

4.4.1 Mobile InfoMap の構成

前節の Java アプレット版の地図インタフェースの場合は、広い範囲にわたる検索結果の分布を比較的広い画面に表示することができたが、携帯情報端末の場合は、表示画面の大きさや解像度の制限からその表示範囲は非常

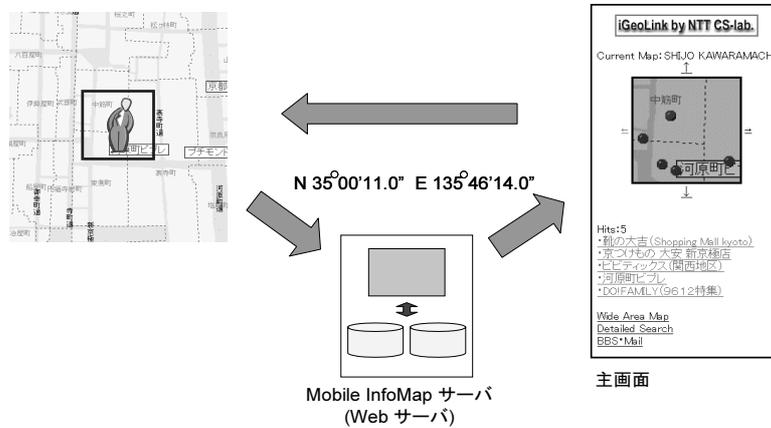


図 4.6: Mobile InfoMap の概要

に制限される．この制限を鑑みて，本節では，測位機能を備えた携帯情報端末用の地図インタフェース Mobile InfoMap (MIM) を構築する．この MIM では，Global Positioning System (GPS) による測位機能を用いて「自分」を中心にした情報検索を自動的に行うことを可能にする．これにより，屋外にいるユーザが携帯情報端末から情報検索を行う場合も，ユーザの現在位置と検索結果の関係が理解しやすくなると考える．

MIMを利用した情報検索は，次のように行われる．まず，街中にいるユーザ（図 4.6 左）が携帯情報端末から MIMを呼び出すと，ユーザの現在位置が測位され，その位置情報を含んだURLにより MIMサーバ（図 4.6 中央下）上の CGIプログラムが呼び出される．そして，CGIプログラムが生成した主画面（図 4.6 右）が MIMサーバからユーザの持つ携帯情報端末に送信され，検索結果として画面に表示される．この主画面には，ユーザの現在位置を中心とした $100m$ 四方の地図と，現在位置から近い順に 5 件の Web ページがリスト形式で表示される．このリストに表示された検索結果のうち地図の範囲内にあるものは，アイコンとして地図上にも表示され，アイコンをクリックすると Web ページが別ウインドウに表示される．

また主画面には，詳細検索，広域地図，BBSやメールなどのコミュニケー

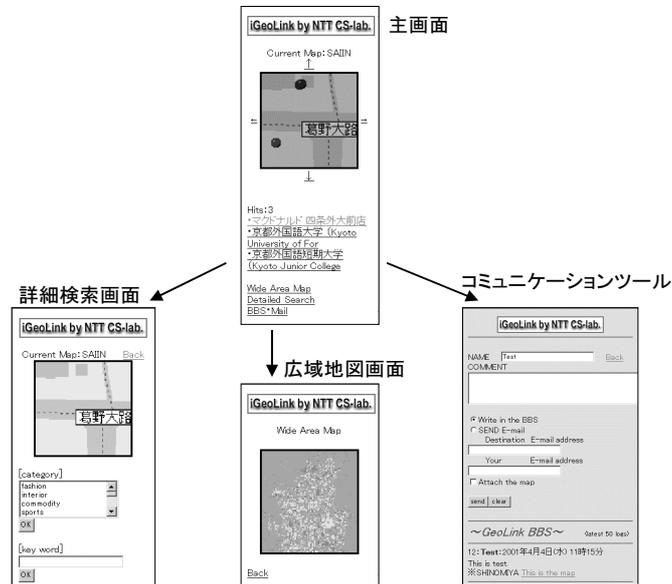


図 4.7: Mobile InfoMap の検索画面

コミュニケーションツールへのリンクも表示される（図 4.7）。このうち詳細検索のページでは、Javaアプレット版の地図インタフェースと同様に、指定範囲内の Web ページを対象にカテゴリ検索とキーワード検索を行うことができる。詳細検索画面の検索結果は、主画面と同様の形式で携帯情報端末の画面上に表示される。なお MIM では、利用する携帯情報端末では、Javaアプレット版の地図インタフェースのようにクライアント側でキャッシュを利用した検索を行うことはできないので、カテゴリ検索とキーワード検索は全て MIM サーバ側で行われる。

コミュニケーションツール（図 4.8）では、メッセージを BBS 形式で残すか、メールとして送信することができる。このメッセージには、ユーザの現在位置を指定した MIM サーバの CGI プログラムへのリンクが添付される。従って、コメントやメールを見たユーザが地図へのこのリンクをクリックすると、発信者の発信時の位置を含む地図が MIM サーバで生成され、ユーザの端末に表示されることになる。

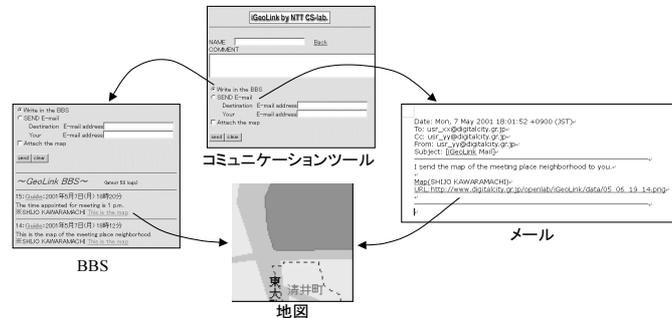


図 4.8: コミュニケーションツールの画面

なお、広域地図へのリンクを選択すると、主画面で表示した範囲を含む広域地図が表示されるが、携帯情報端末の画面の解像度などを考慮し、検索機能等は一切付加していない。

4.4.2 システム構成

図 4.9 に MIM のシステム構成を示す。MIM は、クライアントである携帯情報端末と位置情報サービスセンタ、そして Webサーバを用いた MIMサーバから構成される。

MIMサーバは、4.3.1 節で述べた情報検索システム GeoLinkに、地図イメージ編集機能を付加したものである。MIMサーバは、クライアントの現在位置と検索条件を受け付けると、GeoLinkの機能を利用して、現在位置周辺に存在する Webページをその地理座標とともに検索する。この際、現在位置周辺の地図イメージの切出しもあわせて行う。そして、検索結果の Webページのアイコンを地図イメージに重ね合わせて、図 4.7 の主画面を作成する。

クライアントは、携帯電話網を利用したダイアルアップ接続により、位置情報サービスセンタとインターネット接続プロバイダにアクセスする。このダイアルアップ接続の通信速度は 9600bps である。このうち、クライアントと MIMサーバの間の通信は、インターネット接続プロバイダ経由

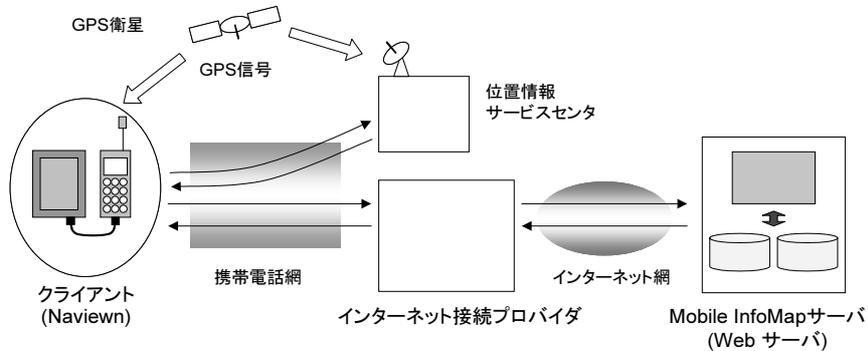


図 4.9: Mobile InfoMap のシステム構成

で行われ、インターネット接続プロバイダと MIMサーバの間は、通常のインターネット網を経由した接続となる。

図 4.10 に MIM のクライアントとして利用した測位機能付き携帯端末 Naviewn を示す。Naviewn は、NTTドコモ社製の WindowsCEベースの携帯情報端末である。端末の大きさは幅 82mm 高さ 153mm 厚さ 18mm、重さは 220g である。表示画面は 240 × 320 ドットサイズの白黒 16階調の液晶画面で、携帯電話接続用のモデムと拡張GPS方式による位置計測機能を持つ。また、組み込みソフトとしてWebブラウザ¹を備えている。

本システムのクライアントの動作を以下に示す。

1. 携帯情報端末の位置の計測

MIM のクライアントは、位置情報センタと通信しながら、拡張GPS方式により自分の位置を取得する。拡張GPS方式は、SnapTrack社 [71]が開発した GPS測位方式であり、GPS衛星からの信号を携帯情報端末と位置情報サービスセンタの双方で受信し、各種データを交換しながら、携帯情報端末の位置を計算する。この際、ディファレンシャルGPSによる測位精度の補正が位置情報サービスセンタで行われるため、GPS受信機だけで測位を行う従来方式よりも、高感度

¹HTMLバージョンは 3.2ベース、HTTPバージョンは 1.1



図 4.10: 測位機能付携帯端末 Naviewn と携帯電話

で高精度の測位が可能となっている。

2. 周辺情報の検索

携帯情報端末は、測位した現在位置を MIMサーバに送り、周辺情報の検索を実施する。この周辺情報の検索機能は、前節で説明した通りである。なお携帯情報端末は、現在位置の測位後に、ダイヤルアップ接続先を位置情報サービスセンタからインターネット接続プロバイダに切り替え、MIMサーバに接続する。

4.4.3 性能評価

本節では、京都市内の新京極近辺で行った MIMの検索実験の結果と、その際に計測した処理時間について説明する。実験は、表 4.1のサーバとデジタルシティ京都で公開しているGeoLink DBを利用して行った。

まず、実験を行った場所を図 4.11左に示す。携帯情報端末の位置として計測された座標は、北緯 $35^{\circ}00'11.0''$ 、東経 $135^{\circ}46'14.0''$ であり、ほぼ正確な座標を示した。また、この時に表示された主画面を図 4.11 右に示す。主画面の地図イメージをみると、ユーザの現在位置を中心とした 100m四方の地図イメージに 5件の Webページがアイコンとして地図上に表示されていることがわかる。従って、この実験を行った場所では、リスト形式で

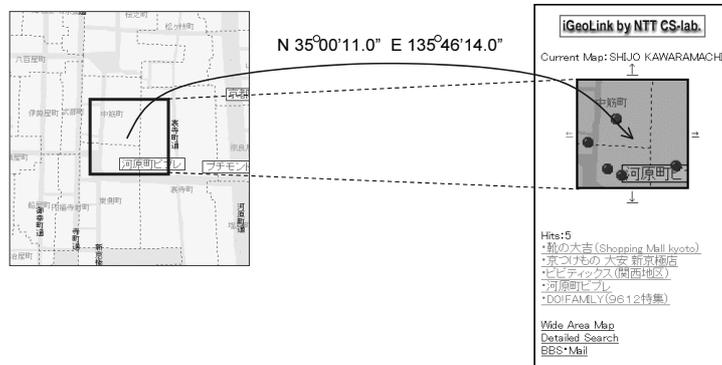


図 4.11: 位置情報に基づいた主画面

表 4.4: 図 4.11の検索結果の詳細

タイトル	緯度	経度
靴の大吉	35°00'10.299"	135°46'13.032"
京つけもの 大安 新京極店	35°00'10.476"	135°46'13.511"
ビビティックス (関西地区)	35°00'10.601"	135°46'15.003"
河原町ビブレ	35°00'10.601"	135°46'13.003"
DO! Family	35°00'10.601"	135°46'13.003"

