

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5247475号
(P5247475)

(45) 発行日 平成25年7月24日(2013.7.24)

(24) 登録日 平成25年4月19日(2013.4.19)

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

請求項の数 18 (全 18 頁)

(21) 出願番号	特願2008-557273 (P2008-557273)	(73) 特許権者	500046438
(86) (22) 出願日	平成19年2月8日(2007.2.8)		マイクロソフト コーポレーション
(65) 公表番号	特表2009-528619 (P2009-528619A)		アメリカ合衆国 ワシントン州 9805
(43) 公表日	平成21年8月6日(2009.8.6)		2-6399 レッドモンド ワン マイ
(86) 国際出願番号	PCT/US2007/003530		クロソフト ウエイ
(87) 国際公開番号	W02007/106269	(74) 代理人	100077481
(87) 国際公開日	平成19年9月20日(2007.9.20)		弁理士 谷 義一
審査請求日	平成22年1月4日(2010.1.4)	(74) 代理人	100088915
(31) 優先権主張番号	60/778, 650		弁理士 阿部 和夫
(32) 優先日	平成18年3月2日(2006.3.2)	(74) 復代理人	100115624
(33) 優先権主張国	米国 (US)		弁理士 濱中 淳宏
(31) 優先権主張番号	11/457, 733	(74) 復代理人	100161908
(32) 優先日	平成18年7月14日(2006.7.14)		弁理士 藤木 依子
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 ウェブ検索の適合性を高めるためにウェブ検索のユーザの振舞いをマイニングすること

(57) 【特許請求の範囲】

【請求項1】

複数のユーザの直接観察された振舞いの特性及び導き出された振舞いの特性に基づいて、問い合わせに対する検索結果の、ユーザのユーザ選択を予測するユーザ振舞いコンポーネントであって、前記直接観察された振舞いの特性は、前記複数のユーザのウェブページ対話を定量化する値から測定され、前記導き出された振舞いの特性は、前記直接観察された振舞いの特性の問い合わせ独立の分布から予測された期待値からの、前記直接観察された振舞いの特性の値の偏差を表し、さらに、各問い合わせURLペア、前記問い合わせの組み合わせである各問い合わせURLペア、及び前記問い合わせから取得される前記検索結果中の対応するウェブページのURL (uniform resource locator) に応じて、前記複数のユーザおよび検索セッションにまたがる前記直接観察された振舞いの特性についての値を平均化することで、前記複数のユーザのユーザ振舞いにおけるバリエーションを軽減するユーザ振舞いコンポーネントと、

前記ユーザおよび前記複数のユーザのウェブブラウジングの対話をキャプチャ及び定量化する表示特性、ブラウジング(走査検索)特性、またはクリックスルー特性の少なくとも一つを備え、前記直接観察された振舞いの特性及び前記導き出された振舞いの特性を含む特性セットと、

検索結果の関連性及び順位のための前記ユーザ選択を組み込む検索エンジンと

であるコンピュータ実行可能なコンポーネントを備えることを特徴とするコンピュータ実装のシステム。

【請求項 2】

前記ユーザ振舞いコンポーネントは、バックグラウンドコンポーネント及び関連性コンポーネントをさらに備えることを特徴とする請求項 1 に記載のコンピュータ実装のシステム。

【請求項 3】

機械学習コンポーネントをさらに備えることを特徴とする請求項 1 に記載のコンピュータ実装のシステム。

【請求項 4】

前記ユーザ振舞いコンポーネントは、ユーザの振舞いのデータドリブンモデルをさらに備えることを特徴とする請求項 1 に記載のコンピュータ実装のシステム。

10

【請求項 5】

前の検索データを含むデータログをさらに備えることを特徴とする請求項 4 に記載のコンピュータ実装のシステム。

【請求項 6】

前記検索エンジンは、検索結果を順位付ける順位付けコンポーネントをさらに備えることを特徴とする請求項 1 に記載のコンピュータ実装のシステム。

【請求項 7】

前記データドリブンモデルを訓練する機械学習コンポーネントをさらに備えることを特徴とする請求項 4 に記載のコンピュータ実装のシステム。

【請求項 8】

検索エンジンとの対話の間、前記検索エンジンと対話を行うユーザ振舞いコンポーネントが、複数のユーザの、ユーザの振舞いを取得するステップであって、前記検索エンジンにユーザの問い合わせを送信する、ステップと、

20

前記問い合わせによって取得される検索結果の、ユーザのユーザ選択を予測するために、前記ユーザ振舞いコンポーネントが、前記ユーザの振舞いの直接観察された特性及び導き出された振舞いの特性の分析についての前記ユーザの振舞いを統合するステップであって、前記直接観察された特性は、前記複数のユーザの対話を定量化する値から測定され、前記導き出された特性は、前記複数のユーザの前記直接観察された特性の問い合わせ独立の分布から予測された期待値からの、前記直接観察された特性の値の偏差を表す、ステップと、

30

前記ユーザ振舞いコンポーネントが、各問い合わせURLペア、前記問い合わせの組み合わせである各問い合わせURLペア、及び前記問い合わせから取得される前記検索結果中の対応するウェブページのURL (uniform resource locator) に応じて、前記複数のユーザおよび検索セッションにまたがる、前記対話の前記直接観察された特性の値を平均化することで、前記複数のユーザのユーザ振舞いに関連するノイズであって、悪意のある又は不合理なブラウジング動作に対応するノイズを軽減するステップと、

前記ユーザ振舞いコンポーネントが、前記ユーザの振舞いを統合するステップ、及び前記ノイズを軽減するステップに基づいて、取り出された結果に対するユーザ選択を予測するステップと

前記検索エンジンが、前記ユーザ選択に基づいて、前記取り出された結果の順位付けを決定するステップと、

40

であるコンピュータ実行可能な動作を備えたことを特徴とするコンピュータ実装の方法。

【請求項 9】

前記ユーザ振舞いコンポーネントが、前記取り出された結果の順位付けのためのモデルを訓練するステップをさらに備えたことを特徴とする請求項 8 に記載のコンピュータ実装の方法。

【請求項 10】

前記ユーザ振舞いコンポーネントが、前記ユーザの振舞いからモデルを自動的に生成するステップをさらに備えたことを特徴とする請求項 8 に記載のコンピュータ実装の方法。

50

【請求項 1 1】

前記ユーザ振舞いコンポーネントが、前記取り出された結果とのユーザの対話に関する特性のセットを作り出すステップをさらに備えたことを特徴とする請求項 8 に記載のコンピュータ実装の方法。

【請求項 1 2】

前記検索エンジンが、前記ユーザの振舞いを組み込むために機械学習を採用するステップをさらに備えたことを特徴とする請求項 8 に記載のコンピュータ実装の方法。

【請求項 1 3】

前記ユーザ振舞いコンポーネントが、前記ユーザの振舞いを予測するステップをさらに備えたことを特徴とする請求項 8 に記載のコンピュータ実装の方法。

10

【請求項 1 4】

前記検索エンジンが、前記取り出された結果の順位付けのために、統合されたユーザの振舞いをマイニングするステップをさらに備えたことを特徴とする請求項 8 に記載のコンピュータ実装の方法。

【請求項 1 5】

前記ユーザ振舞いコンポーネントが、前記ユーザ選択を予測するために、前記取り出された結果との前記ユーザの対話から、直接観察された特性を採用するステップをさらに備えたことを特徴とする請求項 8 に記載のコンピュータ実装の方法。

