

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4035685号
(P4035685)

(45) 発行日 平成20年1月23日(2008.1.23)

(24) 登録日 平成19年11月9日(2007.11.9)

(51) Int. Cl.	F I
G06F 17/30 (2006.01)	G06F 17/30 320C
	G06F 17/30 340B

請求項の数 28 (全 21 頁)

(21) 出願番号	特願2000-560519 (P2000-560519)	(73) 特許権者	397051302
(86) (22) 出願日	平成11年7月8日(1999.7.8)		アマゾン ドット コム インコーポレイ
(65) 公表番号	特表2002-520740 (P2002-520740A)		テッド
(43) 公表日	平成14年7月9日(2002.7.9)		アメリカ合衆国 ワシントン州 9814
(86) 国際出願番号	PCT/US1999/015596		4-2734 シアトル スイート 12
(87) 国際公開番号	W02000/004465		00 トゥウェルフス アベニュー サウ
(87) 国際公開日	平成12年1月27日(2000.1.27)		ス 1200
審査請求日	平成16年1月21日(2004.1.21)	(74) 代理人	100065215
(31) 優先権主張番号	09/115,662		弁理士 三枝 英二
(32) 優先日	平成10年7月15日(1998.7.15)	(74) 代理人	100076510
(33) 優先権主張国	米国 (US)		弁理士 掛樋 悠路
		(74) 代理人	100086427
			弁理士 小原 健志
		(74) 代理人	100090066
			弁理士 中川 博司

最終頁に続く

(54) 【発明の名称】 サーチ照会におけるスペルミスを修正するためのシステムおよび方法

(57) 【特許請求の範囲】

【請求項1】

コンピュータネットワーク上でユーザがアクセス可能なサーチエンジンを実装するコンピュータシステムにおいて、サーチ照会内のミススペルされたサーチタームを処理する方法であって、

(a) 少なくとも同一サーチ照会内でのサーチタームの以前の出現頻度に基づいてサーチターム間の相関関係を示す相関データを生成するように、ある期間中に複数のユーザによってサーチエンジンに提出されたサーチ照会を処理するステップと、

(b) 複数のサーチタームを含み、サーチが行われるように情報データベースに向けられた前記サーチ照会を、コンピュータネットワーク上でユーザから受け取るステップと、

(c) 前記情報データベース内でマッチを生じないマッチしないサーチターム、および前記情報データベース内でマッチを生じる少なくとも1つのマッチするサーチタームを、前記サーチ照会内で識別するステップと、

(d) 前記少なくとも1つのマッチするサーチタームに関係していると考えられる複数の追加のタームを識別するために、少なくとも前記相関データを使用するステップと、

(e) 前記マッチしないサーチタームに対する正しくスペルされた置換候補タームである追加のタームを識別するために、ステップ(d)において識別された前記追加のタームを前記マッチしないサーチタームと比較するステップとをコンピュータ実行ステップとして含む方法。

【請求項2】

10

20

請求項 1 に記載の方法であって、

(f) ユーザ入力を必要とせずに、前記サーチ照会内で前記マッチしないサーチタームを前記置換候補タームで置き換えて、変更されたサーチ照会を生成するステップと、

(g) サーチを実行するために、前記変更されたサーチ照会を前記情報データベースに適用するステップと、

(h) 前記ユーザに、ステップ (g) の照会結果およびステップ (f) で行われた置き換えを通知するステップとをコンピュータ実行ステップとしてさらに含む方法。

【請求項 3】

ステップ (a) が、ユーザからサーチ照会を受け取ると、ほぼリアルタイムで前記関連データを更新するステップを含む、請求項 1 に記載の方法。

10

【請求項 4】

ステップ (a) が、前記サーチエンジンに提出されたサーチ照会を含んだログの構文解析を行うステップを含む、請求項 1 に記載の方法。

【請求項 5】

ステップ (a) が、古くなったサーチ照会提出よりも最新のサーチ照会提出を優先するように、時間ベースのバイアス機能をログに適用するステップをさらに含む、請求項 4 に記載の方法。

【請求項 6】

時間ベースのバイアス機能を適用するステップが、前記ログにウィンドウ機能を適用するステップを含む、請求項 5 に記載の方法。

20

【請求項 7】

ステップ (a) が、ヌル照会結果を生じるサーチ照会を無視するステップを含む、請求項 1 に記載の方法。

【請求項 8】

ステップ (a) が、成功したと考えられるサーチ照会を識別するために、ユーザの照会提出後の行動を評価して、関連データを生成する際に、成功したサーチ照会により多く加重するステップをさらに含む、請求項 1 に記載の方法。

【請求項 9】

ステップ (c) が、前記サーチ照会内で複数のマッチしないサーチタームを識別するステップを含み、マッチしないサーチタームのそれぞれに対する置換候補タームを識別するために、前記マッチしないサーチタームのそれぞれについてステップ (e) を繰り返すステップを含む、請求項 1 に記載の方法。

30

【請求項 10】

請求項 1 に記載の方法であって、ステップ (d) が、

(d 1) 複数のマッチするサーチタームのそれぞれについて、関係するタームのそれぞれのリストを識別するステップと、

(d 2) ステップ (d 1) において識別された関係するタームの前記リストを組み合わせるステップとを含む方法。

【請求項 11】

ステップ (e) が、前記マッチしないサーチタームを前記追加のタームと比較するために、第 1 文字ストリングおよび第 2 文字ストリングをソートして、前記ソートされた第 1 文字ストリングおよび第 2 文字ストリングを文字単位で比較することによって、前記第 1 文字ストリングおよび第 2 文字ストリングを比較するように適合されたスペリング比較機能を使用するステップを含む、請求項 1 に記載の方法。

40

【請求項 12】

前記スペリング比較機能が、前記第 1 文字ストリングおよび第 2 文字ストリングの間の類似の度合いを示すスコアを生成し、ステップ (e) が、対応する追加のタームが置換候補であるかどうかを判断するために、スコア機能によって生成されたスコアを、前記マッチしないサーチタームの文字数に依存する閾値と比較するステップをさらに含む、請求項 11 に記載の方法。

50

【請求項 13】

ステップ(e)が、前記マッチしないサーチタームに対する複数の正しくスペルされた置換候補タームを識別するステップを含み、選択のために前記ユーザに置換候補タームのリストを提示するステップをさらに含む、請求項1に記載の方法。

【請求項 14】

前記ユーザにリストを提示するステップが、変更されたサーチを開始するために、それぞれのハイパーリンク内で前記ユーザによって選択され得る置換候補タームを提示するステップを含む、請求項13に記載の方法。

【請求項 15】

前記コンピュータシステムがウェブサイトの一部分であって、前記情報データベースが、前記ウェブサイトを紹介して購入可能な製品に関する情報を含む、請求項1に記載の方法。

10

【請求項 16】

コンピュータネットワーク上でユーザがアクセス可能なサーチエンジンを実装するコンピュータシステムにおいて、少なくとも1つのミススペルされたサーチタームを含むサーチ照会を処理する方法であって、

(a) 複数のサーチタームを含み、サーチを行うために情報データベースに向けられたサーチ照会を、ネットワーク上でユーザから受け取るステップと、

(b) 前記情報データベース内でマッチを生じないマッチしないサーチターム、および前記情報データベース内でマッチを生じる少なくとも1つのマッチするサーチタームを、前記サーチ照会内で識別するステップと、

