

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5543384号
(P5543384)

(45) 発行日 平成26年7月9日(2014.7.9)

(24) 登録日 平成26年5月16日(2014.5.16)

(51) Int.Cl. F I
G06F 17/30 (2006.01) G O 6 F 17/30 3 4 O B
 G O 6 F 17/30 2 2 O Z

請求項の数 9 (全 24 頁)

(21) 出願番号	特願2011-6229 (P2011-6229)	(73) 特許権者	392026693
(22) 出願日	平成23年1月14日 (2011.1.14)		株式会社NTTドコモ
(65) 公開番号	特開2012-146264 (P2012-146264A)		東京都千代田区永田町二丁目11番1号
(43) 公開日	平成24年8月2日 (2012.8.2)	(74) 代理人	100088155
審査請求日	平成25年8月16日 (2013.8.16)		弁理士 長谷川 芳樹
		(74) 代理人	100113435
			弁理士 黒木 義樹
		(74) 代理人	100121980
			弁理士 沖山 隆
		(74) 代理人	100128107
			弁理士 深石 賢治
		(72) 発明者	赤塚 隼
			東京都千代田区永田町二丁目11番1号
			株式会社エヌ・ティ・ティ・ドコモ内

最終頁に続く

(54) 【発明の名称】 ローカルクエリ抽出装置、ローカルクエリ抽出プログラム、およびローカルクエリ抽出方法

(57) 【特許請求の範囲】

【請求項1】

検索サービスにおいてユーザが入力したクエリ、およびユーザ特定情報を含む検索ログを格納する検索ログ格納手段と、

前記ユーザ特定情報を含むユーザデータを格納するユーザデータ格納手段と、

位置情報、および前記位置情報に関連付けられたメタデータを格納するメタデータ格納手段と、

前記クエリに含まれるメタデータを基に、前記メタデータ格納手段に格納された位置情報を抽出し、当該抽出した位置情報を前記検索ログと関連付ける検索ログ側関連付け手段と、

前記ユーザデータに含まれるメタデータを基に、前記メタデータ格納手段に格納された位置情報を抽出し、当該抽出した位置情報を前記ユーザデータと関連付けるユーザデータ側関連付け手段と、

前記クエリの地域毎の検索頻度を算出する検索頻度算出手段と、

前記検索ログと関連付けられた位置情報と、前記ユーザデータと関連付けられた位置情報とを照合することにより、ローカルクエリを抽出するローカルクエリ抽出手段と、

を備え、

前記ローカルクエリ抽出手段は、ある地域におけるあるクエリの前記検索頻度が、他の地域における前記あるクエリの前記検索頻度より閾値以上高い場合に、前記あるクエリを前記ある地域における前記ローカルクエリとして抽出する、

ことを特徴とするローカルクエリ抽出装置。

【請求項 2】

特定の地域にて特定のクエリが検索された回数を、前記特定の地域での総検索回数で割ることによって得られる第 1 の値を算出する第 1 値算出手段と、

地域の総数を、前記特定のクエリが検索された地域の数で割ることによって得られる第 2 の値を算出する第 2 値算出手段と、

を更に備え、

前記ローカルクエリ抽出手段は、前記第 1 の値および前記第 2 の値に基づき算出される第 3 の値が基準値以上である場合に、前記特定のクエリを前記特定の地域における前記ローカルクエリとして抽出する、

10

ことを特徴とする請求項 1 に記載のローカルクエリ抽出装置。

【請求項 3】

前記検索頻度、前記第 1 の値、前記第 2 の値、および前記第 3 の値のうち何れか 1 つ以上をユニークユーザ数単位で正規化する正規化手段を更に備え、

前記ローカルクエリ抽出手段は、当該正規化後の値を用いて前記ローカルクエリとして抽出する、

ことを特徴とする請求項 2 に記載のローカルクエリ抽出装置。

【請求項 4】

前記メタデータ格納手段に格納されたメタデータは、都道府県名、市町村名、電話番号、郵便番号、地域の固有名詞、施設名、住所、IP アドレス、GPS 測位による緯度経度、およびネットワーク測位による緯度経度のうち何れか 1 つ以上を示す情報である、

20

ことを特徴とする請求項 1 ~ 3 の何れか 1 項に記載のローカルクエリ抽出装置。

【請求項 5】

前記ユーザデータ側関連付け手段は、前記ユーザデータに含まれる前記ユーザにより事前に登録された住所を基に、前記メタデータ格納手段に格納された位置情報を抽出し、当該抽出した位置情報を前記ユーザデータと関連付ける、

ことを特徴とする請求項 4 に記載のローカルクエリ抽出装置。

【請求項 6】

前記ユーザデータ側関連付け手段は、前記ユーザデータに含まれる前記ユーザの IP アドレス、または前記 GPS 若しくはネットワーク測位による緯度経度を基に、前記メタデータ格納手段に格納された位置情報を抽出し、当該抽出した位置情報を前記ユーザデータと関連付ける、

30

ことを特徴とする請求項 4 に記載のローカルクエリ抽出装置。

【請求項 7】

前記ユーザデータ側関連付け手段は、前記ユーザが過去に発行したクエリまたはライブログを基に、前記メタデータ格納手段に格納された位置情報を抽出し、当該抽出した位置情報を前記ユーザデータと関連付ける、

ことを特徴とする請求項 4 に記載のローカルクエリ抽出装置。

【請求項 8】

コンピュータを、

40

検索サービスにおいてユーザが入力したクエリ、およびユーザ特定情報を含む検索ログを格納する検索ログ格納手段と、

前記ユーザ特定情報を含むユーザデータを格納するユーザデータ格納手段と、

位置情報、および前記位置情報に関連付けられたメタデータを格納するメタデータ格納手段と、

前記クエリに含まれるメタデータを基に、前記メタデータ格納手段に格納された位置情報を抽出し、当該抽出した位置情報を前記検索ログと関連付ける検索ログ側関連付け手段と、

前記ユーザデータに含まれるメタデータを基に、前記メタデータ格納手段に格納された位置情報を抽出し、当該抽出した位置情報を前記ユーザデータと関連付けるユーザデータ

50

側関連付け手段と、

前記クエリの地域毎の検索頻度を算出する検索頻度算出手段と、

前記検索ログと関連付けられた位置情報と、前記ユーザデータと関連付けられた位置情報とを照合することにより、ローカルクエリを抽出するローカルクエリ抽出手段、

として動作させ、

前記ローカルクエリ抽出手段は、ある地域におけるあるクエリの前記検索頻度が、他の地域における前記あるクエリの前記検索頻度より閾値以上高い場合に、前記あるクエリを前記ある地域における前記ローカルクエリとして抽出する、

ことを特徴とするローカルクエリ抽出プログラム。

【請求項 9】

検索ログ格納手段に、検索サービスにおいてユーザが入力したクエリ、およびユーザ特定情報を含む検索ログが格納されており、

ユーザデータ格納手段に、前記ユーザ特定情報を含むユーザデータが格納されており、

メタデータ格納手段に、位置情報、および前記位置情報に関連付けられたメタデータが格納されており、

検索ログ側関連付け手段が、前記クエリに含まれるメタデータを基に、前記メタデータ格納手段に格納された位置情報を抽出し、当該抽出した位置情報を前記検索ログと関連付ける検索ログ側関連付けステップと、

ユーザデータ側関連付け手段が、前記ユーザデータに含まれるメタデータを基に、前記メタデータ格納手段に格納された位置情報を抽出し、当該抽出した位置情報を前記ユーザデータと関連付けるユーザデータ側関連付けステップと、

検索頻度算出手段が、前記クエリの地域毎の検索頻度を算出する検索頻度算出ステップと、

ローカルクエリ抽出手段が、前記検索ログと関連付けられた位置情報と、前記ユーザデータと関連付けられた位置情報とを照合することにより、ローカルクエリを抽出するローカルクエリ抽出ステップと、

を備え、

前記ローカルクエリ抽出ステップでは、ある地域におけるあるクエリの前記検索頻度が、他の地域における前記あるクエリの前記検索頻度より閾値以上高い場合に、前記あるクエリを前記ある地域における前記ローカルクエリとして抽出する、

ことを特徴とするローカルクエリ抽出方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ローカルクエリ抽出装置、ローカルクエリ抽出プログラム、およびローカルクエリ抽出方法に関するものである。

【背景技術】

【0002】

ウェブサービスとして提供されている検索サービスを利用することで、ウェブ上に存在する膨大なデータから目的の情報を抽出することが可能である。検索サービスの利用者は、目的の情報を取得するために単語または複数語の文字列から構成されるクエリを検索サービスに入力する。検索サービスを提供する検索サーバは入力されたクエリと検索サーバに保持されているドキュメントとの適合度を算出し、適合度の高い順に列挙されたドキュメントを検索結果として利用者に表示するのが一般的である。

【0003】

パソコンにおいても、携帯電話機などのモバイル端末においても、位置に関連する検索は重要である。地図上での検索を行うことで、レストランや施設などの街の情報を取得したり、道案内の経路情報を取得したりすることが近年では当たり前となっている。高性能モバイル端末の普及により今後、さらに位置連動検索の重要性が増す可能性がある。

【0004】

10

20

30

40

50

ユーザがある特定の地域に関して検索を行う場合は、クエリに地域の名称を含めて検索することが多いと見受けられる。例えば「六本木 レストラン」と検索することで、ユーザは六本木周辺のレストラン情報を取得することができる。このクエリ及びユーザがクリックした検索結果は検索サーバにて検索ログとして蓄積される。