【請求項 1 6】

実行されると、一つまたは複数のプロセッサに、
複数のユーザによるウェブブラウジングの振舞いについてのデータを取得するステップと、

20

前記データからユーザ選択を予測するモデルを形成するステップであって、前記モデルは直接観察された特性及び導き出された特性を含む特性セットを使用し、前記直接観察された特性は観察されたユーザのブラウジングの振舞いを定量化する値によって測定され、前記導き出された特性は、前記直接観察された特性の問い合わせ独立の分布から予測された期待値からの、前記直接観察された特性の値の偏差を表し、前記特性セットは、前記複数のユーザのウェブブラウジングの対話をキャプチャ及び定量化する表示特性、ブラウジング（走査検索）特性、またはクリックスルー特性の少なくとも一つを備え、さらに、ユーザの振舞いを予測することは、ウェブ検索問い合わせ、前記ウェブ検索問い合わせの組み合わせである各問い合わせ URL ペア、及びウェブ検索問い合わせから取得される前記検索結果中の対応するウェブページの URL (uniform resource locator) から取得される検索結果に含まれる複数の問い合わせ URL ペアのそれぞれに対して、前記複数のユーザおよび検索セッションにまたがる前記直接観察された特性についての値を平均化することを含む、ステップと、

30

前記ユーザ選択に基づいて前記ウェブ検索問い合わせの前記検索結果に順位を割り当てるために、前記モデルと対話するステップと、
を含む動作を実行させる、コンピュータが実行可能な命令を格納するコンピュータ可読メモリ。

【請求項 1 7】

実行されると、一つまたは複数のプロセッサに、
直接観察された特性及び導き出された特性の両方を使用した検索問い合わせ結果ペアに応じたユーザの振舞いをモデル化するステップ
をさらに含む動作を実行させる、コンピュータが実行可能な命令を格納する請求項 1 6 に記載のコンピュータ可読メモリ。

40

【請求項 1 8】

実行されると、一つまたは複数のプロセッサに、
問い合わせがウェブドキュメントのテキストにどのくらい近く一致するかを含む問い合わせ独立の特性、及び、ウェブドキュメントのページランク (PageRank) を含む問い合わせ独立の特性、の両方に基づいて前記検索結果を順位付けするステップ

50

をさらに含む動作を実行させる、コンピュータが実行可能な命令を格納する請求項 16 に記載のコンピュータ可読メモリ。

【発明の詳細な説明】

【技術分野】

【0001】

WWW(World Wide Web)及びインターネットの人気により、ユーザは大量の情報ソースからほとんど任意のトピックに関する情報を獲得できる。情報を見つけるためにユーザは一般に、情報取り出しのタスクのために様々な検索エンジンを適用する。検索エンジンは、特定の語又は句を含むインターネット上の情報又は他の題材を含むウェブページをユーザが見つけることを可能にする。

10

【0002】

一般に、キーワード検索は、コンピュータの能力の限りまで、特定された任意のキーワード及び句に関する任意の情報を有するすべてのウェブサイトを見つけることができる。検索エンジンのサイトは、ユーザがキーワードを入力するためのボックス及び検索を開始するために押すボタンを有する。多くの検索エンジンは、効率的に検索するためにどのようにキーワードを使用するかについてのヒント情報(tips)を有する。典型的には、このようなヒント情報は、余計な関係のない情報が返されず、情報取り出し処理が散らからないように、ユーザが検索用語を狭くして定義するのを助ける。このような用語を手動で狭くすることは、特定の情報を探すときにソートする数千のサイトを受信することを軽減することができる。

20

【0003】

このような場合において検索トピックは、トピック領域及びサブトピック領域に前もって配置される。たとえば「Yahoo」は、可能性のあるトピック(例えばビジネス、政府、化学など)の階層的に配置された既定のリストを提供し、ユーザはトピックを選択し、その後リスト内のサブトピックをさらに選択する。トピックの既定のリストの別の例は、デスクトップのパーソナルコンピュータのヘルプユーティリティ上で共通であり、ヘルプトピックのリスト及び関連したサブトピックがユーザに提供される。これらの既定された階層は、いくつかの状況において役に立つであろう一方で、ユーザはしばしばこれらの既定のリストの外部及び/又は含まれていない情報を検索/調べる必要がある。従って検索エンジン及び他の検索システムは、所望の情報を見つけるために直接の問い合わせ(クエリ)をユーザに可能にするためにしばしば採用される。それでもなお、ユーザがどのように特定の問い合わせを作成するか又は構築するかは確かでないかもしれないので、ユーザが検索している間に多くの関係のない結果が取り出される。さらに、このようなシステムは普通、問い合わせを継続的に修正することをユーザに要求し、検査するために適度な数の結果を取得するために取り出された検索結果の精度を高める。

30

【0004】

検索システムの問い合わせ入力フィールドにおいて語又は句を打ち込み、その後可能性のある候補として数百万の結果を取り出すことは一般的ではない。多数の取り出された候補を理解するために、ユーザは他の語の組み合わせでしばしば試して、リストをさらに狭める。

40

【0005】

一般に検索システムは、問い合わせに対する予測された関連性のある結果に従って、結果を順位付ける(rank)であろう。この順位付けは一般に、問い合わせへのウェブページの類似性も、ドキュメントの固有の質も含む多くのパラメータを結合する機能に基づき、しばしばウェブトポロジー情報から推定される。ユーザは一般により低く位置付けられた結果を閲覧しないので、ユーザの検索の経験の質は、位置付け機能の質に直接関係する。

【0006】

一般に検索システムは、「検索した」トピックがユーザが実際に興味を持っていたトピックの領域又はカテゴリに状況的な関係を有さないか有すかどうかに関わらず、ユーザの問い合わせ入力に関係する全てのトピックをマッチ又は見つけるようとする。例として、

50

天文学に興味を持ったユーザが従来の検索システムに「土星」という問い合わせを入力した場合、関係のないすべてのタイプの結果が、「土星」という語を有する車、自動車販売業者、コンピュータゲーム及び他のサイトに関するものを含んで返される可能性が高い。従来の検索の遂行の他の問題は、検索エンジンは異なるユーザのニーズ及び環境に関わらず全てのユーザに同じように動作することである。したがって、2人のユーザが同じ検索問い合わせを入力する場合、彼らの興味又は特徴、過去の検索履歴、現在のコンピューティングの状況（例えばファイルが開かれている）又は環境的な状況（例えば、使用されている機械、時刻、曜日）に関わらず、彼らは通常同じ結果を取得する。

【発明の開示】

【発明が解決しようとする課題】

10

【0007】

関連性のある結果をトップに戻すための検索位置付け機能をチューニングすることは、重大な努力を必要とする。現代の検索エンジンに対する一般的なアプローチは、順位付け機能を訓練し、機能パラメータを設定し、手動で順位付けした検索結果の例に基づいて自動的に重み付けすることである。人の注釈者はとらえた関連性に従って問い合わせに対するページのセットを明確に順位付けることができ、異なる順位付けアルゴリズムがチューニングされ評価されうる「黄金のスタンダード」を作成する。しかし明確な人の順位付けは高価で取得が難しく、訓練が完全でなく、準最適な順位付け機能である。

【課題を解決するための手段】

【0008】

20

以下は、特許請求された主題のいくつかの態様の基本的な理解を提供するために簡略化された概要を示す。この課題を解決するための手段は、広い概要である。これは鍵となる/重要な要素を特定すること、又は特許請求された主題の範囲を描くことを意図しない。この唯一の目的は、以降で示されるより詳細な説明に先立つものとして、簡略化された形式でいくつかの概念を示すことである。