20

(c) 前記少なくとも1つのマッチするサーチタームに関係していると考えられる複数の追加のタームを識別するために、少なくとも履歴上のユーザの照会提出に基づき且つサーチターム間の相関を示すサーチターム相関データを使用するステップと、

(d) 前記マッチしないサーチタームに対する正しくスペルされた置換候補である追加のタームを識別するために、ステップ(c)で識別された前記追加のタームを前記マッチしないサーチタームと比較するステップとをコンピュータ実行ステップとして含む方法。

【請求項 17】

前記サーチターム相関データが、少なくともサーチタームが以前に同一照会内で出現した頻度に基づく、請求項16に記載の方法。

30

【請求項 18】

前記サーチターム相関データを生成するためにログ内の履歴上の照会提出を処理するステップをさらに含む、請求項16に記載の方法。

【請求項 19】

前記履歴上の照会提出を処理するステップが、前記ログにウィンドウ機能を適用するステップを含む、請求項18に記載の方法。

【請求項 20】

前記履歴上の照会提出を処理するステップが、ヌル照会結果を生じた照会を無視するステップを含む、請求項18に記載の方法。

【請求項 21】

請求項16に記載の方法であって、

(e) ユーザ入力を必要とせずに、前記サーチ照会内で前記マッチしないサーチタームを前記置換候補タームで置き換えて、変更されたサーチ照会を生成するステップと、

(f) サーチを実行するために、前記変更されたサーチ照会を前記情報データベースに適用するステップと、

(g) 前記ユーザに、ステップ(f)の前記照会結果およびステップ(e)において行われた置き換えを通知するステップとをコンピュータ実行ステップとしてさらに含む方法。

40

【請求項 22】

ステップ(d)が、前記マッチしないサーチタームを前記追加のタームと比較するため

50

に、第 1 文字ストリングおよび第 2 文字ストリングをソートして、前記ソートされた第 1 文字ストリングおよび第 2 文字ストリングを文字単位で比較することによって、前記第 1 文字ストリングおよび第 2 文字ストリングを比較するように適合されたスペリング比較機能を使用するステップを含む、請求項 16 に記載の方法。

【請求項 23】

スペリング比較機能が、前記第 1 文字ストリングおよび第 2 文字ストリングの間の類似の度合いを示すスコアを生成し、ステップ (d) が、対応する追加のタームが置換候補であるかどうかを判断するために、スコア機能によって生成されたスコアを、前記マッチしないサーチタームの文字数に依存する閾値と比較するステップをさらに含む、請求項 22 に記載の方法。

10

【請求項 24】

コンピュータネットワークのユーザが、項目のデータベースの検索を行えるようにする検索エンジンであって、

少なくともユーザの以前の照会提出に基づいて、サーチターム間の相関関係を示すサーチターム相関データを、そのメモリ内に格納しているコンピュータシステムを備え、

前記コンピュータシステムが、
ユーザから受け取った検索照会を使用して項目データベースを検索するように構成され、

前記コンピュータシステムが、少なくとも、

前記検索照会の、1つまたは複数のマッチするタームに関係していると考えられる複数の追加のタームを識別するために、前記サーチターム相関データにアクセスするステップと、

20

前記追加のタームのどれかが、マッチしないタームに対する置換候補であるかどうか判断するために、前記追加のタームのスペリングを、前記検索照会の、少なくとも1つの
マッチしないタームのスペリングと比較するステップによって、

前記マッチするタームと前記マッチしないタームの両方を含んだ複数ターム検索照会を処理するように構成されている検索エンジン。

【請求項 25】

前記サーチターム相関データが、少なくとも同一の検索照会内で以前にサーチタームが出現した頻度に基づいている、請求項 24 に記載の検索エンジン。

30

【請求項 26】

ユーザ入力を必要とせずに、検索照会内でマッチしないタームを置換候補タームで置き換えて、変更された検索照会を生成し、かつ前記変更された検索照会でデータベースの検索を行うように、前記コンピュータシステムが構成されている、請求項 24 に記載の検索エンジン。

【請求項 27】

第 1 文字ストリングおよび第 2 文字ストリングをソートして、ソートされた前記第 1 文字ストリングおよび第 2 文字ストリングを文字単位で比較することによって、前記第 1 文字ストリングおよび第 2 文字ストリングを比較するように適合されたスペリング比較機能を使用して、前記マッチしないタームを前記追加のタームと比較するように、前記コンピュータシステムが構成されている、請求項 24 に記載の検索エンジン。

40

【請求項 28】

前記スペリング比較機能が、前記第 1 文字ストリングと第 2 文字ストリングの間の類似の度合いを示すスコアを生成し、かつ

前記コンピュータシステムが、対応する追加のタームが置換候補であるかどうか判断するために、前記スコア機能によって生成されたスコアを閾値と比較するように構成されている、請求項 27 に記載の検索エンジン。

【発明の詳細な説明】

【0001】

(添付資料および著作権資料)

50

この明細書は、2つの文字ストリングを比較するのに使用されるスペリング比較機能のC++リストを添付資料として含む。この添付資料の内容は、著作権保護の対象となる。著作権者は、米国特許商標庁または他の特許庁のファイルまたは記録に表われるとおり、特許文献またはその一部分を複製することに異議はないが、それ以外ではすべてのような権利も保持する。

【0002】

(発明の分野)

本発明は、情報サーチおよび検索に関するものであり、より詳細にはサーチ照会を処理するための方法に関係する。

【0003】

(発明の背景)

多くのワールドワイドウェブサイトおよびオンラインサービスは、項目のドメインから興味のある項目をユーザが探し出すのを助けるために、サーチエンジンプログラム(「サーチエンジン」)を提供する。例えば、AltaVistaおよびInfoSeekは、ユーザが他のウェブサイトを探し出すのを助け、LexisおよびWestlawなどのオンラインサービスは、ユーザが論文や法廷の見解を探し出すのを助けるためのサーチエンジンを実装している。さらに、オンラインマーチャントは通常、顧客がオンラインカタログから項目を探し出すのを助けるためのサーチエンジンを提供している。

【0004】

サーチエンジンを使用してサーチを実行するために、ユーザは、1つまたは複数のサーチタームを含んだ照会を提出する。また、照会は明示的または暗黙的に、項目のタイトル、著者、またはサブジェクト分類などのサーチされるべきレコードフィールドを識別する。例えば、オンライン書店のユーザは、本のタイトルで表われるとユーザが信じるタームを含んだ照会を提出することができる。サーチエンジンの照会サーバプログラムは、照会にマッチするどのような項目でも識別するように、その照会を処理する。照会サーバプログラムによって識別された項目のセットは、「照会結果」と呼ばれ、通常、探し出された項目のリストとしてユーザに提示される。書店の例では、照会結果は通常、サーチタームのすべてを含んだ本のタイトルのセットであることになり、通常、これらの項目のハイパーテキストのリストとしてユーザに提示される。