表示される 5件の Webページと、地図イメージ上にアイコンで表示された Webページが一致していたことになる。この表示された 5件の Webページの詳細を表 4.4 に示す。

次に、この検索の際のクライアントの処理時間の内訳を図 4.12に示す。クライアントの処理時間は大きく分けて次の 3つに分けられる。

1. 位置計測セッション

MIMが呼び出されると、クライアントはまず、位置情報サービスセンタにダイヤルアップ接続を行い、相互に通信しながら携帯情報端末の位置計測を行う。このセッションで消費した時間は、約 23秒であった。

2. 内部処理

位置計測セッション後，位置計測の終了処理と情報検索処理の起動処理のためのタスクスイッチが行われる．この内部処理で消費した時間は，約 41 秒であった．

3. 情報検索セッション

内部処理終了後，クライアントはインターネット接続プロバイダにダイヤルアップ接続を行い，計測した位置情報を引数として MIMサーバ上の CGIプログラムを呼び出す．MIMサーバは，位置情報に基づき周辺情報を検索し，検索結果を合成した地図イメージを含む主画面をクライアント上の Webブラウザに返送する．この検索セッションが開始され，携帯情報端末に主画面が表示されるまでにかかった時間は約 36 秒であった．

以上，検索処理時間の合計は約 100 秒であった．このうち，最も消費時間が大きかったのは，タスクスイッチに関連した内部処理である．WindowsCE を OS とした携帯情報端末では，タスクスイッチに必要な時間が携帯情報端末の計算性能に大きく影響をされることが知られている．本実験の場合の内部処理時間も，Naviewn の CPU の性能に影響されたものと考えられる．また，主画面を構成する HTML ファイルのサイズは約 1KB，地図イメージのサイズは約 0.5KB と小さかったが，このダウンロードには約 26 秒かかった．その他，ダイヤルアップ接続を確立するための時間や，測位処理も全体の処理時間に大きな影響を与えていた．この測位処理時間については，GPS 衛星の位置や天候の影響を受ける可能性もある．

4.5 地図を用いた Web 検索行動の分析

本研究の地図インタフェースでは，地図イメージ上に表示された範囲の Web ページ対象に，カテゴリ検索やキーワード検索を行うことができる．またその検索結果は，それぞれの Web ページの内容に対応する位置にアイコンで表示される．この際，地図インタフェースのユーザは，Web ページの内容とその地理的な位置を参照しながら，検索目標を探していると考えられる．従って，地図インタフェース上でのユーザの Web ページの選択，

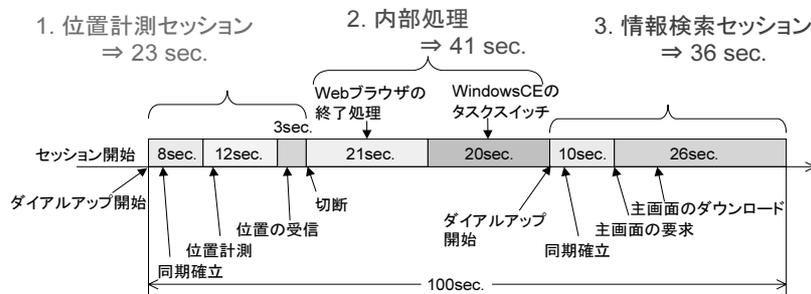


図 4.12: 検索処理時間の内訳

検索されたキーワードや選択されたカテゴリは，単にインターネット上の Web ページの検索過程を表すだけでなく，都市空間を意識したユーザによる都市に関連した Web ページの検索過程を表すと考えられる．

そこで，インターネット上に公開した Java アプレット版地図インタフェースの利用状況の分析を行い，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分析する．この利用状況の分析は，地図インタフェースに関わるアクセスログに基づいて行うが，アクセス頻度などの各種統計情報や時間軸を考慮して可視化されたユーザ検索行動は，地図インタフェース上の地図イメージで表示された都市空間において，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．また，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のフィードバックとしてだけでなく，都市に関連する Web ページの情報発信戦略を決定する際にも有用と考えられる．

4.5.1 検索行動を記録するための設定

地図インタフェースの利用状況は，地図インタフェースのバックエンドである Web サーバのアクセスログ (図 4.13²) に記録される．この際，アク

²図 4.13, 図 4.14 中のホスト名はプライバシー保護のために変更を施している．

```

client.abc.co.jp - - [04/Jun/2002:14:11:49 +0900] "GET
/cgi/map.cgi?url=http://www.city.kyoto.jp/kotsu/tikadia/hyperdia/
021900.htm&map=kyoto2s_585x540.gif&category=new&seq=1020&listbox=yes
HTTP/1.0" 302 0 "-" "Mozilla/4.0 (compatible; MSIE 5.01; Windows NT
5.0)" client.abc.co.jp.119091023085603623

client.abc.co.jp - - [04/Jun/2002:14:12:06 +0900] "GET
/cgi/geolink_content.cgi?TB=geolink_000918&no=2053 HTTP/1.0" 200 204
 "-" "Mozilla/4.0 (compatible; MSIE 5.0; Win32)"
client.abc.co.jp.119091023085603623

client.abc.co.jp - - [04/Jun/2002:14:12:06 +0900] "GET
/cgi/map.cgi?url=http://gnavi.joy.ne.jp/kansai/JP/K023000s.htm
&map=kyoto2s_585x540.gif&category=new&seq=2053&listbox=yes
HTTP/1.0" 302 0 "-" "Mozilla/4.0 (compatible; MSIE 5.01; WindowsNT
5.0)" client.abc.co.jp.119091023085603623

```

図 4.13: Webサーバに記録されるアクセスログの例

```
client.abc.co.jp.119091023085603623
```

図 4.14: Webサーバが発行するクッキーの例

セスログ上でのユーザ識別のために、Webサーバがユーザの利用する Web ブラウザに対してクッキー（図 4.14）を配布するように設定している。

ユーザによる地図インタフェースの操作は、以下のような形でクッキーとともに Webサーバのアクセスログに記録される。ただし以下の方法は、地図インタフェース上での各種操作が Webサーバ上のアクセスログに記録されるようにしたものであるため、検索結果として選択された Web ページを閲覧した後の操作、たとえば、閲覧した Web ページからハイパーリンクをたどって他の Web ページを閲覧する操作等は、アクセスログに記録されない。

- 地図インタフェースの起動
地図インタフェースを起動する際の Java クラスファイルのダウンロードが、アクセスログに記録される。
- 検索機能の利用
地図インタフェースのカテゴリ検索およびキーワード検索は、クラ

クライアント側にキャッシュされている情報によりその実施方法が異なるが、いずれの場合も、Webサーバ上の CGIプログラムを呼び出し、検索条件が CGIプログラムの URLと一緒にアクセスログに記録されるようにした。もし、クライアント側だけで検索処理が可能な場合は、その検索条件を引数に含めたダミー CGIプログラムが呼び出され、その呼び出しがアクセスログに記録される。また、クライアント側だけで検索処理ができない場合には、通常の CGIプログラムの呼び出しがアクセスログに記録される。

- 検索結果の選択

検索結果として地図イメージ上に表示された Webページの選択は、クライアント側の Webブラウザ上で行われるが、リダイレクト情報を返す CGIプログラムを利用することで、Webサーバ上のアクセスログにその選択が記録されるようにした。具体的には、地図インタフェース上のアイコンのリンク先を実際の URLではなく、その URLを引数としたリダイレクト CGIプログラムとし、ユーザがアイコンをクリックして CGIプログラムが呼び出されると、引数として与えられた URLへのリダイレクト情報が設定されたメタタグが、クライアントの Webブラウザに返されるようにした。

上記の設定により、地図インタフェース上の操作は、バックエンドの Webサーバのログファイルに記録されるようになるが、ログファイル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。このため、地図インタフェースに関連した操作を抽出するためには、まずクリーニング処理が必要となる。このクリーニング処理は、ログファイルに含まれるテキストのパターンマッチングにより実施する。

さらに、2人以上のユーザが同時に地図インタフェースを利用している場合には、その操作が同一ログファイル上に並行して記録されるので、ユーザごとに地図インタフェースの起動から終了までの操作をまとめるトランザクション抽出処理が必要となる。なお、地図インタフェースの設定上、明確な終了操作は記録されないので、一定時間³以上操作間隔が空いた時点

³本分析では 2時間とした。

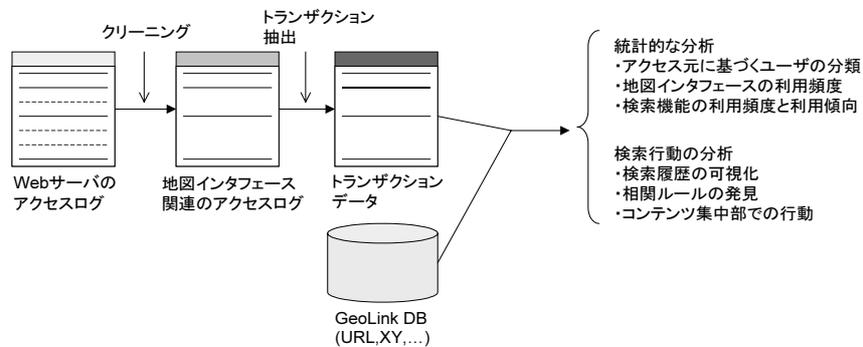


図 4.15: 検索行動分析手順の概要

でトランザクションの終了とした。また、同一ユーザの操作中に地図インタフェースのクラスファイルの再読み込みが記録された場合には、地図インタフェースが再起動されたものとして、トランザクションの分割を行った。なお、このトランザクションの抽出はログファイルに記録された時刻とクッキーに基づいて行うため、Webブラウザの設定等によりクッキーが正しく配布されない場合には、トランザクションの抽出に失敗する場合もある。このため、トランザクションを考慮した利用状況の分析では、トランザクションの抽出が成功した場合だけを対象に分析を行っている。

これらの処理を行ったアクセスログに GeoLink DB中の地理座標を付加することで、地理属性を考慮した検索行動の分析が可能となる。以上の検索行動分析の流れを図 4.15 に示す。

4.5.2 アクセスユーザの特徴

本研究の Java アプレット版地図インタフェースの公開運用は、インターネット上に公開されたデジタルシティ京都の実験サイトにおいて行った。この実験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ったことも確認している。Webサーバのアクセスログに記録されたクッキーにより、