【 0 0 0 5 】

特許文献 1 では、ユーザが検索においてローカル情報をどの程度求めているかを、入力したクエリとユーザ自身のデータから推定する技術が記載されている。具体的には、クエリ、クリックした検索結果、及び関連クエリからクエリのローカル検索意図を推定し、同時にユーザのこれまでの検索行動などからユーザのローカル検索意図を推定した上で、二つの推定値から検索がローカル検索か否か判断している。クエリにローカル検索の意図があるか判断するに当たり、クエリの文字列から位置と共起する特徴的なワードが有るか否かを判断し、クエリの文字列からではローカル検索意図を把握するだけの十分な情報が得られない場合は、クリックした検索結果におけるページの中身を解析することでクエリがローカルクエリか否かを判断している。情報がまだ不足している場合は、さらに関連クエリが位置と共起しているか否かを解析することで、クエリのローカル検索意図を推定している。

10

【先行技術文献】

【特許文献】

【 0 0 0 6 】

【特許文献 1】特開 2 0 0 6 - 1 0 7 5 0 2 号公報

20

【発明の概要】

【発明が解決しようとする課題】

【 0 0 0 7 】

しかし、特許文献 1 の技術では以下のような問題点がある。一つ目の問題点は、クエリの文字列からではローカル検索意図を正確に把握するだけの十分な情報が得られない場合に、クリックされた検索結果や関連クエリなどを解析しないと、ローカルクエリかどうか正確に判断できないところである。例えば、「鎌倉パスタ」というクエリがあった場合、地域名称である「鎌倉」と共起しているのか、または「鎌倉パスタ」という名前のチェーン店のレストランと共起しているのか、クエリの文字列からでは判断することは難しい。特許文献 1 では、このような場合に、クリックされた検索結果のページの内容を解析することでローカル検索意図を読み取っている。しかし、ページ解析は処理的にも時間的にもコストがかかるため実用性に欠ける。

30

【 0 0 0 8 】

二つ目の問題点は、特許文献 1 の技術では、クエリの文字列に含まれている地域名称が二つ以上の地名と共起する場合、どちらの地名と共起しているのかが推定できない点にある。例えば、「日本橋 レストラン」というクエリがあった場合、関東の「日本橋」なのか関西の「日本橋」なのかが特許文献 1 の技術では推定できない。

【 0 0 0 9 】

そこで、本発明は上記に鑑みてなされたもので、低コスト化を図りながらも、精度良くローカルクエリを抽出することが可能なローカルクエリ抽出装置、ローカルクエリ抽出プログラム、およびローカルクエリ抽出方法を提供することを目的とする。

40

【課題を解決するための手段】

【 0 0 1 0 】

上記課題を解決するために、本発明のローカルクエリ抽出装置は、検索サービスにおいてユーザが入力したクエリ、およびユーザ特定情報を含む検索ログを格納する検索ログ格納手段と、前記ユーザ特定情報を含むユーザデータを格納するユーザデータ格納手段と、位置情報、および前記位置情報に関連付けられたメタデータを格納するメタデータ格納手段と、前記クエリに含まれるメタデータを基に、前記メタデータ格納手段に格納された位置情報を抽出し、当該抽出した位置情報を前記検索ログと関連付ける検索ログ側関連付け手段と、前記ユーザデータに含まれるメタデータを基に、前記メタデータ格納手段に格納

50

された位置情報を抽出し、当該抽出した位置情報を前記ユーザデータと関連付けるユーザデータ側関連付け手段と、クエリの地域毎の検索頻度を算出する検索頻度算出手段と、検索ログと関連付けられた位置情報と、ユーザデータと関連付けられた位置情報とを照合することにより、ローカルクエリを抽出するローカルクエリ抽出手段と、を備え、ローカルクエリ抽出手段は、ある地域におけるあるクエリの検索頻度が、他の地域におけるあるクエリの検索頻度より閾値以上高い場合に、あるクエリをある地域におけるローカルクエリとして抽出する、ことを特徴とする。

【0011】

また、本発明のローカルクエリ抽出プログラムは、コンピュータを、検索サービスにおいてユーザが入力したクエリ、およびユーザ特定情報を含む検索ログを格納する検索ログ格納手段と、前記ユーザ特定情報を含むユーザデータを格納するユーザデータ格納手段と、位置情報、および前記位置情報に関連付けられたメタデータを格納するメタデータ格納手段と、前記クエリに含まれるメタデータを基に、前記メタデータ格納手段に格納された位置情報を抽出し、当該抽出した位置情報を前記検索ログと関連付ける検索ログ側関連付け手段と、前記ユーザデータに含まれるメタデータを基に、前記メタデータ格納手段に格納された位置情報を抽出し、当該抽出した位置情報を前記ユーザデータと関連付けるユーザデータ側関連付け手段と、クエリの地域毎の検索頻度を算出する検索頻度算出手段と、検索ログと関連付けられた位置情報と、ユーザデータと関連付けられた位置情報とを照合することにより、ローカルクエリを抽出するローカルクエリ抽出手段、として動作させ、ローカルクエリ抽出手段は、ある地域におけるあるクエリの検索頻度が、他の地域におけるあるクエリの検索頻度より閾値以上高い場合に、あるクエリをある地域におけるローカルクエリとして抽出する、ことを特徴とする。

【0012】

また、本発明のローカルクエリ抽出方法は、検索ログ格納手段に、検索サービスにおいてユーザが入力したクエリ、およびユーザ特定情報を含む検索ログが格納されており、ユーザデータ格納手段に、前記ユーザ特定情報を含むユーザデータが格納されており、メタデータ格納手段に、位置情報、および前記位置情報に関連付けられたメタデータが格納されており、検索ログ側関連付け手段が、前記クエリに含まれるメタデータを基に、前記メタデータ格納手段に格納された位置情報を抽出し、当該抽出した位置情報を前記検索ログと関連付ける検索ログ側関連付けステップと、ユーザデータ側関連付け手段が、前記ユーザデータに含まれるメタデータを基に、前記メタデータ格納手段に格納された位置情報を抽出し、当該抽出した位置情報を前記ユーザデータと関連付けるユーザデータ側関連付けステップと、検索頻度算出手段が、クエリの地域毎の検索頻度を算出する検索頻度算出ステップと、ローカルクエリ抽出手段が、検索ログと関連付けられた位置情報と、ユーザデータと関連付けられた位置情報とを照合することにより、ローカルクエリを抽出するローカルクエリ抽出ステップと、を備え、ローカルクエリ抽出ステップでは、ある地域におけるあるクエリの検索頻度が、他の地域におけるあるクエリの検索頻度より閾値以上高い場合に、あるクエリをある地域におけるローカルクエリとして抽出する、ことを特徴とする。

【0013】

このような本発明のローカルクエリ抽出装置、ローカルクエリ抽出プログラム、およびローカルクエリ抽出方法によれば、検索ログと関連付けられた位置情報と、ユーザデータと関連付けられた位置情報とを照合することにより、ユーザがクリックしたページに対してページ解析をしなくても、ローカルクエリを抽出することができる。これにより、ローカルクエリ抽出処理における低コスト化を図ることができる。

【0014】

また、ユーザデータと関連付けられた位置情報を用いてローカルクエリ抽出処理を行うことにより、複数の地域で存在する同一名称を含むクエリの地域性を区別でき、精度良くローカルクエリを抽出することが可能となる。例えば、関東と関西には同一名称の「日本橋」が存在しており、「日本橋 レストラン」と記載されているクエリの文字情報だけで

は、どちらの「日本橋」を指しているのかが把握できない。しかし、ユーザデータと関連付けられた位置情報により、ユーザが関東にいるのか、または関西にいるのかが把握できるので、これをもって当該ユーザが入力したクエリがどちらの地域を指すのかを推定することができる。

【 0 0 1 5 】

また、本発明においては、前記クエリの地域毎の検索頻度を算出する検索頻度算出手段を更に備え、前記ローカルクエリ抽出手段は、ある地域におけるあるクエリの前記検索頻度が、他の地域における前記あるクエリの前記検索頻度より閾値以上高い場合に、前記あるクエリを前記ある地域における前記ローカルクエリとして抽出してもよい。

【 0 0 1 6 】

この発明によれば、クエリの検索頻度を地域毎に比較することで、そのクエリが地域に特定があるものか否かが判断可能となる。これにより、地域名が例えば人名やタレント名と同一である場合に、適切にフィルタリングをかけることができ、更に精度良くローカルクエリを抽出することが可能となる。

【 0 0 1 7 】

また、本発明においては、特定の地域にて特定のクエリが検索された回数を、前記特定の地域での総検索回数で割ることで得られる第1の値を算出する第1値算出手段と、地域の総数を、前記特定のクエリが検索された地域の数で割ることで得られる第2の値を算出する第2値算出手段と、を更に備え、前記ローカルクエリ抽出手段は、前記第1の値および前記第2の値に基づき算出される第3の値が基準値以上である場合に、前記特定のクエリを前記特定の地域における前記ローカルクエリとして抽出してもよい。

【 0 0 1 8 】

この発明によれば、ローカルクエリ抽出処理の結果に対する重み付けが第1の値、第2の値および第3の値により行われ、更に精度良くローカルクエリを抽出することが可能となる。

【 0 0 1 9 】

また、本発明においては、前記検索頻度、前記第1の値、前記第2の値、および前記第3の値のうち何れか1つ以上をユニークユーザ数単位で正規化する正規化手段を更に備え、前記ローカルクエリ抽出手段は、当該正規化後の値を用いて前記ローカルクエリとして抽出してもよい。

