

付録 B

モバイルインフォサーチのログマイニング

住所（ジオワード）とキーワードを持つログデータから地理的な特性を取り出すジオワードマイニングに関しては第 5 章で述べたが、付録の本章と次章を使って、本編で取り上げていないマイニング手法とその結果を報告する。どちらもジオワードマイニングに先立って行った研究で、ローカルサーチサービスのログデータに対してマイニングを行っている。

まず本章では、モバイル位置情報のポータルサイトであるモバイルインフォサーチ実験のログマイニング方法とその結果について述べる [Iko 2002]。ここではマイニングの対象を利用者の地理位置とサービスという粒度で分析する。なおモバイルインフォサーチ実験（MIS）は著者らが運用していたインターネット上の実験サービスで、第 4 章で述べた技術を用いて構築されている。くわしくは付録 A.3 を参照していただきたい。

B.1 概要

モバイル位置情報ポータルのログマイニングとして留意した点を示す。MIS は任意の地理位置に関する情報提供する検索エンジンで、利用者が検索する地理位置（現在地や好みの場所）を指定することによって、その位置に関する情報を取得することができる。MIS に残される検索ログデータは利用者の地理位置に関する情報取得活動の記録である。ログデータが持つ属性は以下の通り整理できる。

- 検索地理位置
 - － 地理位置指定手段
 - － 地理位置実体（緯度経度）

- 検索目的

- ー サービス名（店情報、地図、天気、等）
- ー 検索ワード（レストラン、ホテル等（店情報の場合））

- 時間

すなわち MIS のログデータは、位置、目的、時間という属性を持つ利用記録の集合である。マイニングを行うことにより、ユーザの位置と検索目的、およびその時系列、あるいはユーザの位置と複数のインターネットサービスの利用、およびその時系列の関係を取り出すことを行った。本取り組みは利用者の位置と複数のインターネットサービスとの関連性をマイニングする方法論を明らかにすることを試みた世界で初めての試みである。

この結果、ローカルサーチにおける地図の必要性が改めて確認できたほか、秋葉原ではレストランが探される、北海道では夜にガソリンスタンドが探されるなどの興味深いルールを発見することができた。本取り組みの結果は、十分な量のログデータが揃わなかったこともあり、モバイル位置情報検索における普遍的な現象の解明にまでは至ったとはいえない。しかし、利用者が地理的に移動し、使うサービスを選択しながら、その中から実際にサービスを受ける過程は、今後より重要となる「ネットワークを使った日常生活」の典型である。その意味でも本取り組みは重要な意味を持つ。

B.2 ログデータの概要

解析の手法として、データマイニング技術の代表的なものである相関ルール (association rule) マイニング、時系列パターン (sequential pattern) マイニングを用いた、相関ルールマイニングはデータの属性間に存在する隠れた相関性を抽出するものである。時系列パターンマイニングは、時間軸に沿った解析を行うことにより時間の持つ意味を含むルールを抽出する手法である。

今回解析したデータは 1999 年 1 月から 5 月までの約 100 万アクセスである。解析には Apache の CookieID をユーザ ID として用い、さらにアクセス時間と MIS リクエストを解析に使った。

MIS リクエストからは MIS から何らかのサービスが実際に呼び出されたもの、および「ここのサーチ」の検索結果からウェブページが閲覧される時に呼び出されるものに該当するデータを使った。このような条件を満たす MIS ログデータは前者が 24,128 アクセス、後者が 1,703 アクセスとなった。また「東京都中央区銀座 4 丁目」は見本としてデフォルト表示されるようにしていたので、この位置は解析から除外した。

MIS リクエストの index.cgi は以下の要素を含んでいる。

- NL: 北緯

- EL: 東経
- from: 地理位置指定の方法
- submit: 何の検索を行ったかを示す部分
- shop_web: お店の検索エンジン
 - shop_cond: お店の検索条件（メニュー）
 - keyword: お店の検索条件（フリーワード）
- map_web: 地図表示検索エンジン
- rail_web: 鉄道経路検索エンジン
- station_web: 時刻表検索エンジン
- tv_web: テレビ番組表表示

地理位置情報の指定手段（from）には以下のものが含まれた。

- lkokonavi: PHS 基地局位置利用クライアント
- kokonogpslink: GPS 位置情報利用クライアント
- address: 住所
- station: 駅
- zip: 郵便番号
- idokeido: 緯度経度
- proatlas: サードパーティークライアント（地図ソフトウェア）
- tranavi95: サードパーティークライアント（交通機関経路探索ソフトウェア）
- pocke: サードパーティークライアント（ポータル）

お店の検索条件（メニュー）（shop_cond）は、固定的に以下のものを使っていた。

- レストラン
- ガソリンスタンド
- 銀行
- ホテル
- 郵便局
- コンビニ
- カラオケボックス
- イタリア料理
- デパート

B.3 ログデータのマイニング

B.3.1 データの変換方式

MIS ログデータをマイニングエンジンに適用するためのデータ変換方式について述べる。

1. MIS ログデータの CID に Apache の CookieID を含むデータを取り出し、Apache の CID、アクセス時間、HTTP リクエストを取り出して、Apache の CID でソートする
2. HTTP リクエスト中から MIS リクエストのみを残す
3. ログからノイズを取り除く。ここでは、HTTP リクエスト中の submit に対応する要素がすべて含まれていないログをノイズとして除去した

相関ルールマイニングの場合

1. HTTP リクエスト中の各要素をアイテムとしてトランザクションを作成する。
2. 住所を (都道府県)→(都道府県, 市)→(都道府県, 市, 町)→(都道府県, 市, 町, 番地)の形に階層化する。

時系列パターンの場合

1. CID の等しいデータをまとめ、ユーザセッション (一連の連続する検索のまとまり) を作成する。
2. ユーザセッションの HTTP リクエスト中の各要素をアイテムとしてシーケンスを作成する。
3. 住所を (都道府県)→(都道府県, 市)→(都道府県, 市, 町)→(都道府県, 市, 町, 番地)の形に階層化する。

B.3.2 ルールの評価方法

相関ルールと時系列パターンの優位性を以下の値により評価する。

- Support: ルールを含むトランザクションの割合

$$Supp(A \Rightarrow C) = Supp(A \cup C) = \frac{|A \text{ と } C \text{ を含むトランザクション数}|}{|\text{トランザクション総数}|}$$

- Confidence: ルールが起きる確からしさ (条件付き確率)

$$Conf(A \Rightarrow C) = \frac{Supp(A \cup C)}{Supp(A)}$$

- User: ルールを含むユーザの割合

$$User(A \Rightarrow C) = \frac{|A \text{ と } C \text{ を含むトランザクションに関わるユーザ数}|}{| \text{ユーザ総数} |}$$

- Repeat: 同一ユーザによる同一パターンでの検索の繰り返し回数の平均値

$$Repeat(A \Rightarrow C) = \frac{|Supp(A \Rightarrow C)|}{|User(A \Rightarrow C)|}$$

- Chi-squared value (Correlation): 関連性を示す値

実際値を f_o , 期待値を f のとき, 以下の式で求められる.

$$\chi^2 = \sum \frac{(f_o - f)^2}{f}$$

この値が 0 のとき独立事象であることを示す. この値が大きい程関連性が高いことを示し, 3.84 で 95% の significance level を示す.

- lift

$$lift(A \rightarrow C) = \frac{Supp(A \Rightarrow C)}{Supp(A) \times Supp(C)}$$

lift の値が 1 のとき, 独立事象となる. 例として, A が生じる確率 (支持度) が $\frac{1}{3}$, C が $\frac{1}{2}$ のとき, A と C が $\frac{1}{6}$ の確率で生じる場合, A と C は独立事象であるため, これらからなるルールは興味深いものではない. lift の値が大きいほど, 興味深いルールと評価される. これは, Interestingness と等しい.

従来の時系列パターンは支持度のみで評価していたが, 本解析では確信度も導入する. 時系列パターン $A_0 \rightarrow A_1 \rightarrow \dots \rightarrow A_m$ のとき, A_0 から A_i まで順に生じ, その後に A_{i+1} から A_m の順に生じる割合を以下のように定義した.

$$Conf(A_0 \rightarrow \dots \rightarrow A_i \Rightarrow A_{i+1} \rightarrow \dots \rightarrow A_m) = \frac{Supp(A_0 \rightarrow A_1 \rightarrow \dots \rightarrow A_m)}{Supp(A_0 \rightarrow \dots \rightarrow A_i)}$$

B.4 基本統計情報

表 B.1 表 B.2 に位置情報の指定手段および利用されたサービスの割合を示す. MIS はモバイル位置情報サービスの可能性を検証すべく, 様々な位置情報指定手段を提供したが, 住所や郵便番号といった指定手段に混じって, 特殊なデバイスとソフトウェアを必要

とする、PHS や GPS による指定も行われていた。また利用されたサービスは地図が多く、ここのサーチが続いている。地図は Repeat が高いが、これはそれぞれの利用者が複数の場所で地図を閲覧したことによる。

表 B.1 位置情報指定方法の割合

	Transactons	Users	Repeat
from=address	6,749(39.6%)	1,386(34.9%)	4.9
from=zip	3,832(22.5%)	982(24.7%)	3.9
from=station	3,418(20.1%)	790(19.9%)	4.3
from=idokeido	887(5.2%)	159(4.0%)	5.6
from=tranavi95	641(3.7%)	209(5.3%)	3.1
from=pocke	586(3.4%)	216(5.4%)	2.7
from=proatlas	247(1.5%)	38(1.0%)	6.5
from=lkokonavi	217(1.3%)	54(1.4%)	4.0
from=kokonogpslink	132(0.8%)	17(0.4%)	7.8

表 B.2 利用されたサービスの割合

	Transactions	Users	Repeat
map	8,572(45.9%)	1,910(28.3%)	4.49
kokono	4,109(22.0%)	2,206(32.7%)	1.9
shop	3,128(16.8%)	1,028(15.3%)	3.0
rail	1,289(6.9%)	539(8.0%)	2.4
hotel	665(3.6%)	397(5.9%)	1.7
weather	375(2.0%)	295(4.4%)	1.3
newspaper	279(1.5%)	183(2.7%)	1.5
tv	260(1.4%)	181(2.7%)	1.4

B.5 ローカル情報サービス間の相関

モーバイルインフォサーチの特徴の一つは複数のローカル情報サービスへのポータルとなっている点であるので、それらの位置情報サービスの関連性に関するルールを調べた(表 B.3)。表の通り、なんらかのサービスを使った後に、地図を見る、あるいは天気予報を見る、という使われ方がされている。

次は個々の検索エンジンに着目してルールを調べる。ここでは「ここのサーチ」と「お店等の情報」に関して調べた結果をまとめる。

B.6 ここのサーチに関する分析

解析に用いたログの中の「ここのサーチ」へのリクエストは 5,760 件あり、そのうち結果表示された URL を見たものは 1,589 件であった。また一回のユーザセッションでの URL を見る平均数は 約 0.3 件であった。

まず「ここのサーチ」(S_kokono) に関するルールを示す。「ここのサーチ」検索後に、検索結果 URL をアクセスした利用者の中に、確信度が高かったルールを表 B.4 に示す。また、その時系列パターンを表 B.5 に示す。前者の結果では単なる位置と動作の相関を取るだけでなく、位置指定を行った手段の関連も含めてルール化できることがこの解析の特徴である。後者の結果はサービス利用の渡り歩きを時系列で捉えることができ、サービス利用状況の把握だけでなく、複数のサービスを組み合わせたサービスを実現する際の画面レイアウト等の設計指針を与えることもできる。

次に「ここのサーチ」の後に何の検索をすることが多いかを表 B.7 に示す。表から確信度が最も高いのは検索結果 URL の閲覧である。これは製作意図としては当然のことであるが、別なサービスにいく場合は地図をリクエストすることが多いことがわかる。この時点での「ここのサーチ」は結果の地図表示インタフェースをもっていなかったため、地図との連動が必要であることを明確に示す結果となった。

B.7 お店等の情報の検索分析

最後に「お店等の情報」(S_shop) に関するルールを表 B.7 に示す。この結果は大変興味深い物で、秋葉原ではパソコンショップではなく、レストランが探されているという結果が導かれた。秋葉原は現在の秋葉原と違いパソコンや電気マニアが部品を買いにくるという色彩が残っていた時代で、著者も昼食にはあまり適切な場所ではないとの印象を持っていたことと符合した。モバイル位置情報サービスでは「見つかりにくいもの」が求められる傾向がある可能性を示したルールである。