【0009】

この主題の新しい部分は、先を争うアイテムに対するユーザの選択を予測するために、ユーザの振舞いの集合について自動的に解釈を容易にするユーザ振舞いコンポーネントを採用することを介して、情報検索システム(information retrieval system)における検索順位付け(ランク)を拡張する。このような選択は、結果の順位付けを改善するなどの様々な目的のためにその後採用できる。ユーザの振舞いコンポーネントは、検索エンジンと対話でき、通常ユーザの振舞い（例えば悪意のある及び/又は不合理なユーザの動作）を伴うノイズを軽減するフィードバック特性を含むことができる。ユーザの振舞いの集合を生かすことにより（例えば各ユーザを個々の専門家として扱うのではなく）、主題の新しい部分は、ノイズを軽減でき、ユーザのフィードバックから関連性のある判断を生成できる。ユーザの振舞いコンポーネントは、ユーザからのフィードバック及び前の問い合わせからの結果との対話を黙示的に又は明確に採用できる。鍵となる振舞いの特性は、結果のタイトル及び説明で見ることにより結果が関連性があるかどうかをユーザが決定するのを助けることが出来る表示特性(presentation feature)；ページ上の滞在時間(dwelling time)などのブラウジング特性(browsing feature)であり、ドメイン上の平均時間からの差で到達する検索結果の方式（例えば他のリンク）など；問い合わせに対する特定の結果でクリック数などのクリックスルー特性(clickthrough feature)を含む。所与の問い合わせ結果ペアの主題の工夫点は、各特性タイプについて観察され導き出された複数の特性値を提供する。

30

40

【0010】

ユーザ振舞いコンポーネントは、ユーザの振舞いのデータドリブン(data-driven: データ駆動)のモデルを採用できる。例えば、2つのコンポーネント: 「バックグラウンド」コンポーネント、(無差別にクリックするユーザなど)及び「関連性のある」コンポーネント、(問い合わせへの結果の関連性により影響される問い合わせ特有の振舞いなど)、により生成されたかのように、ユーザの振舞いコンポーネントは、ユーザのウェブ検索の振舞

50

いをモデル化できる。

【0011】

この主題の工夫点のさらなる態様によれば、ユーザの振舞いコンポーネントは、期待されたユーザの振舞いからの差を生成及び/又はモデル化できる。したがって、導き出された特性をコンピューティング(演算)でき、このような導き出された結果は、問い合わせに寄らない情報で、結果に対する期待値からの所与の検索結果について、観察された特性値の差に明確に向けられる。

【0012】

さらに、主題の工夫点のユーザの振舞いコンポーネントは、ユーザの振舞いを説明するための2つの特性タイプを有し、直接(direct)及び偏差(deviational)という名であり、前者は直接計測された値であり、後者是对応する直接観察された特性についての(問い合わせ独立の)分布全体から予測された期待値からの差である。したがって、問い合わせq及び結果rに対する特性fの観察された値oは、2つのコンポーネントの混合として表現でき、

$$o(q, r, f) = C(r, f) + rel(q, r, f)$$

であり、 $C(r, f)$ はrに対応する全ての問い合わせにまたがって統合されたfの値についての前の「バックグラウンド」分布であり、 $rel(q, r, f)$ は、問い合わせへの結果の関連性により影響された振舞いの「関係性のある」コンポーネントである。例えば、ユーザの振舞いの関連性の予測は、所与の位置で観察されたクリックスルー周波からのバックグラウンド分布の減算を介してクリックスルー特性で取得できる。振舞いにおける個々のユーザのバリエーションの影響を軽減するために、主題の工夫点により、各問い合わせ結果ペアについて、全てのユーザおよび検索セッション(session)にまたがる特性値を平均化できる。このような集合は、さらなる構造安定性を供給でき、「うるさい(noisy)」ユーザの対話に頼らない。

【0013】

したがって、問い合わせ結果ペアに対するユーザの振舞いは、直接観察された特性及び導き出された「修正された」特性値の両方を含む特性ベクトルにより表すことができる。様々な機械学習技術もまた、情報検索システム(information retrieval system)に対する順位付けアルゴリズムの訓練と併せて採用できる。例えば明確な人の関連性の判断を、初期的には様々な検索問い合わせについて提供し、それに続く順位付けアルゴリズムを訓練するために採用できる。

【0014】

関連した態様において、ウェブ検索エンジンと対話するユーザの集約的な振舞いは、未来のユーザの選択を予言するために自動的に解釈でき;従ってシステムをユーザ振舞いパターン、及び最近のユーザ振舞いデータで自動的にシステムを再訓練することにより異なる検索設定を変更するように順応させることができる。

【0015】

上記及びその関連の達成のために、特許請求された主題のある例示的な態様がここで、以下の詳細な説明及び添付の図面と併せて説明される。これらの態様は主題が実施できる様々な方法を示し、これらの全ては特許請求された主題の範囲内であることを意図する。他の利点及び新規の特徴は、以下の図面と併せて考慮されるとき、以下でより詳細な説明から明らかとなるであろう。

【発明を実施するための最良の形態】

【0016】

本発明の様々な態様が添付の図面を参照してここで説明され、全体を通して同様の番号は、同様又は対応する要素を示す。しかし、これに関係する添付の図面及び発明を実施するための最良の形態は、開示された特定の形式で特許請求の範囲の主題を限定することを意図しないことを理解されたい。むしろ、意図は特許請求された主題の精神及び範囲内で全ての修正、均等物及び代替物をカバーすることである。

【0017】

個々で使用される用語である「コンポーネント」「システム」「特性」などは、コンピュータ関連の実体物を意図し、ハードウェア、ハードウェア及びソフトウェアの組み合わせ、ソフトウェア、又は実行におけるソフトウェアである。例えば、コンポーネントは、プロセッサ上で実行中のプロセス、プロセッサ、オブジェクト、実行ファイル、実行のスレッド、プログラム及び/又はコンピュータであることができるが、これらに限定されない。例として、コンピュータ上で実行中のアプリケーション及びコンピュータはコンポーネントであることができる。1又は複数のコンポーネントは、プロセス内及び/又は実行のスレッド内に存在することができ、コンポーネントを1コンピュータ上で及び/又は2以上の分散されたコンピュータ上でローカライズすることができる。

【0018】

10

「例示的な」という語は、例としての提供を意味するためにここで使用され、例えば又は例示である。「例示的な」としてここで説明される任意の態様又は設計は、他の態様又は設計上で好適又は利点として解釈される必要はない。

【0019】

さらに、開示の主題はシステム、方法、装置、又はここで詳細化された態様を実装するために製造するための標準のプログラミング及び/又はエンジニアリング技術を使用する製造項目として実装できる。コンピュータプログラムと言う用語は、ここで使用されるように、任意のコンピュータ可読デバイス、キャリア又はメディアからアクセス可能なコンピュータプログラムを包含することを意図する。例えばコンピュータ可読媒体は、磁気ストレージディスク（例えばハードディスク、フロッピー（登録商標）ディスク、磁気ストライプ）、光ディスク（例えばCD (Compact Disk)、DVD (Digital Versatile disk)...)、スマートカード、フラッシュメモリデバイス（例えばカード、スティック）を含むことができるがこれらに限定はされない。さらに当然のことながら、搬送波は、電子メールを送受信する又はインターネット若しくはLAN (Local Area Network)などのネットワークにアクセスするのに使用される、コンピュータ可読の電子データを搬送するのに使用できる。当然、当業者は、特許請求された主題の範囲又は精神から逸脱することなくこの構成に多くの修正がなされうることを認識されたい。

20

【0020】

最初に図1に戻り、システム100のブロック図が示されており、主題の工夫点の例示的な態様によって検索エンジンと対話するユーザ振舞いコンポーネントを組み込む。検索エンジン102に関連付けられたユーザ振舞いコンポーネント104は、ユーザ101、103、105（1からNであり、Nは整数）の振舞いの集合を自動的に解釈できる。このようなユーザ振舞いコンポーネント104は、ノイズを軽減するフィードバック特性を含むことが出来、一般的にユーザ振舞いに伴う（例えば悪意のある及び/又は不合理なユーザ動作）。ユーザ101、103、105の振舞いの集合を利用することにより（例えば個々の専門家として個々のユーザを扱わず）、システム100はノイズを軽減でき、ユーザのフィードバックから関連性のある判断を生成できる。