【 0 0 2 0 】

この発明によれば、正規化を行うことで処理結果に対する信頼性が高まり、更に精度良くローカルクエリを抽出することが可能となる。

【 0 0 2 1 】

また、本発明において、前記メタデータ格納手段に格納されたメタデータは、都道府県名、市町村名、電話番号、郵便番号、地域の固有名詞、施設名、住所、IPアドレス、GPS測位による緯度経度、およびネットワーク測位による緯度経度のうち何れか1つ以上を示す情報であってもよい。

【 0 0 2 2 】

この発明によれば、メタデータを構成するための具体的な手法が提供される。

【 0 0 2 3 】

また、本発明において、前記ユーザデータ側関連付け手段は、前記ユーザデータに含まれる前記ユーザにより事前に登録された住所を基に、前記メタデータ格納手段に格納された位置情報を抽出し、当該抽出した位置情報を前記ユーザデータと関連付けてもよい。

【 0 0 2 4 】

この発明によれば、位置情報をユーザデータと関連付けるための具体的な手法が提供される。

【 0 0 2 5 】

また、本発明において、前記ユーザデータ側関連付け手段は、前記ユーザデータに含まれる前記ユーザのIPアドレス、または前記GPS若しくはネットワーク測位による緯度

10

20

30

40

50

経度を基に、前記メタデータ格納手段に格納された位置情報を抽出し、当該抽出した位置情報を前記ユーザデータと関連付けてもよい。

【0026】

この発明によれば、位置情報をユーザデータと関連付けるための具体的な手法が提供される。

【0027】

また、本発明において、前記ユーザデータ側関連付け手段は、前記ユーザが過去に発行したクエリまたはライフログを基に、前記メタデータ格納手段に格納された位置情報を抽出し、当該抽出した位置情報を前記ユーザデータと関連付けてもよい。

【0028】

この発明によれば、位置情報をユーザデータと関連付けるための具体的な手法が提供される。

【発明の効果】

【0029】

本発明によれば、低コスト化を図りながらも、精度良くローカルクエリを抽出することが可能なローカルクエリ抽出装置、ローカルクエリ抽出プログラム、およびローカルクエリ抽出方法を提供することができる。

【図面の簡単な説明】

【0030】

【図1】本発明の第1の実施形態に係るローカルクエリ抽出装置1の構成概要図である。

【図2】ローカルクエリ抽出装置1のハードウェア構成図である。

【図3】検索ログ格納部10に格納されている検索ログの一例を示す図である。

【図4】ユーザデータ格納部20に格納されているユーザデータの一例を示す図である。

【図5】メタデータ格納部30に格納されているメタデータの一例を示す図である。

【図6】検索ログ側付加部40が位置情報と検索ログとを関連付けた一例を示す図である。

【図7】ユーザデータ側付加部50が位置情報とユーザデータとを関連付けた一例を示す図である。

【図8】ローカルクエリ抽出部60によるローカルクエリ抽出動作の具体的な手順を示すフローチャートである。

【図9】ローカルクエリ抽出部60によりステップS11が実行された様子の一列を示す図である。

【図10】ステップS13においてエリア毎に共起語が集計された様子の一列を示す図である。

【図11】本発明の第2の実施形態に係るローカルクエリ抽出装置1の構成概要図である。

【図12】検索頻度算出部61およびローカルクエリ抽出部60による、ローカルクエリ抽出動作の具体的な手順を示すフローチャートである。

【図13】ステップS26においてローカルクエリ抽出部60よりローカルクエリが抽出された様子の一列を示す図である。

【図14】本発明の第3の実施形態に係るローカルクエリ抽出装置1の構成概要図である。

【図15】TF値算出部62、IDF値算出部63、およびローカルクエリ抽出部60による、ローカルクエリ抽出動作の具体的な手順を示すフローチャートである。

【図16】ステップS38においてローカルクエリ抽出部60よりローカルクエリが抽出された様子の一列を示す図である。

【発明を実施するための形態】

【0031】

以下、添付図面を参照して本発明にかかるローカルクエリ抽出装置、ローカルクエリ抽出プログラム、およびローカルクエリ抽出方法の好適な実施形態を詳細に説明する。なお

10

20

30

40

50

、図面の説明において同一の要素には同一の符号を付し、重複する説明を省略する。

【 0 0 3 2 】

[第 1 の実施形態]

まず、本発明の第 1 の実施形態に係るローカルクエリ抽出装置 1 の構成について、図 1 を参照しながら説明する。図 1 は、ローカルクエリ抽出装置 1 の構成概要図である。図 1 に示すように、ローカルクエリ抽出装置 1 は、検索ログ格納部 1 0 (特許請求の範囲の「検索ログ格納手段」に相当)、ユーザデータ格納部 2 0 (特許請求の範囲の「ユーザデータ格納手段」に相当)、メタデータ格納部 3 0 (特許請求の範囲の「メタデータ格納手段」に相当)、検索ログ側付加部 4 0 (特許請求の範囲の「検索ログ側関連付け手段」に相当)、ユーザデータ側付加部 5 0 (特許請求の範囲の「ユーザデータ側関連付け手段」に相当)、ローカルクエリ抽出部 6 0 (特許請求の範囲の「ローカルクエリ抽出手段」に相当)、およびローカルクエリ出力部 7 0 を備える。

10

【 0 0 3 3 】

図 2 はローカルクエリ抽出装置 1 のハードウェア構成図である。図 2 に示すように、ローカルクエリ抽出装置 1 は、物理的には、CPU 1 1、ROM 1 2 及び RAM 1 3 等の主記憶装置、キーボード及びマウス等の入力デバイス 1 4、ディスプレイ等の出力デバイス 1 5、外部装置(例えば検索サーバなど)との間でデータの送受信を行うためのネットワークカード等の通信モジュール 1 6、ハードディスク等の補助記憶装置 1 7 などを含む通常のコンピュータシステムとして構成される。後述するローカルクエリ抽出装置 1 の各機能は、CPU 1 1、ROM 1 2、RAM 1 3 等のハードウェア上に所定のコンピュータソフトウェアを読み込ませることにより、CPU 1 1 の制御の元で入力デバイス 1 4、出力デバイス 1 5、通信モジュール 1 6 を動作させると共に、主記憶装置 1 2、1 3 や補助記憶装置 1 7 におけるデータの読み出し及び書き込みを行うことで実現される。

20

【 0 0 3 4 】

図 1 に戻り、検索ログ格納部 1 0 は、検索サービスにおいてユーザが入力したクエリ、および当該ユーザのユーザ ID (特許請求の範囲の「ユーザ特定情報」に相当) を含む検索ログを格納するものである。言い換えれば、検索ログ格納部 1 0 は、検索サービスを提供する検索サーバから利用者が検索行動においてとった検索行動履歴を受け取り、それを記憶する蓄積部である。図 3 は、検索ログ格納部 1 0 に格納されている検索ログの一例を示している。検索ログは、ユーザ ID、ユーザが入力したクエリ、検索サーバが提示した検索結果のうちユーザがクリックしたドキュメントの URL、およびユーザが当該ドキュメントをクリックした日時を示す情報を主な構成要素とする。これらに加え、クリックしたドキュメントの順位、検索結果に表示された上位数件の URL などが検索ログに更に含まれていてもよい。全ての項目のデータが必須として検索ログに含まれている必要はなく、本実施形態においては、例えば、ユーザ ID、クエリ、日時が少なくとも検索ログに含まれているものとする。また、図 1 では、検索ログ格納部 1 0 がローカルクエリ抽出装置 1 内に存在している構成を示しているが、これに限らず、検索ログ格納部 1 0 がローカルクエリ抽出装置 1 の外部に存在し、所定の通信回線によってローカルクエリ抽出装置 1 の本体と接続されていてもよい。

30

【 0 0 3 5 】

ユーザデータ格納部 2 0 は、ユーザデータを格納するものである。ユーザデータはユーザ ID を含む他、図 4 に示されるように、当該ユーザデータが生成された日時、ユーザの IP アドレスを示す情報、ユーザの位置に対する GPS 測位またはネットワーク測位による緯度経度を示す情報を含んでいてもよい。これらに加え、ユーザの年齢、性別、住所を示す情報等を更に含んでいてもよい。全ての項目の情報が必須としてユーザデータに含まれている必要はなく、本実施形態においては、例えば、図 4 に示されているユーザ ID および日時が少なくとも含まれ、更に IP アドレスおよび緯度経度のうち何れか一方がユーザデータに含まれているものとする。

40

【 0 0 3 6 】

メタデータ格納部 3 0 は、位置情報、および当該位置情報に関連付けられたメタデータ

50

を格納するものである。図5はメタデータ格納部30に格納されている情報を示しており、位置情報と種々のメタデータとが紐づけられている。図5の場合、「ディズニーランド」という施設の名称、緯度経度を示す情報等が本実施形態でいうメタデータである。メタデータは、図5に示された名称や緯度経度に限らず、位置情報と紐づけることが可能な種々の情報をいう。例えば、メタデータ格納部30に格納されたメタデータが、都道府県名、市町村名、電話番号、郵便番号、地域の固有名詞、施設名、住所、IPアドレス、GPS測位による緯度経度、およびネットワーク測位による緯度経度のうち何れか1つ以上を示す情報であってもよい。

