

(19) 日本国特許庁 (JP)

(12) 公表特許公報 (A)

(11) 特許出願公表番号

特表2011-516989

(P2011-516989A)

(43) 公表日 平成23年5月26日 (2011.5.26)

(51) Int. Cl.		F 1		テーマコード (参考)
GO6F 17/30	(2006.01)	GO6F 17/30	350C	5B075
GO6F 13/00	(2006.01)	GO6F 17/30	380E	5B084
		GO6F 13/00	540E	

審査請求 未請求 予備審査請求 未請求 (全 23 頁)

(21) 出願番号	特願2011-504031 (P2011-504031)	(71) 出願人	500046438 マイクロソフト コーポレーション アメリカ合衆国 ワシントン州 98052-6399 レッドモンド ワン マイクロソフト ウェイ
(86) (22) 出願日	平成21年3月10日 (2009.3.10)	(74) 代理人	110001243 特許業務法人 谷・阿部特許事務所
(85) 翻訳文提出日	平成22年10月12日 (2010.10.12)	(72) 発明者	ウラディミール タンコビッチ 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 エルシーエー-インターナショナル パテント内
(86) 国際出願番号	PCT/US2009/036597		
(87) 国際公開番号	W02009/126394		
(87) 国際公開日	平成21年10月15日 (2009.10.15)		
(31) 優先権主張番号	12/101, 951		
(32) 優先日	平成20年4月11日 (2008.4.11)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

(54) 【発明の名称】 編集距離および文書情報を使用する検索結果順位付け

(57) 【要約】

クエリ文字列に基づいて検索結果として受信した文書から文書情報を抽出して、データ文字列とクエリ文字列の間の編集距離を計算するためのアーキテクチャ。編集距離は、クエリ全体またはクエリの一部のニアマッチを検出することにより、結果の順序付けの一部として文書の関連度を決定する際に採用される。編集距離は、クエリ文字列が、T A U C (タイトル、アンカーテキスト、URL、クリック) 情報などの文書情報を含む所与のデータストリームまでどのくらい近いかを評価する。アーキテクチャは、クエリ用語をさらに効果的に発見できるように、URLの複合語のインデックス-時間分割を含む。加えて、文書結果のうちの1つまたは複数の最上位N個のアンカーを見出すために、アンカーテキストのインデックス-時間フィルタリングが使用される。T A U C情報は、検索結果を順位付けするための関連度メトリックを向上させるために、ニューラルネットワーク (たとえば、2層) に入力することができる。

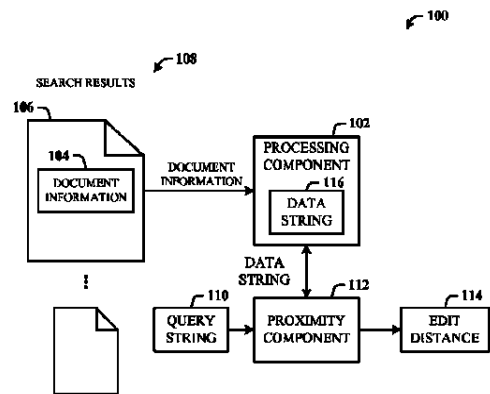


FIG. 1

【特許請求の範囲】

【請求項 1】

コンピュータにより実装される関連度システム（100）であって、
クエリ文字列に基づいて検索結果として受信した文書から文書情報を抽出するための処理コンポーネント（102）と、

データ文字列と前記クエリ文字列の間の編集距離を計算するための近接性コンポーネント（112）であって、前記編集距離は結果順位付けの一部として文書の関連度を決定する際に採用される、コンポーネントと

を備えることを特徴とするシステム。

【請求項 2】

前記データ文字列を生成するために採用される前記文書情報は、タイトル情報、URL 情報、クリック情報、またはアンカーテキストのうちの少なくとも 1 つを含むことを特徴とする請求項 1 に記載のシステム。

【請求項 3】

前記処理コンポーネントは、インデックス時間において前記文書情報の複合語を分割して URL に対する前記編集距離を計算することを特徴とする請求項 1 に記載のシステム。

【請求項 4】

前記処理コンポーネントは、インデックス時間において前記文書情報のアンカーテキストをフィルタリングしてアンカーテキストの上位順位付けセットを計算することを特徴とする請求項 1 に記載のシステム。

【請求項 5】

前記文書情報は、タイトル文字、アンカー文字、クリック文字、または URL 文字のうちの少なくとも 1 つを含むが、その文書情報は、BM25F 関数、クリック距離、ファイルタイプ、言語および URL 深さの生の入力特徴と共にニューラルネットワークに入力されて、前記文書の関連度を計算することを特徴とする請求項 1 に記載のシステム。

【請求項 6】

前記編集距離の計算は、前記データ文字列と前記クエリ文字列の間の近接性を高めるために、用語の挿入および削除に基づいて行われることを特徴とする請求項 1 に記載のシステム。

【請求項 7】

前記編集距離の計算は、前記データ文字列と前記クエリ文字列の間の近接性を高めるために、用語の挿入および削除に関連付けられているコストに基づいて行われることを特徴とする請求項 1 に記載のシステム。

【請求項 8】

関連度を決定するコンピュータ実施の方法であって、
クエリ文字列を検索プロセスの一部として受信するステップ（1000）と、
文書情報を前記検索プロセス中に返された文書から抽出するステップ（1002）と、
データ文字列を前記文書情報から生成するステップ（1004）と、
前記データ文字列と前記クエリ文字列の間の編集距離を計算するステップ（1006）と、

前記編集距離に基づいて関連度スコアを計算するステップ（1008）と
を備えることを特徴とする方法。

【請求項 9】

前記編集距離を計算するステップの一部として用語挿入を採用するステップと、前記データ文字列を生成するために前記クエリ文字列に用語を挿入する挿入コストを計算するステップであって、前記コストは重み付けパラメータとして表される、ステップをさらに備えることを特徴とする請求項 8 に記載の方法。

【請求項 10】

前記編集距離を計算するステップの一部として用語削除を採用するステップと、前記データ文字列を生成するために前記クエリ文字列の用語を削除する削除コストを計算するス

10

20

30

40

50

トップであって、前記コストは重み付けパラメータとして表される、ステップをさらに備えることを特徴とする請求項 8 に記載の方法。

【請求項 1 1】

前記編集距離を計算するステップの一部として位置コストを計算するステップであって、前記位置コストは前記データ文字列内の用語位置の用語挿入および／または用語削除に関連付けられている、ステップをさらに備えることを特徴とする請求項 8 に記載の方法。

【請求項 1 2】

前記編集距離を計算する全体コストを計算するため、前記データ文字列の文字と前記クエリ文字列の文字の間で照合プロセスを実行するステップをさらに備えることを特徴とする請求項 8 に記載の方法。

【請求項 1 3】

インデックス時間において前記データ文字列の URL の複合語を分割するステップをさらに備えることを特徴とする請求項 8 に記載の方法。

【請求項 1 4】

前記文書内の出現度数に基づいてアンカーテキストの上位順位付けセットを見出すために前記データ文字列のアンカーテキストをフィルタリングするステップをさらに備えることを特徴とする請求項 8 に記載の方法。

【請求項 1 5】

前記セット内のアンカーテキストの編集距離を計算するステップをさらに備えることを特徴とする請求項 1 4 に記載の方法。

【請求項 1 6】

前記編集距離を計算するステップから導かれたスコアを、変換関数の適用後に 2 層ニューラルネットワークに入力するステップであって、前記スコアはタイトル情報、アンカー情報、クリック情報、または URL 情報のうちの少なくとも 1 つおよびその他の生の入力特徴に関連付けられている前記編集距離を計算するステップに基づいて生成される、ステップをさらに備えることを特徴とする請求項 8 に記載の方法。

【請求項 1 7】

文書の関連度を計算するコンピュータ実施の方法であって、クエリ文字列を検索プロセスの一部として処理して文書の結果セットを返すステップ (1 1 0 0) と、

前記結果セットの文書から抽出された文書情報に基づいてデータ文字列を生成するステップであって、前記文書情報は前記文書からのタイトル情報、アンカーテキスト情報、クリック情報、および URL 情報のうちの 1 つまたは複数を含む、ステップ (1 1 0 2) と、

用語挿入、用語削除、および用語位置に基づいて前記データ文字列と前記クエリ文字列の間で編集距離を計算するステップ (1 1 0 4) と、

前記編集距離に基づいて関連度スコアを計算するステップであって、前記関連度スコアは前記結果セットの前記文書を順位付けするために使用される、ステップ (1 1 0 6) とを備えることを特徴とする方法。