【0005】

サーチエンジンのユーザが照会内で、タームをミスタイプする、またはそれを覚えていないなどしてサーチタームをミススペルしたとき、そのミススペルされたタームは通常、サーチに含まれるデータベースのタームのどれにもマッチしないことになる。こうした場合、多くのサーチエンジンは単に、ヌル(空の)サーチ結果を返すことになる。しかし、ヌルサーチ結果をユーザに提示することは、かなりのユーザのフラストレーションを引き起こし得る。この問題を低減するため、いくつかのサーチエンジンは事実上、サーチの間、マッチしないタームを無視する。この方法には、ユーザによって指定された潜在的に重要な情報を考慮しないという不都合があり、比較的多数の無関係なタームを含んだ照会結果を生み出す傾向がある。

【0006】

(発明の概要)

本発明は、サーチ照会内でのミススペルされたタームを修正するためのシステムおよび方法を提供することによって、前述の問題に対処する。このシステムは、サーチターム間の相関関係を示す相関データのデータベースを含む。相関データは好ましくは、特定のタームが同一の照会内で履歴上、一緒に表われた頻度に基づき、好ましくは、照会ログファイルから生成される。一実施の形態では、データベース(表として実装される)内の各エントリは、キーワードおよび「関係ターム」リストを含み、そこでの関係タームリストは、高い頻度でキーワードとの組み合わせで表われたタームから成っている。

【0007】

スペリング修正方法は、好ましくは、少なくとも1つのマッチするタームと少なくとも1

10

20

30

40

50

つのマッチしないタームを含んだサーチ照会が提出されたときに呼び出される。関連データベースを使用して、マッチする1つまたは複数のタームに関係していると考えられるタームのリストが、最初に生成される。これは例えば、各マッチするタームごとに関係するタームを抽出することによって、また、照会が複数のマッチするタームを含む場合には、これらのリストを単一の関係タームリストに組み合わせることによって達成することができる。

【0008】

次に、関係するタームは、なんらかの適切な置き換えを識別するため、そのスペリングについて、マッチしない単語と比較される。スペリングの比較は好ましくは、2つの文字ストリングの間の類似の度合いを示すスコアを生成する、アナグラムタイプのスペリング比較機能を使用して実行される。マッチしないタームと十分に類似したスペリングをもつ関係するタームが見つかった場合、そのマッチしないタームは、好ましくは、その関係するタームで自動的に置き換えられる。別法としてユーザは、タームのリストから置き換える1つまたは複数のタームを選択するようにプロンプト指示を受けることができる。マッチしない1つまたは複数のタームが置き換えられると、その変更された照会が、サーチを実行するのに使用される。また好ましくは、ユーザは、照会に加えられた1つまたは複数の変更の通知を受ける。

10

【0009】

従来のスペリング修正方法に勝る前述のスペリング修正方法の重要な利点は、選択された置き換えるタームが、ユーザによって意図されたタームである可能性が相当に高いということである。この利点は、サーチターム関連データの前述の使用、詳細には履歴の照会提出を反映する関連データの使用の結果として生じる。この方法はこれにより、照会結果がユーザの関心のある項目を含むことになる可能性を高める。もう1つの利点は、この方法が、著者やアーティストの固有名およびタイトルや製品名にある空想的なタームなどの、辞書に現れないタームを修正するのによく適しているということである。

20

【0010】

この発明の別の態様によれば、関連データは好ましくは、最新の照会提出を反映し、したがってユーザの現在の好みを強く反映するように生成される。これは例えば、望みの数(例えば、12)の最新デイリー照会ログから関連表を定期的に生成することによって達成される。最新の照会提出を強く反映する関連データを使用することは、スペリング修正処理によって行われる置き換えが、ユーザによって意図されたものとなる可能性をさらに高める。

30

【0011】

次に、この発明のこれらの特徴および他の特徴が、以下に要約される図面を参照して説明される。これらの図面およびそれに関連する説明は、この発明の好ましい実施の形態を例示するために提供されるものであり、発明の範囲を制限するものではない。

【0012】

(好ましい実施の形態の詳細な説明)

本発明は、サーチエンジンに提出される照会のスペリングエラーを訂正する方法を提供する。端的に言えば、この方法は、サーチターム関連データを使用して正確に綴られた照会のサーチタームに関係するサーチタームを識別すること、およびその関係したタームのどれかがミススペルのサーチタームに類似のスペルを有することを評価することを含む。サーチターム関連データは、照会提出履歴に基づくのが好ましく、より具体的には、同じ照会内で以前一緒に実行したサーチタームの頻度に基づくのが好ましい。この方法は、例えばインターネットサーチエンジン、リーガルリサーチエンジン、オンライン業者の提供するサーチエンジンなどを含む様々な異なるサーチエンジン内で実施することができる。

40

【0013】

例示のため、この方法は、本明細書では製品のオンラインカタログから項目(書籍、CDなど)を位置付ける際のAmazon.com Inc.の顧客を補助するために使用されるサーチエンジンという設定で記述する。この記述を通して、Amazon.comの

50

実施態様の様々な実施態様特有の詳細を参照する。この詳細は、本発明の好ましい実施の形態を完全に例示するために提供されるもので、本発明の範囲を限定するものではない。本発明の範囲は、添付の請求の範囲に記述されている。

【0014】

I. ウェブサイトおよびサーチエンジンの概要

図1は、本発明に従ってサーチエンジンを実施するために使用される構成要素を含めて、Amazon.comウェブサイト30を図示している。インターネット商取引の技術で周知の通り、Amazon.comウェブサイトは、ユーザが検索し、ブラウズし、オンラインカタログから書籍タイトル、音楽タイトル、その他の種類の項目を購入するための機能を含んでいる。そのカタログには多数のブラウジング可能な項目が含まれているので、そのサイトでユーザが項目を位置付けるのを補助するための効率的な機構を提供することが重要である。

10

【0015】

図1に例示の通り、ウェブサイト30はウェブサーバアプリケーション32(「ウェブサーバ」)を含み、そのアプリケーションは、ユーザコンピュータ34からインターネットを経て受信した要求を処理する。その要求は、カタログで製品をサーチするためにユーザによって提出されるサーチ照会を含む。ウェブサーバ32は、照会ログ36内での照会提出を含むユーザトランザクションを記録する。図1で示される実施の形態で、照会ログ36は、それぞれが1日のトランザクションを表す日毎照会ログ36のシーケンスを含む。

【0016】

ウェブサイト30は、目録データベース40をサーチすることによってサーチ照会を処理する照会サーバ38も含む。目録データベース40は、そのサイトから購入することが可能な様々な項目に関する情報を含む。この情報は、例えば書籍タイトルの書名、著者名、出版社名、主題の記述、ISBN(国際標準図書番号)、音楽タイトルの曲名、アーティスト、レーベル、音楽の分類を含む。各項目の情報は、フィールド内で調整され(「著者名」フィールドや「タイトル」フィールドなど)、データベース40をフィールド限定ベースでサーチ可能にする。このサイトは、HTML(Hypertext Markup Language)コンテンツのデータベース41を含み、そのコンテンツはとりわけ様々なカタログ項目を示し記述する製品情報ページを含む。

20

【0017】

図2は、目録データベース40で書籍タイトルをサーチするのに使用することができる書籍サーチページの一般形式を例示する。このページは、著者名フィールド、タイトルフィールド、主題フィールド42、43、44とユーザに書籍タイトルのフィールド限定サーチを開始させる関連付けられた制御を含む。ユーザは、アーティスト、タイトル及びレーベルのフィールドを使用して音楽タイトルをサーチするための音楽サーチページ(図示せず)にもアクセスすることができる。このサイトの他の領域は、ユーザがサーチタームを特有のデータベースフィールドに限定することなくサーチ照会を提出することを可能にする。