表 B.3 複数のインターネットローカル情報サービスに関する時系列アクセスパターンの例

Sequential pattern	Conf%	Supp% (#)	User% (#)	Repeat	Lift	Chi-sq
[S.shop=お店等の情報][from=address][shop_cond=blank][shop_web=mapoo] ⇒ [S.map=地図][from=address] → [S.shop=お店等の情報][shop_cond=blank]→[S.rail=駅経路探索] → [S.newspaper=新聞]⇒ [S.weather=天気予報]	76.47	0.23 (13)	0.33 (12)	1.08	3.81	320.70
[S.shop=お店等の情報][shop_cond=blank][shop_web=townpage]→[S.rail=駅経路探索] → [S.newspaper=新聞]⇒ [S.weather=天気予報]	73.68	0.25 (14)	0.38 (14)	1.00	13.22	188.03
[S.shop=お店等の情報][shop_cond=blank][shop_web=townpage]→[S.rail=駅経路探索] → [S.newspaper=新聞]⇒ [S.weather=天気予報]	72.22	0.23 (13)	0.36 (13)	1.00	12.96	172.28
[S.shop=お店等の情報]→[S.rail=駅経路探索] → [S.newspaper=新聞]⇒ [S.weather=天気予報]	71.43	0.27 (15)	0.41 (15)	1.00	12.82	193.80
[S.weather=天気予報]→[S.shop=お店等の情報][shop_cond=blank][shop_web=townpage] → [S.kokono=このサーチ]⇒ [S.map=地図]	71.43	0.18 (10)	0.27 (10)	1.00	1.33	3503.32
[S.shop=お店等の情報][access.hour=22][access.week=Fri] ⇒ [S.map=地図][access.hour=22][access.week=Fri] [S.shop=お店等の情報][shop_cond=gasoline] ⇒ [S.map=地図]	65.00	0.23 (13)	0.36 (13)	1.00	90.69	1166.11
[S.shop=お店等の情報][shop_cond=blank][shop_web=townpage]→[S.weather=天気予報] ⇒ [S.kokono=このサーチ]⇒ [kokono.visit=url] [S.map=地図]→[S.tv=TVガイド] → [S.kokono=このサーチ]⇒ [kokono.visit=url] [S.map=地図]→[S.rail=駅経路探索][rail_web=s.stranavi] → [S.newspaper=新聞]⇒ [S.weather=天気予報]	63.33	0.34 (19)	0.47 (17)	1.12	1.18	3513.70
[S.shop=お店等の情報][shop_cond=blank][shop_web=townpage]→[S.weather=天気予報] ⇒ [S.kokono=このサーチ]⇒ [kokono.visit=url] [S.map=地図]→[S.tv=TVガイド] → [S.kokono=このサーチ]⇒ [kokono.visit=url] [S.map=地図]→[S.rail=駅経路探索][rail_web=s.stranavi] → [S.newspaper=新聞]⇒ [S.weather=天気予報]	63.16	0.22 (12)	0.25 (9)	1.33	4.61	164.07
[S.shop=お店等の情報][shop_cond=blank][shop_web=townpage]→[S.weather=天気予報] ⇒ [S.kokono=このサーチ]⇒ [S.map=地図]	62.50	0.18 (10)	0.22 (8)	1.25	4.57	156.36
[S.kokono=このサーチ]⇒ [kokono.visit=url] [S.map=地図]→[S.rail=駅経路探索][rail_web=s.stranavi] → [S.newspaper=新聞]⇒ [S.weather=天気予報]	61.54	0.29 (16)	0.44 (16)	1.00	11.04	175.65
[S.weather=天気予報]→[S.shop=お店等の情報][shop_web=townpage] → [S.kokono=このサーチ]⇒ [S.map=地図]	61.11	0.20 (11)	0.30 (11)	1.00	1.14	3501.79

表 B.4 このサーチで検索結果の URL をアクセスする確信度の高いルール

Association rule	Conf%	Supp% (#)	User% (#)	Repeat	Lift	Chi-sq
[address=富山県, 富山市,][station=富山口] ⇒ [kokono_visit=url]	43.24	0.08 (16)	0.14 (5)	3.20	5.64	66.32
[address=福岡県,][station=井尻] ⇒ [kokono_visit=url]	39.13	0.04 (9)	0.14 (5)	1.80	5.11	32.23
[access_month=04][address=兵庫県,][from=zip] ⇒ [kokono_visit=url]	37.78	0.08 (17)	0.14 (5)	3.40	4.93	57.81
[access_hour=06][address=神奈川県,] ⇒ [kokono_visit=url]	37.74	0.10 (20)	0.17 (6)	3.33	4.93	67.92
[access_hour=10][access_month=03][access_week=Sun] ⇒ [kokono_visit=url]	37.50	0.06 (12)	0.19 (7)	1.71	4.89	40.33
[address=神奈川県,横須賀市,][from=address][station=京急長沢] ⇒ [kokono_visit=url]	35.00	0.10 (21)	0.17 (6)	3.50	4.57	63.56

表 B.5 このサーチで検索結果の URL をアクセスする時系列パターン

Sequential pattern	Conf%	Supp% (#)	User% (#)	Repeat	Lift	Chi-sq
[S_weather= 天気予報][access_hour=14] → [S_kokono= このサーチ][access_hour=14] ⇒ [kokono_visit=url][access_hour=14]	90.91	0.18 (10)	0.25 (9)	1.11	46.55	458.03
[S_shop=お店等の情報][shop_cond=blank][shop_web=townpage] → [S_weather=天気予報] → [S_kokono= このサーチ] ⇒ [kokono_visit=url]	63.16	0.22 (12)	0.25 (9)	1.33	4.61	164.07
[S_map=地図] → [S_tv= T V ガイ ド][tv_web=tvguide] → [S_kokono= このサーチ] ⇒ [kokono_visit=url]	62.50	0.18 (10)	0.22 (8)	1.25	4.57	156.36
[S_kokono= このサーチ][address=神奈川県,横須賀市] ⇒ [kokono_visit=url][address=神奈川県,横須賀市]	53.57	0.27 (15)	0.33 (12)	1.25	199.32	2982.96

表 B.6 このサーチの後に多い検索パターン

Sequential pattern	Conf%	Supp% (#)	User% (#)	Repeat	Lift	Chi-sq
[submit_kokono=このサーチ] ⇒ [kokono_visit=url]	22.75	10.98 (613)	15.76 (574)	1.07	1.66	746.20
[kokono_visit=url] ⇒ [kokono_visit=url] ⇒ [kokono_visit=url]	35.86	4.91 (274)	7.14 (260)	1.05	2.62	561.03
[submit_kokono=このサーチ] ⇒ [submit_map= 地 図]	34.30	16.56 (924)	21.36 (778)	1.19	0.64	5178.49
[submit_kokono=このサーチ] ⇒ [submit_kokono=このサーチ]	29.99	14.48 (808)	19.38 (706)	1.14	0.62	3780.00
[submit_kokono=このサーチ] ⇒ [submit_shop=お店等の情報]	19.64	9.48 (529)	12.93 (471)	1.12	0.81	739.14

表 B.7 お店情報に関する検索について発見したルール

Rule	Conf	Supp % (#)	User % (#)	Repeat	Lift	Chi-sq
秋葉原にはいいレストランはないのか [address=東京都,][keyword=][station=秋葉原] ⇒ [shop_cond=restaurant]	21.05	0.02 (4)	3.00 (0)	0.00	0.08	25.21
北海道ではガソリンスタンドがよく調べられる [hour=20][address=北海道,][from=address][shop_web=townpage] ⇒ [shop_cond=gasoline]	61.91	0.06 (13)	3.00 (0)	0.00	1.62	2184.15
[hour=20][address=北海道,][shop_web=townpage] ⇒ [from=address][shop_cond=gasoline]	50.00	0.06 (13)	3.00 (0)	0.00	1.92	2723.22
北海道ではコンビニがよく検索される [hour=16][address=北海道,][shop_web=townpage] ⇒ [from=address][shop_cond=conveni]	23.08	0.01 (3)	3.00 (0)	0.00	0.54	217.14

付録 C

i タウンページの時系列ログマイニング

前章に続くローカルサーチサービスのログマイニングの例として、イエローページ検索サービスである i タウンページのログ解析を報告する [Bowo 2002]. i タウンページは住所とキーワードから検索を行うサービスで、典型的なローカルサーチである。

本章では利用者の検索の流れを時系列パターンとして解析した結果を報告する。i タウンページに関しては付録 A.2 を参照していただきたい。

C.1 概要

i タウンページはローカルサーチ検索を提供するサービスである。本サービスから利用者の検索行動を時系列パターンとして捉え、そのマイニングを行った結果を報告する。時系列パターンマイニングとは「事象 A が生じ、次に事象 B が生じたとき、次に事象 C が生ずることが多い」といったパターンを発見することである。分析の結果、利用者が検索条件を入力し結果へたどり着く典型的なパターンが明らかになっただけでなく、利用者が検索結果を得られずにサイトを去ってしまう現象も明らかになった。また結果を可視化することにより、サイト全体の使われ方を直感的に分かりやすくすることができた。

C.2 解析の指針

今回解析したデータは 1999 年 7 月から 2000 年 6 月までの一年間分のものである。i タウンページは通常のブラウザ等からのアクセスを想定した PC 用検索サービスと i モード等の携帯電話からのアクセスを想定した携帯用検索サービスを持っているが、ここではモバイル環境からの情報検索の動向をみるために、携帯用検索サービスのログを解析

した、
データ変換方式は付録前章で述べた MIS のログマイニング方法に順じ、利用者からの一連の連続するリクエストのまとまりをセッションとして作成した。ルールの評価方法は同じく MIS ログマイニングに準じている。

C.3 サイトマップの作成

携帯用サイトの分析を行うために、はじめに行ったのがサイトマップの作成である。利用者の検索行動の時系列パターンを分析する時、端末からリクエストされる各ページが検索条件によってダイナミックに生成されている場合は、ログに記録された URI をそのまま解析単位として利用すると、対象が無数に広がってしまい事実上解析ができない。例えば、「東京でラーメン」を探している URI と「埼玉でラーメン」を探している URI は異なるが、両者を同じ枠内に整理して分析できる必要がある。このようにリクエスト URI に対して抽象化したページクラスを作成し、その関連をサイトマップとして構築する必要がある。作成したサイトマップが図 C.1 である。

サイトマップ構成するページクラスの粒度は、解析する目的に応じて適切に選ぶことが必要である。また選んだ単位は CGI のパラメータ等で定義できるものでなければならない。

C.4 基本統計

サイトマップの各ページクラスごとの滞留時間、ホップ数、検索回数に関する分析結果を述べる。

滞留時間とは利用者があるページから次のページへ遷移するまでの時間（秒）を示したものである。表 C.1 に示す。

表 C.1 代表的なページの滞留時間

class	view-time (s)
T1 (top)	18.5
J1 (local top)	10.3
T2 (region list)	9.8
T3 (prefecture list)	10.7
NR (input form)	43.1
S (search result)	32.2