表 4.5: 国別アクセス
件数

国名	件数
日本	25055
米国	1920
フランス	113
ドイツ	100
スロバキア	56
その他	350

表 4.6: JP第二ドメイ
ン別アクセス件数

ドメイン名	件数
ne	12750
co	3674
ad	2617
or	2539
ac	2472
その他	1003

表 4.7: 県別アクセス
件数

県名	件数
京都	5219
大阪	1037
東京	994
愛知	411
滋賀	375
県名判明分：	11521

地図インタフェースの利用者数と利用回数を集計したところ，1999年 12月
から 2001年 11月までの約 2年間で 32,120人からの利用があった．

クッキーに含まれるクライアントホスト名に基づき，ドメイン別，地域
別にアクセス元を分類したデータを表 4.5, 4.6, 4.7 に示す．まず，国名な
どを表す第一ドメインでアクセス元を分類すると，クライアントホストのド
メインが判明したアクセス 27594件のうち，JPドメインからのアクセスが
25,055件（90.8%），JPドメイン以外からのアクセスが 2,539件（9.2%）
であった．このうち，JPドメイン以外からのアクセスは主に海外からのア
クセスと考えられる．第一ドメインに基づき国別に集計すると，多い方か
ら順に米国⁴，フランス，ドイツとなり，計 51ヶ国からのアクセスが記録
されていた（表 4.5）．

一方，JPドメインからのアクセスを第二ドメインにより分類すると，neド
メインからのアクセスが 12,750件で全体の 50.9%を占めていた（表 4.6）．
また，クライアントホストの所在地が明確な学校，サービス範囲が限定さ
れているケーブルインターネットのプロバイダ，およびダイヤルアップ時
にプロバイダが割り当てるホスト名から地域が限定できるドメインに限定
してアクセス元を調査したところ，JPドメインからのアクセスの 46.0% を
県別に集計することができた．この集計結果を表 4.7に示す．この集計に
よると，京都府内からのアクセスが全体の 45.3%を占め，次いで大阪，東
京，愛知の順となっていた．

⁴この集計では，com, org, edu, net, mil, gov, us ドメインを米国と分類した．

表 4.8: 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年 2月 4日から 2000年 3月 23日

4.5.3 地図インタフェースの利用頻度

まず，2000年 2月 4日から 2000年 3月 23日の 50日間に，地図インタフェースを利用したユーザ数を表 4.8に示す．この期間中に 8,110回のアクセスがあり，2,524人が地図インタフェースを利用した．このうち，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回の利用があった．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．

4.5.4 検索機能の利用傾向

検索機能それぞれの利用人数とその頻度を表 4.8下段に示す．近傍検索の利用者は 411人で頻度が 2.68，カテゴリ検索の利用者は 1,128人で頻度が 5.82，キーワード検索の利用者は 594人で頻度が 3.75 であった．3つの検索機能のうちでは，カテゴリ検索が一番多く利用され，その利用頻度も

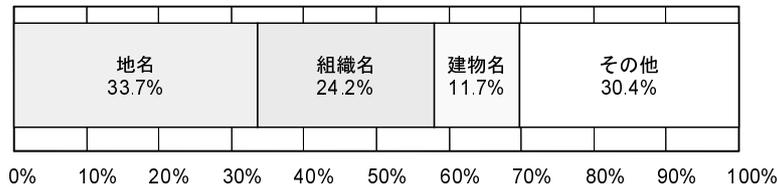


図 4.16: 検索キーワードの傾向

対し、「ホテル」や「ラーメン」といった一般名詞やそれら名詞の一部分など、その他の表現を利用した検索は 30.4%であった。この結果から、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を利用して検索を実施する傾向があったといえる。

2. カテゴリの選択傾向

地図インタフェースで利用されたカテゴリの割合をみると、食事、買物、交通、観光といったカテゴリの利用率が高く、緊急避難、研究、健康、企業といったカテゴリの利用率が低かった。このカテゴリの利用傾向は、Yahoo! JAPANのアンケート結果[21]と同様に、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、実験を行った場所柄の影響により、実際に京都に行ったときに役立つようなカジュアルな情報カテゴリの利用も顕著であった。

また、クライアントホストのドメインに基づき京都のユーザと京都以外の日本国内のユーザに分け、トランザクション中のカテゴリ検索の回数をカテゴリごとに集計したものが図 4.18、選択されたWebページ数をカテゴリごとに集計したものが図 4.19、そして各カテゴリの検索回数と選択された Webページ数の比率を示したのが図 4.20 である。なお、集計対象とした京都のユーザによるトランザクションは 4,049、それ以外の日本国内のユーザによるトランザクションは 4,732 である。まず特徴的なのは、京都のユーザはその他の地域のユーザと比較し、観光情報へのアクセスが少なかった。これはユーザと検索対

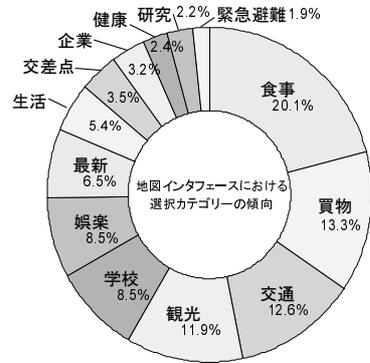


図 4.17: カテゴリの選択傾向

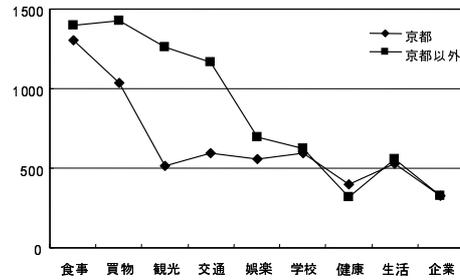


図 4.18: 各カテゴリの検索回数

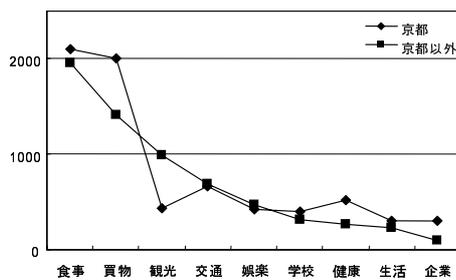


図 4.19: 選択された Web ページ数のカテゴリ

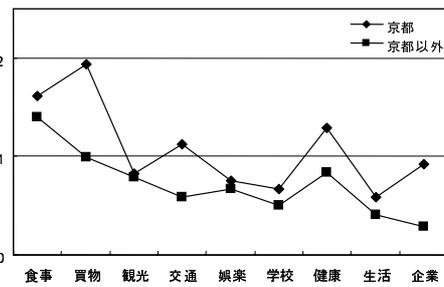


図 4.20: カテゴリ検索 1 回あたりに選択される Web ページ数

象地域の距離や、地図インターフェースの利用目的が反映された結果と考えられる。また京都のユーザは、買物情報に関する Web ページを多く選択していた。買物カテゴリの検索回数と選択した Web ページ数の比率で比較すると、京都のユーザが 1.94、その他のユーザが 0.99 であり、京都のユーザの方が検索結果として得られた情報に興味を持ち、比較を行っていたと考えられる。この傾向は、交通、健康、企業カテゴリにもみられることから、地域住民が日常生活に関連する情報を検索する場合の一つの傾向といえるだろう。

以上の傾向から，地図インタフェースの検索機能では，実際の街と密着した情報が検索できることが重要と考えられる．地図インタフェースと街の密着度は，地理的な網羅性と情報量が大きく影響する．地理的な網羅性を高めるためには，都市に関する情報を蓄積した地理情報システムとの連携が重要となる．また，検索対象となる情報量を増やすためには，Webページを対象にした全文検索システムの技術が有効である．特に，インターネット上に発信される最新情報を地図インタフェース上で検索できるようにするためには，本研究で提案した GeoLinkのように，Webページと地図情報の双方を統合して扱えるような検索システムが有効と考えられる．

4.5.5 アクセス履歴の可視化

地図インタフェースのユーザは，地図インタフェースを起動後，カテゴリ検索やキーワード検索を行い，その結果である Web ページの内容を吟味しながら，都市に関連した Web ページの検索を進めていく．この際行われる検索結果の選択には，検索結果として表示される Web ページの内容だけでなく，その地理的な位置の影響も大きいと考えられる．

そこで，各ユーザの地図インタフェース上における検索結果の選択履歴を図 4.21 のような形式のログビューアで可視化してみる．このログビューアでは，地図イメージの二次元平面に対して垂直方向の時間軸を追加した三次元の表示空間を作り，ユーザが選択した Web ページを，地図イメージ上の位置からアクセス時刻の分だけ垂直上方に球状のアイコンで表示する．この際，アイコンの地理的な位置を把握しやすくするために，Web ページを示すアイコンから地図イメージへの垂線を表示する．そして，連続してアクセスされたアイコン間を線分でつなぎ，ユーザが選択した Web ページの履歴を数珠状につなげた形で表示する．

なお，このログビューアは，Java3D API を利用して構築したものであり，同 API の機能を利用した表示空間の回転，移動，拡大縮小が可能である．また，表示空間の時間軸は，地図イメージに向かう方向を正の方向とすることで，ユーザが最後に選択した Web ページを示すアイコンが，地図イメージの最も近くに表示されるようにしている．この時間軸のスケールは，トランザクションとして記録されたユーザの検索時間に合わせて調整

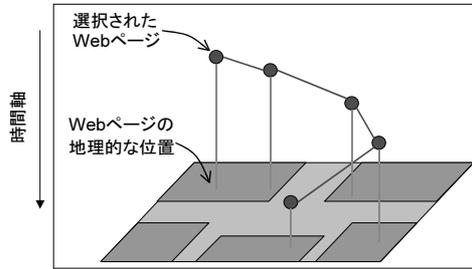


図 4.21: アクセス履歴の可視化手法

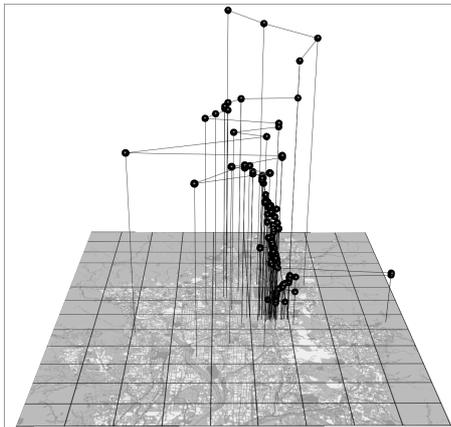


図 4.22: アクセス履歴の可視化 (1)

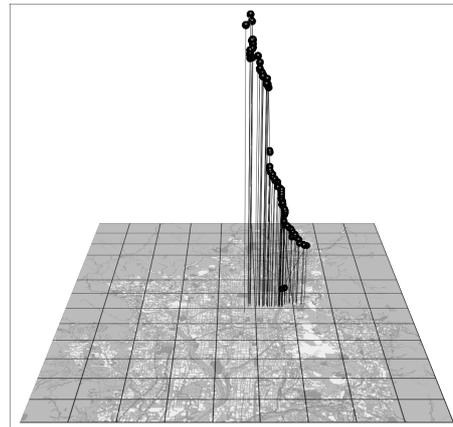


図 4.23: アクセス履歴の可視化 (2)

される。

このログビューアを利用して、アクセスログから抽出されたトランザクションのうち、比較的長時間にわたり検索が行われた特徴的な 2例を示す。

1例目の図 4.22 のトランザクションでは、1時間 42分 31秒の間に 148 件の Webページが選択された。このトランザクションは、検索範囲の地理的な広がりやアクセス時刻の間隔から、大きく前半と後半に分けることができる。図上部に表示された前半では、広い範囲で Webページの選択が行

われ，上下方向の間隔として表示されるアクセス時刻の間隔も大きい．これに対し図下部に表示された後半では，京都市四条河原町周辺の地理的に狭い範囲で Web ページの選択が行われ，アクセス時刻の間隔も短い．従って，このアクセス履歴からは，ユーザが検索範囲をだんだんと絞り込んでいく様子と，検索範囲が絞り込まれていくのに従い，Web ページの選択が地理的にも時間的にも集中して行われた様子を窺うことができる．