30

【0018】

ユーザが書籍サーチページまたはそのサイトの他のページからサーチ照会を提出すると、照会サーバ38(図1)は目録データベース40に照会を適用し、照会内のあらゆるフィールド制限を考慮する。この照会結果が単一項目であれば、その項目の製品情報ページがユーザに提示される。照会結果が複数項目を含む場合、項目のそれぞれの製品情報ページへのハイパーテキストリンクを含む1つまたは複数のサーチエンジン結果ページ(図示せず)を介してその項目のリストがユーザに提示される。

40

【0019】

複数ターム照会については、照会サーバ38は、実際にはサーチタームを一緒に論理積をとってサーチを実行する。たとえば、ユーザが「Java」および「programming」というタームをタイトルフィールドに入力すると、照会サーバ38はタイトル内にそれらの両方のタームを有するすべての項目をサーチし、そのリストを戻す。したがって

50

、この照会のいかなるサーチタームもマッチを生成しなければ（本明細書では「マッチしないターム」と称する）、この照会はヌル照会結果を生成する。この場合、「近似マッチ」とみなされる項目のリストとして、ユーザは提示され得る。

【0020】

本明細書で記述されるサーチエンジンはサーチタームと一緒に論理積をとるが、本発明はサーチタームを結合する他の方法を使用するサーチエンジンに適用することもできることが理解されよう。

【0021】

II. スペリング訂正機能の概要

本発明に従って、マッチするタームとマッチしないタームの両方を含む複数タームサーチ照会をユーザが提出すると、照会サーバ38のスペリング訂正プロセス48（図1）は、サーチターム関連データとあいまってマッチするタームを使用して、マッチしないタームのスペリングの訂正を試みる。これは、マッチするサーチタームに関係するとみなされる追加のタームを識別するサーチターム関連表50を使用し、次いでそのような関係タームのスペリングをマッチしないタームのスペリングと比較することによって達成されるのが好ましい。たとえば、ユーザが、

J a v a A P P I

という照会を提出し、「A P P I」がマッチしないタームであるとする、スペル訂正プロセス48は、関連表を使用して「J a v a」に関係するとみなされるタームのリストを識別する。このリストは次のように、c o f f e e、A P I、m a n a g e m e n t、l a n g u a g eとなるかもしれない。この関係タームのスペリングをマッチしないタームのスペリングと比較すると、スペル訂正プロセス48は、「A P I」を置換候補タームとして識別する。スペリングの比較は、アナグラムタイプのスペリング比較機能を使用して実行するのが好ましいが、この機能は、2種類の文字ストリング間の類似性の度合いを示すスコアを生成するものである。

【0022】

関連表50に含まれるデータは、サーチタームの間関係を示し、同じ照会内に現れる可能性のあるサーチタームを効果的に予見するのに使用される。このような予見をスペリング訂正プロセスに組み込むことによって、そのプロセスによって指示される所与の置換タームがユーザの意図したタームである可能性を著しく高める。

【0023】

前記プロセスが所与のマッチしないタームに対する適切な置換を識別すると、そのマッチしないタームは自動的に関係タームと置き換えられる。所与のマッチしないタームに対する適切な置換が見つからなければ、そのマッチしないタームは照会から削除されるのが好ましい。一旦すべてのマッチしないタームが置換または削除されると、修正された照会が使用されて目録データベース40をサーチし、その結果がユーザに提示される。ユーザには、そのサーチ照会になされた修正も通知される。

【0024】

マッチしないタームの自動置換の別法として、置換候補タームのリストから置換タームを選択するように、ユーザがプロンプト指示されることが可能である。ユーザが修正されたサーチを開始するために選択できるそれぞれのハイパーリンク（図示せず）として各置換候補タームを提示することによって達成されることが好ましい。その結果、ユーザは修正された照会を選択することも、単一の行為で新規サーチを開始することも可能となる。

【0025】

本発明の他の態様では、関連表50は、同じサーチ照会内に一緒に現れた特有のサーチタームの頻度に関する履歴情報を含むかまたは反映することが好ましい。図1に示す通り、このデータはテーブル作成プロセス46を使用して照会ログ36から抽出されるのが好ましい。このような履歴情報をスペル訂正プロセスに組み込むことによって、このプロセスによって位置付けられる置換タームがユーザの意図したタームである可能性がさらに増大する。

10

20

30

40

50

【0026】

一実施の形態（下記参照）では、テーブル作成プロセス46は、M個の最新日毎ログファイル36(1)ないし36(M)から日毎の相関表50を再生成するが、ここでMは10または20などの固定番号である。この「スライディングウィンドウ」の手法は、最新照会提出にのみ基づいて相関表を生成することが有利であり、したがってユーザの現在の好みを反映するものである。たとえば、比較的多数のユーザが、Jon Krakauer 著「Into Thin Air」という書籍を先週一杯サーチしたとすると、ターム「into」、「thin」、「air」、「Krakauer」の間の相関関係が相応して高くなる。その結果、相関が高まることによって、この書籍をサーチする（多重ターム照会の）タームをミススペルしたユーザが、この書籍に誘導される可能性が高まる。同様の結果を達成するために多様な他のタイプのバイアス方法をログに適用することが可能であり、古い照会提出よりも最新の照会提出に重きをおく方法もこれに含まれる。

10

【0027】

図3は、相関表50の一般形式を例示する。図3に示され、本明細書で詳細を記述した実施の形態では、サーチターム間の相関関係は同じ照会内の発生頻度に基づくものである。下記で記述するように、他のタイプのサーチターム相関を追加的または別法として使用することができる。さらに、開示の実施態様では相関データを格納するためにテーブルを使用するが、他のタイプのデータベースも使用され得る。

【0028】

図3に例示するように、相関表内の各エントリ（3つのエントリを示す）は、次の2つの主要な構成要素を有することが好ましい。すなわち、(1)キーワード60および(2)そのキーワードに対する「関係付けられたターム」リスト62である。関係付けられたタームリスト62は、同じ照会内に高頻度のキーワードとして表示されたN（例：20）サーチタームのリストであり、頻度に従って順序付けられる。たとえば、キーワードCOSMOSのエントリが、

20

COSMOS: ASTRONOMY, SAGAN, UNIVERSE, SPACE, CARL...