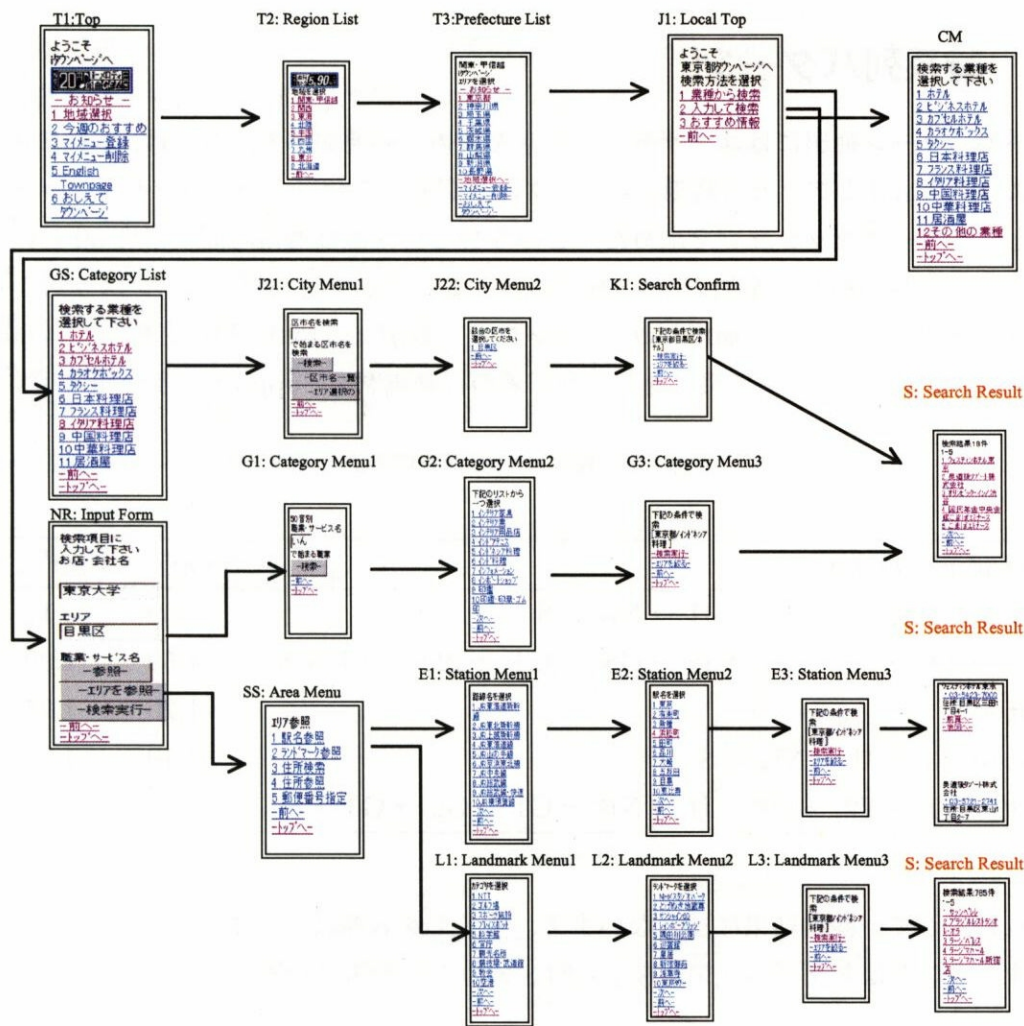


図 C.1 i タウンページ携帯用サービスのサイトマップ

ホップ数とは利用者がページとページの間をたどる回数のことをいう。一般に携帯のアクセスでは長いホップを要するサイトを嫌うといわれている。i タウンページの携帯用サイトを訪れるユーザの 48% が 4 ホップしか費やさず、一方ユーザの 30% が 10 ホップ以上費やしている。

検索が成功するまで、ユーザが平均 9.7 ホップを要する。そのうち、業種リストを経由するユーザは平均 11.4 ホップを要する。検索が成功するまでの最大ホップ数は 85 ホップである。検索結果にたどり着かない利用者の平均ホップ数は 3.1 であり、最大ホップ数は 118 である。

また、検索回数の平均値は 2.1 で、最大回数は 67 であった。

C.5 時系列パターン

時系列パターン抽出にはユーザセッションを求める必要がある。今回は、iモード版アクセスログの uid とアクセス時刻から、アクセス間隔が5分以上の部分セッションの切れ目としてユーザセッションを求めた。この5分という値は滞在時間の平均30秒にその標準偏差140秒の約2倍を加えたものに等しい。ここで示す時系列パターンは5月1日～7日の1週間分のログから抽出された。セッション数が389,890、利用者数が284,003である。ここで表C.2に検索結果に達した利用者の行動履歴をSupportの多い順に示す。

表 C.2 検索結果に達した利用者の代表的なパターン

Sequential Patterns	Supp%	User%	Repeat
start → T1 → T2 → T3 → J1 → NR ⇒ S	4.59	5.17	1.22
start → T1 → T2 → T3 → J1 → GS → J21 → J22 → K1 ⇒ S	1.59	1.99	1.10
start ⇒ S	1.25	1.25	1.37
start → T3 → J1 → NR ⇒ S	1.10	1.27	1.19
start → T1 → T2 → T3 → J1 → NR → G1 → G2 → G3 ⇒ S	1.01	1.17	1.19

それに対して、検索結果に達しないまま、サイトから離れてしまう利用者も存在する。注目すべきことはトップページからでも約3分の1の利用者が逃げる事がわかる。

表 C.3 検索結果に達しなかった利用者の代表的なパターン

Sequential Patterns	Supp%	User%	Repeat
start → T1 ⇒ end	26.33	32.86	1.10
start → T3 ⇒ end	2.29	3.01	1.04
start → T1 → T2 → T3 → J1 ⇒ end	2.26	3.07	1.01

C.6 アクセスログのヴィジュアライザ

以上の結果をグラフ表示したのが、図C.3図C.4である。基本的な考え方はウェブログの中に記録されているページ関係の情報を、相関ルールや時系列の方法でその関係の強さを見出し、有向グラフの形式に可視化している。有向グラフからはページ遷移の流れが明確になり、訪問者の行動パターンも理解できるようになる。

強い関係はグラフの上の部分に配置され太い線で描かれ、弱くなるにつれて下に配置され細い線で描かれるようになる。また関係の信頼性 (Conf) も定義され、高いものは赤い色で、低くなるにつれて青い色で描かれるようになる。

時系列のグラフ化より、以下のことを読み取ることができる。

- ページの訪問者の数
- ページからページへの遷移の流れ
- ページ間の遷移確率
- 訪問者の時間的に連続的なページ遷移の流れ
- 目的ページに辿りつく訪問者の行動パターン
- 目的ページに辿りつかない訪問者の行動パターン

また、アクセスログデータからサイトの構成を示すサイトマップを再現することが可能である。図 C.2 にその例を示す。

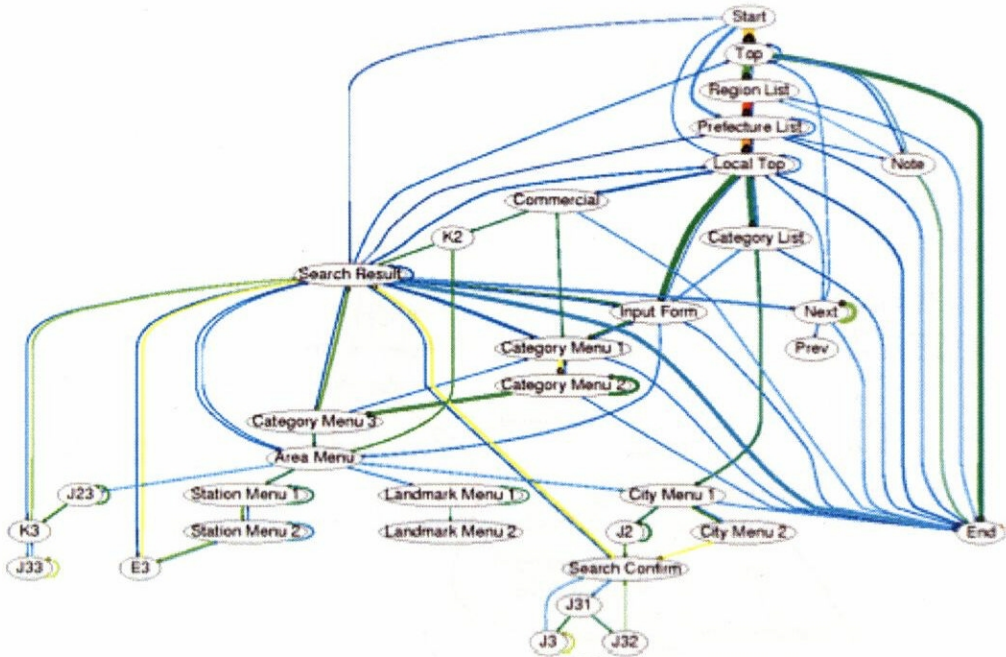


図 C.2 再現されたサイトマップ

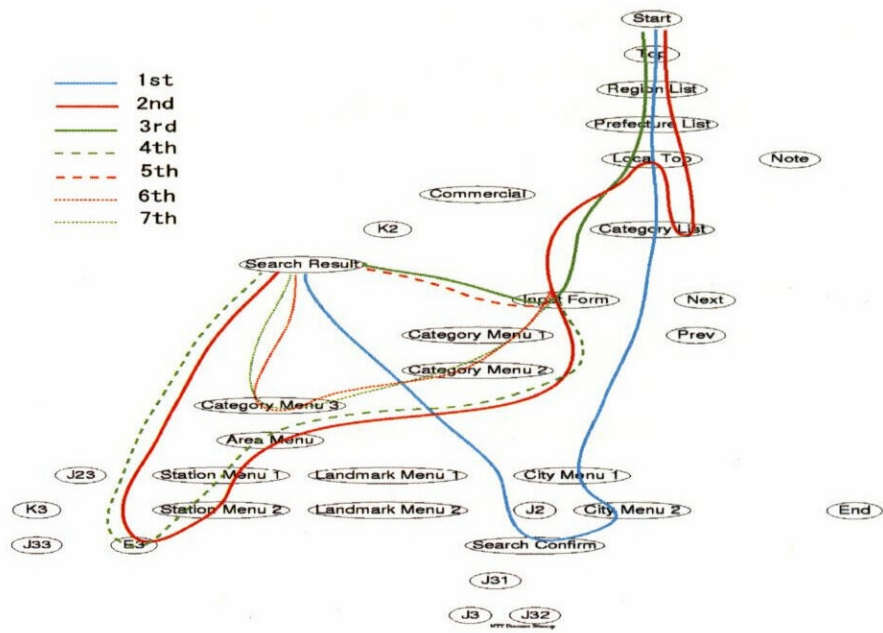


図 C.3 検索結果に達した利用者の代表的なパターン

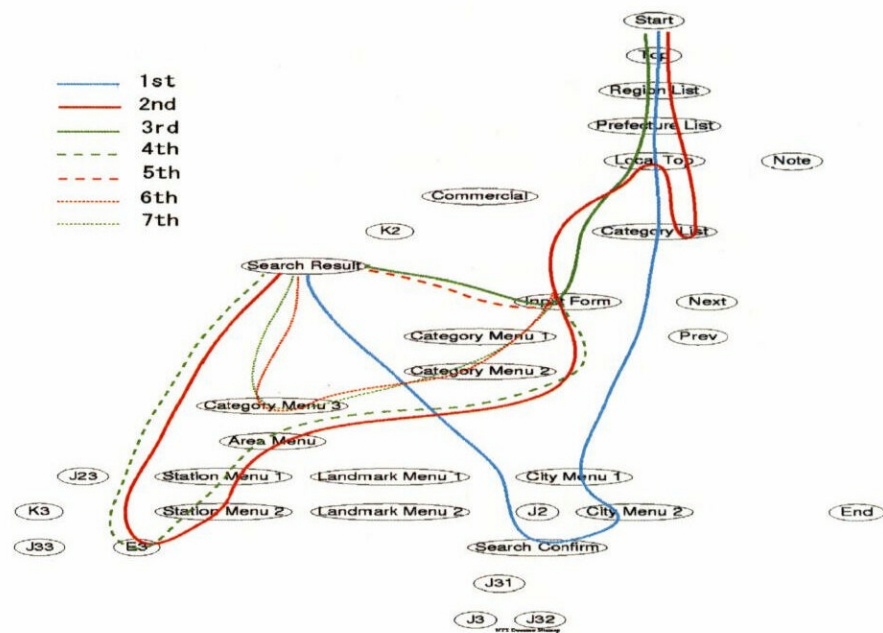


図 C.4 検索結果に達しなかった利用者の代表的なパターン

付録 D

i タウンページの業種のクラスタリング

続いて同 i タウンページのログ解析を報告して、使われている業種のクラスタリングを行った結果を示す [Ohura 2002].