30

【0021】

ユーザ振舞いコンポーネント104は、順位付けコンポーネント対話ができる。所与の問い合わせに対して、ユーザ振舞いコンポーネント104は、この問い合わせに対する前に訓練された振舞いモデルから導き出された予測を取り出し、前のユーザについての関連性が現れた結果がより高く順位付けられるように、問い合わせに対する結果を再順序付けする。例えば、所与の問い合わせ q 、暗黙のスコア $I S_r$ は、利用可能なユーザの対話特性からそれぞれの結果 r について演算でき、各結果に対して暗黙の順位付け I_r という結果となる。合併されたスコア $S M(r)$ は、暗黙のフィードバック I_r から取得された順位を r 、 O_r の元の順位と結合することにより、 r について演算できる。

40

【0022】

【数 1】

$$SM(r, I_r, O_r, w_i) = \begin{cases} w_i \frac{1}{I_r + 1} + \frac{1}{O_r + 1} & r \text{ についての暗黙のフィードバックは存在する場合} \\ \frac{1}{O_r + 1} & \text{さもなければ} \end{cases}$$

【0023】

重み w_i は、暗黙のフィードバックの関連性のある「重要性」を表す発見的問題解決法であるチューニングされたスケール要素である。この問い合わせ結果を、最終的な順位付けを作るために $SM(r)$ の値を減らすことにより順序付けることができる。このようなモデルの一特定のな場合は、 w_i がとても大きい値であるとき m クリックされていない結果よりクリックされた結果がより高くなるよう効率的にさせることが挙げられる - ベースラインとして採用される直感的及び効率的な発見的問題解決法。一般に、上述のアプローチは、元のウェブサイト順位付けを作る基礎の特性及び暗黙のフィードバック特性の間に対話がないと想定する。主題の工夫点の他の態様は、以下の詳細で説明されるように、暗黙のフィードバック特性を順位付けプロセスに直接統合することにより、このような想定を緩和する。さらに、より高度化されたユーザの振舞い及び順位付け組み合わせアルゴリズムが採用され、主題の工夫点の範囲内と同様であることは当然である。

10

【0024】

図 2 は、主題の工夫点のさらなる態様を図示し、検索エンジン 202 は主題の工夫点の態様に従って、訓練モデル 204 をさらに備える。訓練モデル 204 は、観察された振舞い特性 201 及び導き出された振舞い特性 203 という名で、ユーザの振舞いを説明するための追加のモデルタイプをさらに備えることができる。観察された振舞い特性 201 は直接測定された値であり、導き出された振舞い特性 203 は、対応する直接観察された特性に対する全体（問い合わせ独立の）分布から予測された期待された値から導き出される。従って、問い合わせ q 及び結果 r に対する特性 f の観察された値は 2 つのコンポーネントの混合として表現できる：

20

$$o(q, r, f) = C(r, f) + rel(q, r, f)$$

ここで $C(r, f)$ は、 r に対応する全ての問い合わせに渡って統合された f の値の前の「バックグラウンド」分布であり、 $rel(q, r, f)$ は結果の関連性により影響を受けた振舞いのコンポーネントである。例えばユーザの振舞いの関連性の予測は、所与の位置で観察されたクリックスルー周波からバックグラウンド分布（例えばノイズ）の減算を介して、クリックスルー特性で取得できる。振舞いにおける個々のユーザのバリエーションの影響を軽減するために、主題の工夫点により、各問い合わせ URL ペアについて、全てのユーザおよび検索セッション(session)にまたがる直接の特性値を平均化できる。このようなさらなるロバスト性を供給でき、個々の「うるさい」ユーザ対話によらない。したがって問い合わせ URL ペアに対するユーザの振舞いは、直接観察された特性及び導き出された「修正された」特性値の両方を含む特性ベクトルにより表現できる。

30

【0025】

図 3 は、主題の工夫点の例示的な態様によって、動作可能なように接続された順位付けコンポーネント 310 をユーザ振舞いコンポーネント 315 及び検索エンジン 340 に組み込むシステム 300 のブロック図である。典型的には、検索エンジン 340 は、以下の詳細で説明されるように、コンテンツベースの特性（例えば問い合わせがテキスト若しくはタイトル又はドキュメントのアンカーテキストにどのくらい近く一致するか）、及び問い合わせ独立ページ量特性（例えばドキュメント又はドメインのページランク (PageRank)）を含む多数の特性に基づいて、検索結果 350 を順位付けできる。さらに、検索エンジン 340 は、このような特性値を結合する特定の順位付け機能をチューニングするための自動的（又は準自動的）方法を採用できる。例えば、問い合わせ 360 を出すユーザが特定の動作を実行すると仮定する。このような動作は、関連性のあるドキュメントを見つけたらなど、クリックすること、ナビゲートすること、改良した問い合わせを出すこ

40

50

とを含むことができる。関連性のあるドキュメントを見つけた上で、ユーザは満足し、振舞いを変更する(たとえばドキュメントを読むこと)。主題の工夫点は、ユーザが取り出された結果に満足するときの検出を可能にする特性の十分なリッチセットを考え出すことを可能にする。このような特性は、出された問い合わせにより、したがって問い合わせ有である。例えばユーザの特性/動作は、図4を参照して説明されるように、表示特性、ブラウジング特性及びクリックスルー特性に分類できる。

【0026】

図4は、ユーザブラウジング動作を表示する特性400のテーブルである。表示特性410は、振舞いのいくつか又は全ての態様に影響するものとして(例えばユーザは表示特性に基づいて結果の上でクリックを決定することができる)、ユーザの経験を表示するよう、一般的に設計できる。ユーザの経験のこのような態様をモデル化するために、主題の工夫点は、問い合わせのタイトル及び語における語の重複(Title Overlap)及び問い合わせ並びに結果概要により共有される語の断片などの特性を採用でき、完全なドキュメントを閲覧するために結果の概要上でクリックするかどうかの決定をするときにユーザによりしばしば考慮される。

10

【0027】

同様に、ブラウジング特性420は、ユーザのウェブページ対話の対話をキャプチャ及び定量化できる。例えば主題の工夫点は、問い合わせに対する期待されたページの滞在時間から滞在時間の差を演算でき、ページブラウジング振舞いの内部問い合わせの多様性をモデル化することを可能にする。これは以下の詳細で説明されるように、直接の特性及び導き出された特性の両方をさらに含むことができる。同様にクリックスルー特性430は検索エンジン結果とのユーザ対話の例である。例えばクリックスルー特性は問い合わせ結果ペアに対するクリック数又は期待したクリックの可能性からの差を含むことができる。

20

【0028】

図4に図示されるように、クリックスルーは、ウェブ検索エンジンとのユーザの対話の一態様を示す。主題の工夫点は、自動的に導き出された予測ユーザ振舞いモデルを採用できる。従って、所与の問い合わせに対して、各結果を図4のテーブルにおける特性で表現できる。関連性のあるユーザ選択をその後、詳細が上記に説明されたように、学習したユーザ振舞いモデルを使用して予測できる。このようなユーザ振舞いモデルの使用は、検索エンジンが、検索結果と対話する群の堅実さだけでなく、検索結果ページを越えるブラウジング特性を特徴付けるよりリッチな特性からも利益を得ることを可能にする。

30

【0029】

図5は、主題の工夫点の態様によって、機械学習コンポーネント535を採用できる自動情報検索システム500を示す。一般的な暗黙のフィードバック解釈戦略は、ユーザ選択のモデルを自動的に学習することを採用できる(例えば発見的問題解決又は洞察力による代わりに)。システム500は、データログ520又は例えばユーザ振舞いコンポーネント515との対話から訓練できる。ログ520のデータは、ローカル又はリモートのデータソースから集めることができ、複数のユーザからの前の検索データ又は動作530に関する情報を含んでいる。訓練の後、順位付けコンポーネント510は、関連性のある結果550として示される未来の検索結果を容易又は高めるために、検索エンジン540と対話できる。例えば1又は複数の新しい検索問い合わせ560は、前の検索データ530からの訓練の一部に基づく検索エンジン540及び/又はユーザ振舞いコンポーネント515からの情報により処理できる。一般にシステム500は、検索エンジンの関連性を改善するために様々なデータマイニング技術を採用できる。これは実行時間の分類子(classifiers)に対する高品質の訓練データを生成するために、順位付けコンポーネント510において関連性の分類子を採用することを含み、検索結果550を生成するために、検索エンジン540で採用できる。図6は、ユーザ動作を表す複数のシステム特性と対話するユーザ振舞いコンポーネント610を示す。一態様において、主題の工夫点は、「バックグラウンド」コンポーネント(例えば、ユーザの振舞いにおける問い合わせ独立ノイズ及び関連性独立ノイズなど)及び「関連性」コンポーネント(例えば、問い合わせへの結果の