である場合、ASTRONOMYは最高の頻度でCOSMOSと一緒に現れたこと、SAGANは2番目に高い頻度でCOSMOSと一緒に現れたこと、以下同様のことを示している。リスト部分62内に表示された各タームは、そのタームが同じ照会内で比較的高頻度で現れたことにより対応するキーワード60に関係付けられると見なされる。本明細書で記述される実施態様では、キーワードと関係付けられたタームはアルファベットの大文字小文字に関係なくテーブルに格納されるが、別法として大文字小文字の情報を保存することも可能である。

30

【0029】

さらに図3を参照すると、テーブル50内の関係付けられた各タームと各キーワード60は、(ユーザがタームを入力したサーチフィールドに基づいて)そのタームが対応するサーチフィールド42、43、44を示す1文字フィールドのプレフィックス（図示せず）を好適に含んでいる。このプレフィックスは、たとえば、A=著者、T=タイトル、S=主題、R=アーティスト、L=レーベルのようなものでよい。したがって、たとえば図3のキーワードCOSMOSがプレフィックス「T」を有し、関連付けられたタームSAGANがプレフィックス「A」を有する場合、これは、タイトルフィールド43にCOSMOSを含む比較的多数の照会が、著者フィールド42のSAGANと一緒に提出されたことを示す。下記で記述するように、関係付けられたタームのプレフィックスは、スペリング訂正プロセス48によって使用されて、関係付けられたタームリストから非フィールド対応タームを効果的に透過し、その結果、所与のサーチフィールド内のマッチしないタームだけが同じフィールドの関係付けられたタームと比較される。したがって、たとえば著者フィールド42内のマッチしないタームは著者フィールド42に入力された履歴のある他のタームとのみ比較される。スペリング訂正プロセス48は、キーワード60のプレフィックスを同様に使用して、マッチするタームのそれぞれのサーチフィールドに対応する

40

50

テーブルエントリを選択する。たとえば、タイトルフィールド43内にマッチするターム MOUNTAINを含む誤った照会が受信されると、スペリング訂正プロセス48がキーワードT MOUNTAINを有するテーブルエントリをサーチする。

【0030】

図3にさらに示すように、相関表50は、キーワードとあいまって関係付けられた各タームが表示される回数を示す相関スコア64も含むことが好ましい。たとえば、JAV Aのエントリ内にタームPROGRAMMINGが320のスコアを有すると、それはJAV AとPROGRAMMINGが同じ照会内で(それぞれのフィールドのプレフィックスによって示されたフィールド内で)320回表示されたことを示している。下記に記述する通り、ヌル照会結果を生成する紹介は、テーブル作成プロセス64によって無視され、その結果、相関スコア64には反映されない。他の実施態様では、相関スコア64は他のタイプの相関を組み込んでよい。下記に記述する通り、スコア64は、照会が複数のマッチするタームを有するときは、関係付けられたタームリストを組み合わせるために使用されるのが好ましい。

10

【0031】

動作中に、照会サーバ38が、照会がマッチするタームとマッチしないタームの両方を含むと判定すると、スペリング訂正プロセス48は当初、相関表50にアクセスして関連付けられた関係付けられたタームリスト62を抽出する。照会が、マッチする複数のサーチタームを含む場合、プロセス48は、マッチする各タームに対する関係付けられたタームリスト62を獲得し、そのリストを(下記に記述するように)一緒に組み合わせて複合関係タームリストを生成する。前記の通り、プロセスは次いで結果として生じる関係付けられたタームのスペリングをマッチしないタームのスペリングと比較し、マッチしない各タームを照会から置換または削除する。このスペリング比較プロセス中、マッチしない各タームはマッチしないタームのサーチフィールド内に移動した関係付けられたタームのみと比較されるのが好ましい。

20

【0032】

この方法の重要な利点は、辞書に表示されないタームのミススペリングを訂正するのに特に良く適している。したがって、この技法は、辞書にないタームによって特徴付けられる傾向のある項目を識別するのに特に有用である。このような項目は、たとえば、オンライン業者によって販売される製品、(通常、関係者の名称によって識別される)裁判所の意見、ビジネスおよびそのウェブサイトを含む。たとえば、オンライン書籍/音楽商店という状況で、この方法は、著者/アーティストの正規の名称とタイトルに表示されるしゃれたタームのミススペリングを訂正するのに有用である。たとえば、Jon Krakauerの著書をサーチするユーザは、タイトルの数語と「Krakauer」のミススペルした語をタイプしてもその書籍を見つけることが可能である。サーチが所望の書籍を識別する可能性は、比較的多数のユーザが最近同じ書籍をサーチした場合には特に高くなる。

30

【0033】

この方法は、前の照会提出に基づくサーチターム提出を使用するのが好ましいが、相関データは他のソースによって生成することができることを理解されたい。たとえば、同じ項目レコード、タイトル、項目記述、書籍レビュー、他のデータベースフィールド内に一緒に表示されたタームを識別するために目録データベース40に情報を処理することによって、相関データを生成することができる。すなわち、このプロセスの実行において、各項目は、例えば先週一杯に販売された項目のユニット数に比例して重み付けされ得る。

40

【0034】

さらに、この方法は、他のタイプのユーザ行動履歴に基づく相関を含む他のタイプの相関を組み込むために修正されることができる。たとえば、照会ログ36から相関データを抽出するにあたり、ユーザの後続の行動に基づいて、成功した照会結果を生成したとみなすことのできる照会提出により重きをおくことができる。たとえばユーザがサーチによって位置付けられた項目を閲覧したり、購入したり、ショッピングカートに追加したりする場合に、このような成功が想定される。

50

【 0 0 3 5 】

図 1 をさらに参照すると、ウェブサーバ 3 2、照会サーバ 3 8、テーブル作成プロセス 4 6、データベースソフトウェアは、ウェブサイト 3 0 の 1 つまたは複数の U n i x ベースのサーバとワークステーション（図示せず）上で実行することもできる。関連表 5 0 は、照会サーバ 3 8 を実施するために使用される同じワークステーション上の R A M（ランダムアクセスメモリ）に格納される。

【 0 0 3 6 】

I I I . 照会処理方法

次に、図 4 を参照しながらスペリング訂正プロセスをさらに詳細に述べるが、図 4 は、ユーザが照会をサブミットするときに照会サーバ 3 8（図 1）によって実施されるステップのフローチャートである。このプロセスを示すために、ユーザがアパラチア山道のハイキングに関する本をサーチしており、「h i k e A p p a l a t i a n t r a i l」の照会を主題フィールド 4 4（図 2）中にタイプしたと仮定する。また、「A p p a l a t i a n」はマッチしないタームであり（「A p p a l a c h i a n」はマッチするが）、さらに「h i k e」および「t r a i l」は、以下の関係タームリストを有するマッチするタームであると仮定する。

H I K E : C A M P I N G (2 3 5)、W A L K S (1 6 0)、T R A I L (1 5 0)

T R A I L : B I K E (2 0 0)、A P P A L A C H I A N (1 6 5)、W A L K S (1 6 0)

また、上記のキーワードおよび関係タームがすべて、「主題」にちなんだフィールドプレフィックス「S」を有するとも仮定する。

【 0 0 3 7 】

ステップ 7 0 で示すように、照会サーバ 3 8 は、最初に目録データベース 4 0 に照会を適用する。ステップ 7 2 および 7 4 で示すように、1 つまたは複数の項目が検出された場合は、照会サーバはそれらの項目のリストをウェブサーバ 3 2 に返す。ウェブサーバ 3 2 は、そのリストを 1 つまたは複数のサーチ結果ページ中に取り込むか、あるいは項目が 1 つだけ突き止められた場合は、その項目に関する製品情報ページを返す。さらに、ウェブサーバは、検出された項目の数を照会ログ 3 6（図 5 参照）に記録する。この例では、「A p p a l a t i a n」が主題語として目録データベース 4 0 中にないため、項目は検出されないことになる。