2 例目の図 4.23 のトランザクションでは，4 時間 36 分 18 秒の間に 93 件の Web ページが選択された．このトランザクションで選択された Web ページの分布は，地図イメージ上でみるとほぼ直線状になり，時間軸も含めた表示空間中でみると，図中の左上から右下へとつながっていることが分かる．従ってこのトランザクションでは，ユーザが地理的に西から東へと，ほぼ直線的に狭い範囲で Web ページを選択していたことがわかる．この直線は，京都市の四条通にほぼ一致していることから，ユーザが四条通周辺の情報を網羅的に検索していたことが窺える．また，時間軸方向でみると，前半と中盤に大きなクラスタがあり，その間に大きな間隔が空いている．この時間軸方向の間隔は，その期間に Web ページの選択が無かったことを示しているが，4.5.1 節で述べたように，地図インタフェース以外で行われたユーザの操作は，アクセスログに記録されないため，この期間に行われたユーザの行動は不明である．

このようにログビューアを利用すると，ユーザが連続してアクセスした Web ページの地理的な収束や拡散，時間的な集中や周期性など，アクセス履歴の時空間的な特徴を観察することができる．

4.5.6 地理的な集中とアクセスの集中の関係

前節で紹介したアクセス履歴の例では，地理的に狭い範囲に分布する Web ページが連続して選択される様子が観察された．このような検索行動が行われた原因の一つに，連続してアクセスされた Web ページ間の距離や方法など，地理的な位置関係の影響が考えられる．また，前節の例で集中的に Web ページの選択が行われた京都市四条通周辺および四条河原町周辺は，地図インタフェースから検索可能な Web ページが密集する地域でもある．従って，周辺に表示される Web ページの密度が，ユーザの検索行動に対し

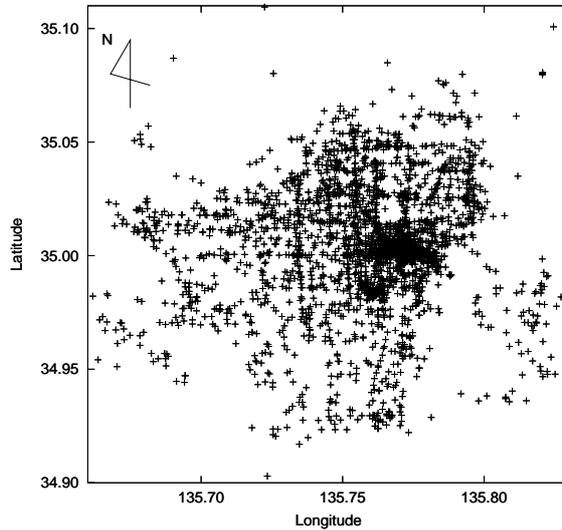


図 4.24: 検索対象の Web ページの分布

て影響を与えていたのではないかと予想される。

そこで本節では、検索対象となる Web ページの地理的な集中とアクセスの集中の関係を分析する。なお、この分析では、1999 年 12 月 10 日から 2001 年 4 月 26 日までの 504 日間を分析対象期間とした。この期間中に、地図インタフェースは 6,232 人のユーザから利用され、のべ 29,637 件の Web ページが地図インタフェース経由でアクセスされた。

まず、公開されていた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 4.24 に示す。この図 4.24 では、'+' 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。各軸の座標値は、十進形式の緯度経度で表示している。図中央やや右、(135.77, 35.00) 付近に見える 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、数多くの店舗が集中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 ページの集中がみられた。この交差点周辺の Web ページの密度は 1km^2 あたり 599 件に達していた。

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

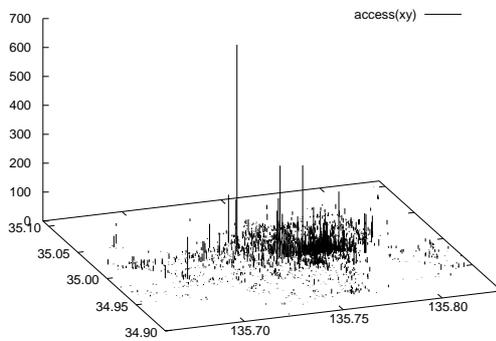


図 4.25: Webページのアクセス数

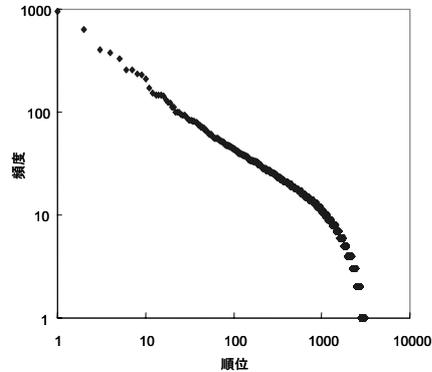


図 4.26: Webページのアクセス数とアクセス順位の関係

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が図 4.25 である。この図 4.25 から、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、図中央やや左 (135.734047, 35.028279) 付近に突出したピークとして現れている人気 Web ページへのアクセスの集中、もう一つは、図中央やや右の四条河原町交差点周辺の Web ページ集中部へのアクセスの集中である。

一つめのアクセスが集中した人気 Web ページは、対象範囲内に数件存在し、周辺の Web ページのアクセス数から突出している。このアクセスの集中は、京都市内に散在していることから、地理的な原因によるものではないと考えられる。そこで、図 4.25 のアクセス数から地図の二次元的な要素を取り除き、各 Web ページへのアクセス数とその順位の間を対数-対数グラフに表示したものが図 4.26 である。図 4.26 左上部のプロットが 100 回以上アクセスされた Web ページを示しているが、その数は非常に少ない。これに対し、グラフ右下部にプロットされたアクセス回数 10 回以下の Web ページが非常に多かった。また、グラフの中間部分是对数-対数グラフ上で直線状になっている。

このような集中傾向は、従来、地図以外の検索方法を提供するインターネットプロバイダや WWW のプロキシシステムのアクセスログにおいて観

察されてきた現象である。これまでに、冨乗則やジップの法則[17]と関連づけて論じられたり、WWWのキャッシュシステムなどの構築の根拠[72]ともされているが、同様の現象が、二次元平面上に検索結果を表示する地図インタフェースにおいても観察されたことになる。しかし、この現象をユーザレベルの検索行動として考えると、地図の有無に関わらず多くのユーザが京都市内に特定の有名な情報を探し、それをユーザが選択した結果と考えることもできるし、そのような選択に十分な検索機能を地図インタフェースが提供していたからだと考えることができる。今回観察された特定 Web ページへのアクセスの集中が、地図インタフェースのユーザのアクセス行動として何を意味しているのかは、今後、様々な視点から慎重に検証する必要があると考えている。

一方、もう一つの集中がみられた四条河原町交差点周辺では、特定の Web ページへのアクセスの集中はみられなかった。従って、特定の範囲に Web ページの集中して存在することが、アクセスの集中の原因の一つとなったと考えられる。そこで、この Web ページの集中とアクセスの集中の関係を分析するために、図 4.24 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のが図 4.27、各区画に属する Web ページへのアクセスを集計したアクセス頻度を示したのが図 4.28 である。

Web ページの密度を示した図 4.27 をみると、四条河原町交差点周辺付近に、Web ページの集中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、アクセス頻度を示した図 4.28 をみると、これまでに指摘したように、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 ページへアクセスが集中した結果であり、そ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アクセスが集中した結果であり、このエリアに属する 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が曖昧で比較的太いものとなっている。また、図

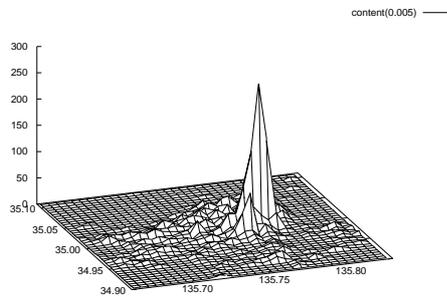


図 4.27: Webページの密度 (1)

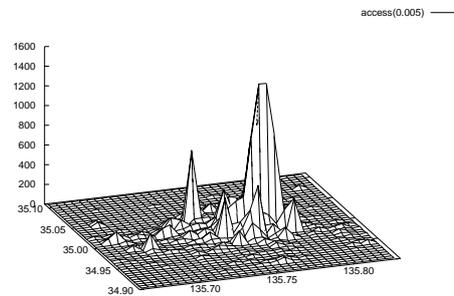


図 4.28: アクセス頻度 (1)

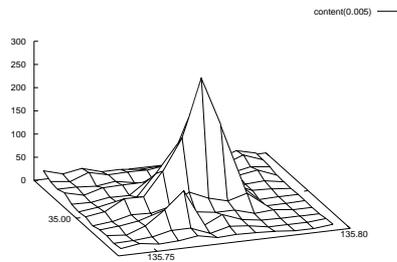


図 4.29: Webページの密度 (2)

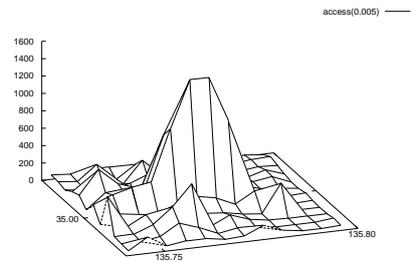


図 4.30: アクセス頻度 (2)

4.27 と図 4.28 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，中腹部も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 4.29 と図 4.30 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 4.30 におけるピークの鈍化と中腹部の広がり，Webページが集中するエリアへのアクセスが，その周辺部にも波及した結果と考えることができる．しかし，図 4.30 のアクセス頻度のピークが，図 4.29 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

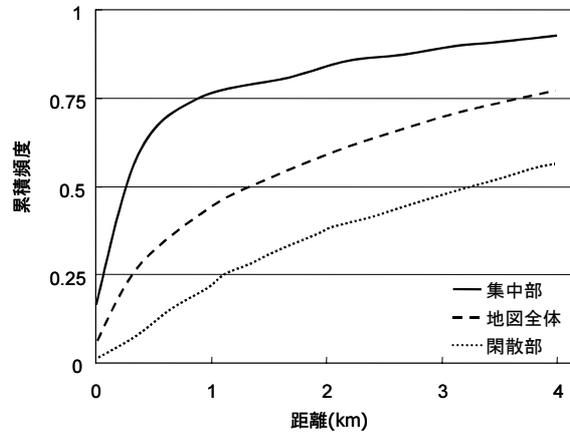


図 4.31: 連続してアクセスされた Web ページ間の距離と周辺密度

かったことを示しており，Web ページ集中部でみられた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 ページ間の距離は，4.5.1 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する。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 ページ集中部に属する場合と，山間部など Web ページが疎らに存在する閑散部に属する場合に分け比較すると，地図全体の平均値が 2.380km であるのに対し，集中部の平均値が 1.035km ，閑散部の平均値が 4.154km ，それぞれ間に有意な差がみられた⁵。この Web ページ間の距離とその累積頻度関係を示した図 4.31 をみると，集中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

⁵Microsoft Excel 2000 による 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 4.10: 連続してアクセスされた Web ページ間の平均距離

	集中部	地図全体	閑散部
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*

スのシミュレーションを行い、集中部もしくは閑散部に属する Web ページを起点とする場合のシミュレーションの結果⁶とアクセスログにおける平均値を比較（表 4.10）したところ、集中部、閑散部、地図全体それぞれで有意な差がみられた⁷。