40

50

関連性を示す問い合わせに特有の振舞い)の組み合わせとして、ウェブ検索の振舞いを考慮する。このような編成は、統合されたユーザの振舞いの利益を得ることができ、特性セットは、直接観察された特性だけでなく、問い合わせ特有の導き出された特性に含まれ(各問い合わせに対する観察から直接演算される)、対応する直接観察された特性値に対する問い合わせ独立の分布からの差として演算される。図6に示されるように、クリックスルー特性612、ブラウジング特性614及び表示特性616などの例示的なシステム特性は、ユーザ振舞いコンポーネント610を通して、ウェブ検索結果とのユーザ体を表すために採用できる。さらに、所与の問い合わせ-所与の位置における結果上での期待されたクリック数からのURLペア、に対する観察されたクリックスルー数の差としての特性も考慮される。さらにブラウジング振舞いはモデル化でき、例えば結果がクリックされた後、その後に所与の問い合わせURLペアに対する平均的なページ滞在時間並びに期待された(平均)滞在時間からの差が、このようなモデルのために採用できる。さらに例えば、ウェブ検索のユーザは、結果タイトル、URL及び概要を見ることにより結果が関連性があるかどうかを決定することができ、-多くの場合において元のドキュメントを見ることは一般的に必要である。ユーザ経験のこの態様をモデル化するために、タイトルにおける語及びクエリにおける語を重複するなどの特徴も採用できる。

10

【0030】

図7は主題の工夫点の態様によって、ユーザ選択を予測するためにユーザの振舞いを解釈する例示的な方法論700を示す。例示的な方法が様々なイベント及び/または動作の表現の一連のブロックとしてここで例示され説明されている一方で、主題の工夫点はこのようなブロックの例示された順に限定されない。例えば、このような動作又はイベントは、工夫点によるここで例示された順とは別個で、異なる順及び/又は他の動作又はイベントと共に起こり得る。さらに、全ては示されていないがブロック、イベント又は動作は、主題の工夫点による方法論を実装するのに必要とされる。さらに工夫点による例示的な方法及び他の方法が、ここで例示及び説明された方法と関連付けてだけでなく、例示又は説明されていない他のシステム及び装置と関連付けても実装できることは当然である。最初に710で、過去の検索ユーザ振舞いなどの検索エンジンとのユーザ対話に関連したデータが取得できる。続いて720でユーザの振舞いは統合でき、例えば統計の分析技術を採用することによってである。730で機械学習をその後、ユーザ選択モデルを訓練するために採用できる。続いて740でユーザ選択の予測を、未来の問い合わせの結果として供給できる。

20

30

【0031】

図8は主題の工夫点の態様による順位付けの一部としてユーザの振舞いを実装する方法論800を示す。最初に810で、ユーザの振舞いに関連するデータを収集する。このようなユーザの振舞いはその後、820で振舞いモデルを訓練及び/又は自動的に生成するために採用できる。このようなモデル(例えば予測振舞いモデル)はその後、順位付け結果への検索エンジンの一部として組み込まれ、及び/又は830でユーザのフィードバックからの暗黙の関連性のある判断を生成することができる。続いて検索エンジンにより取り出された生成された及び/又は訓練された振舞いのモデル情報の一部に基づく830を、その後に順位付けることができる。

40

【0032】

開示された主題の工夫点の様々な態様に対する状況を提供するために、図9及び図10並びに以下の議論は、開示された主題の様々な態様が実装できる適切な環境の簡単で一般的な説明を提供することを意図する。主題がコンピュータ及び又は複数のコンピュータ上で実行するコンピュータプログラムのコンピュータ実行可能命令の一般的な状況において説明されてきた一方で、当業者は工夫点が他のプログラムモジュールと組み合わせることもできることを理解されたい。一般にプログラムモジュールは、ルーチン、プログラム、コンポーネント、データ構造などを含み、特定のタスクを実行、及び/又は特定の抽象データ型を実装する。さらに、当業者にとって、工夫した方法が、シングルプロセッサ若しくはマルチプロセッサコンピュータシステム、ミニコンピューティングデバイス

50

、メインフレームコンピュータ、並びにパーソナルコンピュータ、ハンドヘルド（携帯）コンピューティングデバイス（PDA(Personal digital assistant)、電話、時計...）、マイクロプロセッサベース又はプログラマブル家庭用電化製品若しくは工業電化製品などを含む他のコンピュータシステム構成で実行できることは当然である。例示された態様は、タスクが通信ネットワークを通してリンクされたりリモート処理デバイスにより実行される、分散コンピューティング環境においても実施できる。しかし、工夫点の態様の全てではないかもしれないが一部は、スタンドアロンコンピュータで実施できる。分散コンピューティング環境において、プログラムモジュールを、ローカル及びリモートメモリストレージデバイスの両方に位置付けることができる。

【0033】

図9を参照して、主題の工夫点の様々な態様を実装するためのコンピュータ912を含む例示的環境910が説明されている。コンピュータ912は、処理ユニット914、システムメモリ916及びシステムバス918を含む。システムバス918は、システムメモリ916を含むがこれに限定されないシステムコンポーネントを、処理ユニット914に結合する。処理ユニット914は様々な利用可能なプロセッサの任意であることができる。デュアルマイクロプロセッサ及び他のマルチプロセッサアーキテクチャが処理ユニット914として採用できる。

【0034】

システムバス918は、11ビットバス、ISA(Industrial Standard Architecture)、MCA(Micro Channel Architecture)、EISA(Extended ISA)、IDE(Intelligent Drive Electronics)、VLB(VESA Local Bus)、PCI(Peripheral Component Interconnect)、USB(Universal Serial Bus)、AGP(Advanced Graphics Port)、PCMCIA(Personal Computer Memory Card International Association Bus)及びSCSI(Small Computer Systems Interface)を含むがこれらに限定されない様々な利用可能なバスアーキテクチャの任意を使用する、メモリバス若しくはメモリコントローラ、周辺バス又は外部バス及び/又はローカルバスを含む多くの種類のバス構造の任意であることができる。

【0035】

システムメモリ916は、揮発性メモリ920、不揮発性メモリ922を含む。スタートアップの間などのコンピュータ912内の要素間で情報を伝えるための基本ルーチンを含むBIOS()は、不揮発性メモリ922に格納される。例示としてであり限定ではなく、不揮発性メモリ922は、ROM(read only memory)、PROM(programmable ROM)、EPROM(electrically programmable ROM)、EEPROM(electrically erasable ROM)又はフラッシュメモリを含むことができる。揮発性メモリ920は、RAMを含み、外部のエコメモリとして動作する。例としてであり限定ではなく、RAMは、SRAM(synchronous RAM)、DRAM(dynamic RAM)、SDRAM(synchronous DRAM)、DDR SDRAM(double data rate SDRAM)、ESDRAM(enhanced SDRAM)、SLDRAM(Synch link DRAM)及びDRRAM(direct Rambus RAM)などの多くの形式で利用可能である。