D.1 概要

i タウンページへの利用者のリクエストから、リクエストに含まれる業種間の関連性を見いだすことを試みた。見いだされた関連性はシステムが提供している業種分類上類似のものだけでなく、同種に分類されていない業種間でも関連が発見できることがわかった。この関連性は利用者の検索意図（コンテキスト）を反映していると考えられる。

D.2 解析の方針

i タウンページは検索条件として住所（ジオワード）と業種または企業名が指定できる。業種の指定では、直接入力することが可能である他、50 音別の索引や業種リストを選択することによっても指定することが出来る。この業種リストは 4 層の階層構造をなしており、約 15 のトップカテゴリの下に約 80, 500, 3,000 の子階層カテゴリを有している（ここで業種総数は約 2,000 であり、これはいくつかの階層に重複して登録されている業種があるためである）。業種リスト選択の例を図 D.1 に示す。例えば、「ホテル」をリストから選択して指定するには、図のように「レジャー産業」→「旅館・ホテル業」→「ホテル」と階層をたどることによって指定することが出来る。

本解析ではシステムが提供している業種分類に着目しながら、ログデータから取り出すことのできる業種間の関係を分析する。

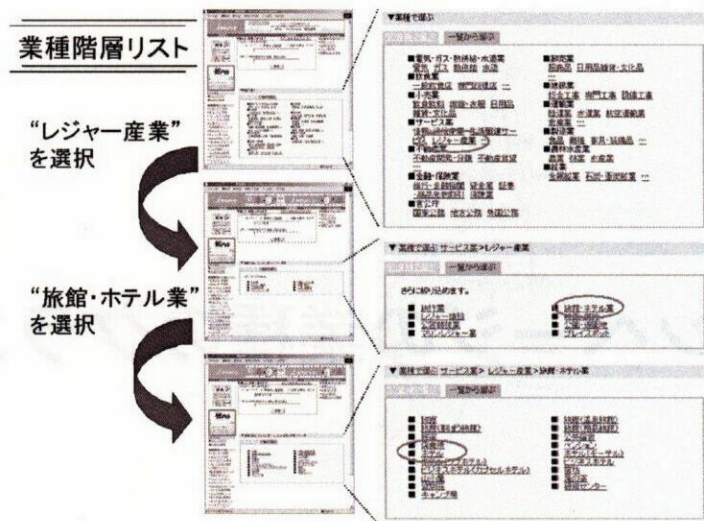


図 D.1 業種リストからの業種選択の例

D.3 検索リクエストにおける業種の性質

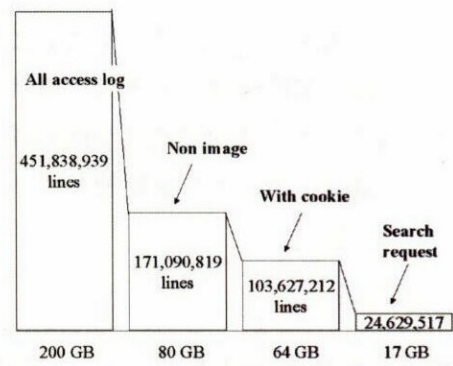


図 D.2 アクセスログのサイズ

今回、分析に使用したデータは 2000 年 2 月 1 日から 6 月 30 日までのアクセスログであり、図 D.2 に示すように総ログ数は約 4.5 億、容量約 200GB である。総ログ数の約 62% は画像データのリクエストが占め、さらに利用者 ID として用いる Cookie ID が取得出来るものはそのうちの約 60% であり、結果として分析対象はログ総数 24,629,517、容量約 17GB の検索リクエストである。

分析に当たって、ユーザセッションを作成した。セッションの判断基準は広く使用されている 30 分の閾値を使用し、各リクエスト間隔が 30 分以内のものをセッションと定義

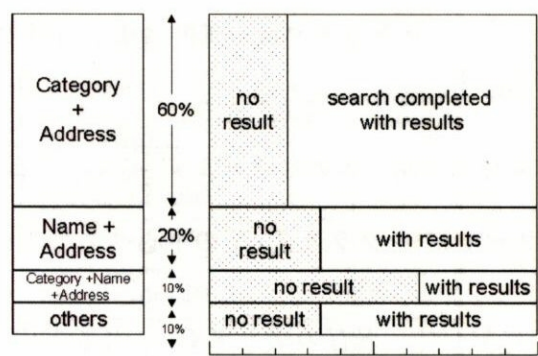


図 D.3 検索リクエストの解析結果

した。

検索リクエストを解析した結果を図 D.3 に示す。解析の結果、主たる検索方法は業種と場所の指定による検索であり、総リクエスト数の約 60% を占めることがわかった。ここで 2 つ以上の異なる業種を入力しているユーザセッションを調べたところ、業種による検索を含むセッション総数のうちの 27.2% 存在していることがわかった。これらのセッションに関して階層構造上異なる階層にある業種が入力されたセッション数を調べたところ、75.2% 存在することが判明した。このことから、設計者によって規定された業種階層構造と利用者が意識している業種分類が必ずしも一致していないか、あるいは複数目的の利用者が多いことが推測できる。一般に職業分類は時代と共に大きく変化するものであり、任意の時点で最適な業種分類を維持することは極めて困難と言えよう。以上の分析から、業種のクラスタリングを行い、業種間の関連性を調べることにする。

D.4 クラスタリングによるアクセスログの分析

利用者のリクエストによって業種の分類を行うために、解析の単位となるユーザセッションを利用者が入力した業種によるベクトルで表した。ここで業種総数を N_g とすると、 i 番目のセッション S_i のセッションベクトル s_i が業種 j を入力したかを下式によって表す。

$$s_{ij} = \begin{cases} 1 & \text{業種 } j \text{ が入力された場合.} \\ 0 & \text{業種 } j \text{ は入力されていない場合.} \end{cases} \tag{D.1}$$

クラスタリングアルゴリズムとして、一般的に広く用いられている K-means アルゴリズムを採用した。K-means は最初に K 個の代表点を選び、全てのデータ点を類似度が高いクラスに割当てて分割型クラスタリングである。しかしながら、ここでは生成すべきクラス数が未知であるため、初期値としてクラス数 K を与えるのではなく、類似

度閾値 TH_{sim} を与えることによりクラスタ数を動的に決定可能であるように改良を施した。以下にアルゴリズムを示す。

入力セッション数が N である時，入力ベクトル $\vec{s}_1, \vec{s}_2, \vec{s}_3, \dots, \vec{s}_N$ に対して，

1. 最初の入力ベクトル \vec{s}_1 をクラスタ C_1 の中心ベクトル \vec{c}_1 とし， S_1 を C_1 のメンバとする。
2. 以降，入力 \vec{s}_i に対して，既存のクラスタ $C_1 \dots C_k$ との類似度を式 (D.2) によって計算し，どのクラスタとの類似度も閾値 TH_{sim} 未満の場合は，新たなクラスタを生成してそのクラスタ中心とし，類似度が閾値 TH_{sim} 以上の場合は，最も類似度が高いクラスタのメンバとする。この時，メンバが新たに増減したクラスタはその中心ベクトルを式 (D.3) にて再計算する。
3. 割当てが収束するまで繰り返す。

$$SIM(i, j) = \frac{\vec{s}_i \cdot \vec{c}_j}{|\vec{s}_i| |\vec{c}_j|} \quad (D.2)$$

$$\vec{c}_j = \frac{\sum_{S_i \in C_j} \vec{s}_i}{M_j} \quad (D.3)$$

ここで M_j はクラスタ C_j のメンバ数である。

D.5 クラスタリング結果

ユーザセッションのクラスタリングを行うにあたって，2つ以上の業種が入力されたユーザセッションを対象とした。この時，該当するセッション数は 564,355 であった。得られたクラスタには，メンバ (セッション) 数が 1 であるクラスタも含まれる。得られた結果を解釈するにあたって，小さいクラスタは考慮しないこととし，最小クラスタサイズ MIN_{cl} よりも少ないメンバ数のクラスタはクラスタとしてカウントしないことにする。今回の実験では，この値を対象セッション数の 0.01% ($MIN_{cl} = 56.4$ 人) 未満とし，それに相当する 56 人以下のユーザセッションで構成されるクラスタは切り捨てた。また，いくつかのパラメータで試験した結果，閾値 TH_{sim} には 0.10 を用いた。その結果，得られたクラスタ数は 826 個であり，クラスタのサイズは，表 D.1 に示すように 57 セッションで構成される小さいクラスタから 21,029 セッションで構成される巨大なクラスタまで抽出された。この時，構成セッション数の平均値は 678.8，メディアンは 330 であった。

クラスタリングを行った結果のうち，「ホテル」が含まれるいくつかのクラスタを例と

表 D.1 クラスタサイズ

メンバ (セッション) 数	クラスタ数
[57 , 1000)	667
[1000 , 2000)	106
[2000 , 3000)	34
[3000 , 4000)	8
[5000 ,)	11

して表 D.2 に掲載する（その他のクラスタ例も付録に掲載してある）。結果のクラスタは、そのクラスタを構成するセッション（メンバ）数、およびクラスタを構成するユーザセッションにおいて入力された業種とその業種を入力したセッション（メンバ）数で表されている。この時、考慮する業種は、新たに設定したクラスタ内閾値 TH_{cat} 以上のセッションが入力したものに限定してある。ここでは、クラスタ内閾値 TH_{cat} をクラスタ構成セッション数の 10% とした。例えば、表 D.2 において、クラスタ 1 は 15,318 セッションから構成されるクラスタであり、そのうち「ホテル」を入力したセッション数が 15,318, 「ビジネスホテル」が 13,654, その他の業種は入力したのがクラスタ構成セッション数 15,318 の 10% 未満、つまり 1,531 以下であることを表している。結果には様々な利用目的やコンテキストを持ったユーザセッションのクラスタが得られた。例えば表 D.2 において、クラスタ 1, 2, 3, 5 は宿泊施設を探しているユーザセッションのクラスタであるが、それぞれの利用目的やコンテキストに相違がある。また、「貸会議室」を探している人や「結婚式場」を探している人が同時に「ホテル」を検索条件に入力している等、宿泊施設を探す目的以外の様々なユーザ挙動が把握出来た。

得られた結果のクラスタと i-Townpage で用いられている業種階層とを比較すると、業種階層において同一階層にない業種が同一クラスタに存在する結果が数多くみられた。例えば、図 D.4 には左側に i タウンページにおける業種階層の一部が、右側にクラスタリング結果の一部が示してある。この図において、右下のクラスタ 1 は「ホテル」と「ビジネスホテル」で構成されるクラスタであり、これら 2 つの業種は、カテゴリ階層において両者とも「旅館・ホテル業」というカテゴリの下に位置する同一階層の業種であるが、右上のクラスタ 5 では、その構成業種である「旅館・ホテル・民宿予約センター」は「旅行業」に属する業種であり、その他の構成業種である「ホテル」、「ビジネスホテル」、「旅館」、「温泉旅館」の属する「旅館・ホテル業」とは異なっている。業種「ホテル」が含まれるいくつかのクラスタ例（表 D.2）においても、「結婚式場」、「レンタカー」、「ゴルフ場」、「貸会議室」等、「ホテル」とは業種階層において離れている業種とともに検索に用いられて

表 D.2 クラスタリング結果例

#	構成メンバ数	入力された業種	入力メンバ数	#	構成メンバ数	入力された業種	入力メンバ数
1	15,318	ホテル ビジネスホテル	15,318 13,654	6	1,258	結婚式場 ホテル ▲ 会館 ▲	1,258 609 303
2	3293	旅館（温泉旅館） 温泉浴場 ▲ ホテル 温泉供給 ▲	3,293 1,153 1,145 549	7	1,158	レンタカー ホテル ▲	1,158 120
3	2,847	民宿 旅館 ホテル ビジネスホテル 旅館（温泉旅館）	2,847 2,448 899 700 295	8	811	ゴルフ場 ホテル ▲	811 155
4	1,805	会館 貸会議室 ▲ ホテル ▲ 公会堂・会館 ▲*	1,805 719 331 211	9	799	スキー場 ホテル ▲	799 126
5	1,628	旅館・ホテル・民宿予約センター ホテル △ ビジネスホテル △ 旅館 △ 旅館（温泉旅館） △	1,628 1,387 733 721 193	10	732	貸会議室 ホテル ▲ 公民館・集会場 ▲	732 215 94
				11	346	結婚披露宴演出 結婚式場 会館 ▲ 祝儀用品 ▲ ホテル ▲	346 300 104 51 42

いることがわかる。表 D.2 および付録に掲載したクラスタには、クラスタを構成するトップの業種（上記クラスタ 5 では「旅館・ホテル・民宿予約センター」）と比較して、その他の業種に関して、その業種が同一階層でない場合、「旅行業」と「旅館・ホテル業」のように親階層まで考慮すれば同一階層である場合は、その業種名の右に△を記し、1 つ上の親階層まで考慮しても同一階層にない場合にはその業種名の右に▲を記した。また、全ての業種が階層リストに分類されているわけではないため、リストに含まれていない業種に関してはその業種名の右に * を記した。

結果より、同じ「ホテル」を入力するにしても、宿泊場所を探している人もあれば、「結婚式場」を探している人、「貸会議室」を探している人と、ユーザの要求やコンテキストが異なっており、多様な要求が存在していることが把握出来た。このように、Web アクセスログのクラスタリングは、ユーザの行動理解に有効であると言える。また、カテゴリ階層では同一カテゴリに属していない業種が同一セッションで入力される場合が多いこともクラスタリングによって確認された。同一階層の業種のみで構成されるクラスタは 826 クラスタ中 134 個 (16.2%)、異なる階層の業種を含むクラスタは 692 個 (83.8%) に上った。