【 0 0 3 8 】

この例のように、検出された項目の数が 0 であることがステップ 7 2 で分かった場合、照会サーバ 3 8 は、その照会がマッチするタームとマッチしないタームを両方とも含むかどうかを決定する（ステップ 7 6）。それらを含む場合、照会サーバ 3 8 は、そのスペリング訂正プロセス 4 8（図 4 のステップ 8 0 ~ 9 0）を呼び出して、マッチしないタームの訂正を試みる。それらを含まない場合、完全なマッチが検出されなかったことを示すメッセージをユーザに返す。この場合、照会サーバ 3 8 はまた、マッチするタームのサブセットしか含まない項目を含めることのできる「近いマッチ」のリストも生成して返す。照会サーバ 3 8 は、追加でまたは別法として、代替スペリング訂正方法（図示せず）を呼び出して、いずれかのマッチしないタームの訂正を試みるように構成することもできる。この例では、照会がマッチするタームとマッチしないタームの両方を含むため、スペル訂正プロセス 4 8 が呼び出されることになる。

【 0 0 3 9 】

スペル訂正プロセスはステップ 8 0 で開始し、マッチする各タームに対する関係タームリストを関連表 5 0 から検索する。ステップ 8 0 の間に関係タームリスト（表エントリ）が検出されない場合、プロセスは、好ましくはヌル照会結果メッセージを返し、終了する（図示せず）。あるいは、プロセスは、代替スペリング訂正方法を使用してマッチしないタームの訂正を試みるか、または「近いマッチ」のリストを生成して返すこともできる。

【 0 0 4 0 】

この例のように、照会が複数のマッチするタームを含む場合、これらのタームに対する関

10

20

30

40

50

係タームリストはステップ80で共にマージされて、複合関係タームリストが形成される。これらのリストは、例えば、いずれかの交差ターム（複数のリスト中にみられるターム）の相関スコアを合計しながらリストを結合し、次いで最高頻度スコアから最低頻度スコアの順に複合リストをソートすることによって、マージすることができる。この例では、この方法は以下の関係タームリストを作成することになる。すなわちCAMPING（235）、WALKS（210）、BIKE（200）、APPALACHIAN（165）、TRAIL（150）である。ソートされた複合リストは、後続のステップの処理負担を低減するために切り捨てることができる。

【0041】

次いで、スペル訂正プロセスはループ（ステップ84～90）に入り、マッチしないタームとその関係タームのスペリングが比較される。このループの各パスで、プロセスは、マッチしないタームを関係タームのリストと1つずつ比較し（ステップ84）、マッチしないタームのサーチフィールドに対応しないフィールドプレフィックスを有する関係タームがあればそれを排除する。これらの比較は、2つの文字ストリングを比較して数値的な類似性スコアを返すアナグラムタイプの関数を使用して行われる。類似性スコアは、2つのストリングのスペリング間における類似性の度合いを示し、したがって、所与の関係タームが目下のマッチしないタームに適した代用である可能性を示す。類似性スコアがゼロなのは類似性の度合いが最高であることを示し、スコアが漸進的に上昇するのは類似性の度合いが漸進的に低下することを示す。

【0042】

STRING1をSTRING2と比較するためにスペリング比較関数によって実施されるステップを以下に述べるが、この場合、変数RESULTは、その関数によって返されるスコアを表す。この方法のC++実施例のコードリスト（2つの別々の関数として実施する）は付録として添付されている。

【0043】

ステップ1：STRING1とSTRING2をアルファベット順にソートする。

【0044】

ステップ2：POINTER1をソート済みSTRING1の第1文字に、POINTER2をソート済みSTRING2の第1文字に、RESULTを0に初期化する。

【0045】

ステップ3：POINTER1およびPOINTER2によって指されたそれぞれの文字を、大文字、小文字を考慮せずに比較する。等しい場合は、両方のポインタをストリングの次の文字に進める。そうでない場合は、アルファベット順でより若い順の文字を指している方のポインタを進め、RESULTを1つインクリメントする。

【0046】

ステップ4：ポインタがそれぞれのストリングの最後の文字を越えるまでステップ3を繰り返す。

【0047】

ステップ5：ポインタの一方がまだ文字を指している場合は、そのストリング中に残っている（指されている文字を含めた）文字数をRESULTに加える。

【0048】

この例では、ステップ84の結果、マッチしないタームAPPALACHIANがCAMPING、WALKS、BIKE、APPALACHIAN、およびTRAILと比較される。これらの比較によって生成されたスコアが下の表に列挙されている。

【表1】

10

20

30

40

関係ターム	「AAAAILNPPT」と比較したソート済みストリング	類似性スコア
CAMPING	ACGIMNP	9
WALKS	AKLSW	11
BIKE	BEIK	12
APPALACHIAN	AAAACHILNPP	3
TRAIL	AILRT	7

表

10

【0049】

ステップ86で、スコアが評価されて、関係タームのいずれかが候補代用タームとして十分に類似するスペリングを有するかどうか決定される。好ましい実施の形態では、関係タームのスコアがマッチしないタームの長さの半分（類似性しきい値）以下である場合、その関係タームは類似する（したがって置換候補と）と考えられる。この例では、ターム「Appalachian」は類似性テストを満たし（3/5なので）、残りの関係タームは満たさない。

20

【0050】

スペリング比較関数の効率を向上させるために、RESULTが類似性しきい値を超えた後でステップ3の処理を停止することができる。また、十分に低いスコア（例えば0）を有するタームが識別された後で追加のタームの処理を停止することによっても効率を向上させることができる。

【0051】

少なくとも1つの関係タームがステップ86で類似性テストを満たす場合、マッチしないタームは最低スコアを有する関係タームで置換される（ステップ88）。ステップ88で複数の関係タームが最低スコアを共有する場合、関係タームリスト中で最初にくる（したがって最高の頻度スコアを有する）タームが代用として使用される。

30

【0052】

ステップ86でどの関係タームも類似性テストを満たさない場合、マッチしないタームは照会から削除される。あるいは、この時点でヌル照会結果メッセージをユーザに返してもよい。マッチしないすべてのタームが置換されるか削除された後、修正された照会を使用してサーチが再試行され、その結果がユーザに返される（ステップ94）。

【0053】

ステップ94で再試行されたサーチが1つまたは複数の項目を生み出す場合、それらの項目は、照会に加えられた修正を示すメッセージと共に呈示される。この例では、このメッセージは次のように書かれる。

主題語「Appalatian」を含む本は見つかりませんでした。しかし我々は、ターム「Appalatian」の代わりに「Appalachian」を使用してあなたのサーチを再試行し、以下のタイトルを見つけました。

40

サーチ結果ページは、ユーザが効果的にサーチタームの代用を拒否し、かつ/またはそうでない場合に照会を変更できるように、修正済み照会を表示し、ユーザが修正済み照会を編集できるようにすることが好ましい。

【0054】

ステップ94の再試行サーチが厳密なマッチを何も生み出さない場合、近いマッチのリストがユーザに呈示されるか、または単に、マッチが検出されなかったことがユーザに通知される。あるいは、追加の置換およびサーチを試行することもできる。

【0055】

50

以上のことから理解されるであろうが、図4のプロセスは、所望の目的を達成する様々な方式のいずれにも改変することができる。例えば、別々の関連表を生成して、異なるタイプの項目（例えば本と音楽）および/または異なるタイプのフィールド（例えばタイトルと主題）に使用することもできる。さらに、スペル訂正プロセスは、特定のフィールド（著者フィールド42など）内のミススペリングを訂正するためだけに使用したり、タームの辞書内に出ていないマッチしないタームだけに適用したりすることもできる。