【0036】

コンピュータ912もまた、リムーバブル/リムーバブルでない、揮発性/不揮発性のコンピュータストレージメディアを含む。図9は例えばディスクストレージ924を示すディスクストレージ924は、磁気ディスクドライブ、フロッピー（登録商標）ディスクドライブ、テープドライブ、Jazドライブ、Zipドライブ、LS-60ドライブ、フラッシュメモリカード又はメモリスティックなどのデバイスを含むがこれらに限定されない。さらにディスクストレージ924は、CD-ROM(compact disk ROM)、CD-Rドライブ(CD recordable drive)、CD-RWドライブ(CD rewritable drive)又はDVD-ROM(digital versatile disk ROM)などの光ディスクドライブを含むがこれらに限定されない他のストレージと分離して又は組み合わせて、ストレージメディアを含むことができる。ディスクストレージデバイス924のシステムバス918への接続を容易にするために、リムーバブル又はリムーバブルでないインターフェースは、インターフェース92

10

20

30

40

50

6などで一般に使用できる。

【0037】

当然のことだが図9は、適切なオペレーティング環境910で説明されるユーザ及び基本のコンピュータリソースの間の中継として動作するソフトウェアを説明する。このようなソフトウェアはオペレーティングシステム928を含む。ディスクストレージ924上に格納できるオペレーティングシステム928は、コンピュータシステム912のリソースを制御及び割り当てるために動作する。システムアプリケーション930は、システムメモリ916又はディスクストレージ924のいずれかに格納されたプログラムモジュール932及びプログラムデータ934を通して、オペレーティングシステム928によりリソースの管理を活用する。当然のことながら、ここで説明される様々なコンポーネントは、様々なオペレーティングシステム又はオペレーティングシステムの組み合わせで実装できる。

10

【0038】

ユーザは入力デバイス936を通してコンピュータ912にコマンド又は情報を入力する。入力デバイス936は、マウス、トラックボール、スタイラス(stylus)、タッチパッド、キーボード、マイクロフォン、ジョイスティック、ゲームパッド、衛星放送受信アンテナ(satellite dish)、スキャナ、TVチューナーカード、デジタルカメラ、デジタルビデオカメラ、ウェブカメラなどのポインティングデバイスを含むがこれらに限定されない。これら又は他の入力デバイスは、インターフェースポート938を介してシステムバス918を通して処理ユニット914に接続する。インターフェースポート938は、例えばシリアルポート、パラレルポート、ゲームポート、USB(universal serial bus)を含む。出力デバイス940は、入力デバイス936として同じタイプのポートのいくつかを使用する。従って例えば、USBポートはコンピュータ912への入力、及びコンピュータ912から出力デバイス949への情報の出力を提供するのに使用されるかもしれない。出力アダプタ942は、特別なアダプタを必要とする他の出力デバイス940の間で、モニタ、スピーカー及びプリンタのようないくつかの出力デバイス940があることを示すために提供される。出力アダプタ942は出力デバイス940及びシステムバス918の間の接続手段を提供するビデオカード及びサウンドカードを例示としてであり限定ではなく含む。他のデバイス及び/又はデバイスのシステムは、リモートコンピュータ944などの入力機能及び出力機能の両方を提供する。

20

30

【0039】

コンピュータ912は、リモートコンピュータ944などの1又は複数のリモートコンピュータへの論理的接続を使用するネットワーク環境において動作できる。リモートコンピュータ944は、パーソナルコンピュータ、サーバ、ルータ、ネットワークPC、ワークステーション、電化製品ベースのマイクロプロセッサ、ピアデバイス又は他の一般ネットワークノードなどであることができ、一般にコンピュータ912に関連して説明される多く又は全ての要素を含む。簡潔さの目的のために、メモリストレージデバイス946はリモートコンピュータ944で図示される。リモートコンピュータ944は、ネットワークインターフェース948を通してコンピュータ912に論理的に接続され、その後通信接続950を介して物理的に接続される。ネットワークインターフェース948は、LAN(local-area networks)及びWAN(wide-area networks)などの通信ネットワークを包含する。LAN技術はFDDI(Fiber Distributed Data Interface)、CDDI(Copper Distributed Data Interface)、イーサネット(登録商標)/IEEE802.3、トークンリング/IEEE802.5などの通信ネットワークを含む。WAN技術はポイント・ツー・ポイントリンク、ISDN(Integrated Services Digital Networks)及びそのバリエーション、パケットスイッチングネットワーク、DSL(Digital Subscriber Lines)などのサーキットスイッチングネットワークを含むがこれらに限られない。

40

【0040】

通信接続950は、ネットワークインターフェース948をバス918に接続するのに採用されるハードウェア/ソフトウェアのことを言う。通信接続950が内部コンピュー

50

タ 9 1 2 内に明確に図示するために示される一方で、コンピュータ 9 1 2 の外部であることもできる。ネットワークインターフェース 9 4 8 への接続のためのハードウェア/ソフトウェア要件は、例としての目的のためであり、通常の電話回線用モデム、ケーブルモデム、DSL モデム、ISDN アダプタ及びイーサネットカードを含むモデムなどの内部及び外部の技術を含む。

【 0 0 4 1 】

ここで使用されるように「コンポーネント」「システム」などの用語は、コンピュータ関連の実体物を指すことを意図し、ハードウェア、ハードウェア及びソフトウェアの組み合わせ、ソフトウェア、又は実行におけるソフトウェアのいずれかである。例えば、コンポーネントは、プロセッサ上で実行中のプロセス、プロセッサ、オブジェクト、実行ファイル、実行のスレッド、プログラム及び/又はコンピュータであることができるが、これらに限定されない。例として、コンピュータ上で実行中のアプリケーション及びコンピュータはコンポーネントであることができる。1 又は複数のコンポーネントは、プロセス内及び/又は実行のスレッド内に存在することができ、コンポーネントを 1 コンピュータ上で及び/又は分散された 2 以上のコンピュータ上でローカライズすることができる。「例示的な」という語は、例えば又は例示としての提供を意味するためにここで使用される。「例示的な」としてここで説明される任意の態様又は設計は、他の態様又は設計上で好適又は利点として解釈される必要はない。

【 0 0 4 2 】

さらに、開示された主題はシステム、方法、装置、又はここで詳細化された態様を実装するために製造するための標準のプログラミング及び/又はエンジニアリング技術を使用する製造項目として実装できる。コンピュータプログラムという用語は、ここで使用されるように、任意のコンピュータ可読デバイス、キャリア又はメディアからアクセス可能なコンピュータプログラムを包含することを意図する。例えばコンピュータ可読媒体は、磁気ストレージディスク（例えばハードディスク、フロッピー（登録商標）ディスク、磁気ストライプ）、光ディスク（例えば CD (Compact Disk)、DVD (Digital Versatile disk) ...）、スマートカード、フラッシュメモリデバイス（例えばカード、スティック）を含むことができるがこれらに限定はされない。さらに当然のことながら、搬送波を、電子メールを送受信するか又はインターネット若しくは LAN (Local Area Network) などのネットワークにアクセスするのに使用される、コンピュータ可読の電子データを搬送するために採用できる。当然、当業者は、特許請求された主題の範囲又は精神から逸脱することなくこの構成に多くの修正がなされうることを認識されたい。