同一セッションにおいて異なる業種が検索に用いられているのは、ユーザの要求の多様性およびサイト構造の欠陥が考えられる。しかしながら、業種が同一で明らかに変更した方が良い場合はサイトの再構築に反映させることが出来るが、そうでない場合はサイト構造へ反映させると逆にユーザを混乱させてしまう可能性がある。例えば、表 D.2 のクラスタ 7 では、「レンタカー」を 1,158 セッションが入力しており、その約 10% の 120 セッションが「ホテル」を入力しているが、サイトに変更を加えて「レンタカー」と「ホテル」

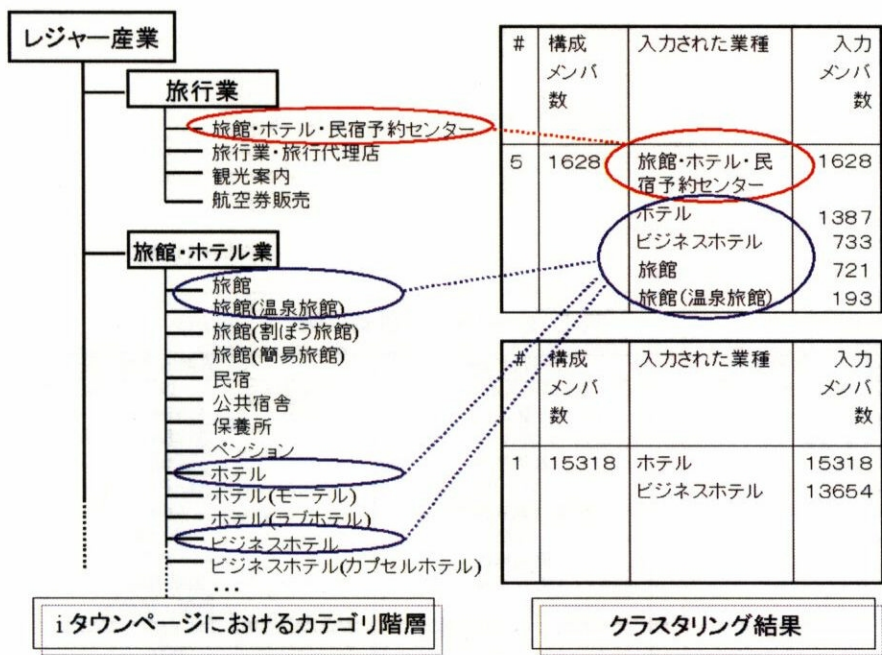


図 D.4 i-Townpage の業種階層とクラスタリング結果

を同一階層にした場合、分類の仕方によっては却ってユーザが分類のポリシーを理解し難くなる可能性がある。このように利用者の状況により、関連する業種が異なってくる場合は、固定的なメニュー提供ではなく、可変的なメニュー提供が有効である。可変的なメニュー方法を実現する一つの方法が、入力に対するキーワードの推薦である。

利用者の問い合わせ業種を含むクラスタを探し、そのクラスタに属する業種を推薦することでキーワード推薦は実現できる。推薦する業種をシステムが提供するメニューによる分類体系と照らし合わせてみる。

例えば「ホテル」が入力された際の拡張を表D.2を使用して考えた場合、「ホテル」の割合が高いクラスタ1, 5, 6, 2, 3, 10,, の順にそのメンバである業種から「ホテル」の同一階層に分類される業種を選んでみると、「ビジネスホテル」、「温泉旅館」、「民宿」などがある。一方、同一階層に分類されない業種を同様に選んでみると、「旅館・ホテル・民宿予約センター」、「結婚式場」、「貸会議室」などが存在する。表D.3に本結果を使って可能な業種拡張の例を示す。また表D.4には、その他のクラスタリング結果を示す。

表 D.3 問合せ拡張結果例

#	入力業種	同階層業種	異階層業種
1	ホテル	ビジネスホテル 旅館 ラブホテル 公共宿舍 ペンション 旅館（温泉旅館） 民宿 簡易旅館 モーテル 保養所	旅館・ホテル・民宿予約センター 宿泊施設 結婚式場 食会議室 ゴルフ場 会館 スキー場 レンタカー 温泉浴場 公民館・集会場
2	葬祭業	ペット霊園・葬祭 霊園	寺院 神社
3	ホームセンター	金物店 刃物店 日用品雑貨店 鍵	スーパーストア・マーケット 園芸店 ディスカウントショップ 生花店 園芸用機械器具 種苗商
4	バス（路線）	バス（観光） サロンバス	鉄道業 交通情報提供サービス
5	駐車場	駐車場（月極）	不動産取引
6	公園	遊園地	スポーツ施設
7	スイミング教室	フラワーデザイン教室 テニス教室 陶芸教室 スポーツ教室 絵画教室 料理教室 ゴルフ教室 手芸教室 ダンス教室 編物教室	スポーツクラブ プール スポーツ施設
8	旅行業	航空券販売 観光案内	バス（観光） 観光ホテル
9	エステティックサロン	美容院 ネイルサロン ビューティーアドバイザー 理容店 クリーニング 温泉浴場 銭湯	あん摩マッサージ指圧 各種療法（カイロプラクティック・整体） 化粧品販売

表 D.4 クラスタリング結果例

#	構成 メンバ 数	入力された業種	入力 メンバ 数	#	構成 メンバ 数	入力された業種	入力 メンバ 数
1	21,029	病院・療養所 総合病院 救急病院・救急医療センター ▲*	21,029 19,811 10,055	17	1,610	旅行業 航空券販売	1,610 367
2	15,192	飲食店 レストラン	15,192 2,320	18	1,609	携帯電話 電気通信業 通信用機械器具 ▲	1,609 545 314
3	8,886	居酒屋 飲食店 △ ろばた焼	8,886 6,041 913	19	1,385	眼科 コンタクトレンズ ▲ 病院・療養所 医院 ▲*	1,385 299 297 185
4	6,305	不動産取引 不動産鑑定 ▲ アパート・マンション ビル管理 △	6,301 2,413 1,583 769	20	1,355	寺院 神社 葬祭業 ▲	1,338 396 238
5	4,250	ビジネスホテル カプセルホテル	4,246 1,068	21	1,352	エステティックサロン 美容院 ビューティアドバイザー 化粧品販売 ▲	1,346 417 229 185
6	3,924	市区町村機関 町村役場 ▲* 都道府県機関	3,894 1,911 634	22	1,330	図書館 市区町村機関 ▲	1,330 142
7	3,522	美容院 理容店 ヘアデザイナー	3,522 1,528 415	23	1,327	あん摩マッサージ指圧 各種療法（カイロプラクティック・整体） エステティックサロン ▲	1,327 428 338
8	2,887	書店 古本 書籍販売取次業 ▲ レンタルビデオ ▲	2,887 752 563 309	24	1,252	テニスコート テニス教室 △ テニス練習場（オートテニス） スポーツ施設	1,252 635 433 323
9	2,675	おもちゃ店 ゲームソフト	2,675 375	25	1,080	駐車場 駐車場（月極） 不動産取引 ▲	1,070 513 246
10	2,410	ラーメン 中華料理店	2,410 1,048	26	1,067	葬祭業 ペット霊園・葬祭（ペット／ペット） 霊園	1,055 622 118
11	2,216	酒屋 酒類卸 ▲ コンビニエンスストア ▲* スーパーストア・マーケット ▲* ディスカウントショップ ▲*	2,216 349 293 249 239	27	904	スイミング教室 スポーツクラブ △ プール △ スポーツ教室 スポーツ施設 △	887 569 361 264 104
12	2,215	ホームセンター スーパーストア・マーケット ▲* ディスカウントショップ ▲* 日用品雑貨店	2,215 756 637 243	28	572	バス（路線） 鉄道業 ▲ 交通情報提供サービス ▲	572 376 62
13	2,131	スポーツクラブ スポーツ施設 プール	2,131 1,530 502	29	480	司法書士 行政書士 土地家屋調査士 △ 不動産取引 ▲ 測量 △	454 236 189 95 65
14	1,936	中国料理店 中華料理店	1,936 1,869	30	274	金物店 ホームセンター 刃物店 鍵	274 54 31 28
15	1,928	サウナぶろ サウナ設備 ▲ 銭湯 温泉浴場 カプセルホテル ▲	1,923 870 532 500 386	31	181	水族館 動物園 遊園地 ▲ 博物館 公園 ▲	181 44 35 27 19
16	1,883	ペンション 民宿 ホテル 旅館 旅館（温泉旅館） ビジネスホテル	1,883 952 769 687 353 191				

謝辞

本研究を進めるにあたって、数多くの方々の御指導や御協力をいただきました。ここの感謝の意を表します。

指導教官である東京大学生産技術研究所の喜連川優教授には高い目標を与えていただき、また辛抱強く指導して下さったことに深く感謝の意を表します。

主査である国立情報学研究所の安達淳教授を始め、東京大学の石塚満教授、中川裕志教授、相田仁教授、京都大学の黒橋禎夫教授には論文審査を通じて大変有益な御指導を賜りました。厚く御礼を申し上げます。

本研究は著者が日本電信電話株式会社で開始し、継続して東京大学生産技術研究所で行った研究を集大成したものである。

6章で述べた近似文字列照合の研究はNTT 情報通信処理研究所で、同基礎研究所の梅村恭司氏（現豊橋技術科学大学教授）で行ったもので、プログラミングの手ほどきのみならず、ネットワークで情報を共有する検索システムのアイデアを授けていただいた。この時期には情報通信処理研究所で番号案内システムの開発に従事し、橋田幸雄氏、武藤信夫氏、戸部美春氏他各位から電話帳検索の基礎的な技術を学んだ。

3章で述べたタウンページの研究はNTT ソフトウェア研究所で島健一氏、三浦信幸氏（ともに現NTT ドコモ）で行ったもので、短期間で企画から開発を行い、自分たちの手で、サービス導入までを密な連携で行った。さらにこの時NTT 研究所の伊土誠一氏、市川晴久博士、伊藤正樹氏、後藤厚弘博士、高橋健司博士をはじめとする多くの方々の手により研究環境の実現と、成果の実サービス実現がなされた。

4章で述べたモバイルインフォサーチの研究は、タウンページのチームに横路誠司氏（現NTT ドコモ）が加わり実現することができた。またタウンページからモバイルインフォサーチを実現するに至る時期には高田敏弘氏、坂本仁明氏らからインターネットに関する貴重な議論をしていただいた。

3章で述べたエージェントをベースとした研究は、NTT 情報通信研究所で西部喜康氏で行った。この研究はNTT と関西の4大学の共同研究プロジェクトを通じて生まれてい

る。この研究のみならずモバイルインフォサーチを含むいくつかのアイデアの元が本共同研究プロジェクトで鍛えられた。京都大学石田亨教授、奈良先端科学技術大学西田豊明教授（現東京大学）、神戸大学田中克己教授（現京都大学）、大阪大学西尾章治郎教授の各教授と研究室の各位、ならびに NTT の服部文夫氏（現立命館大学教授）、森原一郎氏と同所属グループの各位からは指導ならびに多くの示唆をいただいた。

5 章で述べたジオワード・マイニングの研究は東京大学生産技術研究所喜連川研究室のイコ プラムディオノ博士（現 NTT）と行った。またログマイニングを一緒に行った、同研究室の大浦亮勇氏、Bowo Prasetyo 氏、NTT の市川俊一氏とは楽しく議論をさせていただいた。さらに喜連川研究室の相良毅博士、豊田正史特任助教授、大塚真吾博士、合田和生博士、中野美由紀助手をはじめとする各位からは常に貴重な助言と支援をいただいた。

i タウンページの実現ならびにデータマイニングに関しては、NTT 番号情報（株）の関係各位には多大なる支援をいただいた。特に佐伯秀男氏、小畠伸和氏の両氏には i タウンページ実現の当初より永きにわたり研究活動に理解いただくだけでなく、この電話帳事業の面白さを教えていただいた。