【請求項 1 8】

前記用語挿入、用語削除、用語位置のそれぞれに関連付けられているコストを計算するステップと、前記コストを前記関連度スコアの計算に組み込むステップとをさらに備えることを特徴とする請求項 1 7 に記載の方法。

【請求項 1 9】

インデックス時間において前記 URL 情報の複合語を分割するステップと、インデックス時間において前記アンカーテキスト情報をフィルタリングして前記文書内のアンカーテキストの出現度数に基づいて前記アンカーテキストの上位順位付けセットを見出すステップとをさらに備えることを特徴とする請求項 1 7 に記載の方法。

【請求項 2 0】

ソース URL 文字列内の出現の順序でクエリ用語の文字列を構築するために前記クエリ

10

20

30

40

50

文字列の用語の出現を読み取るステップと、前記用語の間のスペースを単語マークで充填するステップとをさらに備えることを特徴とする請求項17に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、編集距離および文書情報を使用する検索結果順位付けに関する。

【背景技術】

【0002】

標準的な検索エンジンサービスにおいて、ユーザは、クエリと一致するURL (universal resource locator) のインデックス付きコレクションから最上位の関連文書を選択することにより、クエリを入力することができる。クエリに迅速に応じるために、検索エンジンは、キーワードを文書にマップする1つまたは複数の方法 (たとえば、逆インデックスデータ構造体) を使用する。たとえば、エンジンにより実行される第1のステップは、ユーザのクエリによって指定されたキーワードを含む候補文書のセットを識別することであってもよい。これらのキーワードは、文書本文もしくはメタデータ、または (アンカーテキストのような) 他の文書もしくはデータストアに実際に格納されるこの文書に関する追加のメタデータにあってもよい。

【0003】

大規模なインデックスコレクションにおいて、候補文書セットの基数は、クエリ用語の共通性に応じて大きく (たとえば、場合によっては数百万に) なる可能性もある。検索エンジンは、候補文書のセット全体を返すのではなく、関連度に関して候補文書を順位付けする第2のステップを実行する。通常、検索エンジンは、特定のクエリに対する文書の関連度の度合いを予測するために、順位付け関数を使用する。順位付け関数は、文書から複数の特徴を入力として取り出し、検索エンジンが予測された関連度により文書をソートできるようにする数値を計算する。

【発明の概要】

【発明が解決しようとする課題】

【0004】

関数がいかに正確に文書の関連度を予測するかに関する順序付け関数の品質は、最終的に、検索結果に対するユーザの満足度、またはユーザが提示した質問に回答を得られる平均回数によって決まる。システムに対する全般的なユーザ満足度は、単一の数値 (またはメトリック) によって概算されてもよいが、それは順位付け関数を変更することで数値を最適化できるからである。通常、メトリックは、クエリログの無作為抽出により事前に選択されるクエリの代表的なセットにわたり計算され、評価クエリのそれぞれに対してエンジンによって返される結果ごとに関連度ラベルを割り当てるステップを伴う。しかし、文書の順序付けおよび関連度のためのこれらのプロセスは、望ましい結果をもたらすには依然として非効率的である。

【課題を解決するための手段】

【0005】

本明細書において説明される一部の新規な実施形態の基礎的な理解をもたらすために、簡略化した要約 (課題を解決するための手段) を以下に提示する。この要約 (課題を解決するための手段) は、広範に及ぶ概要ではなく、主要/重要な要素を特定すること、またはその範囲を詳細に記述することを意図するものではない。その唯一の目的は、後段において提示されるさらに詳細な説明の前置きとして簡略化された形態で一部の概念を提示することである。

【0006】

アーキテクチャは、クエリ文字列に基づいて検索結果として受信した文書から文書情報を抽出して、データ文字列とクエリ文字列の間の編集距離を計算するための機構を提供する。データ文字列は、たとえば、TAUC (タイトル、アンカーテキスト、URL (uniform resource locator)、およびクリック) のような、文書情

10

20

30

40

50

報から得られた文書の簡略で正確な記述であってもよい。編集距離は、結果の順序付けの一部として文書の関連度を決定する際に採用される。機構は、クエリ全体またはクエリの一部のニアマッチを検出するために近接性に関連する特徴のセットを採用し、検索結果の順位付けの関連度を向上させる。

【0007】

編集距離は、クエリ文字列が、文書情報を含む所与のデータストリームまでどのくらい近いかを評価するために処理される。アーキテクチャは、クエリ用語をさらに効果的に発見できるように、URLに複合語のインデックス-時間分割を含む。加えて、1つまたは複数の文書結果の最上位N個のアンカーを見出すために、アンカーテキストのインデックス-時間フィルタリングが使用される。TAUCを使用して、情報はニューラルネットワーク（たとえば、2層）に入力され、検索結果を順位付けするための関連度メトリックを向上させることができる。

10

【0008】

前述の目的および関連する目的を達成するため、本明細書において特定の例示の態様が、以下の説明および付属の図面と共に説明される。しかし、それらの態様は、本明細書において開示される原理が採用されうるさまざまな方法のほんの一部を示すに過ぎず、すべてのそのような態様および等価物を含むことが意図される。その他の利点および新規の特徴は、図面と併せて検討すれば以下の詳細な説明から明らかとなる。

【図面の簡単な説明】

【0009】

【図1】コンピュータにより実施される関連度システムを示す図である。

【図2】編集距離を計算するための例示の照合アルゴリズムを示す流れ図である。

【図3】修正された編集距離および照合アルゴリズムを使用し、クエリ文字列およびデータ文字列に基づいて編集距離値を処理して生成するステップを示す図である。

【図4】修正された編集距離および照合アルゴリズムを使用し、クエリ文字列およびデータ文字列に基づいて編集距離値を処理して生成するステップのもう1つの例を示す図である。

【図5】ニューラルネットワークを採用して文書の関連度スコアの生成を補助するコンピュータ実施の関連度システムを示す図である。

【図6】クエリ文字列とデータ文字列の間の編集距離を決定するために文書情報に採用されるデータの型を示す図である。

【図7】インデックス-時間処理データフローを示す図である。

【図8】結果を順位付けするための図7のインデックスプロセスからニューラルネットワークへの入力を示すブロック図である。

【図9】検索結果の生成を計算するためのニューラルネットワーク、編集距離入力、および生の特徴入力の例示的なシステムの実施態様を示す図である。

【図10】文書結果セットの文書関連度を決定する方法を示す図である。

【図11】文書の関連度を計算する方法を示す図である。

【図12】開示されるアーキテクチャによる、TAUC特徴を使用して検索結果を順位付けするために編集距離処理を実行するよう動作可能なコンピューティングシステムを示すブロック図である。

【発明を実施するための形態】

【0010】

開示されるアーキテクチャは、クエリ全体のニアマッチ、またはタイトル、アンカー、URL、もしくはクリックのような、文書に関する正確なメタデータとの一致を検出するために近接性に関連する特徴のセットを実施することにより、検索結果順序付けの関連度を向上させる。たとえば、クエリ「company store」、第1の文書の文書タイトル「company store online」および第2の文書の文書タイトル「new NEC LCD monitors in company store」を考察する。その他の特性は、第1の文書および第2の文書のいずれも同じであると仮定し

20

30

40

50

て、アーキテクチャは、選択されたストリームをクエリと一致させるためにどれほどの編集努力が注がれるかに基づいて、文書のスコアを割り当てる。この例において、文書タイトルは、評価のために選択される。完全一致させるため、第1の文書のタイトルには（「online」という用語を削除する）1つの削除操作だけが必要であるが、第2の文書のタイトルには（「new」、「NEC」、「LCD」、「monitors」、および「in」という用語を削除する）5つの削除が必要である。したがって、第1の文書は、より関連度が高いと計算される。

【0011】

タイトルは、クエリ用語が複合語から見出されるようにデータの一部のストリーム（たとえば、URL）に処理を適用することができるTAUC（タイトル、アンカー、URL、およびクリック）文書情報の1つの要素である。たとえば、ここでもまた、クエリ「company store」を考察し、URLが「www.companystore.com」とする。その結果、URLが「www」、「company」、「store」、および「com」という4つの部分（または用語）に分割される。