【0037】

図5では住所を位置情報とし、名称や緯度経度等をメタデータとしているが、それに限らず、名称や緯度経度等を位置情報とし、住所をメタデータとしてもかまわない。また、図5に示す「エリア」という項目は、位置情報を上位概念化または抽象化したものである。本実施形態においては、「エリア」もそれ自体位置情報の一種である。例えば、ある住所に対して、当該住所における最寄の駅名を「エリア」項目の値とすることができる。一例として、品川駅周辺のある住所に対して、「エリア」項目の値を「品川駅」とすることができる。もちろん、駅名に限らず、都道府県名、市町村名等を「エリア」項目の値としてもかまわない。つまり、上記例において、「エリア」項目の値を「東京」や「関東」としてもかまわない。以上説明した位置情報およびメタデータは、メタデータ格納部30に予め格納されている情報であってもよい。

【0038】

検索ログ側付加部40は、クエリに含まれるメタデータを基に、メタデータ格納部30に格納された位置情報を抽出し、当該抽出した位置情報を検索ログと関連付けるものである。具体的に、検索ログ側付加部40は、検索ログ格納部10に格納されているクエリに含まれるワードと、メタデータ格納部30に格納されている位置情報またはメタデータとを照合し、それらが一致した場合に、対応する位置情報をメタデータ格納部30より抽出し、検索ログ格納部10の上記ワードを含む検索ログに付加する。図6は、位置情報に加え、エリアもが抽出され、検索ログに付加された一例を示す。図3のクエリq1に仮にワード「ディズニーランド」が含まれている場合に、検索ログ側付加部40は、図5に示されるメタデータのテーブルの中で名称が「ディズニーランド」で一致している位置情報「
県××市 町 」およびエリア「
」を抽出する。そして、検索ログ側付加部40は、図6に示すように、位置情報「
県××市 町 」およびエリア「
」を、図3でクエリq1を含んでいた検索ログに付加する。以上により、位置情報と検索ログとの関連付けが行われる。なお、図6に示した例においては、位置情報およびエリアの両方を抽出して付加しているが、これに限らず、位置情報のみを抽出して付加してもよく、エリアのみを抽出して付加してもよい。また、クエリに含まれるワードとメタデータ格納部30の位置情報またはメタデータとの照合方法としては、完全一致に限らず、部分一致による照合でもかまわない。更に、上記例では関連付けの手法としてデータそのものを付加しているが、これに限らず、その他の方法で関連付けを行ってもよい。

【0039】

ユーザデータ側付加部50は、ユーザデータに含まれるメタデータを基に、メタデータ格納部30に格納された位置情報を抽出し、当該抽出した位置情報をユーザデータと関連付けるものである。具体的に、ユーザデータ側付加部50は、ユーザデータ格納部20に格納されているユーザデータに含まれるIPアドレス、緯度経度等の値と、メタデータ格納部30に格納されているメタデータとを照合し、それらが一致した場合に、対応する位置情報をメタデータ格納部30より抽出し、ユーザデータ格納部20の上記IPアドレス、緯度経度等を含むユーザデータに付加する。緯度経度はGPSまたはネットワーク測位により得られるものである。図7は、位置情報に加え、エリアもが抽出され、ユーザデータに付加された一例を示す。図4のユーザデータのうち緯度経度の値が「
×」であるユーザIDがu1のユーザデータの場合、ユーザデータ側付加部50は、図5に示されるメタデータのテーブルの中で緯度経度の値が「
×」で一致している位置情報「
県××

10

20

30

40

50

市町」およびエリア「」を抽出する。そして、ユーザデータ側付加部50は、図7に示すように、位置情報「県××市町」およびエリア「」を、図4のユーザIDがu1のユーザデータに付加する。以上により、位置情報とユーザデータとの関連付けが行われる。なお、図7に示した例においては、位置情報およびエリアの両方を抽出して付加しているが、これに限らず、位置情報のみを抽出して付加してもよく、エリアのみを抽出して付加してもよい。また、ユーザデータに含まれる緯度経度とメタデータ格納部30の緯度経度との照合方法としては、完全一致に限らず、前方一致による照合でもかまわない。更に、上記例では関連付けの手法としてデータそのものを付加しているが、これに限らず、その他の方法で関連付けを行ってもよい。

【0040】

付加された位置情報やエリアは、ユーザの活動区域を示す情報であり、複数の活動地域がある場合は位置情報やエリアを複数付加してもよい。活動範囲が複数ある場合は、その活動日時を示すタイムスタンプを用意し、それぞれの活動範囲を区別してもよい。このタイムスタンプとしては、例えば図4に示した「日時」項目の値を用いることができる。図7の例は、一つのユーザIDに位置情報やエリアが複数付加されている例（ユーザIDがu1の場合）を示している。

【0041】

図7で示した上記例においては、緯度経度の値を照合して位置情報とユーザデータとの関連付けを行ったが、これに限らず、IPアドレスの値を照合して位置情報とユーザデータとの関連付けを行ってもよい。また、ユーザのIPアドレス、緯度経度等を取得することが不可能または難しい場合には、当該ユーザが過去に入力したクエリから逆引きすることで、ユーザの活動範囲を推定することもできる。つまり、ユーザデータ側付加部50は、ユーザが過去に発行したクエリまたはライフログを基に、メタデータ格納部30に格納された位置情報やエリアを抽出し、当該抽出した位置情報やエリアをユーザデータと関連付けることができる。例えば横浜エリアに住んでいる人は横浜周辺のスーパー及びレストランなどを検索する傾向があるため、当該ユーザの過去のクエリから当該ユーザの活動範囲を推定することができる。推定方法の一例としては、各エリアで頻繁に検索されるワードと、ユーザがこれまで入力したクエリとの類似度とを例えば強調フィルタリングを用いて算出し、類似度が最も近いエリアをそのユーザの活動範囲とすることができる。

【0042】

更に、ユーザデータ側付加部50は、ユーザデータに含まれる当該ユーザにより事前に登録された住所を基に、メタデータ格納部30に格納された位置情報やエリアを抽出し、当該抽出した位置情報やエリアをユーザデータと関連付けるようにしてもよい。この場合には、例えば、住所に含まれたワードと、位置情報やメタデータとを照合し、それらが一致した場合に、対応する位置情報やエリアをメタデータ格納部30より抽出し、上記ワードを含む住所のユーザデータに付加する。なお、住所に含まれたワードと、位置情報またはメタデータとの照合方法としては、完全一致に限らず、部分一致による照合でもかまわない。

【0043】

ローカルクエリ抽出部60は、検索ログと関連付けられた位置情報またはエリアと、ユーザデータと関連付けられた位置情報またはエリアとを照合することにより、ローカルクエリを抽出するものである。図8はローカルクエリ抽出部60による動作の具体的な手順を示すフローチャートである。最初に、ローカルクエリ抽出部60は、図6に示した位置情報およびエリア付き検索ログと、図7に示した位置情報およびエリア付きユーザデータとの結合を行う（ステップS11）。結合は、ユーザIDをキーとして行われる。また、ユーザデータに複数の活動拠点が存在する場合には、図6の日時項目と図7の日時項目を比較した上で、両時刻が一致またはある程度の範囲内である場合に限り結合を行う。図9は、ローカルクエリ抽出部60より、図6の位置情報およびエリア付き検索ログ（図9の（A））と、図7の位置情報およびエリア付きユーザデータ（図9の（B））とが結合された様子（図9の（C））を示す。図9の（A）におけるエリア項目は図9の（C）にお

10

20

30

40

50

いて検索ログ側エリア項目として結合され、図9の(B)におけるエリア項目は図9の(C)においてユーザデータ側エリア項目として結合されている。なお、図9の例においては、記載の便宜上、IPアドレス、緯度経度、位置情報等、一部の項目についてはその表示を省略している。

【0044】

次に、ローカルクエリ抽出部60は、図9の(C)のテーブルにおいて、検索ログ側エリア項目の値とユーザデータ側エリア項目の値とが一致するデータを抽出する(ステップS12)。両エリア項目の値が一致しない場合には、抽出を行わない。よって、図9において、検索ログ側エリア項目の値が「
」で、ユーザデータ側エリア項目の値が「×××」のデータは抽出対象外となる。なお、この動作は、位置情報をもとに行われてもかまわない。

10

【0045】

次に、ローカルクエリ抽出部60は、抽出対象となったデータからエリア毎に共起語を集計する(ステップS13)。「共起語」とは、図9の(C)のテーブルの各クエリにおいて、各エリアと共起するワードをいう。例えば、クエリq1が「品川 レストラン」である場合、「レストラン」はエリア「品川」に対する共起語である。図10は、ステップS13においてエリア毎に共起語が集計された様子の一例を示す。「品川 ブログ」、「品川 レストラン」、「品川 水族館」等のクエリに対して集計が行われている。なお、図10では、各クエリの検索回数毎に集計が行われているが、検索人数を基準にして集計を行ってもかまわない。

20

【0046】

次に、ローカルクエリ抽出部60は、上記集計されたクエリをローカルクエリとして抽出する(ステップS14)。つまり、「品川 ブログ」、「品川 レストラン」、「品川 水族館」等のクエリがローカルクエリとして抽出される。なお、ローカルクエリ抽出部60が、図10で示した検索回数が一定の閾値以上となるクエリをローカルクエリとして抽出してもよい。

【0047】

図1に戻り、ローカルクエリ出力部70は、ローカルクエリ抽出部60が抽出したローカルクエリを出力するものである。ローカルクエリ出力部70は、図10で示されているような形式で出力してもよく、共起語とエリアとの相関関係を示す値を更に追加して出力してもよい。

30

【0048】

[第2の実施形態]