【0056】

IV. 関連表の生成

次に、図5および6を参照しながらテーブル作成プロセス46（図1）について述べる。

【0057】

テーブル作成プロセス46は、一日一度など定期的に行われるオフラインプロセスとして実行されて、新しい関連表50を生成する。上記のように、このプロセスは、M個の最新デイリー照会ログファイル36(1)~36(M)から表を生成する。相対的に小さいM（例えば5）を使用すると、短期の購買動向（例えば新刊、習慣ベストセラーなど）を強く反映する関連を生成する傾向があり、より大きいM（例えば100）を使用すると、より広い範囲にわたるデータベースを生成する傾向がある。別法としてハイブリッド手法を使用することもでき、その場合は、表は多数のログファイルから生成されるが、最新ログファイルにより重みが与えられる。例えば、関連スコア64を生成するとき、先週サブミットされた照会を3倍に数え、1週間前から1箇月前までにサブミットされた照会を1倍に数えることができる。

【0058】

図5に、照会ログファイルの一般的なフォーマットを示す。ログ中の各エントリ（4つのエントリを示してある）は、特定のHTTP（Hypertext Transfer Protocol）トランザクションに関する情報を含む。例えば、エントリ100は、1998年2月13日金曜日午前2時23分にユーザ29384719287が、書籍検索ページから照会{著者=Seagal, タイトル=Human Dynamics}をサブミットし、照会にマッチした2つの項目が検出されたことを示す。ログ中のITEMS_FOUNDの値は、元の照会に厳密にマッチした項目の数を示すことが好ましく、したがって「近いマッチ」もスペリング訂正から得られたマッチも反映しないことが好ましい。

【0059】

エントリ102は、同じユーザが20秒後に1883823064のISBNを有する項目を選択し、この選択が検索結果ページから行われた（「HTTP_REFERENCE」の行から明白である）ことを示す。ショッピングカートに項目を入れる要求や、項目を購入する要求など、その他のタイプのユーザアクションも同様にログファイル内に反映される。上記の例で示したように、ログファイル36内のエントリを比較することによって、所与のユーザのナビゲーションパスを決定することができる。

【0060】

図6に、テーブル作成プロセス46によって実施される一連のステップを示す。この例では、プロセスは一日一度、最新デイリーログファイルが閉じられたすぐ後の真夜中に実行されると仮定する。また、M-1個の最新デイリーログファイルがプロセスのステップ110~114を使用してすでに処理されて、それぞれのデイリー結果ファイルが生成されたとも仮定する。

【0061】

ステップ110で、プロセスは、新しいログファイルを解析して、ITEMS_FOUND>0であるすべての照会サブミッションを抽出する。ヌル照会結果（ITEMS_FOUND=0）を生じた照会サブミッションを無視することにより、（1）マッチしないタームがキーワードまたは関連タームとして関連表に追加されることが回避され、（2）マッチするターム間の潜在的に「弱い」相関関係が考慮から除外されるという重要な利点をもたらされる。

10

20

30

40

50

【0062】

ステップ112において、ステップ110において抽出されたエントリが処理されて、同じ照会内で発生する頻度によってサーチタームが相関付けられる。このステップは、同じ照会内に少なくとも1回出てくるサーチタームの対のそれぞれについて、その2つのタームが一日を通して一緒に出てくる回数をカウントすることを含む。このプロセスの間、異なるサーチフィールド中でサブミットされた同一のタームは、異なるタームとして扱われる。例えば、フィールドプレフィックス「T」を有するタームTRAILは、プレフィックス「S」を有するTRAILとは異なるものとして扱われることになる。

【0063】

上に示したように、他の様々なファクタが、ステップ112プロセスの間に考慮され得る。例えば、ユーザが引き続きサーチ結果ページから項目を選択した場合に、所与の照会サブミッションを2倍に数えることができ、次いでユーザがその項目を購入したか、またはその項目を個人ショッピングカートに追加した場合には、3倍に数えることができる。外部からの相関データもまた、プロセスに組み込むことができる。ステップ112の結果は、図3の一般的な相関表の形をとり、デイリー結果ファイルとして保存される。

10

【0064】

ステップ116で、ステップ114で作成されたデイリー結果ファイルがその前のM-1個のデイリー結果ファイルとマージされて、相関表50が作成される。このプロセスの一部として、関係タームリストが固定長Nに切り捨てられ、得られる相関データは、効率的な参照のためにBツリーデータ構造で格納される。次いで、新しい相関表50が、既存の相関表の代わりにRAMに書き込まれる。

20

【0065】

本発明をいくつかの好ましい実施の形態から述べたが、当業者に明らかなその他の実施の形態もまた本発明の範囲内である。したがって、本発明の範囲は、添付の特許請求の範囲を参照することによってのみ定義されるものである。

【0066】

特許請求の範囲では、プロセスステップを示すのに使用する参照文字を設けるが、これは記述の便を図るためでしかなく、ステップを実施する特定の順番を意味するためではない。

【0067】

【表2】

30

添付資料

```
char *sort_string(char *string_to_sort) {  
    qsort(string_to_sort, strlen(string_to_sort),  
          sizeof(char), qsort_char_compare);  
    return string_to_sort;  
}
```

10

```
int score_sorted_strings(char* string1, char* string2) {  
    int result = 0;  
    int finished = FALSE;  
    int compare = 0;  
  
    unsigned char* str1_ptr = (unsigned char*) string1;  
    unsigned char* str2_ptr = (unsigned char*) string2;
```

20

【表3】


```
while (!finished) {  
  
    if (*str1_ptr == 0 || *str2_ptr == 0)  
        finished = TRUE;  
  
    /* The scoring function compares two characters; if they are the  
       same, both pointers are incremented to move on, otherwise  
       only the lesser pointer is incremented and compared again */  
  
    if (!finished) {  
        compare = qsort_char_compare((void*)str1_ptr, (void*)str2_ptr);  
  
        if (compare < 0) {  
            str1_ptr++;  
            result++;  
        } else if (compare > 0) {  
            str2_ptr++;  
            result++;  
        } else if (compare == 0) {
```

【表 4】

```
    str1_ptr++;
    str2_ptr++;
}
}
}
```

```
/* Add to the score any remaining results */
```

```
while (*str1_ptr++ != NULL) {
```

```
    result++;
```

```
}
```

```
while (*str2_ptr++ != NULL) {
```

```
    result++;
```

```
}
```

```
return result;
```

```
}
```