10

【0012】

これ以降、図面が参照されるが、全体を通じて類似する要素を参照するために類似する番号が使用される。以下の説明において、その十分な理解をもたらすために、説明を目的として多数の具体的な詳細が示される。しかし、新規の実施形態がそれらの特定の詳細なくして実施されうることは明らかとなろう。その他の例において、よく知られた構造体およびデバイスが、その説明を容易にするためにブロック図の形態で示される。

20

【0013】

図1は、コンピュータにより実施される関連度システム100を示す図である。システム100は、クエリ文字列110に基づいて検索結果108として受信した文書106から文書情報104を抽出するための処理コンポーネント102を含む。システム100はまた、文書情報104から導かれたデータ文字列116とクエリ文字列110の間の編集距離114を計算するための近接性コンポーネント112を含むこともできる。編集距離114は、検索結果108の一部として文書106の関連度を決定する際に採用される。

【0014】

データ文字列116を生成するために採用される文書情報104は、たとえば、タイトル情報（または文字）、リンク情報（たとえば、URL文字）、クリックストリーム情報、および／またはアンカーテキスト（または文字）を含むことができる。処理コンポーネント102は、インデックス時間において文書情報104の複合語を分割して、編集距離114を計算する。処理コンポーネント102はまた、インデックス時間において、アンカーテキストのような文書情報をフィルタリングして、アンカーテキストの上位順位付けセットを計算する。

30

【0015】

編集距離114の計算は、データ文字列116とクエリ文字列110の間の近接性を高める（近づける）ために、用語の挿入および削除に基づいて行われる。編集距離114の計算はまた、データ文字列116とクエリ文字列110の間の近接性を高める（近づける）ために、用語の挿入および削除に関連付けられているコストに基づいて行われてもよい。

40

【0016】

クエリ文字列110からの用語の挿入および／または削除に基づいてデータ文字列116（たとえば、TAUC）を生成するシナリオを考察する。この用語の処理は、非クエリの単語をクエリ文字列110に挿入する、クエリ用語をクエリ文字列110に挿入する、クエリ文字列110からTAUC用語を削除する、および／または、非TAUCの用語をクエリ文字列110から削除する、という4つの操作に従って実行されてもよい。

【0017】

編集距離114は、挿入操作および削除操作に基づくが、置換には基づいていない。挿入に対して定義されるコストには2つのタイプがあってもよい。データ文字列116をク

50

クエリ文字列 110 から生成するシナリオを考察する。生成する際、単語はクエリ文字列 110 に挿入することができるが、これが元のクエリ文字列 110 に存在する場合、コストは 1 として定義され、それ以外の場合、コストは w_1 (≥ 1) として定義される。ここで、 w_1 は、調整される重み付けパラメータである。たとえば、クエリ文字列 110 が AB である場合、ABC のデータ文字列を生成するコストは、データ文字列 ABA を生成するコストよりも高い。「不適切な単語」をデータ文字列 116 に挿入することでデータ文字列 116 (たとえば、T A U C) 全体をさらに不適切にすることは直感的に分かる。

【0018】

削除のコストには 2 つのタイプがあってもよい。ここでもまた、データ文字列 116 をクエリ文字列 110 から生成するシナリオを考察する。クエリ文字列 110 の用語を削除する場合、その用語が元のデータ文字列 116 に存在すれば、コストは 1 として定義され、それ以外の場合、コストは w_2 (≥ 1) として定義される。

10

【0019】

コストのもう 1 つのタイプは、位置コストである。削除または挿入がデータ文字列 116 の第 1 の位置において行われた場合、追加のコストがある ($+w_3$)。2 つの文字列 (クエリ文字列 110 およびデータ文字列 116) の最初の部分の一致は、文字列の後半部分の一致よりも重要視されることは直感的に分かる。クエリ文字列 110 が「cnn」であり、データ文字列 116 がタイトル = 「cnn. com-blur blur」である以下の例を考察する。挿入または削除が第 1 の位置において行われる場合、それはソリューションの有効性を大幅に低下させる可能性がある。

20

【0020】

図 2 は、編集距離を計算するための例示の修正済み照合アルゴリズム 200 を示す流れ図である。説明を簡潔にするために、本明細書において、たとえば流れ図またはフローダイアグラムの形態で示される 1 つまたは複数の方法論は、一連の動作として示され説明されるが、一部の動作はそれに従って、本明細書において示され説明される他の動作と異なる順序および/または他の動作と同時に生じることもあるので、方法論は、動作の順序に限定されないことを理解および認識されたい。たとえば、当業者は、方法論が代替として、状態図におけるように、一連の相互関係のある状態またはイベントとして表されてもよいことを理解し認識するであろう。さらに、方法論に示される動作のすべてが、新規の実施形態に必須とされるわけではない。

30

【0021】

200 において、クエリ文字列およびデータ (またはターゲット) 文字列の要素が列挙される。これは、クエリ文字列の長さを n に設定し (クエリ文字列内の各用語は $s[i]$ である)、ターゲット (またはデータ) 文字列の長さを m に設定する (ターゲット文字列内の各用語は $t[j]$ と表される) ことで達成される。202 において、 $0 \dots m$ 行および $0 \dots n$ 列を含む行列が構築される (ここで、行列内の各用語は $d[j, i]$ と表される)。204 において、第 1 行は、削除のさまざまなコストに依存する値で初期化され、第 1 列は、挿入のさまざまなコストに依存する値で初期化される。206 において $n = 0$ である場合、208 に示されるように $d[m, 0]$ を返して終了し、206 において $M = 0$ である場合、208 に示されるように $d[0, n]$ を返して終了する。210 において、クエリ文字列の各文字は検査される (1 から n までの i)。212 において、ターゲット文字列の各文字は検査される (1 から m までの j)。214 において、クエリ文字列内の文字列がデータ文字列内の文字と等しい場合、フローは、コストがゼロであり次の行列セルが処理される 216 に進む。言い換えれば、 $s[i]$ が $t[j]$ と等しい場合、コストは 0 であり、 $d[j, i] = d[j-1, i-1]$ である。

40

【0022】

クエリ文字列セル内の文字がデータ文字列セル内の文字と等しくない場合、フローは 214 から 218 に進み、そこで現在のセルは真上のセルまたはすぐ左のセルに挿入コストまたは削除コストを加えた値に設定される。言い換えれば、 $s[i]$ が $t[j]$ と等しくない場合、行列のセル $d[j, i]$ を、対応する挿入コストを加えた真上のセル ($d[j$

50

- 1, i] + cost_insertionで表される) または対応する削除コストを加えたすぐ左のセル (d [j, i - 1] + cost_deletionで表される) の最小値と等しく設定する。220において、ステップ210、212、214、216、および218は、完了するまで繰り返される。222において、セルd [m, n] で得られた最終コストが出力される。例におけるcost_insertionおよびcost_deletionは、たとえば挿入コストにw = 1、w3 = 4、および削除コストにw2 = 1、w4 = 26というように、2種類の値を有することに留意されたい。

【0023】

言い換えれば、d [j, i] は、文字列s [0...i] とt [0...j] の間の編集距離を含む。定義によりd [0, 0] = 0 (空文字列を空文字列と等しくするために編集は必要ない)。d [0, y] = d [0, y - 1] + (w2またはw4)。文字列d [0, y - 1] を作成するために編集が使用される回数が既知である場合、d [0, y] はd [0, y - 1] + ターゲット文字列から現在の文字を削除するコスト (そのコストはw2またはw4となりうる) として計算されてもよい。現在の文字がs [0...n]、t [0...m] の両方に存在する場合、コストw2が使用され、それ以外の場合w4が使用される。d [x, 0] = d [x - 1, 0] + (w1またはw3)。文字列d [x - 1, 0] を作成するために編集が使用される回数が既知である場合、d [x, 0] はd [x - 1, 0] + s から t までの現在の文字の挿入のコスト (そのコストはw1またはw3となりうる) として計算されてもよい。現在の文字がs [0...n]、t [0...m] の両方に存在する場合、コストw1が使用され、それ以外の場合w3が使用される。

10

20

【0024】

各 (j, i) について、s [i] = t [j] である場合、d [j, i] はd [j - 1, i - 1] と等しくなりうる。編集距離は文字列t [j - 1]、s [i - 1] の間で計算されてもよく、s [i] = t [j] である場合、文字列を等しくするために、編集を使用せずに、共通の文字を両方の文字列に付加することもできる。したがって、3つの移動が採用されており、現在のd [j, i] の最小編集距離をもたらす移動が選択される。言い換えれば、