また大学での研究と職務を両立させるために、NTT 情報流通プラットフォーム研究所の工藤明彦所長をはじめとする多くの方々に寛容かつ暖かく見守っていただいた。

以上の各位に関して、改めて深く感謝を申し上げる。

最後に常に暖かく励まし支えてくれた妻高橋啓子に感謝をする次第である。

参考文献

- [Agrawal 1994] R. Agrawal and R. Srikant. Fast Algorithms for Mining Association Rules. In Proc. of the 20th Int. Conf. on Very Large Data Bases(VLDB), 1994
- [Anitay 2004] E. Amitay, N. Har'El, R. Sivan, A. Soffer. Web-a-Where: Geotagging Web Content. Proceedings of the ACM SIGIR2004 2004
- [Ask 2004] Ask Jeeves Press Release. Ask Jeeves Teams With Citysearch to Create Leading Local Search Experience. http://www.irconnect.com/askjinc/pages/news_releases.html?d=61820 2004.8
- [Baeza-Yates 2004a] Baeza-Yates. Query Recommendation using Query Logs in Search Engines. In International Workshop on Clustering Information over the Web (ClustWeb, in conjunction with EDBT). 2004
- [Baeza-Yates 2004b] R. Baeza-Yates. Query Usage Mining in Search Engines. Web Mining: Applications and Techniques, Anthony Scime, editor. Idea Group, 2004
- [Beeferman 2000] D. Beeferman and A. Berger. Agglomerative Clustering of a Search Engine Query Log. The 6th ACM SIGKDD Intl Conf on Knowledge Discovery and Data Mining (KDD2000), 2000
- [Bowo 2002] P. Bowo, P. Iko, K. Takahashi, M. Kitsuregawa. Naviz: Website Navigational Behavior Visualizer. Advances in Knowledge Discovery and Data Mining 6th Pacific-Asia Conference, PAKDD2002. 2002
- [Brin 1998] S. Brin, R. Page. The Anatomy of a Large-scale Hypertextual Web Search Engine. Computer Networks, Vol.30, No.1-7, pp.161-117. 1998
- [Buyukkokten 1999] O. Buyukkokten, J. Cho, H. Garcia-Molina, L. Gravano, and N. Shivakumar. Exploiting geographical location information of web pages. In Proceedings of the ACM SIGMODWorkshop on the Web and Databases, WebDB. 1999
- [Chen 2006] Y.Y. Chen, T Suel, A. Markowetz. Efficient Query Processing in Geographic Web Search Engines. ACM SIGMOD 2006. 2006

- [Cheong 1996] F. C. Cheong. Internet Agents: Spiders, Wanderers, Brokers, and Bots. New Riders Publishing 1996
- [Childs 2001] Colin Childs. Migrating Coverages to Geodatabases. ArcUser. <http://www.esri.com/news/arcuser/0701/migrating.html> 2001
- [Cleventon 1962] C. W. Cleventon. Report on the Testing and Analysis of an Investigation into the Comparative Efficiency of Indexing Systems. Aslib Cranfield Research Report, Cranfield. 1962
- [Cowie 1996] J. Cowie, W. Lehnert. Information Extraction. Communication of CM, Vol.39, No.1. 1996
- [Ding 2000] Junyan Ding, Luis Gravano, and Narayanan Shivakumar. Computing geographical scopes of web resources. In 26th International Conference on Very Large Databases, VLDB 2000, Cairo, Egypt. 2000
- [Doorenbos 1996] R. Doorenbos, O. Etzioni, D. Weld. A scalable comparison-shopping agent for the World Wide Web. Technical Report UW-CSE-96-01-03 University of Washington. 1996
- [Ester 2001] M. Ester, H.-P. Kriegel, and J. Sander. Algorithms and Applications for Spatial Data Mining . In Geographic Data Mining and Knowledge Discovery, Research Monographs in GIS, Taylor and Francis. 2001
- [Everitt 1993] B. S. Everitt. Cluster Analysis. Edward Arnold, third edition. 1993
- [Fielding] R. Fielding. libwww-perl: WWW Protocol Library for Perl. <http://www.ics.uci.edu/pub/websoft/libwww-perl/> .
- [Gaihua 2005] Gaihua Fu, Christopher B. Jones and Alia I. Abdelmoty. Ontology-based Spatial Query Expansion in Information Retrieval. Int'l Conf on Ontologies, Databases and Applications of SEmanantics (ODBASE 2005). 2005
- [Genesereth 1994] M.R. Genesereth, S.P. Ketchpel. Software Agents. Communications of the ACM Vol.37 No.7. 1994
- [Geosearch 2000] Vicinity, Northern Light Partner to Offer Geographically Oriented Web-Search Capability. Information Today, Volume 17, Number 6. <http://www.infotoday.com/it/jun00/news5.htm> 2000
- [Guting 1994] R. Guting, An Introduction to Spatial Database Systems, VLDB Journal, Vol.3, No.4. 1994
- [Guttman 1984] Antonin Guttman. R-Trees: A Dynamic Index Structure for Spatial Searching, SIGMOD'84, Proceedings of Annual Meeting, Boston, Massachusetts, ACM Press. 1984
- [Hall 1980] P.A.V. Hall, G. Dowling. Approximate String Matching. ACM Computing

- Survey, Vol.12, No.4 pp.381-402. 1980
- [Han 1997] J. Han, K. Koperski and N. Stefanovic. Geominer: A system prototype for spatial data mining. In Proc. of the ACM SIGMOD Conf. on Management of Data. 1997
- [Han 2000] J. Han, J. Pei and Y. Yin. Mining Frequent Pattern without Candidate Generation. In Proc. of the ACM SIGMOD Conf. on Management of Data 2000
- [Himmelstein 2005] Mark Himmelstein. Local Search: The Internet is the Yellow Pages. Computer, IEEE Computer Society, 0018-9162/05, pp.26-34. (Feb 2005)
- [Iko 2002] P. Iko, T. Shintani, K. Takahashi, M. Kitsuregawa. User Behavior Analysis of Location Aware Search Engine. In Proc. of International Conference On Mobile Data Management (MDM02). 2002
- [Jones 2004] C.B. Jones, A. I. Abdelmoty, D. Finch, G. Fu, S. Vaid. The SPRIT Spatial Search Engine: Architecture, Ontologies and Spatial Indexing. Proceedings of the Intl Conference on Geographic Information Science (GIScience). 2004
- [Jones 2006] R. Jones, B. Rey, O Madani, W. Greiner. Generating Query Substitutions. World Wide Web Conference (WWW 2006), Edinburgh. 2006
- [Kitsuregawa 2001] M. Kitsuregawa, M. Toyoda, I. Pramudiono. WEB community mining and WEB log mining: Commodity Cluster based Execution. ADC 2002
- [Knuth 1970] D.E. Knuth, P.G. Bendix. Simple word problem in universal Algebras. Computational problems in abstract algebra. pp.263-297, Pergamon Press. 1970
- [Koperski 1995] K. Koperski, and J. Han. Discovery of Spatial Association Rules in Geographic Information Databases. In Proc. of the Int. Symposium on Spatial Databases (SSD), 1995
- [Krulwich 1995] B. Krulwich. Bergen Finder agent prototype. Technical Report. Andersen Consulting. 1995
- [Lewis 1991] J.W. Lweis. Wrappers:integration utilities and services for the DICE architecture. In proceedings of the Second National Symposium on Concurrent Enginnerig, pp.445-457, Concurrent Enginnering Research Center. 1991
- [Markowetz 2005] A. Markowetz, Y.Y. Chen, T Suel, X. Long, B. Seeger. Design and Implementation of a Geographic Search Engine. Proceedings of the 8th Intl Workshop on the Web and Databases (WebDB 2005). 2005
- [McCurley 2001] K. McCurley. Geospatial Mapping and Navivagtion of the Web. Prcceedings of the 10th World Wide Web Conference, WWW10. 2001
- [Neches 1991] R. Neches, R. Fikes, T. Finin, T. Gruber, R. Patil, T. Senator, and W. Swartout. Enabling technology for knowledge sharing. AI Mag. Vol. 12, No. 3.

1991

- [Odell 1918] M.K. Odell, R.C. Russell: U.S. Patent, 1-261-167. 1918
- [Odell 1922] M.K. Odell, R.C. Russell: U.S. Patent, 1-435-663. 1922
- [Ohura 2002] Y. Ohura, K. Takahashi, P. Iko, M. Kitsuregawa. Experiments on Query Expansion for Internet Yellow Page Services Using Web Log Mining. 28th International Conference on Very Large Data Bases (VLDB2002). 2002
- [Oracle 2005] Oracle 10g, Oracle Spatial and Oracle Locator. <http://www.oracle.com/technology/products/spatial/> 2005
- [Rijsbergen 1979] C. van Rijsbergen. INFORMATION RETRIEVAL, 2nd Edition, Chapter 7, Butterworths, London. 1979
- [Robbins 1991] C.J. Robbins, S.E.Kille. The ISO Development: User's Manual Volume 5: QUIPU, p.35. 1991
- [Sagara 2004] T.Sagara, M.Kitsuregawa. Yellow Page driven Methods of Collecting and Scoring Spatial Web Documents. in SIGIR Workshop on Geographical Information Retrieval. 2004
- [Sanderson 2004] M. Sanderson, and J. Kohler, Analyzing Geographic Queries. Workshop on Geographic Information Retrieval, SIGIR 2004
- [Srikant 1995] R. Srikant, and R. Agrawal. Mining Generalized Association Rules. In Proc. of the Int. Conf. on Very Large Data Bases(VLDB). 1995
- [Srikant 1996] R. Srikant, and R. Agrawal. Mining Sequential Patterns: Generalizations and Performance Improvements. In Proc. of 5th Int. Conf. on Extending Database Technology(EDBT). 1996
- [Takahashi 1997] K. Takahashi, Y. Nishibe, I. Morihara, F. Hattori, Intelligent Pages: Collecting Shop and Service Information with Software Agents, Journal of Applied Artificial Intelligence, Vol 11, pp.489-499, Taylor & Francis. 1997
- [Tezuka 2006] Taro Tezuka, Takeshi Kurashima, Katsumi Tanaka. Toward Tighter Integration of Web Search with a Geographic Information System. WWW 2006, Edinburgh, 2006
- [Toyoda 2001] M. Toyoda, M. Kitsuregawa. Creating a web community chart for navigating related communities. Hypertext 2001
- [Vaid 2005] S. Vaid, C. Jones, H. Joho, M. Sanderson. Spatio-Textual Indexing for Geographic Search on the Web. Proceedings of the 9th Intl Symposium on Spatial and Temporal Database, STTD2005. 2005
- [Wen 2002] J. Wen, J. Nie, and H. Zhang. Query Clustering using User Logs. ACM Transactions on Information Systems (ACM TOIS), Vol.20, No.1, pp.59-81. 2002

- [White 1994] J. White. Telescript Technology-The foundation for the electronic marketplace, White Paper, General Magic Corp. 1994
- [Wiederhold 1994] G. Wiederhold. Interoperation, mediation, and ontologies. FFCS '94 Workshop on Heterogeneous Cooperative Knowledge-Base. 1994
- [Yahoo 2004] Yahoo Press Release. YAHOO! LAUNCHES LOCAL SEARCH BETA: AN INNOVATIVE WAY FOR PEOPLE TO SEARCH FOR LOCAL INFORMATION ONLINE. <http://docs.yahoo.com/docs/pr/release1175.html> 2004.8
- [Zhang 2005] J. Zhang, Y. Ishikawa, S. Kurokawa, H. Kitagawa. LocalRank: Ranking Web Pages Considering Geographical Locality by Integrating Web and Databases. In Proc. DEXA 2005, LNCS 3588, pp. 145-155. 2005
- [google 2002] 2002 Google Programming Contest Winner: Daniel Egnor, Project title: Geographic Search. <http://www.google.com/programming-contest/winner.html> . 2002
- [google 2004] Google Press Release. Google Connects Searchers With Local Information. <http://www.google.com/press/pressrel/local.html> 2004
- [sheth 1990] A. Sheth and J. Larson. Federated database systems for managing distributed, heterogeneous, and autonomous databases, ACM Trans. Database System, Vol. 22, No.3. 1990
- [Wang 2005] L. Wang, C. Wnag, X. Xie, J. Forman, Y. Lu, W.Y. Ma, Y. Li. Detecting Dominant Locations from Search Queries. SIGIR 05 Brazil. 2005
- [whois] WHOIS Protocol Specification. RFC 3912. <http://www.rfc-editor.org/rfc/rfc3912.txt> 2004
- [wikipedia] <http://www.wikipedia.org/>
- [セジウィック 1993] R. セジウィック: アルゴリズム, Vol.3, pp.21-21, 近代科学社 1993
- [タウンページ] インターネットタウンページ. <http://itp.ne.jp/>
- [マピオン] マピオン. <http://www.mapion.co.jp/>
- [宇土 1993] 宇土行良, 遠山潤, 森茂樹. 日本人著者の名前についての調査, 図書館学会年報, Vol.39, No.1, pp.13-24 1993
- [横路 2000] 横路誠司, 高橋克巳, 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
- [岩爪 1997] 岩爪道昭, 白神謙吾, 畑谷和右, 武田英明, 西田豊明. オントロジーに基づく広域ネットワークからの情報収集・分類・統合化. 情報処理学会論文誌, Vol.38, No.3.
- [宮部 1983] 宮部博, 大山実, 本郷郁夫. 名義検索システムー電話番号案内業務への適用ー, 情報処理学会論文誌 Vol.24, No.4. pp.841-850. 1983