引き続き、本発明の第2の実施形態について説明する。第2の実施形態は、第1の実施形態(図1を参照)と比べ、検索頻度算出部61(特許請求の範囲の「検索頻度算出手段」に相当)を更に備え、ローカルクエリ抽出部60の動作に相違点がある。以下、第2の実施形態における相違点を中心に説明し、共通する部分においては説明を省略する。

【0049】

図11は、本発明の第2の実施形態に係るローカルクエリ抽出装置1の構成概要図である。図11に示すように、ローカルクエリ抽出装置1は、図1の第1の実施形態に比べて、検索頻度算出部61を更に設けている。検索頻度算出部61は、クエリの地域(エリア)毎の検索頻度を算出するものである。第2の実施形態におけるローカルクエリ抽出部60は、検索頻度算出部61が算出した検索頻度を用いて、ある地域におけるあるクエリの検索頻度が、他の地域における当該あるクエリの検索頻度より閾値以上高い場合に、当該あるクエリを当該ある地域におけるローカルクエリとして抽出する。

40

【0050】

以下、第2の実施形態におけるローカルクエリの抽出方法について詳細に説明する。図12は検索頻度算出部61およびローカルクエリ抽出部60による動作の具体的な手順を示すフローチャートである。最初に、第1の実施形態のステップS11に相当する動作が行われる(ステップS21)。ただし、この動作は検索頻度算出部61により行われるも

50

のとする。すなわち、検索頻度算出部 6 1 が、図 6 に示した位置情報およびエリア付き検索ログと、図 7 に示した位置情報およびエリア付きユーザデータとの結合を行う。結合は、ユーザ ID をキーとして行われる。また、ユーザデータに複数の活動拠点が存在する場合には、図 6 の日時項目と図 7 の日時項目を比較した上で、両時刻が一致またはある程度の範囲内である場合に限り結合を行う。図 9 は、検索頻度算出部 6 1 より、図 6 の位置情報およびエリア付き検索ログ (図 9 の (A)) と、図 7 の位置情報およびエリア付きユーザデータ (図 9 の (B)) とが結合された様子 (図 9 の (C)) を示す。図 9 の (A) におけるエリア項目は図 9 の (C) において検索ログ側エリア項目として結合され、図 9 の (B) におけるエリア項目は図 9 の (C) においてユーザデータ側エリア項目として結合されている。なお、図 9 の例においては、記載の便宜上、IP アドレス、緯度経度、位置情報等、一部の項目についてはその表示を省略している。

10

【 0 0 5 1 】

次に、第 1 の実施形態のステップ S 1 2 に相当する動作が行われる (ステップ S 2 2)。ただし、この動作は検索頻度算出部 6 1 により行われるものとする。すなわち、検索頻度算出部 6 1 が、図 9 の (C) のテーブルにおいて、検索ログ側エリア項目の値とユーザデータ側エリア項目の値とが一致するデータを抽出する。両エリアが一致しない場合には、抽出を行わない。よって、図 9 において、検索ログ側エリア項目の値が「」で、ユーザデータ側エリア項目の値が「x x x」のデータは抽出対象外となる。なお、この動作は、図 9 には表示を省略している位置情報をもとに行われてもかまわない。

【 0 0 5 2 】

20

次に、第 1 の実施形態のステップ S 1 3 に相当する動作が行われる (ステップ S 2 3)。ただし、この動作は検索頻度算出部 6 1 により行われるものとする。すなわち、検索頻度算出部 6 1 が、抽出対象となったデータからエリア毎に共起語を集計する。「共起語」とは、図 9 の (C) のテーブルの各クエリにおいて、各エリアと共起するワードをいう。例えば、クエリ q 1 が「品川 レストラン」である場合、「レストラン」はエリア「品川」に対する共起語である。図 1 0 は、ステップ S 2 3 においてエリア毎に共起語が集計された様子の一例を示す。「品川 ブログ」、「品川 レストラン」、「品川 水族館」等のクエリに対して集計が行われている。なお、図 1 0 では、各クエリの検索回数毎に集計が行われているが、検索人数を基準にして集計を行ってもかまわない。

【 0 0 5 3 】

30

次に、検索頻度算出部 6 1 が、各エリアにて共起語が検索される頻度を算出する (ステップ S 2 4)。例えば図 1 0 に示した検索回数の例において、「品川」を含むクエリの総検索回数が 1 万であるとすれば、共起語「ブログ」の検索頻度は 0 . 0 2 5 である。このような検索頻度算出部 6 1 の動作を数式で表現すると、下記の式 (1) となる。式 (1) によれば、各エリアでの共起語の検索頻度 $Frequency_{\text{エリア}, \text{共起語}}$ を求めることができる。式 (1) において、 $SearchNum_{\text{エリア}, \text{共起語}}$ は、エリアで検索された共起語の回数、つまり検索回数を示す。検索頻度算出部 6 1 は、式 (1) を用いて、全てのエリアと共起語の組合せについて $Frequency_{\text{エリア}, \text{共起語}}$ を求める。

【 数 1 】

$$Frequency_{\text{エリア}, \text{共起語}} = \frac{SearchNum_{\text{エリア}, \text{共起語}}}{\sum_{\text{共起語}} SearchNum_{\text{エリア}, \text{共起語}}} \dots (1)$$

40

【 0 0 5 4 】

次に、検索頻度算出部 6 1 が、共起語毎の検索頻度の期待値を全てのエリアを押し並べて求める (ステップ S 2 5)。期待値の求め方としては、エリア毎の共起語の検索頻度に対してその平均値を期待値として推定してもよい。またはエリア毎の人口密度の偏りを考慮してより正確な平均値を計算し、それを期待値として推定してもよい。このような検索頻度算出部 6 1 の動作を数式で表現すると、下記の式 (2) となる。式 (2) によれば、

50

エリア毎の共起語の検索頻度の平均値が共起語の検索頻度の期待値として算出される。式(2)において、 $AvgFrequency_{共起語}$ は共起語の検索頻度の期待値を示す。

【数2】

$$AvgFrequency_{共起語} = \frac{\sum_{エリア、共起語} Frequency_{エリア、共起語}}{\sum_{エリア} 1} \dots (2)$$

【0055】

次に、ローカルクエリ抽出部60が、上記式(1)で検索頻度算出部61より求められた共起語の検索頻度 $Frequency_{エリア、共起語}$ と、上記式(2)で検索頻度算出部61より求められた共起語の検索頻度の期待値 $AvgFrequency_{共起語}$ との比率を算出し、当該比率(以下「リフト値」という。)が所定の閾値を越えた場合に、当該共起語をローカルクエリとして抽出する(ステップS26)。このようなローカルクエリ抽出部60の動作を数式で表現すると、下記の式(3)となる。式(3)において、ローカルクエリ抽出部60は、リフト値が閾値を越えた場合に、当該共起語をローカルクエリとして抽出する。

【数3】

$$\frac{Frequency_{エリア、共起語}}{AvgFrequency_{共起語}} > \beta \dots (3)$$

【0056】

例えば、図10で示したような場合において、「品川」エリアにおける共起語「水族館」の検索頻度 $Frequency_{エリア、共起語}$ が0.069であり、共起語「水族館」の検索頻度の全国平均値 $AvgFrequency_{共起語}$ が0.015であり、閾値は3であるとする。この場合に、リフト値は $0.069 / 0.015 = 4.6$ であり、閾値より大きく、上記式(3)を満たす。このような場合に、ローカルクエリ抽出部60は共起語「水族館」を「品川」エリアに対するローカルクエリとして抽出する。一方で、「品川」エリアにおける共起語「ブログ」の検索頻度 $Frequency_{エリア、共起語}$ が0.025であり、共起語「ブログ」の検索頻度の全国平均値 $AvgFrequency_{共起語}$ が0.023であり、閾値は3であるとする場合には、リフト値が $0.025 / 0.023 = 1.087$ である。この場合には、リフト値が閾値より小さく、上記式(3)を満たさないため、ローカルクエリ抽出部60は共起語「ブログ」を「品川」エリアに対するローカルクエリとして抽出しない。以上の説明において、閾値の値は実装にあたり適宜設定される値であり、全てのエリアにおいて固定の値であってもよく、エリア毎の検索回数や人口に応じて値が変動するものであってもよい。

【0057】

図13は、ステップS26においてローカルクエリ抽出部60よりローカルクエリが抽出された様子の一例を示す。「品川 レストラン」、「品川 水族館」等のクエリがローカルクエリとして抽出されているが、「品川 ブログ」はローカルクエリとして抽出されていない。

【0058】

上記例においては、ある地域におけるあるクエリの検索頻度と、全国における当該あるクエリの検索頻度の平均値との比率をリフト値として算出しているが、これに限らず、ある地域におけるあるクエリの検索頻度と、他の地域における当該あるクエリの検索頻度、または複数の他の地域における当該あるクエリの検索頻度の平均値との比率をリフト値として算出してもよい。ポイントは「品川 ブログ」のように全国的にどこでも調べられていそうなクエリはリフト値が低くなり、ローカルクエリとしては除外されることである。一方で、「品川 水族館」のように、水族館がある場所でしか調べられないようなワード

10

20

30

40

50

は、その地域固有の共起語である可能性が高いため、リフト値が高く、ローカルクエリとして抽出される。

【 0 0 5 9 】

更に、上記ステップ S 2 3 においては検索回数をもとにエリア毎の共起語を集計し、上記ステップ S 2 4 においては上記ステップ S 2 3 にて集計した検索回数をもとに検索頻度を算出しているが、これに限らず、検索回数または検索頻度をユニークユーザ数単位で正規化する手段（特許請求の範囲の「正規化手段」に相当）を更に備え、ローカルクエリ抽出部 6 0 が、当該正規化後の値を用いてローカルクエリを抽出するようにしてもよい。