【0025】

【表1】

```
d[j,i] = min(
    d[j-1,i-1] if s[i]=t[j];
    d[j-1,i] + (w1, if s[j] is present in both strings; else, w3);
    d[j,i-1] + (w2, if t[i] is present in both strings; else, w4)
)
```

30

【0026】

図3は、修正された編集距離および照合アルゴリズムを使用し、クエリ文字列およびデータ文字列に基づいて編集距離値を処理して生成するステップを示す。プロセスは、左右、上下、および対角の計算のうちの1つまたは複数に伴う。用語「ABC」のクエリ文字列は、用語「CBAX」(Xはクエリ文字列内にはない用語を表す) のターゲットデータ文字列に対して処理される。編集距離を計算するためのプロセスは、さまざまな方法で実行されてもよいが、編集距離の修正バージョンを実行するための特定の詳細は、開示されるアーキテクチャにより計算される場合とは異なっている。4x5行列300は、n x mに基づいて構築され、クエリ文字列の場合n = 3であり、データ文字列の場合m = 4である。クエリ文字列302は、行列300の横軸に沿って配置され、ターゲットデータ文字列304は縦軸に沿って配置される。

40

【0027】

説明では、4列(0~3)および5行(0~4)で表される行列300を使用する。図2で説明される編集距離照合アルゴリズムを行0、列0で始めて左から右に適用すると、クエリ文字列ABCの空セルとターゲットデータ文字列CBAXの空セルを比較してもクエリ文字列をターゲットデータ文字列と同じにするために用語の挿入または削除を生じな

50

いので、交差するセル $d[0, 0]$ は「0」を受信する。「用語」は同じなので、編集距離はゼロである。

【0028】

クエリ文字列 302 の用語 A と行 0 の空セルを比較するように右へ移動すると、文字列を同じにするために 1 回の削除を使用するので、セル $d[0, 1]$ は「1」の値を受信する。再度右に列 2 まで移動すると、クエリ文字列 302 の用語 AB とターゲットデータ文字列の空セルの間で比較が行われる。したがって、文字列を同じにするためにクエリ文字列 302 で 2 回の削除が使用され、その結果「2」の編集距離がセル $d[0, 2]$ に入れられる。同じプロセスは列 3 に適用され、ここでクエリ文字列 302 の用語 ABC がターゲット文字列の空セルと比較され、文字列を同様にするために 3 回の削除を使用して、その結果「3」の編集距離がセル $d[0, 3]$ に入れられる。

10

【0029】

行 1 に下り、左から右へと続行すると、クエリ文字列行の空セルがターゲットデータ文字列 304 の第 1 の用語 C と比較される。文字列を同じにするために 1 回の削除が使用され、「1」の編集距離が $d[1, 0]$ に入れられる。右に列 1 まで移動すると、クエリ文字列 302 の用語 A とターゲットデータ文字列 304 の用語 C の間で比較が行われる。文字列を同様にするために削除および挿入が使用され、したがって、「2」の値がセル $d[1, 1]$ に挿入される。最後のセル $d[1, 3]$ までスキップすると、ABC を C に照合するための照合プロセスは、結果として 2 回の削除を使用して「2」の編集距離が $d[1, 3]$ に入れられる。簡略にするため、および全体の編集距離を求めるために行 4 および列 3 に移動すると、用語 ABC の用語 CBA X との照合は、結果として「8」の編集距離をセル $d[4, 3]$ にもたますが、ターゲット文字列の第 1 の用語 C に挿入／削除を使用して「2」の値、用語 B の照合に「0」の値、第 3 の用語 C および A の照合のための挿入／削除に「2」の値、用語 X の挿入に「1」の値、および位置コストに「3」の値で、結果として「8」の最終編集距離値がセル $d[4, 3]$ に入れられる。

20

【0030】

図 4 は、修正された編集距離および照合アルゴリズムを使用し、クエリ文字列およびターゲットデータ文字列に基づいて編集距離値を処理して生成するステップのもう 1 つの例を示す。ここで、行列 400 は、挿入コストの $w_1 = 1$ 、 $w_3 = 4$ の `cost_insertion`、および削除コストの $w_2 = 1$ および $w_4 = 26$ の重み付けに基づいて、ABC のクエリ文字列 402 を AB のターゲットデータ文字列 404 と比較するために生成される。言い換えれば、行 0 を左から右に作業し、クエリ文字列 402 の用語 A をターゲット文字列 404 の前の空セルに照合すると、結果として用語 A のターゲット文字列 404 への 1 回の挿入で「1」の値がセル $d[0, 1]$ に入れられる。クエリ文字列 402 の用語 AB をターゲット文字列 404 の前の空セルに照合すると、結果として用語 AB のターゲット文字列 404 への 2 回の挿入で「2」の値がセル $d[0, 2]$ に入れられ、クエリ文字列 402 の用語 ABC をターゲット文字列 404 の前の空セルに照合すると、結果として用語 AB のターゲット文字列 404 への 2 回の挿入の値に用語 C に対する値 $w_4 = 26$ を加えて（用語 C は両方の文字列にはないので）「28」の値がセル $d[0, 3]$ に入れられる。

30

40

【0031】

行 1 を左から右に作業し ($d[1, 0] = 1$ であることを理解して)、クエリ文字列 402 の用語 A をターゲット文字列 404 の用語 A に照合すると、結果としてターゲット文字列 404 とクエリ文字列 402 の等価で ($d[j-1, i-1] = d[0, 0] = \text{「0」}$ から値を取ることにより) 「0」の値がセル $d[1, 1]$ に入れられる。クエリ文字列 402 の用語 AB をターゲット文字列 404 の用語 A に照合すると、結果として用語 B のターゲット文字列 404 への 1 回の挿入で「1」の最小値がセル $d[1, 2]$ に入れられる。セル $d[1, 3]$ について、クエリ文字列 402 の用語 ABC をターゲット文字列 404 の用語 A に照合すると、結果として $d[j-1, i] = d[0, 3]$ に w_3 を加えた値に関連付けられている最小値で「28」の値がセル $d[1, 3]$ に入るが、（用語 C は

50

両方の文字列にはないので) $d[j, i-1] = d[1, 2]$ の値 1 に 26 を加えた値 27 と比較して、結果として「27」の最小値がセル $d[1, 3]$ に入れられる。

【0032】

行 2 を左から右に作業し、クエリ文字列 402 の用語 A をターゲット文字列 404 の用語 AB に照合すると、結果としてターゲット文字列 404 の 1 回の削除で「1」の値がセル $d[2, 1]$ に入れられる。セル $d[2, 2]$ の距離についてクエリ文字列 402 の用語 AB をターゲット文字列 404 の用語 AB に照合すると、結果として等価となり、それにより $d[j-1, i-1] = d[1, 1]$ から値を取りセル $d[2, 2]$ の値「0」とする。セル $d[2, 3]$ について、クエリ文字列 402 の用語 ABC をターゲット文字列 404 の用語 AB に照合すると、結果として $d[j-1, i] = d[1, 3]$ に $w_3 = 1$ を加えた値に関連付けられている最小値で「28」の値となるが、比較して、用語 C はターゲット文字列にはないので (また、用語 C は両方の文字列にないので、 $d[i, j-1] = d[2, 2] = 0$ に 26 を加えた値に基づいて)、 $d[2, 3]$ の最小値が $d[2, 3]$ に入れられる。

10

【0033】

図 5 は、ニューラルネットワーク 502 を採用して文書 106 の関連度スコア 504 の生成を補助するコンピュータ実施の関連度システム 500 を示す。システム 500 は、クエリ文字列 110 に基づいて検索結果 108 として受信した文書 106 から文書情報 104 を抽出するための処理コンポーネント 102、および文書情報 104 から導かれたデータ文字列 116 とクエリ文字列 110 の間の編集距離 114 を計算するための近接性コンポーネント 112 を含む。編集距離 114 は、検索結果 108 の一部として文書 106 の関連度を決定する際に採用される。

20

【0034】

ニューラルネットワーク 502 は、文書 106 の関連度スコアを計算する際の入力として文書情報 104 を受信するために採用されてもよい。一部またはすべての検索結果 108 の関連度スコアのみに基づくかまたはその一部に基づいて、検索結果 108 内の文書は順位付けられてもよい。システム 500 は、ニューラルネットワーク 502 およびコードベースを採用して、検索結果 108 内の関連する文書を順序付けるための関連度スコアを生成する。