- [戸部 1990] 戸部美春, 武藤信夫, 山本康二. 高付加価値型番号案内システム (CUPID) の電話帳検索方式. NTT R&D, Vol.39, No.6, pp.841-850. 1990
- [高橋 1992] 高橋克巳, 岩瀬成人: 人名の読みからの検索法, 情報処理学会自然言語処理研究会報告, 92-NL-91-4, pp.25-32. 1992
- [高橋 1996] 高橋克巳, 高田敏弘, 三浦信幸, 島健一. 電話帳情報のハイパーテキスト化のデザイン. 第7回データ工学ワークショップ (DEWS96), 電子情報通信学会 データ工学専門委員会. 1996
- [高橋 1997] 高橋克巳, 三浦信幸, 島健一. インターネット・タウンページの構築 (2) ～あいまい検索～. 情報処理学会第54回全国大会. 1997
- [高橋 1998] 高橋克巳, 三浦信幸, 横路誠司, 島健一. Mobile Info Search: Information Integration for Location Aware Computing. 情報処理学会 モバイルコンピューティング研究会 (インターネット研究会 合同研究会) 論文集. pp.115-122. 1998
- [高橋 2000] 高橋克巳, 三浦信幸, 横路誠司, 島健一. Mobile Info Search: Information Integration for Location-Aware Computing. 情報処理学会論文誌, Vol.41, No.4, pp.1192-1201. 2000
- [高橋 2002] 高橋克巳. インターネットの知的情報技術 第1巻 情報検索とエージェント 第4章モバイル情報検索. 東京電機大学出版局. 2002
- [国語課 1992] 文化庁国語課 国語研究会編集. 国語表記事務必携 [改定版], ぎょうせい. 1992
- [国語研 1955] 国立国語研究所. 読みの実験的研究－音読にあらわれた読みあやまりの分析－. 国立国語研究所報告 9, 秀英出版 1955
- [国語研 1983] 国立国語研究所. 現代表記のゆれ, 国立国語研究所報告 75, 秀英出版 1983
- [国語研 1985] 国立国語研究所: 語彙の研究と教育 (上)(下), 大蔵省印刷局 1985
- [国語研 1988] 国立国語研究所: 文字・表記の教育, 大蔵省印刷局 1988
- [国土情報整備室 2003] 国土交通省 国土計画局 国土情報整備室: 街区レベル位置参照情報 (表) 平成13年 (1.0) および平成15年 (2.0a)
- [佐藤 1989] 佐藤大和. 複合語におけるアクセント規則と連濁規則, 日本語と日本語教育, 第2巻 日本語の音声・音韻 (上), pp.223-265. 明治書院. 1989
- [佐藤 1995] 佐藤円, 佐藤理史, 篠田陽一. 電子ニュースのダイジェスト自動生成, 情報処理学会論文誌, Vol.36, No.10. 1995
- [三浦 1997] 三浦 信幸, 高橋克巳, 坂本仁明, 島健一. モバイルインフォサーチ: 移動環境下でのユーザ指向型 WWW 検索. 情報処理学会 モバイルコンピューティング研究会論文集, pp.131-136. 1997
- [三浦 1998] 三浦信幸, 高橋克巳, 横路誠司, 島健一. 情報分布を考慮した外部リソースの位置指向情報検索. 情報処理学会 全国大会論文集, Vol.57, pp.165-166. 1998

- [三浦 1999] 三浦信幸, 横路誠司, 井上香織, 高橋克巳, 高橋健司, 島健一. 位置指向の情報構造化と情報分類 ～モバイルインフォサーチ 3 実験～. 情報処理学会 第 11 回 モバイルコンピューティング研究会. 愛媛大学. 1999
- [獅々堀 1994] 獅々堀正幹, 津田和彦, 青江順一. 片仮名異表記の生成, 電子情報通信学会論文誌, Vol.J77-D-II, No.2. pp.380-387, 1994
- [小松 1981] 小松英雄. 日本語の音韻, 日本語の世界 7, 中央公論社 1981
- [相良 2003] 相良毅. 非構造・半構造空間情報の高度利用に関する研究. 東京大学学位論文, p.20. 2003
- [大久保 1998] 大久保雅, 杉崎正行, 井上孝史, 田中一男. WWW 検索ログに基づく情報ニーズの抽出. 情報処理学会論文誌, Vol.39, No.7, pp.2250-2258. 1998
- [大塚 2005] 大塚真吾, 豊田正史, 喜連川優. 大域ウェブアクセスログを用いた関連語の発見法に関する一考察. 情報処理学会論文誌 データベース, Vol.46, No.SIG18(TOD26), pp. 82-92. 2005
- [丹羽 1985] 丹羽基二. 日本ユニバック：日本姓氏大辞典, 角川書店. 1985
- [中井 1994] 中井検裕. 地域統計 人文・社会科学の統計学 東京大学教養学部統計学教室編, 東京大学出版会. 1994
- [田中 1975] 田中康仁. 姓名のカナ漢字変換システム, 情報処理, Vol.16, No.3, pp.230-238. 1975
- [田中 1977] 田中康仁. カナ漢字のデータチェッカー特に名前のエラーチェックについて, ユニバック研究会, SYSTEMS No.122, pp.41-47. 1977
- [島 1997a] 島健一, 高橋克巳, 三浦信幸. インターネット・タウンページの構築 (1) ～概要～. 情報処理学会第 54 回全国大会. 1997
- [島 1997b] 島健一, 高橋克巳, 三浦信幸. あいまいな単語からも地図からも電話帳検索が可能 -インターネットタウンページの構築-. NTT 技術ジャーナル Vol. 9, No. 5, pp. 72-5. 1997
- [湯沢 1987] 湯沢質幸. 漢字の慣用音, 漢字講座 3, 漢字と日本語, pp.82-111, 明治書院. 1987
- [藤堂 1987] 藤堂明保編. 漢和大辞典, 学習研究社 1978
- [藤澤 1993] 藤澤浩道, 絹川博之. 情報検索における自然言語処理, 情報処理, Vol.34, No.10, pp.1259-1265. 1993
- [梅村 1990] 梅村恭司, 山崎憲一. サーバクライアントモデルを使った広域データベースの試み. 第 41 回情報処理学会全国大会. 1990
- [武田 1996] 武田英明. ネットワークを利用した知的情報統合. 人工知能学会, Vol.11, No.5. 1996
- [平松 2000] 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

[林 1996] 林良彦, 菊井玄一郎, 鷺崎誠司. インターネット情報ナビゲーションサービス TITAN (タイタン). NTT 技術ジャーナル, Vol.8, No.8, pp.20-23. 1996

[鷺坂 1997] 鷺坂 光一, 山崎 憲一, 廣津 登志夫, 尾内 理紀夫. 情報検索のための高速日本語形態素解析システム「すもも」. 情報処理学会 全国大会論文集, Vol.54, pp.2, 59-60. 1997

発表文献

論文誌

1. 高橋克巳, Iko Pramudiono, 喜連川優. ウェブアクセスログからの地理的な相関ルール発見法, 情報処理学会論文誌 (投稿中). 2006
2. 高橋克巳, 三浦信幸, 横路誠司, 島健一. Mobile Info Search: Information Integration for Location-Aware Computing. 情報処理学会論文誌, Vol.41, No.4, pp.1192-1201. 2000
3. 横路誠司, 高橋克巳, 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 (平成 12 年度情報処理学会論文賞受賞)
4. 高橋克巳, 三浦信幸, 西部喜康, 島健一. 不均一で分散した情報の構造情報集との関連付けによる統合 —情報統合ディレクトリー. 電子情報通信学会論文誌, ソフトウェアエージェントとその応用論文特集, Volume J81-D1 No.5, pp.443-450. 1998
5. 三浦信幸, 高橋克巳, 島健一. 個人適応型 WWW におけるユーザモデル構築法. 情報処理学会論文誌 Vol. 39, No. 5, pp. 1523-1535. 1998
6. Katsumi Takahashi, Yoshiyasu Nishibe, Ichiro Morihara, Fumio Hattori. Intelligent Pages: Collecting Shop and Service Information with Software Agents. Special Issue on the Best of PAAM'96, Journal of Applied Artificial Intelligence, Vol 11, pp489-499, Taylor & Francis. 1997
7. 高橋克巳, 梅村恭司. 人名のかな表記のゆれに基づく近似文字列照合法. 情報処理学会論文誌, Vol. 36, No. 8, pp. 1906-1915. 1995

国際会議

1. Katsumi Takahashi, Iko Pramudiono, Masaru Kitsuregawa. Geo-word Centric Association Rule Mining. 6th International Conference on Mobile Data Management (MDM2005). pp.273-280. Cyprus. 2005.5
2. Yusuke Ohura, Katsumi Takahashi, Iko Pramudiono, Masaru Kitsuregawa. Experiments on Query Expansion for Internet Yellow Page Services Using Web Log Mining. 28th International Conference on Very Large Data Bases (VLDB2002), Hong Kong 2002.8.20-23
3. Iko Pramudiono, Takahiko Shintani, Katsumi Takahashi, Masaru Kitsuregawa. User Behavior Analysis of Location Aware Search Engine. International Conference On Mobile Data Management (MDM'02), IEEE Computer Society Press, pp. 139-145. 2002.1
4. Bowo Prasetyo, Iko Pramudiono, Katsumi Takahashi, Masaru Kitsuregawa. Naviz: Website Navigational Behavior Visualizer. Advances in Knowledge Discovery and Data Mining 6th Pacific-Asia Conference, PAKDD2002. Taipei, Taiwan. 2002.5
5. Yokoji Seiji, Katsumi Takahashi, Nobuyuki Miura. Kokono Search: A Location Based Search Engine. 10th International World Wide Web Conference. <http://www10.org/cdrom/posters/p1146/index.htm> Hong Kong, 2001.5.1
6. Rie Ohtsubo, Katsumi Takahashi, Yoshiyasu Nishibe, Ichiro Morihara. Action Navigator: An Information Service Based on Agent-Communication for Supporting Decision Making. The Third International Conference and Exhibition on the Practical Application of Intelligent Agents and MultiAgents (PAAM 98). London. 1998.3.23
7. Katsumi Takahashi, Yoshiyasu Nishibe, Ichiro Morihara, Fumio Hattori. Intelligent Pages: Collecting Shop and Service Information with Software Agents. The First International Conference and Exhibition on the Practical Application of Intelligent Agents and MultiAgents (PAAM 96). London, 1996.4.22

書籍・学会誌

1. 高橋克巳. インターネットの知的情報技術 第1巻 情報検索とエージェント 第4章 モバイル情報検索. 東京電機大学出版局. 2002.3

2. Katsumi Takahashi, Nobuyuki Miura, Seiji Yokoji. Location Oriented Integration of Internet Information - Mobile Info Search. Toru Ishida and Katherine Isbister (Eds): Digital Cities, Technologies, Experiences, and Future Perspectives [the book is based on an international symposium held in Kyoto, Japan, in September 1999]. LNCS Vol. 1765/2000, Springer. 2000
3. 高橋克巳. モバイル環境下で情報収集を支援するエージェント. 人工知能学会誌, Vol.14, No.4, pp.590-597, 1999
4. 高橋克巳, 寺岡文男, 小橋喜嗣. 高機能化携帯電話によって提案される新規社会情報基盤. 情報処理, Vol.40, No.6, pp.610-614, 1999

国際ワークショップ

1. Katsumi Takahashi. Location-based search services. Keynote Speech of Mobile Location Workshop - MLW 2001. Espoo, Finland. 2001.6.7-8
2. Katsumi Takahashi, A Mobile Portal Service to Provide Location Dependent Information. The Joint W3C-WAP Forum workshop on Position dependent information services. <http://www.w3.org/Mobile/posdep/> INRIA Sophia Antipolis, France. 2000.2.15
3. Katsumi Takahashi, Nobuyuki Miura, Seiji Yokoji. Location Oriented Integration of Internet Information - Mobile Info Search. Kyoto Workshop on Digital Cities. 於 京都大学. 1999.09.16-18