【 0 0 6 0 】

[第 3 の実施形態]

引き続き、本発明の第 3 の実施形態について説明する。第 3 の実施形態は、第 1 の実施形態（図 1 を参照）と比べ、TF 値算出部 6 2（特許請求の範囲の「第 1 値算出手段」に相当）および IDF 値算出部 6 3（特許請求の範囲の「第 2 値算出手段」に相当）を更に備え、ローカルクエリ抽出部 6 0 の動作に相違点がある。以下、第 1 の実施形態における相違点を中心に説明し、共通する部分においては説明を省略する。

【 0 0 6 1 】

図 1 4 は、本発明の第 3 の実施形態に係るローカルクエリ抽出装置 1 の構成概要図である。図 1 4 に示すように、ローカルクエリ抽出装置 1 は、図 1 の第 1 の実施形態に比べて、TF 値算出部 6 2 および IDF 値算出部 6 3 を更に設けている。TF 値算出部 6 2 は、特定の地域にて特定のクエリが検索された回数を、当該特定の地域での総検索回数で割ることによって得られる TF 値（特許請求の範囲の「第 1 の値」に相当）を算出するものである。IDF 値算出部 6 3 は、地域の総数を、当該特定のクエリが検索された地域の数で割ることによって得られる IDF 値（特許請求の範囲の「第 2 の値」に相当）を算出するものである。第 3 の実施形態におけるローカルクエリ抽出部 6 0 は、TF 値および IDF 値に基づき算出される TF IDF 値（特許請求の範囲の「第 3 の値」に相当）が基準値以上である場合に、当該特定のクエリを当該特定の地域におけるローカルクエリとして抽出する。以上により、第 3 の実施形態においては、共起語の出現頻度に加え、エリアとの相関の高さが考慮され、よりローカライズされたローカルクエリの抽出が可能となる。

【 0 0 6 2 】

以下、第 3 の実施形態におけるローカルクエリの抽出方法について詳細に説明する。図 1 5 は、TF 値算出部 6 2、IDF 値算出部 6 3、およびローカルクエリ抽出部 6 0 による動作の具体的な手順を示すフローチャートである。最初に、第 1 の実施形態のステップ S 1 1 に相当する動作が行われる（ステップ S 3 1）。すなわち、ローカルクエリ抽出部 6 0 が、図 6 に示した位置情報およびエリア付き検索ログと、図 7 に示した位置情報およびエリア付きユーザデータとの結合を行う。結合は、ユーザ ID をキーとして行われる。また、ユーザデータに複数の活動拠点が存在する場合には、図 6 の日時項目と図 7 の日時項目を比較した上で、両時刻が一致またはある程度の範囲内である場合に限り結合を行う。図 9 は、ローカルクエリ抽出部 6 0 より、図 6 の位置情報およびエリア付き検索ログ（図 9 の（A））と、図 7 の位置情報およびエリア付きユーザデータ（図 9 の（B））とが結合された様子（図 9 の（C））を示す。図 9 の（A）におけるエリア項目は図 9 の（C）において検索ログ側エリア項目として結合され、図 9 の（B）におけるエリア項目は図 9 の（C）においてユーザデータ側エリア項目として結合されている。なお、図 9 の例においては、記載の便宜上、IP アドレス、緯度経度、位置情報等、一部の項目についてはその表示を省略している。

【 0 0 6 3 】

次に、第 1 の実施形態のステップ S 1 2 に相当する動作が行われる（ステップ S 3 2）。すなわち、ローカルクエリ抽出部 6 0 が、図 9 の（C）のテーブルにおいて、検索ログ側エリア項目の値とユーザデータ側エリア項目の値とが一致するデータを抽出する。両エリアが一致しない場合には、抽出を行わない。よって、図 9 において、検索ログ側エリア項目の値が「
」で、ユーザデータ側エリア項目の値が「×××」のデータは抽出対

10

20

30

40

50

象外となる。なお、この動作は、図9には表示を省略している位置情報をもとに行われてもかまわない。

【0064】

次に、第1の実施形態のステップS13に相当する動作が行われる(ステップS33)。すなわち、ローカルクエリ抽出部60が、抽出対象となったデータからエリア毎に共起語を集計する。「共起語」とは、図9の(C)のテーブルの各クエリにおいて、各エリアと共起するワードをいう。例えば、クエリq1が「品川 レストラン」である場合、「レストラン」はエリア「品川」に対する共起語である。図10は、ステップS33においてエリア毎に共起語が集計された様子の一例を示す。「品川 ブログ」、「品川 レストラン」、「品川 水族館」等のクエリに対して集計が行われている。なお、図10では、各クエリの検索回数毎に集計が行われているが、検索人数を基準にして集計を行ってもかまわない。

10

【0065】

次に、後述するステップS35~S38の手順を行うためのエリアが選択される(ステップS34)。ステップS34は、例えば処理対象のエリアを順番に選択することで行われる。ステップS34を行うための別途の手段を設けてもよく、TF値算出部62、IDF値算出部63、ローカルクエリ抽出部60のそれぞれがステップS34を行ってもよい。

【0066】

次に、TF値算出部62が、ステップS34にて選択されたエリアにて特定のクエリが検索された回数を、当該特定のエリアでの総検索回数で割ることで得られるTF値を算出する(ステップS35)。TF値は、共起語の出現頻度を表す値である。TF値算出部62の動作を数式で表現すると、下記の式(4)となる。式(4)において、 $tf_{\text{エリア、共起語}}$ は特定のエリアにおける共起語のTF値を示し、 $n_{\text{エリア、共起語}}$ はステップS34にて選択されたエリアにてクエリが検索された回数を示し、 $n_{\text{エリア}}$ は当該特定のエリアでの総検索回数を示す。例えば、「渋谷」エリアにおける共起語「レストラン」の出現頻度であるTF値は、当該「渋谷」エリアにてクエリ「渋谷 レストラン」が検索された回数 $n_{\text{エリア、共起語}}$ を、当該「渋谷」エリアでの総検索回数 $n_{\text{エリア}}$ で割ることで得られる。

20

【数4】

$$tf_{\text{エリア、共起語}} = \frac{n_{\text{エリア、共起語}}}{\sum_{\text{共起語}} n_{\text{エリア、共起語}}} \dots(4)$$

30

【0067】

次に、IDF値算出部63が、エリアの総数を、特定のクエリが検索されたエリアの数で割ることで得られるIDF値を算出する(ステップS36)。IDF値は、共起語の逆出現頻度を表す値である。逆出現頻度は、どこのエリアでも検索されるような共起語の場合は1に近い値となるが、特定のエリアとしか共起しないような共起語の場合は1以上の値となるものである。よって逆出現頻度が高いほど、エリアに対してユニークで相関が高い共起語であることがいえる。IDF値算出部63の動作を数式で表現すると、下記の式(5)となる。式(5)において、 $idf_{\text{エリア、共起語}}$ は選択されたエリアにおけるIDF値を示し、 $| \text{エリア} |$ はエリアの総数を示し、 $| \{ \text{エリア} : \text{共起語} \text{ エリア} \} |$ は共起語を含むエリアの数を示す。例えばエリア数が10で、そのうちレストランという共起語を含むエリア数が2だった場合、 $| \text{エリア} |$ は10、 $| \{ \text{エリア} : \text{共起語} \text{ エリア} \} |$ は2となる。

40

【数5】

$$idf_{\text{エリア、共起語}} = 1 + \log \frac{\left| \sum_{\text{エリア}} 1 \right|}{|\text{エリア:共起語} \in \text{エリア}|} \dots(5)$$

【0068】

次に、ローカルクエリ抽出部60が、ステップS35で算出されたTF値およびステップS36で算出されたIDF値に基づき、TFIDF値を算出する(ステップS37)。ローカルクエリ抽出部60がTFIDF値を算出する方法の一例を、下記の式(6)に示す。式(6)では、 $tf_{\text{エリア、共起語}}$ と $idf_{\text{エリア、共起語}}$ とを積算することにより、 $tfidf_{\text{エリア、共起語}}$ が算出される。

$$tfidf_{\text{エリア、共起語}} = tf_{\text{エリア、共起語}} \cdot idf_{\text{エリア、共起語}} \dots(6)$$

【0069】

次に、ローカルクエリ抽出部60が、ステップS37で算出したTFIDF値が基準値以上である場合に、当該特定のクエリを当該特定の地域におけるローカルクエリとして抽出する(ステップS38)。図16は、ステップS38においてローカルクエリ抽出部60よりローカルクエリが抽出された様子の一例を示す。この例においては基準値が1.0として設定されており、「品川 レストラン」、「品川 水族館」等のクエリはTFIDF値が基準値以上であるため、ローカルクエリとして抽出されている。一方、「品川 ブログ」はTFIDF値が基準値以上でなかったため、ローカルクエリとして抽出されていない。以上のような処理により、特定のエリアでよく検索され、且つユニークな共起語の抽出が可能となる。なお、抽出したローカルクエリを出力する際には、少なくともエリアおよび共起語を表示するものとし、必要に応じてエリア毎にTFIDF値を降順にソートし、検索頻度が高く地域固有の共起語を上位に出力するようにしてもよい。