【0035】

以下で、クエリ文字列と各データ文字列の間の編集距離を計算して各ペアの TAUC スコアを取得するための編集距離アルゴリズムを説明する。

30

【0036】

文書にはタイトルが 1 つしかないので、TAUC スコアは、以下のようにタイトルに関して計算されてもよい。

$$TAUC(Title) = ED(Title)$$

ここで、 $TAUC(Title)$ は、変換関数の適用後にニューラルネットワークへの入力として後に使用され、 $ED(Title)$ はタイトルの編集距離である。

【0037】

文書のアンカーテキスト、ならびに URL およびクリック (ここでクリックはその文書がクリックされた以前実行されたクエリである) に複数のインスタンスがあってもよい。これは、その文書が、類似するクエリについてより関連が高いということである。インデックス時間において、最高頻度を有する N 個のアンカーテキストが選択される。次いで、選択されたアンカーごとに ED スコアが計算される。最後に、アンカーに対して TAUC スコアが、以下のように決定される。

40

$$TAUC(Anchor) = \text{Min} \{ ED(Anchor_i) \} \quad i: \text{上位 } N \text{ 個のアンカー}$$

アンカーのうちの 1 つとの良好な一致が存在する場合、それで十分であることが直感的に分かる。 $TAUC(Anchor)$ は、変換関数を適用した後にニューラルネットワークの入力として使用される。

50

【0038】

特殊処理は、URL文字列のEDを計算する前に使用される。インデックス時間において、URL文字列は、分離記号として文字のセットを使用して部分に分割される。次いで、用語は、タイトルおよびアンカー用語のディクショナリから各部分に見出される。ディクショナリからの用語の出現はそれぞれ、URL文字列の最初から文字で位置を測定されてインデックスに格納される。

【0039】

クエリ時間において、クエリ用語のすべての出現は、インデックス時間に格納されたインデックスから読み込まれ、中断は「非クエリ」用語で充填される。この処理の後に、EDは計算される。ED処理の結果は、変換関数の適用後のニューラルネットワークの入力である。

10

【0040】

処理されるもう1つの特性は、ユーザが所与の文書コンテンツに対して入力する「クリック」の数である。ユーザが文書をクリックするごとに、ストリームがデータベースに入力されて、文書に関連付けられる。このプロセスはまた、データの短いストリームのような、文書情報テキスト内のストリームデータに適用されてもよい。

【0041】

インデックス時間URL処理アルゴリズムは、URL全体を、分離記号として文字のセットを使用して部分に分割する。分割関数はまた、`urlpart.startpos`をソース文字列内の部分の位置に設定する。分割関数は、URLの重要ではない部分のフィルタリングを行う。

20

【0042】

たとえば、「`http://www.companymeeting.com/index.html`」は「`companymeeting/index`」にフィルタリングされ、「`companymeeting`」および「`index`」に分割される。

【0043】

【表2】

Startpos: 0

30

```

Urlparts = split(url, dictionary)

// find terms in different url parts.
For each (term in dictionary)
{
    Int pos = 0;
    For each(urlpart in urlparts)
    {
        pos = urlpart.Find(term, pos);
        while (pos >= 0)
        {
            // parts_separator is used to distinguish different parts at
            query time
            storeOccurrence(term, pos +
            urlpart.startpos*parts_separator);
            pos = urlpart.Find(term, pos + term.length);
        }
    }
    setIndexStreamLength(parts_separator * urlparts.Count);
}

```

40

【0044】

ディクショナリが「`company meeting comp`」を含むと仮定すると、`Company: 0`、`Meeting: 7`、および`Comp: 0`というキーが生成され

50

る。文字列の合計長さは、`parts_separator*2`である。

【0045】

ED前のクエリ時間処理に関し、クエリ時間においてクエリ用語の出現が読み込まれ、ソースURL文字列内の出現の順序でクエリ用語の文字列が構築され、用語の間のスペースが「非クエリ」単語マークで充填される。たとえば、「`company policy`」のクエリ文字列、および結果として生じる「`company`」「非クエリ用語」「非クエリ用語」の文字列を考察する。

【0046】

`parts_separator`、クエリ用語位置、およびストリーム長さは、元のURL文字列に部分が何個あるか、およびどの部分が所与のクエリを含むかを知るために決定される。用語のない各部分は、「非クエリ用語」を含むと見なされる。部分がクエリ用語から始まらない場合、「非クエリ用語」がその用語の前に挿入される。クエリ用語の間のすべてのスペースは、「非クエリ用語」で充填される。

10

【0047】

図6は、クエリ文字列とデータ文字列の間の編集距離を決定するために文書情報104に採用されうるデータの型を示す。文書情報104は、たとえば、処理コンポーネント102による処理およびデータ（またはターゲット）文字列116の生成のために、タイトルテキスト604、アンカーテキスト606、URL608テキストまたは文字、およびクリック情報610のような、TAUCデータ602を含むことができる。文書情報104はまた、ユーザが文書コンテンツをクリックする回数、ユーザが（クリックを介して）選択するコンテンツのタイプ、コンテンツ、一般的な文書のクリック回数などに関連するクリック情報610も含むことができる。

20

【0048】

図7は、インデックス時間処理データフロー700を示す。最上部において、文書情報は、タイトル604、文書アンカー606、クリック情報610などの形態で、文書分析および抽出に基づいて受信される。タイトル604は、用語分割アルゴリズム704を通じて処理され、次いでディクショナリ706に進む。ディクショナリ706は、タイトル604、アンカー606、クリック情報610などに見出されたさまざまな用語の一次記憶域である。ディクショナリ706は、URL分割アルゴリズム708を介してURL608を分割するために使用される。URL分割アルゴリズム708の出力は、関連度および順位付け処理のためにインデックス付けプロセス710に送信される。文書アンカー606はまた、上位N個のアンカーのフィルター712を通じて処理されてもよい。クリック情報610は、インデックス付けプロセス710を介して直接処理されてもよい。その他の文書情報は、（たとえば、用語分割、フィルタリングなど）それに応じて処理されてもよい。

30

【0049】

図8は、結果を順位付けするための図7のインデックスプロセス710からニューラルネットワークへの入力を示すブロック図800である。インデックス付けプロセス710は、クエリ文字列110に対するURL編集距離（ED）802、クエリ文字列110に対する上位N個のアンカーED804、クエリ文字列110に対するタイトルED806、クエリ文字列110に対するクリックED808、および編集距離に関連しないその他の特徴810を計算するために使用されてもよいが、これらの一部または全部（URL ED802、上位N個のアンカーED804、タイトルED806、クリックED808、およびその他の特徴810）は、最終的に関連文書の関連度スコア、次いで他の文書の検索結果の中の順位付けを見出すために、ニューラルネットワーク502への入力として採用されてもよい。ニューラルネットワーク502は、文書の関連度を識別することに寄与する生の入力特徴として少なくともTAUC特徴を受信する2層モデルであってもよい。ニューラルネットワークは、それらの特徴が、検索エンジンによりソートするために使用できる単一の数字にどのように結合されるかを決定する。

40

【0050】

50

ニューラルネットワーク502は、関連度および順位付け処理に採用されうる数学モデルまたは計算モデルの一例に過ぎないことを理解されたい。単純ベイズ、ベイジアンネットワーク、決定ツリー、ファジー論理モデル、および独立のさまざまなパターンを表すその他の統計分類モデルのような、統計的回帰のその他の形態を採用することもできるが、ここで分類は順位および／または優先度を割り当てるために使用される方法を含む。

【0051】

図9は、検索結果の生成を計算するためのニューラルネットワーク502、編集距離入力、および生の特徴入力の例示的なシステム900の実施態様を示す。ニューラルネットワーク502の入力にある生の順位付け特徴のセット810は、BM25関数902（たとえば、BM25F）、クリック距離904、URL深さ906、ファイルタイプ908、および言語一致910を含むことができる。BM25コンポーネントは、たとえば、本文、タイトル、著者、アンカーテキスト、URL、表示名、および抽出されたタイトルを含むことができる。

10

【0052】

図10は、関連度を決定する方法を示す。1000において、クエリ文字列は、検索プロセスの一部として受信される。1002において、文書情報は、検索プロセス中に返された文書から抽出される。1004において、データ文字列は、文書情報から生成される。1006において、編集距離は、データ文字列とクエリ文字列の間で計算される。1008において、関連度スコアは、編集距離に基づいて計算される。