10

20

Copyright 1998 Amazon.com

Protected as an unpublished work

30

【図面の簡単な説明】

【図1】 この発明に係るサーチエンジンを実装するウェブサイトを示す図である。

【図2】 ウェブサイトの書籍検索ページを示す図である。

【図3】 図1の関連表の一般的形式を示す図である。

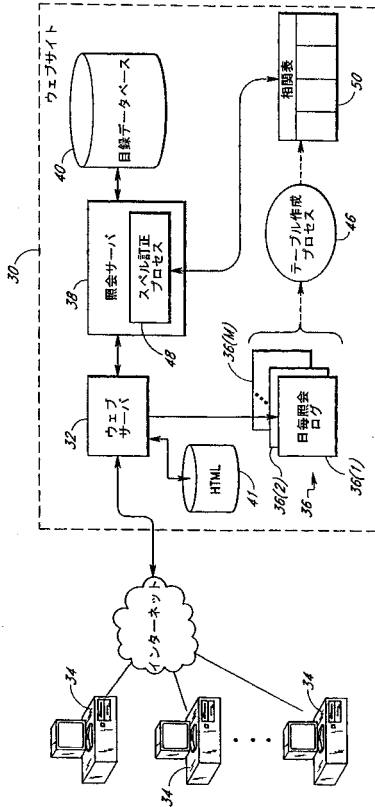
【図4】 照会提出を処理するために照会サーバによって実行されるステップを示す図である。

【図5】 図1の照会ログの一般的形式を示す図である。

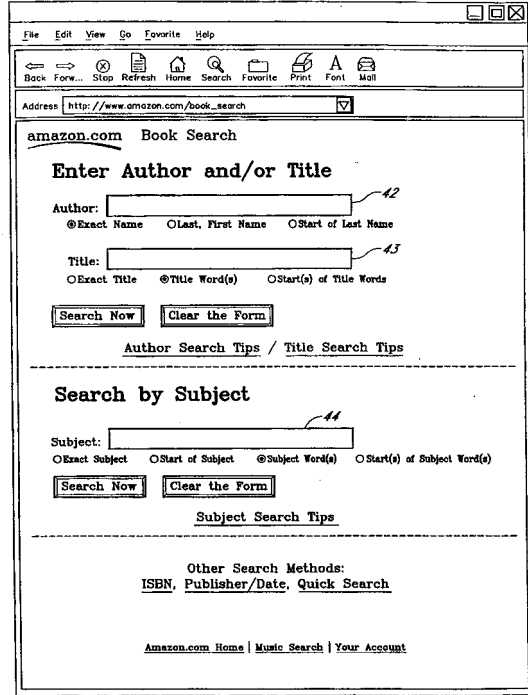
【図6】 照会ログから関連表を定期的に生成するために利用される処理を示す図である。

40

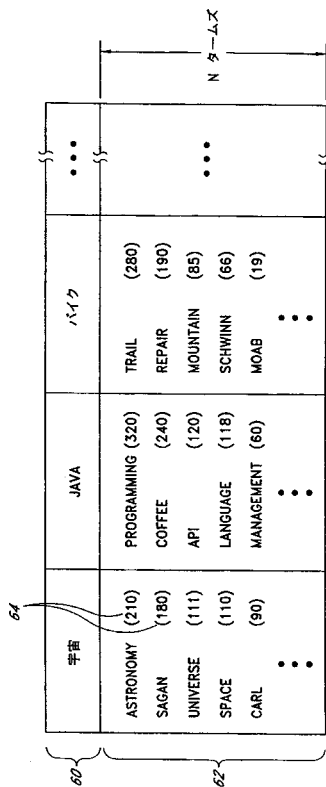
【 図 1 】



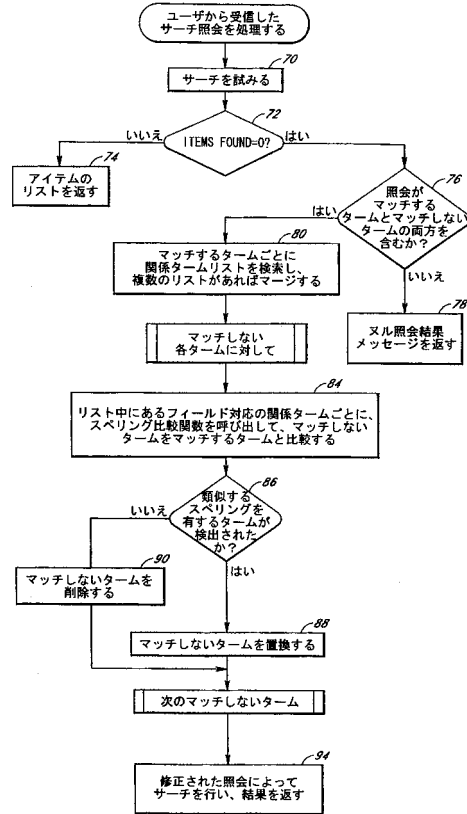
【 図 2 】



【 図 3 】



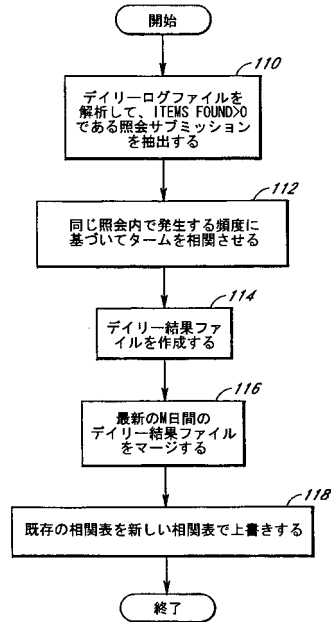
【 図 4 】



【 図 5 】

100 { Friday, 13-Feb-98 02:23:52
User Identifier = 29384719287
HTTP_REFERER= http://www.amazon.com/book_search_page
PATH_INFO=/book_search
author = Seagal
title = Human Dynamics
items_found = 2
⋮
102 { Friday, 13-Feb-98 02:24:11
User Identifier = 29384719287
HTTP_REFERER= http://www.amazon.com/book_search
PATH_INFO=/ISBN = 1883823064
⋮
Friday, 13-Feb-98 06:15:03
User Identifier = 54730543261
HTTP_REFERER= http://www.amazon.com/book_search_page
PATH_INFO=/book_search
subject = biking China
items_found = 0
⋮
Friday, 13-Feb-98 10:07:34
User Identifier = 027385918272
HTTP_REFERER= http://www.amazon.com/music_search_page
PATH_INFO=/music_search
artist = Miles Davis
items_found = 22
⋮

【 図 6 】



フロントページの続き

- (74)代理人 100094101
弁理士 舘 泰光
- (74)代理人 100099988
弁理士 斎藤 健治
- (74)代理人 100105821
弁理士 藤井 淳
- (74)代理人 100099911
弁理士 関 仁士
- (74)代理人 100108084
弁理士 中野 睦子
- (72)発明者 オルテガ ルベン アーネスト
アメリカ合衆国 9 8 1 0 5 ワシントン シアトル ノース イースト 3 3 番 アベニュー
4 7 1 2
- (72)発明者 バウマン ドウェイン エドワード
アメリカ合衆国 9 8 0 7 2 ワシントン ウッディンビル ノース イースト 2 1 4 番 ウェ
イ 1 4 2 4 4

審査官 深津 始

- (56)参考文献 米国特許第0 5 6 9 4 5 9 2 (U S , A)
特開平9 - 3 1 9 7 6 7 (J P , A)
Peter G. Anick , Adapting a Full-text Information Retrieval System to the Computer Trou-
bleshooting Domain , Proceedings of the 17th annual international ACM SIGIR conference
on Research and development in information retrieval , 英国 , 1 9 9 4 年 7 月 , pp.349-3
58
- (58)調査した分野(Int.Cl. , D B 名)
G06F 17/30