以上の比較から、地図イメージ上における Web ページの集中がユーザの検索行動に何らかの影響を及ぼしていたと考えられる。また、多くのユーザのアクセスが集中し、そこに留まるという意味で、地図イメージ上における情報の集中が街として機能し、インターネット経由でアクセスするユーザの行動に影響を与えていたともいえるだろう。

4.5.7 人気 Web ページに関連する相関ルール

前節で、ユーザの検索行動における Web ページの密度とアクセス頻度の関係と連続してアクセスされた Web ページ間の距離について分析を行ったが、この他にも、前後に利用した検索機能や選択した Web ページから何らかの影響を受けていると考えられる。そこで本節では、高い頻度でアクセスされた人気 Web ページ 5 件に注目し、地図インタフェースに関するトランザクションから相関ルールの抽出を試みる。なお、この相関ルールの抽出には、Agrawalらが提案したアルゴリズム[73]による apriori[74]を利用した。5 件の人気 Web ページがそれぞれ頭部となる相関ルールで、ユーザがトランザクション中に検索したカテゴリとの相関ルールと、ユーザがトラン

⁶それぞれの場合について 10,000 回の試行を実施。

⁷Microsoft Excel 2000 による 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 4.11: 人気 Webページに関連する相関ルール

	頭部: 2688 (2.55%)			頭部: 7465 (2.74%)			頭部: 930 (1.32%)		
	本体	支持度	信頼度	本体	支持度	信頼度	本体	支持度	信頼度
カテゴリ	買物	0.84	2.98	買物	0.73	2.62	買物	0.73	2.57
	食事	0.62	1.86	食事	0.51	1.53	食事	0.51	1.53
	健康	0.38	4.75	健康	0.36	4.45	健康	0.40	4.98
Web ページ	930	0.82	62.59	930	0.82	61.87	2688	0.82	32.40
	7465	0.82	29.71	2688	0.82	32.03	7465	0.82	29.71
	4639	0.51	56.02	4639	0.55	60.21	4639	0.51	56.54
	頭部: 1643 (1.44%)			頭部: 1660 (1.26%)			支持度，信頼度は 百分率で表示		
	本体	支持度	信頼度	本体	支持度	信頼度			
カテゴリ	観光	0.51	2.31	観光	0.45	2.01			
	食事	0.32	0.97	交通	0.23	1.05			
	交通	0.26	1.20	買物	0.16	0.55			
Web ページ	1660	0.12	12.03	1643	0.15	10.53			

表 4.12: 相関ルールに出現した Webページ

ID	Webページのタイトル	(アクセス回数)/(利用者数)
2688	フルーツパーラー クリケット	32.03
7465	京都青果卸売株式会社	39.38
930	小坂歯科診療所	16.40
4639	楽しい農業	7.74
1643	京都御所	1.39
1660	二条城	1.28

ザクシオン中に選択した Webページとの相関ルールのうち，支持度 0.1% 以上の上位 3件を表 4.11 に示す．また，表 4.11 中の Webページの IDと対応する Webページのタイトルを表 4.12 に示す．

まず，カテゴリ検索で指定されるカテゴリと人気 Webページの相関をみると，いずれのルールも信頼度が低い．カテゴリ検索で観光が指定された

場合に、京都御所や二条城の Web ページが選択されるというルールは発見されたものの、特に相関が強いとはいえない。

一方、人気 Web ページと同一のトランザクション中にアクセスされた Web ページとの相関をみると、カテゴリ検索の場合より信頼度の高い相関ルールが抽出された。相関ルールを詳しくみると、まず、表 4.11 上段のルールに含まれる 4 件の Web ページが、相互に関連してアクセスされていたことが分かる。このグループに属する Web ページは、Web ページのアクセス回数を利用者数で割った反復利用回数（表 4.12）が多いことから、特定ユーザから繰り返しアクセスされ、しかも、高い確率でこのグループに属する Web ページがセットで選択されていたことがわかる。この 4 件の Web ページのタイトルをみると、内容的には関連が無いようにみえることから、同じ Web ページを習慣的に選択するような検索行動が特定ユーザの間に生まれていたと考えられる。

また、表 4.11 下段の 2 件の Web ページも、相互にアクセスされているが、それぞれの Web ページの反復利用回数は上段のグループに比べ少ない。しかし、下段のグループの 2 件の Web ページは、京都を代表する観光地を紹介するものであり、観光情報の検索を目的としたユーザが、同じトランザクションの中で検索する可能性は高い。従って、下段のグループの Web ページがセットでアクセスされたのは、同様の目的をもった数多くのユーザが、同様の検索行動をとった結果と考えられる。

このように、抽出された相関ルールは、その外部要因によって様々な解釈が可能となる。従って、ユーザの検索行動や情報空間の状況に基づき、相関ルールを選択することにより、地図インタフェース上の検索結果の推薦や、検索のナビゲーションが可能になると考えられる。

4.6 まとめ

本章では、都市に関する大量の Web ページを効率的かつインタラクティブに検索できるようにするために、地図を利用した多角的な Web ページ検索機能に組み入れた地図インタフェースを提案した。また、デジタルシティ京都における地図インタフェースの公開運用環境において、検索要求に対する即応性とキャッシュデータを活用するシステム構成の有効性を確認し

た。

キーワード検索だけで都市に関連する Web ページを検索する場合には、地理的な検索条件をキーワードで指定する必要があり、検索結果として得られた Web ページ間の地理的な関係もあらためて判断する必要があった。これに対し、本研究の地図インタフェースと GeoLink では、利用者のインタラクティブな地図操作も情報検索システムへの入力として捉えることで、都市に関連する大量の Web ページを効率良く検索できるようにしている。具体的には、表 4.8 のように利用される検索機能とあわせて、画面上に表示された地図の範囲を情報検索の範囲として扱うことで、地理的な条件をキーワード等で明示的に入力するのを省くと同時に、検索結果を地図上に表示することで結果間の位置関係の把握が容易になったと考えている。

また本章では、Java アプレット版の地図インタフェースの利用状況を分析し、サービスの利用頻度、検索機能の利用状況、コンテンツの集中とアクセスの集中の関係を明らかにした。このユーザの検索行動に関する分析は、地図インタフェースを改善するためのフィードバックとしてだけでなく、都市に関連する Web ページの情報発信戦略を決定する際にも有用と考えている。

地図インタフェースおよび GeoLink は、社会情報基盤の一つとして開発を行っているデジタルシティ京都の要素技術として開発したものであり、現在システムをインターネット上に公開し、運用実験を行っている。今後は、Web ページだけでなく都市のセンサ情報や各種コミュニケーションツールとのリンクによる表示情報の拡充と、系列的な検索を実施するための情報検索システムとの連携について検討する予定である。

第 5 章

考察

5.1 都市空間と情報のマッピング

本研究で提案した拡張Web空間は、デジタルシティ京都に関わる活動で生じた実際のニーズに基づいているが、Webページと地理オブジェクトを関連づける際には、以下の課題を考慮する必要がある。

1. アドレスマッピングに関する課題

3.2.2 節で述べたように、Webページと地理オブジェクトを関連づけるためには、Webページからの地理属性の抽出と、抽出した地理属性の表記と GIS中の住所表記の対応付けを行う必要がある。この際、実際の Webページや地理オブジェクトでは表記が完全に一致しない場合があるため、本研究で構築した情報検索システムでは、Webページ中の地理属性と GIS中の住所表記の間の一致条件を順次緩めながら対応付けを行うようにしている。

また現在の実装では、GIS内のオブジェクトを点として扱い、Webページとマッピングしているが、検索言語、インタフェースおよび Webページからの情報抽出方法を拡張することにより、道路や街区のような地理的な線や面によるマッピングが可能となる。この拡張により、より広範なオブジェクトの多様な地理的關係に基づいた検索ができるようになると考えている。

2. 検索対象となる情報の質

拡張Web空間に基づいた情報検索システムで検索結果の質を向上させるためには、地理的なオブジェクトと Web ページの対応付けの精度だけではなく、対応づけられる Web ページの質も考慮する必要がある。本研究で実装した情報検索システムや地図インタフェースでは、検索される Web ページを全て一様なものとして扱っているが、同じ地理オブジェクトに対応付ける Web ページでも、以下のような分類やランク付けが可能であろう。

- Web ページの内容と発信者の関係による分類
たとえば、街角の店舗に関する情報を記載した Web ページの場合、以下のような分類が可能である。
 - (a) 店舗のオーナーが自ら作成し発信している Web ページ。
 - (b) 商店街等、店舗が所属する団体が発信している Web ページ。
 - (c) 取材情報に基づいた店舗紹介を掲載した Web ページ。
 - (d) 店舗に訪れた人による紹介を掲載した Web ページ。
- 情報鮮度によるランク付け
インターネット上に Web ページが発信された時刻により、優先順位をつけることが可能である。
- 検索サービス提供者によるランク付け
情報提供者から情報配信者への報酬により、検索結果中の順位を調整する。この手法は、検索サービスを提供する Web サイトのビジネスモデルの一つとして、現実化しつつある。

このような分類やランク付けは、大量の検索結果の中で質の高い情報を選択したり、ユーザからの問い合わせに適した結果を出す際に有効に機能すると考えられる。しかし、ユーザが設定する検索条件の複雑化や、検索結果の不公平な偏向をもたらす可能性もあるため、実際の運用にあたってはコンセンサスを得るための議論が必要と考えている。また拡張Web空間に基づいた情報検索システムの場合には、最終的に検索結果として得られる Web ページと同様に、Web ページ

間のリンクや検索途中に経由する Web ページについても、その質を考慮する必要があると考えている。

3. 情報発信者との協調

デジタルシティ京都で公開した情報検索システムおよび地図インタフェースでは、プライバシーや著作権など、検索結果に関する社会的問題への配慮から、検索できる Web ページに一部制限を加えている。具体的には、Web ページの所有者から利用許諾のとれた Web ページだけが検索されるように、地図インタフェースの検索機能や地理的ジェネリックリンクの生成に制限を加えている。特に地図インタフェースの場合には、Web ページの示す情報の実際の都市での位置が明示されてしまうことから、プライバシーや犯罪等の危険性に配慮した情報提供を行う必要があると考えている。

また、4.5.4 節で示したように、地図インタフェースのキーワード検索では、地図から連想されるような住所や建物名などを用いて検索が行われる傾向が強かった。このような検索に対して十分な結果を出すためには、地図インタフェースから検索可能な情報の網羅性が重要になると考えられる。一方、カテゴリー検索では、街歩きで役立つようなカジュアルな情報がより多く検索される傾向があることから、都市に関する最新の情報とのリンクも重要と思われる。このような情報検索は、従来の Web ページだけを対象とした全文検索システム、もしくは GIS を利用した地理的属性の検索だけで実現することは困難であり、本研究で提案した拡張 Web 空間のように、インターネット上の情報と地理情報が統合して扱える情報検索システムが今後重要になると予想される。

5.2 問い合わせ処理の最適化に関わる課題

本研究では、拡張 Web 空間に基づく情報検索システムを試作し、実際の京都に関する Web 情報と地理情報を利用した複数の最適化手法の比較実験を通して、拡張 Web 空間における問い合わせ処理に適した最適化手法を明らかにした。今後の課題としては、情報検索システムの分散運用による処