20

【0053】

方法のその他の態様は、編集距離を計算するステップの一部として用語挿入を採用するステップ、およびデータ文字列を生成するためにクエリ文字列に用語を挿入する挿入コストを計算するステップを含むことができ、コストは重み付けパラメータとして表される。方法はさらに、編集距離を計算するステップの一部として用語削除を採用するステップ、およびデータ文字列を生成するためにクエリ文字列で用語を削除する削除コストを計算するステップを含むことができ、コストは重み付けパラメータとして表される。位置コストは、編集距離を計算するステップの一部として計算されてもよく、位置コストはデータ文字列内の用語位置の用語挿入および／または用語削除に関連する。加えて、照合プロセスは、編集距離を計算する全体コストを計算するため、データ文字列の文字とクエリ文字列の文字の間で実行される。

30

【0054】

データ文字列のURLの複合語を分割するステップは、インデックス時間に生じてもよい。方法はさらに、文書内の出現度数に基づいてアンカーテキストの上位順位付けされたセットを見出すためにデータ文字列のアンカーテキストをフィルタリングするステップと、セット内のアンカーテキストの編集距離を計算するステップとを備えることができる。編集距離を計算するステップから導かれた編集距離スコアは、変換関数の適用後に2層ニューラルネットワークに入力されてもよく、そのスコアはタイトル情報、アンカー情報、クリック情報、またはURL情報のうちの少なくとも1つに関連付けられている編集距離を計算するステップに基づいて生成される。

40

【0055】

図11は、文書の関連度を計算する方法を示す。1100において、クエリ文字列は、検索プロセスの一部として処理されて、結果の文書のセットを返す。1102において、データ文字列は、結果セットの文書から抽出された文書情報に基づいて生成され、文書情報は、タイトル情報、アンカーテキスト情報、クリック情報、および文書からのURL情報のうちの1つまたは複数を含む。1104において、編集距離は、用語挿入、用語削除、および用語位置に基づいて、データ文字列とクエリ文字列の間で計算される。1106において、関連度スコアは編集距離に基づいて計算され、関連度スコアは結果セットの文書を順位付けするために使用される。

【0056】

方法はさらに、用語挿入、用語削除、および用語位置のそれぞれに関連付けられている

50

コストを計算するステップと、関連度スコアの計算にコストを組み込むステップと、インデックス時間においてURL情報の複合語を分割するステップと、インデックス時間においてアンカーテキスト情報をフィルタリングして文書内のアンカーテキストの出現度数に基づきアンカーテキストの上位順位付けセットを見出すステップとを備えることができる。クエリ文字列の用語の出現の読み取りは、ソースURL文字列内の出現の順序でクエリ用語の文字列を構築して、用語の間のスペースを単語マークで充填するように実行されてもよい。

【0057】

本明細書において使用されるように、用語「コンポーネント」および「システム」は、ハードウェア、ハードウェアとソフトウェアの組み合わせ、ソフトウェア、または実行中のソフトウェアのいずれかの、コンピュータ関連エンティティを示すことを意図している。たとえば、コンポーネントは、プロセッサ上で稼働するプロセス、プロセッサ、ハードディスクドライブ、複数の（光および／または磁気記憶媒体の）ストレージドライブ、オブジェクト、実行可能ファイル、実行のスレッド、プログラム、および／またはコンピュータであってもよいが、これらに限定されることはない。例として、サーバ上で稼働するアプリケーションおよびサーバはいずれもコンポーネントであってもよい。1つまたは複数のコンポーネントは、プロセスおよび／または実行のスレッド内に常駐することができ、コンポーネントは1つのコンピュータでローカライズおよび／または2つ以上のコンピュータ間で分散されてもよい。

【0058】

これ以降図12を参照すると、開示されるアーキテクチャによるTAUC特徴を使用して検索結果を順位付けするために編集距離処理を実行するように動作可能なコンピューティングシステム1200のブロック図が示される。図12および以下の説明は、そのさまざまな態様の追加のコンテキストを提供するため、さまざまな態様が実施されうる適切なコンピューティングシステム1200の簡潔で一般的な説明をもたらすことを意図している。上記の説明は、1つまたは複数のコンピュータ上で実行できるコンピュータ実行可能命令の一般的なコンテキストにおけるものであるが、当業者であれば、新規の実施形態はまた、その他のプログラムモジュールとの組み合わせおよび／またはハードウェアとソフトウェアの組み合わせとして実施されてもよいことを理解するであろう。

【0059】

一般に、プログラムモジュールは、特定のタスクを実行するか、または特定の抽象データ型を実施するルーチン、プログラム、プログラム、コンポーネント、データ構造などを含む。さらに、本発明の方法が、それぞれが1つまたは複数の関連するデバイスに動作可能に結合されうる、シングルプロセッサまたはマルチプロセッサコンピュータシステム、ミニコンピュータ、メインフレームコンピュータ、およびパーソナルコンピュータ、ハンドヘルドコンピューティングデバイス、マイクロプロセッサベースまたはプログラマブル消費者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施されてもよいことを当業者は理解するであろう。

【0060】

示されている態様はまた、特定のタスクが通信ネットワークを通じてリンクされるリモート処理デバイスによって実行される分散コンピューティング環境において実施されてもよい。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は、ローカルおよびリモートの記憶装置に配置されてもよい。

【0061】

コンピュータは通常、各種のコンピュータ可読媒体を含む。コンピュータ可読媒体は、コンピュータによってアクセスされてもよく、揮発性および不揮発性媒体、取り外し可能および固定式の媒体を含む任意の使用可能な媒体であってもよい。一例として、コンピュータ可読媒体は、コンピュータ記憶媒体および通信媒体を備えることができるが、これらに限定されることはない。コンピュータ記憶媒体は、コンピュータ可読命令、データ構造、プログラムモジュールおよびその他のデータなど、情報を格納するための任意の方法ま

10

20

30

40

50

たは技術において実施される揮発性および不揮発性の、取り外し可能および固定式の媒体を含む。コンピュータ記憶媒体は、RAM、ROM、EEPROM、フラッシュメモリまたはその他のメモリ技術、CD-ROM、DVD (digital video disk) またはその他の光ディスクストレージ、磁気カセット、磁気テープ、磁気ディスクストレージまたはその他の磁気記憶装置、もしくは望ましい情報を格納するために使用できる、コンピュータによりアクセスできる任意の他の媒体を含むが、これらに限定されることはない。

【0062】

図12を再度参照すると、さまざまな態様を実施するための例示のコンピューティングシステム1200は、処理ユニット1204、システムメモリ1206、およびシステムバス1208を有するコンピュータ1206を含む。システムバス1208は、システムメモリ1206を含む（これに限定されることはない）システムコンポーネントの処理ユニット1204へのインターフェイスを提供する。処理ユニット1204は、さまざまな市販のプロセッサのいずれであってもよい。デュアルマイクロプロセッサおよびその他のマルチプロセッサアーキテクチャもまた、処理ユニット1204として採用されてもよい。

10

【0063】

システムバス1208は、メモリバス（メモリコントローラの有無にかかわらず）、周辺バス、およびさまざまな市販のバスアーキテクチャのいずれかを使用するローカルバスにさらに相互接続することができるいくつかのタイプのバス構造のいずれかであってもよい。システムメモリ1206は、不揮発性メモリ（NON-VOL）1210および/または揮発性メモリ1212（たとえば、ランダムアクセスメモリ（RAM））を含むことができる。基本入出力システム（BIOS; basic input/output system）は、不揮発性メモリ1210（たとえば、ROM、EPROM、EEPROMなど）に格納されてもよいが、BIOSは起動時などにコンピュータ1202内の要素間の情報の転送を補助する基本ルーチンである。揮発性メモリ1212はまた、データをキャッシュに入れるための静的RAMのような、高速RAMを含むことができる。

20

【0064】

コンピュータ1202はさらに、適切なシャーシでの外部使用のために構成されてもよい内部ハードディスクドライブ（HDD）1214（たとえば、EIDE、SATA）、（たとえば、取り外し可能ディスク1218との間の読み取りまたは書き込みを行う）磁気フロッピー（登録商標）ディスクドライブ（FDD）1216、および（たとえば、CD-ROMディスク1222を読み取るか、またはDVDのようなその他の大容量光媒体との間の読み取りまたは書き込みを行う）光ディスクドライブ1220を含む。HDD1214、FDD1216、および光ディスクドライブ1220は、それぞれHDDインターフェイス1224、FDDインターフェイス1226、および光ドライブインターフェイス1228によってシステムバス1208に接続されてもよい。外部ドライブ実施のためのHDDインターフェイス1224は、USB (Universal Serial Bus) およびIEEE1394インターフェイス技術のうちの少なくとも1つまたは両方を含むことができる。