【0070】

次に、全てのエリアに対して上記ステップS34～S38の手順が行われたかが判断される(ステップS39)。全てのエリアに対して上記ステップS34～S38の手順が行われた場合には、処理はそのまま終了する。一方で、上記ステップS34～S38の手順が行われていないエリアがある場合には、上記ステップS34に戻りステップS39までの処理が繰り返して行われる。なお、ステップS39の判断手順を行うための別途の手段を設けてもよく、ローカルクエリ抽出部60がこの手順を行ってもよい。

【0071】

なお、上記ステップS33、S35、S36、S37においては検索回数をもとに集計または算出が行われているが、これに限らず、検索回数、TF値、IDF値、およびTFIDF値のうち何れか1つ以上をユニークユーザ数単位で正規化する手段(特許請求の範囲の「正規化手段」に相当)を更に備え、ローカルクエリ抽出部60が、当該正規化後の値を用いてローカルクエリを抽出するようにしてもよい。

【0072】

〔ローカルクエリ抽出プログラム〕

以上、本発明の第1～第3の実施形態をローカルクエリ抽出装置1として説明した。以上の説明は、コンピュータをローカルクエリ抽出装置1として動作させるローカルクエリ抽出プログラムに係る説明としても捉えることができる。重複する説明は省略するが、ローカルクエリ抽出プログラムは、コンピュータを、以上で説明した検索ログ格納部10、ユーザデータ格納部20、メタデータ格納部30、検索ログ側付加部40、ユーザデータ側付加部50、ローカルクエリ抽出部60、およびローカルクエリ出力部70として機能させる。ローカルクエリ抽出プログラムは、例えば、記録媒体に格納されて提供される。なお、記録媒体としては、フレキシブルディスク、CD、DVD等の記録媒体、ROM等の記録媒体、または半導体メモリ等が例示される。

【0073】

〔ローカルクエリ抽出方法〕

また、以上の説明は、ローカルクエリ抽出装置 1 より行われるローカルクエリ抽出方法に係る説明としても捉えることができる。重複する説明は省略するが、ローカルクエリ抽出方法では、検索ログ格納部 10 に、検索サービスにおいてユーザが入力したクエリ、およびユーザ ID を含む検索ログが格納されており、ユーザデータ格納部 20 に、ユーザ ID を含むユーザデータが格納されており、メタデータ格納部 30 に、位置情報、および当該位置情報に関連付けられたメタデータが格納されている。そして、ローカルクエリ抽出方法は、検索ログ側付加部 40 が、クエリに含まれるメタデータを基に、メタデータ格納部 30 に格納された位置情報を抽出し、当該抽出した位置情報を検索ログと関連付ける検索ログ側関連付けステップと、ユーザデータ側付加部 50 が、ユーザデータに含まれるメタデータを基に、メタデータ格納部 30 に格納された位置情報を抽出し、当該抽出した位置情報をユーザデータと関連付けるユーザデータ側関連付けステップと、ローカルクエリ抽出部 60 が、検索ログと関連付けられた位置情報と、ユーザデータと関連付けられた位置情報とを照合することにより、ローカルクエリを抽出するローカルクエリ抽出ステップと、を備える。

10

【0074】

〔本実施形態の作用及び効果〕

続いて、本実施形態にかかるローカルクエリ抽出装置 1、コンピュータをローカルクエリ抽出装置 1 として動作させるローカルクエリ抽出プログラム、ローカルクエリ抽出装置 1 より行われるローカルクエリ抽出方法の作用及び効果について説明する。

20

【0075】

（第 1 の実施形態の作用及び効果）

第 1 の実施形態では、検索ログと関連付けられた位置情報と、ユーザデータと関連付けられた位置情報とを照合することにより、ローカルクエリを抽出することができる。検索ログは、ユーザが検索サービス使用時に入力した検索クエリと、検索結果から遷移し閲覧したウェブサイトの URL 情報とを蓄積したログである。検索ログ側付加部 40 は検索ログに格納されているクエリに含まれるメタデータを基に、検索ログに位置情報を追加するものである。メタデータは位置または地名に関連する抽象的なデータであり、数値または文字列から構成されている。例えば、都道府県名、市町村名、電話番号、郵便番号、あるいは東京ドーム、ディズニーランド、上野動物園などといった固有名詞が一例として含まれる。ユーザデータに含まれるメタデータからユーザの位置情報をユーザデータ側付加部 50 にて推定する。ユーザデータは、ユーザの IP アドレス、GPS またはネットワーク測位による位置データ（緯度経度）、ユーザがこれまで入力したクエリやライフログのことを指し、これらのデータから位置情報が推定される。ローカルクエリ抽出部 60 は、クエリの位置情報とユーザの位置情報とを照合し、地域毎にローカルクエリの抽出を行う。地域（エリア）とは、都道府県、市町村、またはある特定の位置から一定の範囲の領域のことを指すが限定はしない。ローカルクエリ出力部 70 における出力形式は固定していないが、クエリの検索回数を基準に規則的にソートされていてもよいし、またランダムな順序でもよい。また集計においては、クエリ単位の集計に限定されない。

30

【0076】

このような第 1 の実施形態によれば、低コストによるローカルクエリ抽出が可能となる。特許文献 1 の技術では、クリックされた検索結果のページを解析することで、クエリが位置と共起しているか否かの判断を行っていた。このようなページ解析には処理的にも時間的にも莫大なコストがかかる。一方、本実施形態では、ページ解析は不要であり、コストの低減が見込める。

40

【0077】

また、第 1 の実施形態によれば、ユーザデータと関連付けられた位置情報を用いてローカルクエリ抽出処理を行うことにより、複数の地域で存在する同一名称を含むクエリの地域性を区別でき、精度良くローカルクエリを抽出することが可能となる。例えば、関東と関西には同一名称の「日本橋」が存在しており、「日本橋 レストラン」と記載されてい

50

るクエリの文字情報だけではどちらの「日本橋」を指しているのかが把握できない。しかし、ユーザデータと関連付けられた位置情報により、ユーザが関東にいるのか、または関西にいるのかが把握できるので、これをもって当該ユーザが入力したクエリがどちらの地域を指すのかを推定することができる。

【0078】

また、第1の実施形態によれば、それぞれの地域のローカルクエリから、その地域の特徴を把握することが可能となる。例えば、映画館がある地域では映画関連クエリが常時ローカルクエリとして抽出される傾向にある。地域の特徴を把握することで、その地域にて活動している人達の趣味趣向や属性などの分析にも活用することができ、地域のプロファイリング及びユーザのプロファイリングへの応用が可能となる。

10

【0079】

(第2の実施形態の作用及び効果)

第2の実施形態では、クエリの検索頻度と他地域における同一クエリの検索頻度とを比較してから、ローカルクエリを抽出する。検索頻度はその地域で特定のワードがどの程度の割合で検索されたかを示すものである。例えば地域名と共起するクエリのうち「ラーメン」が検索された割合が1割ならば、「ラーメン」の当該地域名における検索頻度は0.1となる。検索頻度の地域毎の比較を行うことで、検索ログからローカルクエリの抽出をさらに洗練することが可能である。一例として、地域名と共起するクエリを母集団とし、「映画」というワードにおいて、それぞれの地域でワード「映画」が検索される平均の割合がXだとした場合を考える。この場合、例えば「六本木」エリアで「映画」と検索される割合がYだった場合、YとXを比較し、YがXを一定の基準値以上上回っていた場合、「映画」というクエリは全国の地域を押し並べて比較した結果、地域「六本木」の特徴を表しているローカルクエリであると言えることになる。対象となる地域の検索頻度と全国平均頻度を必ず比較するとは限らない。地域毎の検索頻度を比較し、ある地域における検索頻度が他の地域における検索頻度に比べて一定の基準を超えた場合、当該クエリを当該地域におけるローカルクエリとしてもよい。また基準値は固定値であってもよいが、地域毎の人口や地域の特徴に応じて変動する値であってもかまわない。

20

【0080】

このような第2の実施形態によれば、クエリの検索頻度を地域毎に比較することで、そのクエリが地域に特定性があるものか否かが判断可能となる。これにより、地域名が例えば人名やタレント名と同一である場合に、適切にフィルタリングをかけることができ、第1の実施形態に比べてより精度良くローカルクエリを抽出することが可能となる。つまり、キャラクタ、人物名などの地域と関連性のないワードを誤って地域の名称と判断し、当該クエリをローカルクエリとして抽出した場合においても、地域毎の検索頻度を比較することでフィルタリングをかけることができ、当該誤って判断され抽出されたものを除外することが可能となる。例えば、「品川 ドロップ」というクエリのみからでは、「ドロップ」が地域名称である「品川区」あるいは「品川駅」に関連しているのか、または芸能人の「品川庄司」が監督を勤める「ドロップ」という映画を指しているのかが分からない。この場合、各地域での「ドロップ」の検索頻度を考慮することで、「ドロップ」が「品川」という地域のローカルクエリであるか否かが把握できる。例えば、「品川区」あるいは「品川駅」における「ドロップ」の検索頻度と、全国における「ドロップ」の検索頻度が同じくらいであれば、「品川 ドロップ」は「品川」地域の固有のクエリではない。このように検索頻度をパラメータとして用いて地域毎に比較を行うことで、非ローカルクエリの判別が可能となり、フィルタリング後のより洗練されたローカルクエリの抽出が可能となる。

30

40

【0081】

また、第2の実施形態によれば、正規化を行うことで処理結果に対する信頼性が高まり、更に精度良くローカルクエリを抽出することが可能となる。例えば、ユニークユーザ単位で正規化することで、単一のユーザの行動により全体の計算結果が偏ることを防止できる。例えば、ある地域にて一人のユーザが同じクエリを何度も投入した場合、投入した回