理の効率化と、情報検索システムで利用するインデックスの効率的な構築があげられる。

3.5.2 節の情報検索システムの評価は、単独の計算機上で行ったが、システムを構成するモジュールを複数の計算機上に分散させ動作させることにより、負荷分散やレスポンスの改善が期待できる。ただ、最近の携帯電話や携帯情報端末の普及により、システムやネットワークの形態は多様化し、情報検索システムの利用環境は複雑化しつつある。今後は、これらの変化にも柔軟に適応できるような最適化手法の検討を行っていきたい。

本研究で構築した情報検索システムでは、Webページ中のキーワードや地理座標など、あらかじめ作成した各種属性に関するインデックスを問い合わせ処理に利用している。このインデックスは、インターネット上に発信されるWebページの変化だけでなく、建物の新築や改築、家主の交代や移転など、都市の地理情報の変化に応じて更新する必要がある。これらの変化に対応していくためには、最新の地理情報の収集と、Webページを収集するロボット、プロキシシステムやアドレスマッチングなどを用いた自動更新処理が必要となる。特に実際に情報検索システムをインターネット上で運営する場合には、最新の情報を検索できるようにしたり、検索結果中に含まれる「リンク切れ」をなくすことが重要であることから、短い周期でインデックスを再構築できる手法の検討が必要と考えている。

また、本研究の情報検索システムでは、検索結果として関連性のあるWebページの列が得られ、必要に応じて検索結果に含まれるWebページの地理属性も参照できることから、WWWを用いた経路検索や歩行者ナビゲーションなどのアプリケーションへも応用可能と考える。こうしたアプリケーションでは、問い合わせに対する最終結果だけでなく、問い合わせ条件を入力する際の補助や、問い合わせ処理中の情報提供も重要になるだろう。今後は、処理の高速化だけでなく、アプリケーションとの密接な関係動作も考慮に入れて情報検索システムの改善を図っていきたい。

5.3 ユーザインタフェースの機能拡張

本研究で構築した地図インタフェースでは、インターネット上に発信されている都市情報を検索するために、キーワード検索やカテゴリ検索など

の検索機能を実装した。今後の課題としては、本研究で提案した拡張Web空間に対する問い合わせのような系列的な情報検索の実現をはじめ、以下のような拡張が必要と考えている。

1. 地図の操作性の向上

地図イメージの表示範囲の移動や拡大・縮小などの操作は、ユーザが行う検索行動の一部と考えられる。従って、ユーザの検索行動を妨げないように、その操作性には十分に注意を払わなければならない。地図イメージのスムーズな操作を実現するために、本研究で構築した地図インタフェースではスクロールに工夫を加えている。具体的には、Javaアプレット上で表示する地図イメージに画面サイズより大きなものを利用し、オフスクリーンと組み合わせた座標管理を行うことで、転送したイメージ内で東西南北方向へ自由に移動できるようにしている。また、ユーザが操作するクライアントで検索結果をキャッシュすることにより、操作におけるインタラクティブ性の向上も図っている。

一方、インターネットにおいて地図を利用したサービスが普及するのに従い、ネットワーク経由で配信する地図の扱い方も変化してきている。従来は二次加工の困難なラスタ形式の地図イメージが主流であったが、ネットワーク環境の改善とユーザが利用するPCの性能向上により、拡大縮小や視点の移動を連続的に行うことができるベクトル形式の地図の配信も行われつつある。

2. 検索条件のインタラクティブな拡張

本研究の地図インタフェースでは、キーワード検索やカテゴリ検索などを実装することにより、インターネット上に発信されている都市に関連したWebページの検索を可能にした。この地図インタフェースの検索機能を拡張する手法の一つとして、より多くの条件を指定して詳細な検索を可能にする手法が考えられる。この際、一度に多数の条件を指定するアプローチと、インタラクティブに条件を変更していくアプローチが考えられるが、地図インタフェースの場合は、一般の画像処理ツールなどとの操作の類似性から、後者のアプローチの方が適していると考えられる。特に、大量の検索結果の分布を見な

がらフィルタリングを実施したり，個々の検索結果を確認しながらインタラクティブに条件を取捨選択する方が，検索結果にユーザの意図が反映されやすいと考える．

3. 検索結果の表示方法の拡張

地図インタフェースでは，検索結果が地図イメージ上の対応する位置にアイコンで表示されるので，検索結果の地理的な分布や位置関係，周辺情報の把握が可能になる．背景として表示される地図にも，様々な周辺情報が含まれるので，地図上にアイコンで表示される Web ページに関する検索だけでなく，都市の地理情報も関連づけて検索可能となることが望ましいと考える．

また，本研究の地図インタフェースでは，検索結果が平面上に表示されるのでその一覧性は高いが，紙に印刷された地図と比較すると，コンピュータの画面に表示される地図は，イメージの解像度や一覧できる表示領域の広さで大きく劣っている．たとえば，検索結果が地理的に集中している場合には，検索結果を表すアイコンが地図イメージ上で重なってしまい，検索結果が判別できなくなる場合もある．こうした場合への対処法としては，検索結果を表示するための地図の拡大縮小や検索条件の追加が一般的であるが，表示する地図の変形や略地図の利用も有効と考えている．

この他，インターネット経由で利用する地図アプリケーションの場合には，地図の拡大縮小や表示領域の移動など，多くの場面で地図データの転送が必要となるため，ネットワークの速度が遅い場合には，インタラクティブな地図の操作が困難となる場合がある．この問題は，表示領域以外の地図データの先読みなど，サーバ，クライアント双方の負荷分散を考慮した適切なデータ転送により改善される部分もあるが，ネットワークの転送速度やシステムの処理速度など，今後の技術の進歩により解決されていく部分が大きいと考えている．

また，拡張 Web 空間に基づいた情報検索システムを用いて地図インタフェースの機能拡張を行うと，Web ページを列として検索することが可能となる．しかし，地図インタフェース上で可能な直接操作だけで，その検索条件を入力することは困難である．たとえば，市内の大学とその最寄り駅の組み

合わせを地図インタフェースで検索する場合、まず起点となる大学を検索してその結果を画面上で記録し、その後で最寄り駅を検索するために距離検索などを行わなくてはならない。このような問い合わせの作成には、自然言語インタフェースやインタフェースエージェントが有効と考えられる。

- 自然言語インタフェースの導入

地図インタフェース上の直接操作だけでは、複数の目標に関する条件とその間の関係条件を同時に指定するような、系列的な問い合わせ条件の指定は困難であるが、自然言語インタフェースを導入すると、こうした条件指定を容易にできる可能性がある。ただ、自然言語による入力自由度が高く、様々な表現が可能となるため、入力された条件の意味解釈や、問い合わせへの変換には、広範な知識と効率的な処理が必要となる。

- インタフェースエージェントの導入

情報検索システムのフロントエンドに Extempo[75]のインタフェースエージェントなどを利用すると、対話を通した利用者の検索意図の抽出が期待できる。この検索意図と、ユーザの検索行動のログやシステム側で蓄積した知識やそれまでの利用傾向等を考慮することにより、より効果的な問い合わせの生成が可能になると考えている。より多くの情報を参照しながら効果的な問い合わせを生成することで、より地域コミュニティにおける活動に役立つ情報検索が可能になるだろうと予想している。

本研究では、地図を用いた検索結果の表示を中心に考えたが、実際の都市空間をメタファとした仮想的な三次元空間の利用や、現実空間と情報空間を融合させた仮想空間の利用など、都市情報の検索結果を可視化には様々なアプローチが考えられる。また、ユーザの検索を支援する手段としても、ユーザの操作順序に基づいた検索手順の提示や、あらかじめ用意された検索テンプレートの活用など、様々な機能が考えられる。こうした機能拡張も含めて、今後、検索インタフェースの改善を図っていきたい。

5.4 コミュニケーションツールとの連携

インターネットの普及により，電子メールやインスタントメッセージ，BBSやチャットなど，様々なコミュニケーションツールが使われるようになってきている．本研究で構築したシステムには，こうしたツールを組み込まなかったが，都市の情報空間では人々の間のコミュニケーションは重要な役割を持つと考えられる．

本研究で構築した地図インタフェースでは，建物や店舗，企業など，都市に存在するオブジェクトに関するインターネット上の Web ページが検索可能であるが，同様の手法により掲示板やチャットサイトなど，情報交換のための Web ページも関連づけることができる．こうしたコミュニケーションツールを実際の都市と関連づけて利用できるようにすると，新たなコミュニケーションの場として機能する可能性がある．たとえば，都市に住む人々が自分の「近所」の掲示板に書き込みを行い，日々の情報交換を行う．そして，交換された情報がその場所に蓄積されていけば，その都市への移住を考えてる人々や，観光やビジネスでその都市を訪れようとしている人々にとって，貴重な情報となるだろう．また，交換される情報に含まれる情報間のリンクが，自動的に抽出されるようになれば，都市の人々の思考に近い，情報空間が構築できるかも知れない．このような情報空間を構築することで，都市における情報検索も多様化できると考えている．

一方，コミュニケーションに利用されるユーザの端末も変化してきており，机上のコンピュータだけでなく，携帯電話や小型の携帯情報端末など，様々な種類の機器が利用されるようになってきている．特に携帯電話は，街中でのコミュニケーションやインターネット利用の機会を増やす上で大きな役割を果たしており，位置情報検索など，街中での利用を考慮した情報検索サービスも展開されつつある．

また，携帯情報端末を持つユーザは，自らの情報をネットワーク上に発信することにより，情報空間へ積極的に参加することができる．この際，自分の地理的な位置も一緒に発信すると，電子的な都市の情報空間の中に没入した形でのコミュニケーションも可能となる．こうした都市に参加するユーザも含んだ多様な情報空間における検索技術や，情報空間の可視化も含めたユーザインタフェースの構築が今後の課題と考えている．

第 6 章

結論

本研究では、都市の日常生活に密着した地域情報サービスを実現することを目的として、地理的関連性を用いた Web 検索技術を構築した。この中で、都市に関する Web 情報空間を Web ページ間の地理的關係によって拡張した拡張 Web 空間の提案するとともに、実際のインターネット上の Web ページと京都の地理情報を用いて、試作した拡張 Web 空間に基づく情報検索システムに適する最適化手法を検討した。また、地域情報検索のための地図インタフェースをデジタルシティ京都の実験サイトにて公開運用し、Web サーバに記録されたアクセスログの分析を通して、地理的に情報が集中する地域でのユーザの検索行動を明らかにした。

本研究の成果をまとめると、以下ようになる。

- 地理的関連性を用いた Web 情報空間の拡張

電子化された地理情報を利用してインターネット上の Web ページ間に地理的なリンクを動的に生成し、Web 情報空間を拡張する手法を提案した。この拡張 Web 空間では、Web ページの属性に関する条件とハイパーリンクと地理的ジェネリックリンクに関する条件を組み合わせ、Web ページの列を検索することができる。また、拡張 Web 空間に対する問い合わせを SELECT-FROM-WHERE 形式で記述できる検索言語を定義し、複数の検索例を用いてその記述性を確認した。

一方、拡張 Web 空間における検索は、半構造データの検索と捉えることができるが、その効率化を図るためには検索空間に適した手法

が必要となる．そこで，拡張Web空間に基づく情報検索システムを試作し，実際の京都に関する Web情報と地理情報を利用した複数の最適化手法の比較実験を行い，拡張Web空間における問い合わせ処理に適した最適化手法の検討を行った．その結果，動的計画法を用いた事前最適化では，検索に含まれるリンク数が増加すると最適な検索プランの作成が困難となる場合があり，その原因がリンク関係の不規則さに起因する予測の誤差の累積にあることがわかった．これに対し，最適な条件の選択とその処理を繰り返す実行時最適化では，処理状況を反映した正確な予測が可能で，手順的に予測誤差の累積も無い．実装した結果も事前最適化より良い性能を示し，最大で二倍の性能向上を観測した．