30

【0065】

ドライブおよび関連するコンピュータ可読媒体は、データの揮発性ストレージ、データ構造、コンピュータ実行可能命令などを提供する。コンピュータ1202の場合、ドライブおよび媒体は、適切なデジタル形式の任意のデータの格納に適応する。上記のコンピュータ可読媒体の説明は、HDD、取り外し可能磁気ディスク（たとえば、FDD）、およびCDまたはDVDのような取り外し可能光媒体を参照するが、zipドライブ、磁気カセット、フラッシュメモリカード、カートリッジなどのような、コンピュータによって読み取ることができるその他の種類の媒体も例示の動作環境において使用されてもよく、さらに、任意のそのような媒体は、開示されるアーキテクチャの新規の方法を実行するためのコンピュータ実行可能命令を含むことができることが、当業者には理解されるべ

40

50

きである。

【0066】

オペレーティングシステム1230、1つまたは複数のアプリケーションプログラム1232、その他のプログラムモジュール1234、およびプログラムデータ1236を含む多数のプログラムモジュールは、ドライブおよび揮発性メモリ1212に格納されてもよい。1つまたは複数のアプリケーションプログラム1232、その他のプログラムモジュール1234、およびプログラムデータ1236は、システム100および関連するブロック、システム500および関連するブロック、文書情報104、T A U Cデータ602、クリック情報610、データフロー700（ならびにアルゴリズム）、およびブロック図800（ならびに関連するブロック）を含むことができる。

10

【0067】

オペレーティングシステム、アプリケーション、モジュール、および／またはデータの全部または一部は、揮発性メモリ1212にキャッシュされてもよい。開示されるアーキテクチャは、さまざまな市販のオペレーティングシステムまたはオペレーティングシステムの組み合わせにより実施されうることを理解されたい。

【0068】

ユーザは、たとえばキーボード1238、およびマウス1240のようなポインティングデバイスなどの1つまたは複数の有線／無線入力デバイスを通じてコンピュータ1202にコマンドおよび情報を入力することができる。その他の入力デバイス（図示せず）は、マイクロフォン、I Rリモートコントロール、ジョイスティック、ゲームパッド、スタイラスペン、タッチスクリーンなどを含むことができる。上記およびその他の入力デバイスは、システムバス1208に結合される入力デバイスインターフェイス1242を通じて処理ユニット1204に接続されることが多いが、パラレルポート、I E E E 1394シリアルポート、ゲームポート、U S Bポート、I Rインターフェイスなどの他のインターフェイスにより接続することもできる。

20

【0069】

モニタ1244またはその他の種類の表示デバイスもまた、ビデオアダプタ1246などのインターフェイスを介してシステムバス1208に接続される。モニタ1244に加えて、コンピュータは通常、スピーカー、プリンタなどの、他の周辺出力デバイス（図示せず）を含む。

30

【0070】

コンピュータ1202は、リモートコンピュータ1248のような、1つまたは複数のリモートコンピュータへの有線および／または無線通信を介する論理接続を使用してネットワーク環境において動作することができる。リモートコンピュータ1248は、ワークステーション、サーバコンピュータ、ルータ、パーソナルコンピュータ、ポータブルコンピュータ、マイクロプロセッサベースのエンターテイメント電気製品、ピアデバイスまたはその他の共通ネットワークノードなどであってもよく、通常はコンピュータ1202に関連して説明される要素の多くまたはすべてを含むが、簡略にするため、メモリ／ストレージデバイス1250のみが示される。示される論理接続は、ローカルエリアネットワーク（L A N）1252および／または、たとえばワイドエリアネットワーク（W A N）1254などのより大規模なネットワークへの有線／無線接続を含む。そのようなL A NおよびW A Nネットワーク環境は、オフィスおよび企業において一般化しており、イントラネットのような企業規模のコンピュータネットワークを容易にするが、それらはすべて、たとえばインターネットなどのグローバル通信ネットワークに接続することができる。

40

【0071】

L A Nネットワーク環境に使用される場合、コンピュータ1202は有線および／または無線通信ネットワークインターフェイスまたはアダプタ1256を介してL A N1252に接続される。アダプタ1256は、L A N1252への有線および／または無線通信を容易にすることができるが、L A N1252はアダプタ1256の無線機能と通信するために配置された無線アクセスポイントも含むことができる。

50

【0072】

WANネットワーク環境に使用される場合、コンピュータ1202は、モデム1258を含むことができるか、またはWAN1254の通信サーバに接続されるか、またはインターネットなどによってWAN1254にわたる通信を確立するための他の手段を有する。モデム1258は、内蔵または外付けであっても、有線および／または無線デバイスであってもよく、入力デバイスインターフェイス1242を介してシステムバス1208に接続される。ネットワーク化された環境において、コンピュータ1202に関連して示されるプログラムモジュール、またはその部分は、リモートメモリ／ストレージデバイス1250に格納されてもよい。示されるネットワーク接続は例示的なものであり、コンピュータ間の通信リンクを確立する他の手段が使用されうることを理解されたい。

10

【0073】

コンピュータ1202は、IEEE802ファミリーの標準を使用して、たとえばプリンタ、スキャナ、デスクトップおよび／またはポータブルコンピュータ、携帯情報端末(PDA)、通信衛星、無線により検出可能なタグに関連付けられた機器または位置(たとえば、kiosk、新聞売店、トイレ)、および電話などとの無線通信(たとえば、IEEE802.11無線変調技法)で動作するように配置された無線デバイスのような、有線および無線デバイスまたはエンティティと通信するように動作可能である。これは、少なくともWi-Fi(つまり、Wireless Fidelity)、WiMax、およびブルートゥース(商標)無線技術を含む。したがって、通信は、従来のネットワークと同様に定義済みの構造であってもよいが、または単に少なくとも2つのデバイス間の臨時の通信であってもよい。Wi-Fiネットワークは、IEEE802.11x(a、b、gなど)と呼ばれる無線技術を使用して、安全で信頼性の高い高速無線通信をもたらす。Wi-Fiネットワークは、コンピュータを、相互、インターネット、および有線ネットワーク(IEEE802.3関連の媒体および機能)に接続するために使用されてもよい。

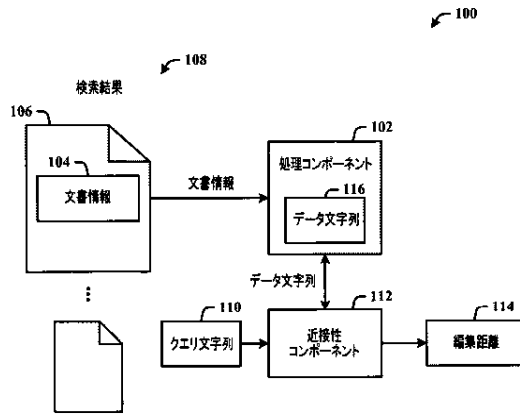
20

【0074】

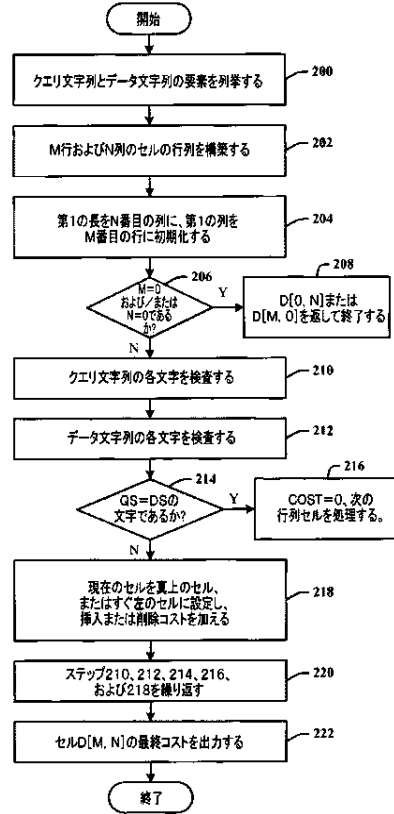
上記で説明された内容は、開示されるアーキテクチャの実施例を含む。すべての考えられるコンポーネントおよび／または方法論の組み合わせについて説明することはもちろん不可能であるが、さらに多くの組み合わせおよび置換が可能であることを当業者であれば理解するであろう。したがって、新規のアーキテクチャは、添付の特許請求の範囲の精神および範囲内に含まれるそのようなすべての変更、修正および変形を網羅することが意図される。さらに、「含む(include)」という用語が詳細な説明または特許請求の範囲で使用される限りにおいて、そのような用語は、「備える(comprising)」が特許請求の範囲において暫定的な用語として採用される場合に解釈される際の「備える」と同様の方法で包括的であることが意図される。