【 0 0 4 3 】

図 1 0 は、主題の工夫点の態様によって、ユーザ振舞いコンポーネントを介してユーザの選択を予測するために採用できるサンプルのコンピューティング環境 1 0 0 0 の概念ブロック図である。システム 1 0 0 0 は、1 又は複数のクライアント 1 0 1 0 を含む。クライアント 1 0 1 0 はハードウェア及び/又はソフトウェア（例えばスレッド、プロセス、コンピューティングデバイス）であることができる。システム 1 0 0 0 は 1 又は複数のサーバ 1 0 3 0 を含むこともできる。サーバ 1 0 3 0 も又はハードウェア及び/又はソフトウェア（例えばスレッド、プロセス、コンピューティングデバイス）であることができる。サーバ 1 0 3 0 例えここで説明されるコンポーネントを採用することにより変換を実行するためのスレッドを収容することができる。クライアント 1 0 1 0 及びサーバ 1 0 3 0 の間の 1 つの可能性のある通信は、2 以上のコンピュータプロセス間で送信されるように適用されたデータパケットの形式であるかもしれない。システム 1 0 0 0 は、クライアント 1 0 1 0 及びサーバ 1 0 3 0 の間の通信を容易にするために採用できる通信フレームワーク 1 0 5 0 を含む。クライアント 1 0 1 0 は、クライアント 1 0 1 0 にローカルに情報を格納するために採用できる 1 又は複数のクライアントデータストア 1 0 6 0 に接続可能である。同様にサーバ 1 0 3 0 はサーバ 1 0 3 0 にローカルに情報を格納するのに採用できる 1 又は複数のサーバデータストア 1 0 4 0 に接続可能である。

【 0 0 4 4 】

上述されてきた事柄は、様々な例示的な態様を含む。当然、これらの態様を説明する目的のために、すべてのあり得るコンポーネント又は方法論の組み合わせを説明することは不可能だが、当業者は、多くのさらなる組み合わせ及び置換が可能であることを理解されたい。したがって、ここで説明された態様は、添付の特許請求の範囲の精神及び範囲内にある全てのこのような代替、修正及び変形を包含することを意図する。

【0045】

さらに、明細書又は特許請求の範囲のいずれかで「含む(include)」という用語が使用される限り、特許請求の範囲で暫定的な語として採用されるときに「備える」は解釈されるので、このような用語は「備える(comprising)」という用語と同様の方式で含まれることになることを意図する。

【図面の簡単な説明】

【0046】

【図1】この主題の工夫点の例示的な態様によるユーザ振舞いコンポーネントのブロック図を示す。

【図2】この主題の工夫点の態様によって、ユーザ振舞いコンポーネントを組み込み、検索エンジンの訓練モデルと対話するシステムのブロック図を示す。

【図3】この主題の工夫点の例示的な態様によって、ユーザ振舞いコンポーネントに動作可能なように接続された順位付けコンポーネント及び検索エンジンを組み込むシステムのブロック図を示す。

【図4】この主題の工夫点の態様によって、ユーザのブラウジング動作を表す特性の表を示す。

【図5】この主題の工夫点の態様によって、機械学習コンポーネントを採用できる、自動化された情報検索システムを示す図である。

【図6】この主題の工夫点の特定の態様によって、ユーザ動作を表す、複数のシステム特性と対話するユーザ振舞いコンポーネントを示す図である。

【図7】この主題の工夫点の態様によって、ユーザ選択を予測するためのユーザの振舞いの解釈の例示的な方法論を示す図である。

【図8】この主題の工夫点の態様によって、値の順位付けの一部としてユーザの振舞いを実装する方法論を示す図である。

【図9】この主題の工夫点の様々な態様を実装するための例示的な環境を示す図である。

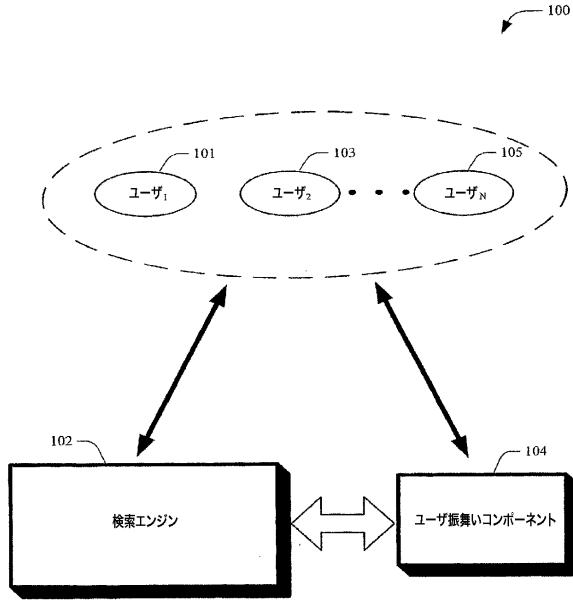
【図10】この主題の工夫点の様々な態様を実装するために採用できる追加のコンピューティング環境の概念ブロック図を示す。