- 地域情報検索のための地図インタフェース
特定都市に関連する Webページを検索するためのユーザインタフェースとして，地域情報検索のための地図インタフェースを構築した．この地図インタフェースには，検索機能として，近傍検索，カテゴリ検索，およびキーワード検索を実装し，検索結果を地図上に重ね合わせて表示することによって，実世界に即した Webページ検索を可能にするとともに，クライアントにキャッシュしたデータを検索に再利用することにより，インタラクティブな検索処理とサーバの負荷の軽減を実現した．また，この地図インタフェースをデジタルシティ京都の実験サイトにて公開運用し，Webサーバのアクセスログの分析を通して，地図インタフェースに実装したカテゴリ検索とキーワード検索の利用傾向と，地理的に情報が集中している地域でのユーザの検索行動を明らかにした．

本研究は，学術的な研究成果やインターネット上で提供されているサービスに関する調査から多くの知見を得て進めてきた．拡張Web空間のモデルは，半構造データと Web検索言語に関わる研究を発展させたものであるし，その問い合わせ処理の効率化手法は，情報検索やデータベースシステムにおける問い合わせ処理の最適化や，XMLデータベースシステムのアイデアに基づいたものである．また，地図を用いた検索インタフェースは，

従来からインターネット上で提供されているシステムの問題点を鑑みて、設計構築したものである。

また、実際にインターネット上で公開運用するためのテストデータを構築する際の試行錯誤や、Webページの所有者から利用許諾を頂く際の手続きに関する議論などからも多くの知見が得られた。特に、インターネット上の情報検索システムでは、Webページへのリンク情報を編集して提供する場合の著作権や、実際の地理的な位置とともに情報を提供する場合のプライバシーの問題などは、各国の法律や地域文化により解釈の揺れがある問題であり、これからも議論が必要であることがわかった。また、これらの問題は実際のデータ構築やシステム作成を通じて初めて明らかになる部分も大きいことがわかった。

都市に関わる情報検索システムの発展には、システムの設計に必要な理論の構築や実際のシステムの実装だけでなく、実験的な公開運用やその利用状況の分析が重要と考えている。本研究では、地図インタフェースの利用状況の分析の第一歩として、実装した検索機能の利用傾向と、地図イメージ上に表示されるWebページの集中とアクセスの集中の関係に関する分析を行った。今後は、Webページの属性と選択順序の関係や、ユーザの居住地やアクセス日時とWebページのアクセス傾向など、より踏み込んだ形でユーザの検索行動を分析していきたい。

本研究の成果は、デジタルシティ京都の情報検索システムを実現する過程で生まれたものであり、実際のWebサイトにおける公開運用を通して具体化してきたものである。ただ本研究で実装した現段階のシステムは研究段階のものであり、ユーザ数が増えたときの同時接続性や負荷分散などには、システムの的にも、研究的にもまだ検討の余地があると考えられる。また、本研究では触れなかったが、デジタルシティを考える上では、人々の間のコミュニケーションのサポートが必要となるだろう。今後は、情報検索システムとコミュニケーションツールの連携も視野にいれ、実際の都市における地域情報サービスの有効性について検証を続けていきたい。

謝辞

まずはじめに京都大学大学院情報学研究科石田亨教授に心から感謝致します。先生の発想により始まったデジタルシティ京都に関する研究プロジェクトに参加できたこと、研究の進め方や論文の書き方に関して非常に親切にご指導頂いたことを幸せに思っております。また、論文の内容について多くのご指導、ご助言を頂きました京都大学大学院情報学研究科田中克己教授、酒井徹教授、奥乃博教授に感謝致します。田中先生には、先生が神戸大学に所属されていた頃より、拡張Web空間に関する研究について貴重なご意見を頂くとともに、アドバイザとしても様々な研究の指針を与えて頂きました。酒井先生には、アドバイザとして研究の進め方に関する数々のご助言を頂きました。厚くお礼申し上げます。

本研究は、NTTコミュニケーション科学研究所オープンラボ、京都大学大学院情報学研究科社会情報学専攻石田研究室、デジタルシティ京都実験フォーラム、NTTコミュニケーション科学基礎研究所社会情報研究部の多くの方々のご指導のもと行われました。NTTコミュニケーション科学基礎研究所にて研究の機会を与えて頂くとともに、日頃からご指導を頂く石井健一郎所長、管村昇企画部長、大山芳史社会情報研究部長、小暮潔グループリーダーに感謝します。また、オープンラボ発足以降の本研究に対して様々なサポートを頂きました東倉洋一先端技術総合研究所長、NTTソフトウェア株式会社 服部文夫博士、(株)国際電気通信基礎技術研究所 下原勝憲博士、エヌ・ティ・ティ・コミュニケーションズ株式会社 北寿郎博士に感謝致します。

オープンラボの常駐メンバーとして、デジタルシティ京都の立ち上げに共に携わった Katherine Isbister博士、Stefan Lisowski氏、Ben Benjamin氏、NTTコムウェア石川賢二氏、小林堅治氏、NTTコミュニケーション科学基礎研究所赤埴淳一主幹研究員に感謝致します。小林堅治氏には、Javaアプレット版地図インタフェースの構築について、多大のご協力を頂きました。また、NTTコムウェア長門裕司氏には、携帯情報端末版地図インタフェースの構築について、多大のご協力頂きました。深く感謝致します。

京都大学大学院情報学研究科八槨博史講師、中西英之助手、小山聡助手、岡本昌之氏、野村早恵子氏、(株)国際電気通信基礎技術研究所 高田司郎博

士には，石田研究室で開催されたドクターセミナーで，研究に関する貴重なご意見を頂きました．深く感謝致します．

デジタルシティ京都実験フォーラムの皆様には，実験サイトにて公開したJavaアプレット版地図インタフェースについて，様々なご意見を頂きました．深く感謝致します．また，フォーラムの運営に携わった三菱電機エンジニアリング（株）山田幸一氏，小山聡助手に感謝致します．そして，京都に関するコンテンツを本研究のために提供して頂いた皆様に感謝致します．

最後に，本研究中に著者を支援して頂いた皆様に感謝致します．

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] ビデオリサーチネットコム. Video Research Information. <http://www.videor.co.jp/info/010308.html>.
- [3] 石田亨. デジタルシティの現状. 情報処理, Vol. 41, No. 2, pp. 163–168, 2000.
- [4] Toru Ishida. Digital city Kyoto. *Communication of the ACM*, Vol. 45, No. 7, pp. 76–81, 2002.
- [5] 石田亨, 野村早恵子編. 特集 デジタルシティ. *bit*, Vol. 33, No. 4, 2001.
- [6] Toru Ishida and Katherine Isbister, editors. *Digital Cities*, Vol. 1765 of *Lecture Notes in Computer Science*. Springer-Verlag, 2000.
- [7] Makoto Tanabe, Peter van den Besselar, and Toru Ishida, editors. *Digital Cities II*, Vol. 2362 of *Lecture Notes in Computer Science*. Springer-Verlag, 2002.
- [8] マピオン. <http://www.mapion.co.jp/>.
- [9] マップファンウェブ. <http://www.mapfan.com/>.
- [10] MapQuest.com. <http://www.mapquest.com/>.

- [11] Helsinki Arena 2000. <http://www.arenanet.fi/>.
- [12] Risto Linturi (訳 野村早恵子). ヘルシンキ・アリーナ2000. *bit*, Vol. 33, No. 4, pp. 17–20, 2001.
- [13] Frederic Le Diberder (訳 小久保卓). ル・ドゥージェム・モンド. *bit*, Vol. 33, No. 4, pp. 21–24, 2001.
- [14] Sheng Huanye, Ding Peng (訳 松原啓明). インフォポート上海. *bit*, Vol. 33, No. 4, pp. 25–27, 2001.
- [15] Stefan Lisowski and Kaoru Hiramatsu. Laying 3DML Interface Groundwork for a Distributed Digital City Kyoto. In *Proceedings of Interaction'99*, pp. 41–42, 1999.
- [16] Peter van den Besselaar (訳 森村まり). アムステルダム・デジタルシティの興亡. *bit*, Vol. 33, No. 4, pp. 13–16, 2001.
- [17] George Kingsley Zipf. *Human Behavior and the Principle of Least Effort*. Addison Wesley, 1949.
- [18] Bernardo A. Huberman, Peter L. T. Pirollo, James E. Pitkow, and Rajan M. Lukose. Strong Regularities in World Wide Web Surfing. *SCIENCE*, Vol. 280, pp. 95–97, 1998.
- [19] Andrei Broder, Ravi Kumar, Farzin Maghout, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener. Graph structure in the web. In *The Ninth International World Wide Web Conference*, 2000.
- [20] Lycos. Lycos 50. <http://50.lycos.com/>.
- [21] Yahoo! JAPAN. 第6回ウェブ・ユーザー・アンケート, 1999. <http://www.yahoo.co.jp/docs/result/result6/>.
- [22] 平松薫. インターネットに浮かぶデジタルシティ. *bit*, Vol. 33, No. 4, pp. 3–7, 2001.

- [23] 平松薫. WWW・実時間センサ情報の統合によるデジタルシティの構築. 人工知能学会誌, Vol. 15, No. 1, pp. 65–68, 2000.
- [24] De Digital Stad. <http://www.dds.nl/>.
- [25] Digital City, Inc. <http://www.digitalcity.com/>.
- [26] 小山聡, 平松薫, 山田幸一. デジタルシティ京都. *bit*, Vol. 33, No. 4, pp. 8–12, 2001.
- [27] Katherine Isbister. A Warm Cyber-Welcome: Using an Agent-Led Group Tour to Introduce Visitors to Kyoto. In *Digital Cities*, Vol. 1765 of *Lecture Notes in Computer Science*, pp. 391–400, 2000.
- [28] 小暮潔, 赤埴淳一, 平松薫, 松原繁夫, 境野英明, 杉山聡, 柳沢豊. 実世界情報統合に向けて. 情報処理学会研究会報告 SIG-MBL/ITS/ICII, pp. 189–196, 2001.
- [29] Toru Ishida, Jun-ichi Akahani, Kaoru Hiramatsu, Katherine Isbister, Stefan Lisowski, Hideyuki Nakanishi, Masayuki Okamoto, Yasuhiko Miyazaki, and Ken Tsutsuguchi. Digital City Kyoto: Towards A Social Information Infrastructure. In *Cooperative Information Agents III*, Vol. 1652 of *Lecture Notes in Computer Science*, pp. 34–46, 1999.
- [30] Serge Abiteboul. Querying Semi-Structured Data. In *Database Theory – ICDT '97, 6th International Conference*, pp. 1–18, 1997.
- [31] 田島敬史. 半構造データのためのデータモデルと操作言語. 情報処理学会論文誌：データベース, Vol. 40, No. SIG 3(TOD 1), pp. 152–170, 1999.
- [32] Daniela Florescu, Alan Levy, and Alberto Mendelzon. Database Techniques for the World-Wide Web: A Survey. *SIGMOD Record*, Vol. 27, No. 3, pp. 59–74, 1998.