50

数分で通常カウントされるが、特定の期間のデータを対象に1ユーザ1回とカウントすることにより、ユニークユーザ単位の正規化が行われ、処理結果に対する信頼性が高まり、更に精度良くローカルクエリを抽出することが可能となる。

【0082】

(第3の実施形態の作用及び効果)

第3の実施形態では、ローカルクエリ抽出処理の結果に対する重み付けがTF値、IDF値およびTFIDF値により行われ、更に精度良くローカルクエリを抽出することが可能となる。また、TF値、IDF値およびTFIDF値を基にローカルクエリの並び変えを行うことで、どのクエリがよりその地域と相関が高いかを把握することが可能となる。例えば、その地域にしかない人気レストランなどは高いTF値、IDF値およびTFIDF値で重み付けされる。また、病院といったこの地域にもあるような一般的なクエリには、低いTF値、IDF値およびTFIDF値で重み付けされる。このように、第3の実施形態では、地域毎にその他の地域と比較してより地域特定性の高いクエリや検索頻度の高いクエリに高い値で重み付けが行われるので、クエリの更なるローカライズが可能となる。また、正規化に関しては、上記「第2の実施形態の作用及び効果」と同様のことが言える。

10

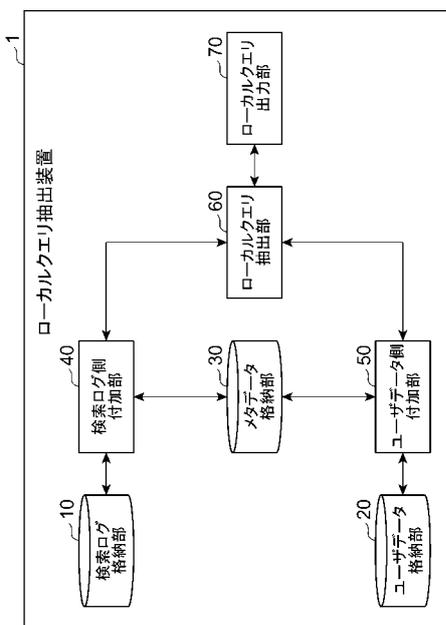
【符号の説明】

【0083】

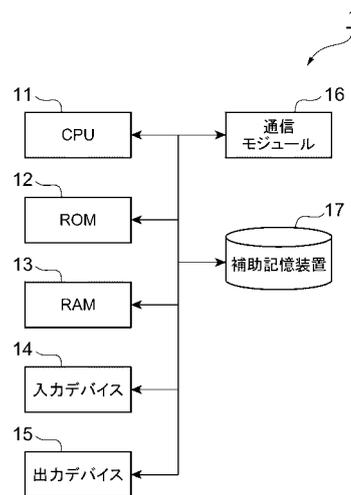
1...ローカルクエリ抽出装置、10...検索ログ格納部、20...ユーザデータ格納部、30...メタデータ格納部、40...検索ログ側付加部、50...ユーザデータ側付加部、60...ローカルクエリ抽出部、61...検索頻度算出部、62...TF値算出部、63...IDF値算出部、70...ローカルクエリ出力部。

20

【図1】



【図2】



【 図 3 】

ユーザID	クエリ	URL	日時	...
u1	q1	s1	○月×日■時◆分◇秒	...
u1	q1	s2	○月×日■時◆分◇秒	...
u2	q2	s3	○月○日■時◆分◇秒	...
∴	∴	∴	∴	∴

【 図 4 】

ユーザID	日時	IPアドレス	緯度経度	...
u1	○月×日■時◆分◇秒	△△	△ x	...
u1	○月×日■時◆分◇秒	△□	□ x	...
u2	○月○日■時◆分◇秒	x x	○○	...
∴	∴	∴	∴	∴

【 図 5 】

位置情報	名称	緯度経度	エリア	...
○○県××市□□町△△△	チイズニークラウド	△ x	○○○	...
∴	∴	∴	∴	∴

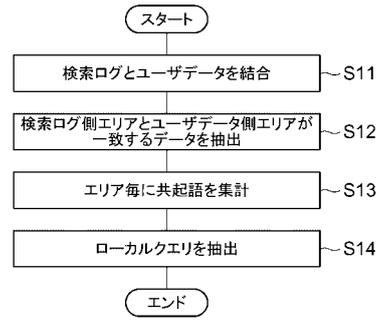
【 図 6 】

ユーザID	クエリ	URL	日時	エリア	位置情報
u1	q1	s1	○月×日■時◆分◇秒	○○○	○○県××市□□町△△△
u1	q1	s2	○月×日■時◆分◇秒	△△△	○○県××市■町△△ x
u2	q2	s3	○月○日■時◆分◇秒	○○○	●●県××市□□町△△△
∴	∴	∴	∴	∴	∴

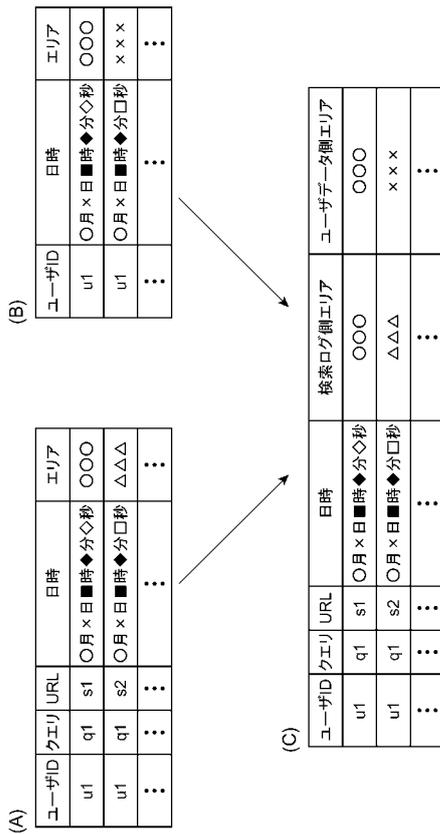
【 図 7 】

ユーザID	日時	IPアドレス	緯度経度	エリア	位置情報
u1	○月×日■時◆分◇秒	△△	△×	○○○	○○県××市□□町△△△△
u1	○月×日■時◆分◇秒	△□	□×	×××	○○県××市■町△△×
u2	○月○日■時◆分◇秒	××	○○	○○○	●●県××市□□町△△△△
...

【 図 8 】



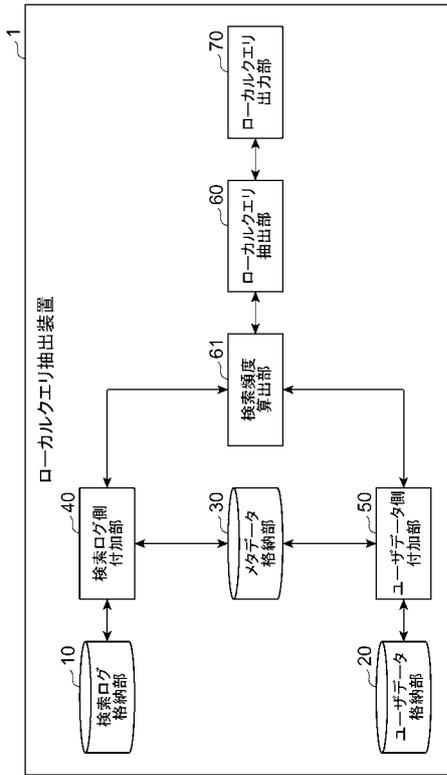
【 図 9 】



【 図 10 】

エリア	共起語	検索回数
品川	ブログ	250
品川	レストラン	480
品川	水族館	694
⋮	⋮	⋮

【図 1 1】



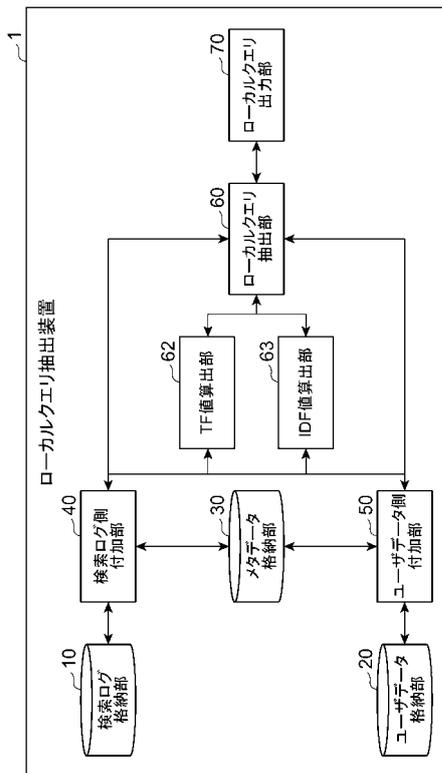
【図 1 2】



【図 1 3】

エリア	共起語	検索回数
品川	レストラン	480
品川	水族館	694
⋮	⋮	⋮

【図 1 4】



【図15】



【図16】

エリア	共起語	TFIDF
品川	水族館	1.5
品川	レストラン	1.0
⋮	⋮	⋮

フロントページの続き

(72)発明者 吉村 健

東京都千代田区永田町二丁目11番1号 株式会社エヌ・ティ・ティ・ドコモ内

(72)発明者 藤本 拓

東京都千代田区永田町二丁目11番1号 株式会社エヌ・ティ・ティ・ドコモ内

審査官 吉田 誠

(56)参考文献 特表2008-527504(JP,A)

特開2010-117824(JP,A)

(58)調査した分野(Int.Cl., DB名)

G06F 17/30