10

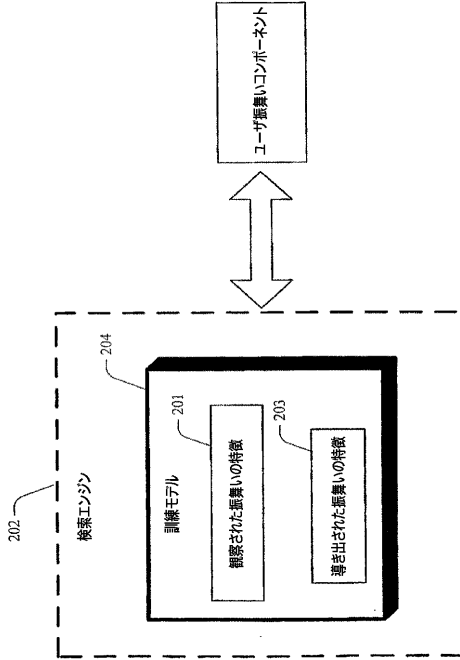
20

30

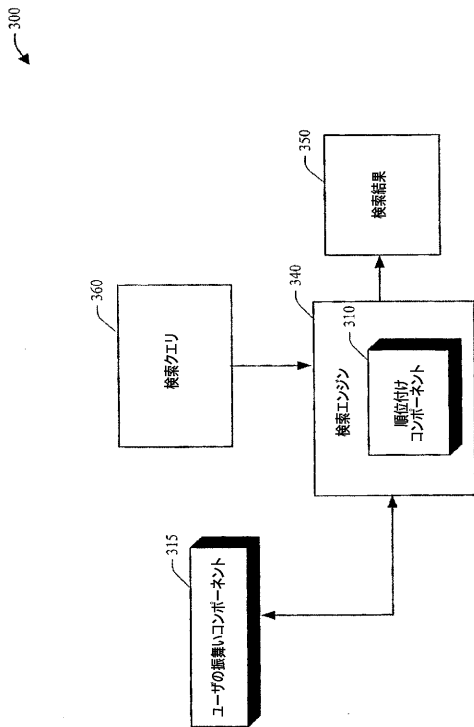
【図1】



【図2】



【図3】

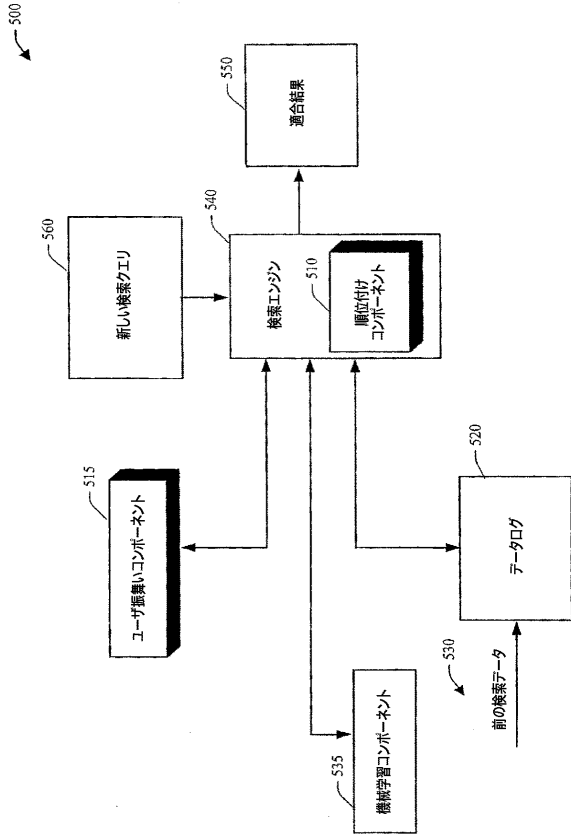


【図4】

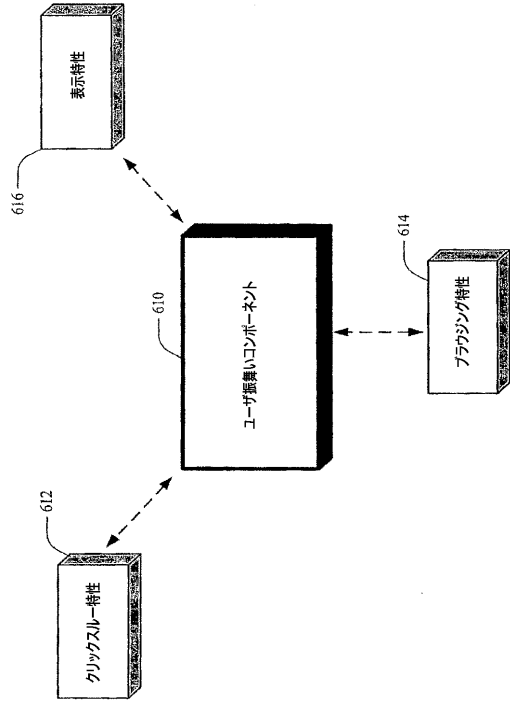
Figure 4 is a table listing various features used in the system, categorized into three groups: Click through features, Browsing features, and Presentation features.

Click through features	
Position	Position of the URL in Current ranking
ClickFrequency	Number of clicks for this query, URL pair
ClickProbability	Probability of a click for this query and URL
ClickDeviation	Deviation from expected click probability
IsNextClicked	1 if there is a click on next position, 0 otherwise
IsPreviousClicked	1 if there is a click on previous position, 0 otherwise
IsClickAbove	1 if there is a click above, 0 otherwise
IsClickBelow	1 if there is a click below, 0 otherwise
Browsing features	
TimeOnPage	Page dwell time
CumulativeTimeOnPage	Cumulative time for all subsequent pages after search
TimeOnDomain	Cumulative dwell time for this domain
TimeOnShortUrl	Cumulative time on URL, prefix, dropping parameters
IsFollowedLink	1 if followed link to result, 0 otherwise
IsExactUrlMatch	0 if aggressive normalization used, 1 otherwise
IsRedirected	1 if initial URL same as final URL, 0 otherwise
IsPathFromSearch	1 if only followed links after query, 0 otherwise
ClicksFromSearch	Number of hops to reach page from query
AverageDwellTime	Average time on page for this query
DwellTimeDeviation	Deviation from overall average dwell time on page
CumulativeDeviation	deviation from average cumulative time on page
DomainDeviation	deviation from average time on domain
ShortURLDeviation	deviation from average time on short URL
Presentation features	
TitleOverlap	Fraction of shared words between query and title
SummaryOverlap	Fraction of shared words between query and summary
QueryURLOverlap	Fraction of shared words between query and URL
QueryDomainOverlap	Fraction of shared words between query and domain
QueryLength	Number of tokens in query
QueryNextOverlap	Average fraction of words shared with next query

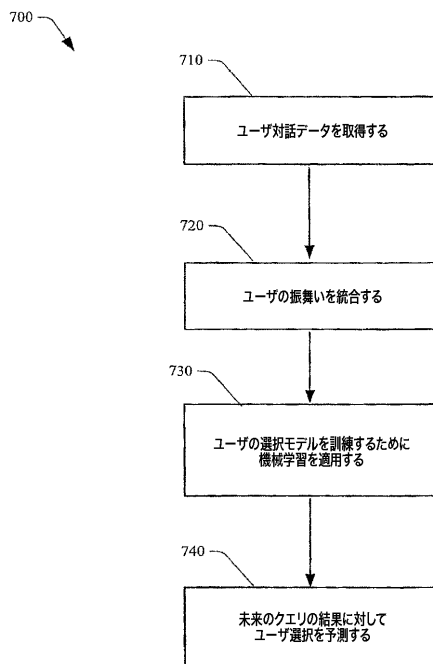
【図5】



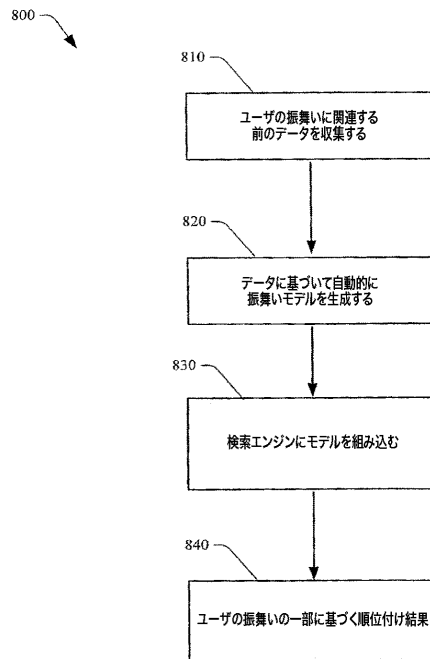
【図6】



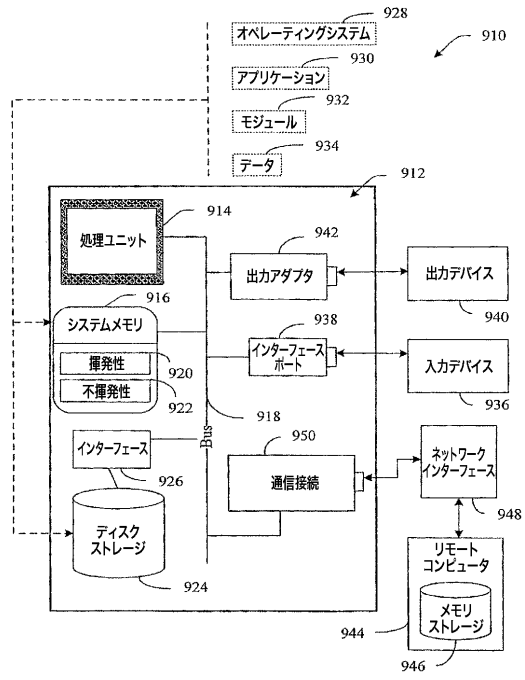
【図7】



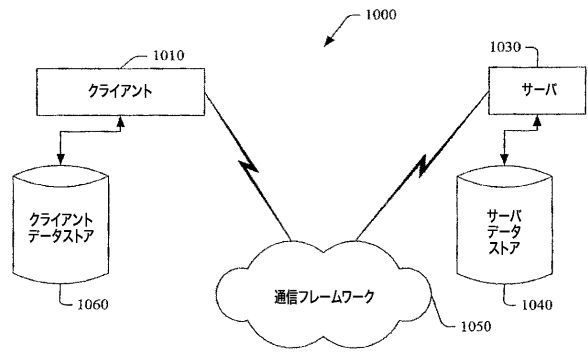
【図8】



【図9】



【図10】



フロントページの続き

- (72)発明者 エフゲニー イー・アジチテイン
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ
イクロソフト コーポレーション インターナショナル パテント内
- (72)発明者 エリック デー・プリル
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ
イクロソフト コーポレーション インターナショナル パテント内
- (72)発明者 スーザン ティー・デュマイス
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ
イクロソフト コーポレーション インターナショナル パテント内
- (72)発明者 ロバート ジェー・ラグノ
アメリカ合衆国 98052 ワシントン州 レッドモンド ワン マイクロソフト ウェイ マ
イクロソフト コーポレーション インターナショナル パテント内

審査官 野崎 大進

- (56)参考文献 米国特許出願公開第2005/0071328(US, A1)
特開2002-032401(JP, A)
特開2005-208943(JP, A)

- (58)調査した分野(Int.Cl., DB名)
G06F 17/30