国内会議等

1. 高橋克巳, Iko Pramudiono, 喜連川優. 位置指向の Web ログマイニング. 情報処理学会 情報処理研究報告 (夏のデータベースワークショップ DBWS2004) Vol.2004, No.71, p.9. 松山. 2004.7.13-15
2. 高橋克巳. 情報検索と異種情報の統合. 情報処理学会東海支部専門講習会「インターネット／イントラネットにおけるスマートシステム」. 名古屋工業大学. 2002.3.6
3. 高橋克巳. サーチエンジンの技術動向. 日本ディレクトリ学会第 6 回全国大会. 於 日本大学国際関係学部. 2002.7.12
4. Bowo Prasetyo, Iko Pramudiono, Katsumi Takahashi, Masashi Toyoda, Masaru Kitsuregawa. Naviz: User Behavior Visualization of Dynamic Page. 電子情報通信学会 第 13 回データ工学ワークショップ (DEWS2002), 倉敷, C5-3, 2002.3

5. Bowo Prasetyo, Iko Pramudiono, Katsumi Takahashi, Masashi Toyoda, Masaru Kitsuregawa. Naviz: User Behavior Visualization System using Web Access Log. 情報処理学会 第 64 回全国大会, 東京電機大学鳩山キャンパス, 6X-01, 2002.3
6. 高橋克巳. インターネットコンテンツと「ここの」サーチ. データベースと Web 情報システムに関する IPSJ DBS/ACM SIGMOD Japan Chapter/JSPS-RFTF AMCP 合同シンポジウム (DBWeb2000), 空間メディアワークショップ. 2000.12.6
7. 高橋克巳. インターネットにおける位置指向の情報検索. ACM SIGMOD 日本支部第 15 回大会. 於 東大生産研 2000.7.5.
8. 三浦信幸, 横路誠司, 井上香織, 高橋克巳, 高橋健司, 島健一. 位置指向の情報構造化と情報分類 ～モバイルインフォサーチ 3 実験～. 情報処理学会 第 11 回 モバイルコンピューティング研究会. 愛媛大学. 1999.11.25-26 (優秀賞受賞)
9. 高橋克巳. インターネットにおけるタウン情報の展開. HMC 地図情報フォーラム, ハイパーメディアコンソーシアム. 1999.2.16
10. 井上香織, 横路誠司, 高橋克巳. 不定形な広告情報の自動構造化. 情報処理学会第 132 回自然言語処理研究科. 1999.7
11. 高橋克巳. 情報検索と異種情報の統合. 情報処理学会東海支部 チュートリアル講演会. 名古屋工業大学. 1998.3.6
12. 大坪理恵, 高橋克巳, 三浦信幸, 島健一. WWW サーバの利用記録に基づく検索行動の分析. 信学会 データ工学研究会. 1998.10.13
13. 三浦信幸, 高橋克巳, 横路誠司, 島健一. 位置指向の情報統合 ～モバイルインフォサーチ 2 実験～. 情報処理学会 第 57 回 全国大会 (第 3 分冊 pp.637～638). 於 名古屋大学工学部. 1998.10.5-7
14. 三浦信幸, 横路誠司, 高橋克巳, 島健一. 情報分布を考慮した外部リソースの位置指向情報検索. 情報処理学会 第 57 回 全国大会 (第 3 分冊 pp.165～166). 於 名古屋大学工学部. 1998.10.5-7
15. 三浦信幸, 横路誠司, 高橋克巳, 島健一. GIS を用いた位置指向の WWW サーチエンジン ～モバイルインフォサーチ 2 実験～. 地理情報システム学会 第 7 回研究発表大会. 於 工学院大学. 1998.10.15-17
16. 三浦信幸, 高橋克巳, 島健一. 個人適応型 WWW のためのユーザの行動モデリング. 情報処理学会. 情報処理学会第 56 回全国大会. 1998.3.17
17. 高橋克巳. モバイルコンピューティングとエージェント技術の応用. 計測自動制御学会 第 14 回ヒューマン・インタフェース・シンポジウム (HIS'98) チュートリアル 「モバイルコンピューティングとエージェント」. 於 東京農工大. 1998.9.28
18. 高橋克巳. モバイルコンピューティングと AI —モバイルインフォサーチ—. 1998 年度人工知能学会全国大会 (第 12 回) チュートリアル, JSAI98, Vol. 1,

- pp.10-17. 於 早稲田大学. 1998.5
19. 横路誠司, 高橋克巳, 鷺坂光一, 三浦信幸, 島健一. 情報内容を考慮した情報収集方法. 情報処理学会 第 56 回 全国大会. 於 中央大学理工学部. 1998.03.17
 20. 横路誠司, 高橋克巳, 三浦信幸, 島健一. 特定分野のリソース収集を行う WWW ロボットの性能評価. 情報処理学会 第 57 回 全国大会 (第 3 分冊 pp.163~164). 於 名古屋大学工学部. 1998.10.5-7
 21. 井上香織, 高橋克巳. 検索のための広告文書構造化. 情処学会第 57 回全国大会. 1998.10.5
 22. Katsumi Takahashi, Nobuyuki Miura, Seiji Yokoji, Kenichi Shima. Mobile Info Search: Information Integration for Location Aware Computing. 情報処理学会 第 6 回 モバイルコンピューティング研究会, 情報処理学会 インターネット研究会 合同研究会. 於 静岡大学. 1998.9.17 日-18 (研究会推薦論文)
 23. 高橋克巳, 三浦信幸, 坂本仁明, 島健一. 位置指向の情報統合. Japan WWW Conference '97. パシフィコ横浜. 1997.12.19
 24. 高橋克巳, 三浦信幸, 大坪理恵, 島健一. モバイル環境における消費行動の支援. 第 6 回マルチ・エージェントと協調計算ワークショップ (MACC'97). 神戸ワイン城. 1997.12.18-20
 25. 三浦信幸, 高橋克巳, 坂本仁明, 島健一. モバイルインフォサーチ: 移動環境下でのユーザ指向型 WWW 検索. 第 3 回 情報処理学会 モバイルコンピューティング研究会. 於 倉敷芸術科学大学. 1997.12.4-5
 26. 三浦信幸, 高橋克巳, 島健一. WWW における検索ランキングとその Social Filtering としての効果. 情報処理学会第 55 回全国大会. 1997.9.24
 27. 高橋克巳, 三浦信幸, 西部喜康, 島健一. 不均一で分散した情報の構造化と統合 -情報統合ディレクトリ-. ソフトウェアエージェントとその応用シンポジウム講演論文集, 電子情報通信学会情報・システムソサイアティ大会併催. 1997.9.3
 28. 大坪理恵, 高橋克巳, 西部喜康, 森原一郎. 意思決定を支援する情報案内システム -Action Navigator-. 情処学会第 1 回モバイルコンピューティング研究会. 情報処理学会 (芝浦) 1997.5.30
 29. 大坪理恵, 高橋克巳, 西部喜康, 森原一郎. コミュニティ形成を支援する情報案内システム -Action Navigator-. 情報処理学会第 54 回全国大会 1997.3.12
 30. 島健一, 高橋克巳, 三浦信幸. インターネット・タウンページの構築 (1) ~概要~. 情報処理学会第 54 回全国大会. 1997.3.12
 31. 高橋克巳, 三浦信幸, 島健一. インターネット・タウンページの構築 (2) ~あいまい検索~. 情報処理学会第 54 回全国大会. 1997.3.12
 32. 三浦信幸, 高橋克巳, 島健一. インターネット・タウンページの構築 (3) ~個人適応

- 型 WWW の試み～. 情報処理学会第 54 回全国大会. 1997.3.12
33. 三浦 信幸, 高橋 克巳, 島 健一. 個人適応型 WWW のためのユーザモデル構築. 情報処理学会 インタラクション '97. 1997.2
 34. モバイル環境における情報案内サービス方式の検討. 西部喜康, 篠原章夫, 高橋克巳, 森原一郎. 情処学会 モバイルコンピューティング研究グループ第 1 回研究報告会. 於 NTT データ. 1996.7.25
 35. 高橋克巳. エージェントを使ったイエローページ検索. 1996 年度人工知能学会全国大会 (第 10 回) S2-01. 早稲田大学. 1996.6.24
 36. 高橋克巳, 高田敏弘, 三浦信幸, 島健一. 電話帳情報のハイパーテキスト化のデザイン. 第 7 回データ工学ワークショップ (DEWS96), 電子情報通信学会 データ工学専門委員会. 神戸市立フルーツ・フラワーパーク. 1996.3.15
 37. 島 健一, 高橋 克巳, 三浦 信幸. オブジェクト指向データベースを用いたインターネット版マルチメディア電話帳. 電子情報通信学会 データ工学研究会. 1995.12.11
 38. 島健一, 高橋克巳, 三浦信幸. インターネット版マルチメディア電話帳の構築. 第 1 回 Japan WWW Conference. 神戸国際会議場. 1995.11.28
 39. 篠原章夫, 高橋克巳, 森原一郎, 服部文夫. PHS&携帯情報端末を利用した通信サービス. 情処学会 夏のプログラミングシンポジウム. 1995
 40. 高橋克巳, 岩瀬成人. 人名の読みからの検索法. 第 91 回情報処理学会自然言語処理研究会. 1992.9.17
 41. 千田昇一, 高橋克巳, 山本康二, 武藤信夫, 福村好美. 高度検索のためのディレクトリ情報構成法の検討. 1992 年電子情報通信学会春季大会. 1992.3.24
 42. 岩瀬成人, 高橋克巳. 企業名の検索方式の高度化. 情報処理学会第 45 回全国大会. 1992.10.11
 43. 高橋克巳, 岩瀬成人. ネットワーク上の電話帳システム. 情報処理学会第 44 回全国大会. 1992.3.17
 44. 高橋克巳, 藤原進. 職業名の自動分割法. 情報処理学会第 42 回全国大会. 1991.3.12

技術誌 (NTT)

1. 坂井博, 村上幸司, 中津留毅, 高橋 克巳. 利用者状況に合わせた携帯電話待受画面への情報配信. NTT 技術ジャーナル Vol.18 No.8. 2006
2. 島健一, 高橋克巳, 三浦信幸. あいまいな単語からも地図からも電話帳検索が可能 -インターネットタウンページの構築-. NTT 技術ジャーナル Vol. 9, No. 5, pp. 72. 1997

3. 川邊秀樹, 唐沢裕明, 高橋克巳, 早坂秀雄. 104 番号案内オペレーション検索支援方式. NTT-R&D Vol.45, No.05, pp.433. 1996
4. 武藤信夫, 高橋克巳. 高付加価値型番号案内システム (CUPID) の広告・伝言情報処理方式. NTT R&D Vol. 39, No. 6, pp. 0857-0864. 1990
5. 戸部美春, 武藤信夫, 川辺秀樹, 高橋克巳, 井川敏彦. 高付加価値型番号案内システム (CUPID) の利用者端末の開発. NTT R&D Vol.39, No.06, pp.0865-0874. 1990

特許

1. 横路誠司, 高橋克巳.
位置関連情報提供装置、方法、プログラム、及びそのプログラムを記録した記録媒体.
特許第 3 5 1 9 3 6 9 号. 日本国特許庁.
平成 1 3 年 1 月 1 2 日出願. 平成 1 6 年 2 月 6 日登録.
2. 井上 香織, 高橋克巳.
意味属性辞書作成方法及び装置並びに意味属性辞書作成プログラムを記録した記録媒体.
特許第 3 5 1 8 9 9 8 号. 日本国特許庁.
平成 1 0 年 9 月 2 1 日出願. 平成 1 6 年 2 月 6 日登録.
3. 高橋克巳, 三浦信幸, 坂本仁明, 島健一.
情報蓄積装置および情報蓄積検索装置並びに情報蓄積方法および情報蓄積検索方法.
特許第 3 5 6 6 0 4 1 号. 日本国特許庁.
平成 1 6 年 6 月 1 8 日登録.
4. 三浦信幸, 高橋克巳, 坂本仁明, 島健一.
位置対応情報提供システム.
特許第 3 5 5 8 8 3 8 号. 日本国特許庁.
平成 9 年 8 月 2 9 日出願. 平成 1 6 年 5 月 2 8 日登録.
5. 高橋克巳, 梅村恭司.
情報検索装置及び情報検索方法.
特許 3 4 6 3 2 7 7 号. 日本国特許庁.
平成 6 年 7 月 2 2 日出願. 平成 1 5 年 8 月 2 2 日登録.