- [33] David Konopnicki and Oded Shmueli. W3QS: A Query System for the World-Wide Web. In *Proceedings of the 21st International Conference on Very Large Data Bases*, pp. 54–65, 1995.
- [34] David Konopnicki and Oded Shmueli. Information Gathering in the World Wide Web: The W3QL Query Language and the W3QS System. *ACM Transactions on Database Systems*, Vol. 23, No. 4, pp. 369–410, 1998.
- [35] Alberto O. Mendelzon, George A. Mihaila, and Tova Milo. Querying the World Wide Web. *International Journal on Digital Libraries*, Vol. 1, No. 1, pp. 54–67, 1997.
- [36] Alberto O. Mendelzon and Tova Milo. Formal Models of Web Queries. In *Proceedings of the 16th ACM Symposium on Principles of Database Systems*, pp. 134–143, 1997.
- [37] Gustavo O. Arocena and Alberto O. Mendelzon. WebOQL: Restructuring Documents, Databases, and Webs. In *Proceedings of International Conference Data Engineering '98*, pp. 24–33, 1998.
- [38] Mary Fernandez, Daniela Florescu, Alon Levy, and Dan Suciu. A Query Language for a Web-Site Management System. *SIGMOD Record*, Vol. 26, No. 3, pp. 4–11, 1997.
- [39] 相良毅, 有川正俊, 坂内正夫. ジオリファレンス情報を用いた空間情報抽出システム. *情報処理学会論文誌 : データベース*, Vol. 41, No. SIG6(TOD7), pp. 69–80, 2000.
- [40] Google. 2002 Google Programming Contest Winner, 2002. <http://www.google.com/programming-contest/winner.html>.
- [41] Open GIS Consortium, Inc. OpenGIS Simple Features Specification For SQL Revision 1.0.

- [42] W3C Recommendation. *Extensible Markup Language (XML) Version 1.0*.
- [43] XML/SGMLサロン. 標準XML完全解説. 技術評論社, 1998.
- [44] W3C. XML Linking Language (XLink) Version 1.0.
- [45] G-XML実用化協議会. G-XML 2.0版.
- [46] Katsumi Tanaka, N. Nishikawa, S. Hirayama, and K. Nanba. Query Pairs As Hypertext Links. In *Proceedings of 7th IEEE Data Engineering Conference*, pp. 456–463, 1991.
- [47] Qing Qian, Masaaki Tanizaki, and Katsumi Tanaka. Abstraction and Inheritance of Hyperlinks in an Object-Oriented Hypertext Database System TextLink/Gem. *IEICE Transactions on Information and Systems*, Vol. E78-D, No. 11, pp. 1343–1353, 1995.
- [48] Jenny Edwards, Kevin McCurley, and John Tomlin. An Adaptive Model for Optimizing Performance of an Incremental Web Crawler. In *The Tenth International World Wide Web Conference*, pp. 106–113, 2001.
- [49] Soumen Chakrabarti, Martin van den Berg, and Byron Dom. Focused crawling: a new approach to topic-specific Web resource discovery. In *The Eighth International World Wide Web Conference*, pp. 545–562, 1999.
- [50] Charu Aggarwal, Fatima Al-Garawi, and Philip Yu. Intelligent Crawling on the World Wide Web with Arbitrary Predicate. In *The Tenth International World Wide Web Conference*, pp. 96–105, 2001.
- [51] Jason McHugh and Jennifer Widom. Query Optimization for XML. In *Proceedings of the 25th International Conference on Very Large Data Bases*, pp. 315–326, 1999.

- [52] Jason McHugh and Jennifer Widom. Query Optimization for Semistructured Data. Technical report, Stanford University Database Group, 1999.
- [53] Serge Abiteboul, Dallan Quass, Jason McHugh, Jennifer Widom, and Janet L. Wiener. The Lorel Query Language for Semistructured Data. *Journal of Digital Libraries*, Vol. 1, No. 1, pp. 68–88, 1997.
- [54] Elisa Bertino and Won Kim. Indexing Techniques for Queries on Nested Objects. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 1, No. 2, pp. 196–214, 1989.
- [55] Jason McHugh and Jennifer Widom. Compile-Time Path Expansion in Lore. In *Workshop on Query Processing for Semi-structures Data and Non-Standard Data Formats*, 1999.
- [56] Paul F. Dietz. Maintaining Order in a Linked List. In *Proceedings of the Fourteenth Annual Symposium on Theory of Computing*, pp. 122–127, 1982.
- [57] Quanzhong Li and Bongki Moon. Indexing and Querying XML Data for Regular Path Expression. In *Proceedings of the 27th International Conference on Very Large Data Bases*, pp. 361–370, 2001.
- [58] Wang-Chien Lee and Dik Lun Lee. Path Dictionary: A New Access Method for Query Processing in Object-Oriented Databases. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 10, No. 3, pp. 371–388, 1998.
- [59] 石塚満. 知識の表現と高速推論. 丸善, 1996.
- [60] Christos Faloutsos, Bernhard Seeger, Agma Traina, and Caetano Traina. Spatial Join Selectivity Using Power Laws. In *Proceedings*

of *ACM SIGMOD International Conference on Management of Data*, pp. 177–188, 2000.

- [61] Patricia G. Selinger, Morton M. Astrahan, Donald D. Chamberlin, Raymond A. Lorie, and Thomas G. Price. Access Path Selection in a Relational Database Management System. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 23–34, 1979.
- [62] Alon Levy (Ed.). Special issue on adaptive query processing. *Data Engineering*, Vol. 23, No. 2, 2000.
- [63] Navin Kabra and David J. DeWitt. Efficient Mid-Query Re-Optimization of Sub-Optimal Query Execution Plans. In *Proceedings of ACM SIGMOD International Conference on Management of Data*, pp. 106–117, 1998.
- [64] Zachary G. Ives, Daniela Florescu, Marc Fiedman, Alon Levy, and Daniel S. Weld. An Adaptive Query Execution System for Data Integration. In *Proceedings of ACM SIGMOD International Conference on Management of Data*, pp. 299–310, 1999.
- [65] PostgreSQL. <http://www.postgresql.org/>.
- [66] 高木悟. 分散型電子地図プラットフォーム JaMaPS. 画像ラボ, Vol. 9, No. 12, pp. 27–32, 1998.
- [67] 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.
- [68] Ang S. Cheong, Peter Brantley, and Michael D. Doyle. Polymap: A Versatile Client-Side Image Map for the Web. In *The Fourth International World Wide Web Conference*, 1995.

- [69] 高木悟, 松本一則. 地図情報を利用した情報検索. 情報処理, Vol. 41, No. 4, pp. 357–362, 2000.
- [70] 馬場肇. 日本語検索システムの構築と活用. ソフトバンク, 1998.
- [71] SnapTrack, Inc. <http://www.snaptrack.com/>.
- [72] Lee Breslau, Pei Cao, Li Fan, Graham Phillips, and Scott Shenker. Web Caching and Zipf-like Distributions: Evidence and Implications. In *Proceedings of the IEEE Infocom'98 Conference*, pp. 126–134, 1999.
- [73] Rakesh Agrawal, Tomasz Imielinski, and Arun Swami. Mining Association Rules between Sets of Items in Large Databases. In *Proceedings of ACM SIGMOD International Conference on Management of Data*, pp. 207–216, 1993.
- [74] Christian Borgelt. apriori. <http://fuzzy.cs.uni-magdeburg.de/~borgelt/index.html>.
- [75] Extempo Systems Inc. <http://www.extempo.com/>.

発表論文

論文誌

1. 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.
2. 平松薫, 小林堅治, Ben Benjamin, 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.
3. Satoshi Oyama, Kaoru Hiramatsu, and Toru Ishida. Operative Information Agents for Digital Cities. *International Journal of Cooperative Information Systems*, Vol. 10, No. 1&2, pp. 197–215, 2001.

国際会議

1. Toru Ishida, Jun-ichi Akahani, Kaoru Hiramatsu, Katherine Isbister, Stefan Lisowski, Hideyuki Nakanishi, Masayuki Okamoto, Yasuhiko Miyazaki, and Ken Tsutsuguchi. Digital City Kyoto: Towards A Social Information Infrastructure, Cooperative Information Agents III, pp. 34–46, 1999.
2. Kaoru Hiramatsu, Kenji Kobayashi, Ben Benjamin, Toru Ishida, and Jun-ichi Akahani. Map-based User Interface for Digital City

- Kyoto. *INET2000 The Internet Global Summit*, in CD-ROM, 2000.
3. Kaoru Hiramatsu, and Toru Ishida. An Augmented Web Space for Digital Cities. *The 2001 Symposium on Application and the Internet (SAINT2001)*, pp. 105–112, 2001.
 4. Kaoru Hiramatsu, Yuji Nagato, Jun-ichi Akahani, and Kiyoshi Kogure. Location Related Information Retrieval in Digital City Kyoto. *INET2001 The Internet Global Summit*, in CD-ROM, 2001.
 5. Kaoru Hiramatsu, Yuji Nagato, Jun-ichi Akahani, and Kiyoshi Kogure. Location Related Web Page Search on Digital City Kyoto. *AsiaGIS2001*, in CD-ROM, 2001.
 6. Kaoru Hiramatsu. Log Analysis of Map-based Web Page Search on Digital City Kyoto. *Digital Cities II*, Lecture Notes in Computer Science, Vol. 2362, pp. 233–245, 2002.
 7. Jun-ichi Akahani, Kaoru Hiramatsu, Yoshikazu Furukawa, and Kiyoshi Kogure. Agent-based Coordination of Regional Information Services. *Digital Cities II*, Lecture Notes in Computer Science, Vol. 2362, pp. 283–291, 2002.
 8. Jun-ichi Akahani, Kaoru Hiramatsu, and Kiyoshi Kogure. Approximate Ontology Translation and its Application to Regional Information Services. *The International Semantic Web Conference (ISWC-2002)*, 2002.

その他

1. Stefan Lisowski, and Kaoru Hiramatsu. Laying 3DML Interface Groundwork for a Distributed Digital City Kyoto, *Interaction'99*, pp. 41–42, 1999.

2. 平松薫. 地域情報サービスに向けたデジタルシティ京都の構築, 情報処理学会研究会報告 SIG-GW 36-14, pp. 79–84, 2000.
3. 平松薫, 石田亨. WWW・実時間センサ情報の統合によるデジタルシティの構築, 人工知能学会誌, Vol. 15, No. 1, pp. 65–68, 2000.
4. 平松薫. デジタルシティにおける都市型情報検索, 人工知能学会研究会資料 SIG-Challenge-0111, pp. 15–22, 2001.
5. 平松薫. インターネットに浮かぶデジタルシティ – 特集 デジタルシティ –, bit (2001年4月号), Vol. 33, No. 4, pp. 3–7, 2001.
6. 小山聡, 平松薫, 山田幸一. デジタルシティ京都 – 特集 デジタルシティ –, bit (2001年4月号), Vol. 33, No. 4, pp. 8–12, 2001.
7. 赤埴淳一, 平松薫, 古川仁一, 小暮潔. エージェントによる地域情報サービスの連携, 2001年度 人工知能学会全国大会(第15回)論文集, 2001.
8. 小暮潔, 赤埴淳一, 平松薫, 松原繁夫, 境野英明, 杉山聡, 柳沢豊. 実世界情報統合に向けて, 情報処理学会研究会報告 SIG-MBL/ITS/ICII, pp. 189–196, 2001.