30

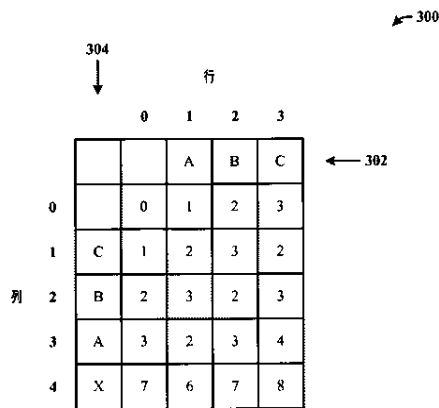
【図1】



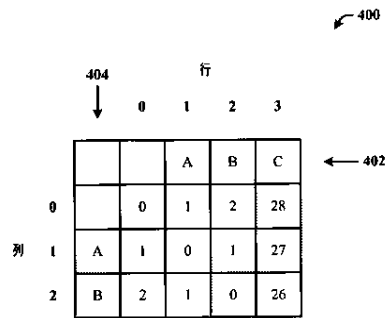
【図2】



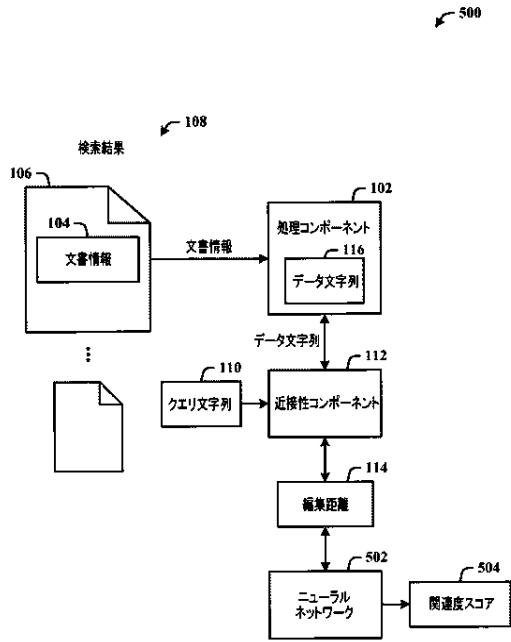
【図3】



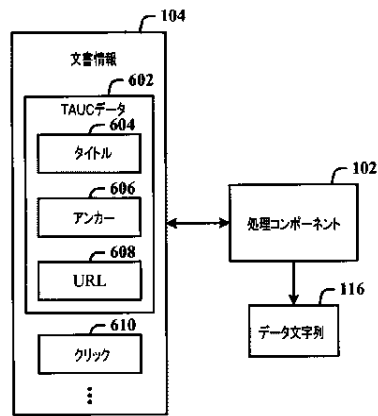
【図4】



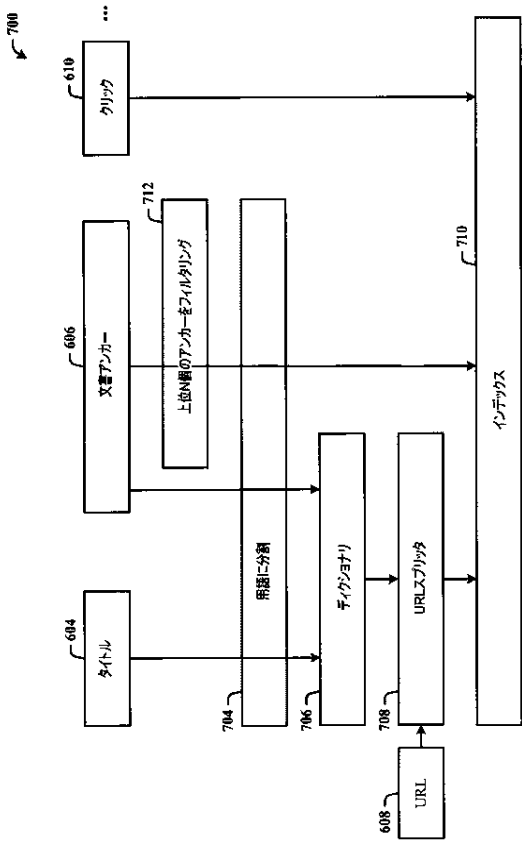
【図5】



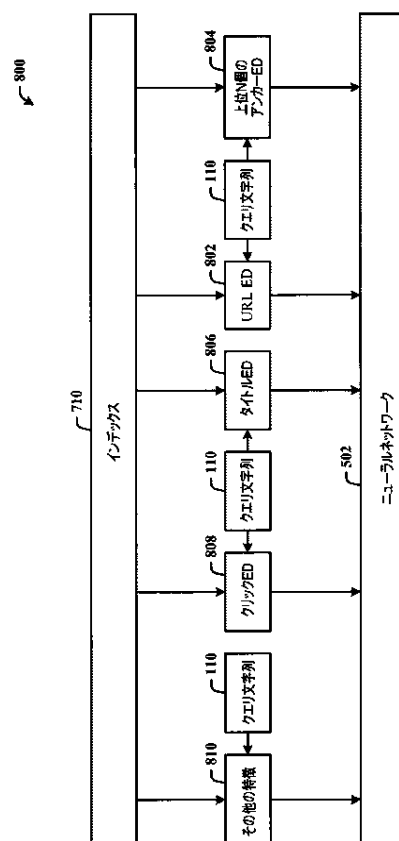
【図6】



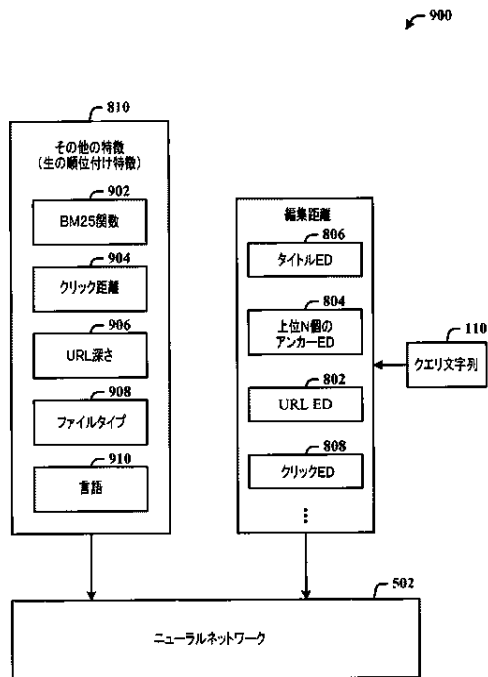
【図7】



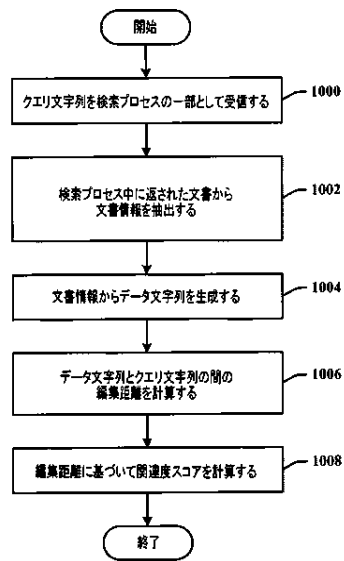
【図8】



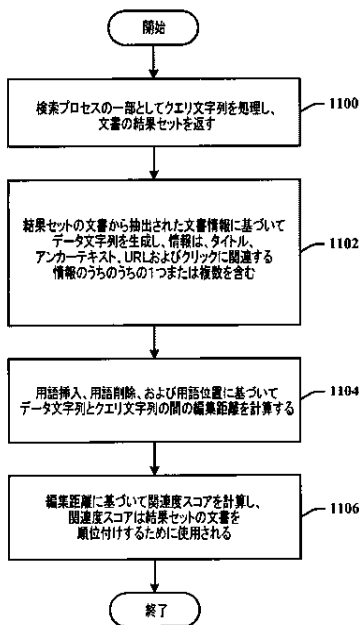
【図9】



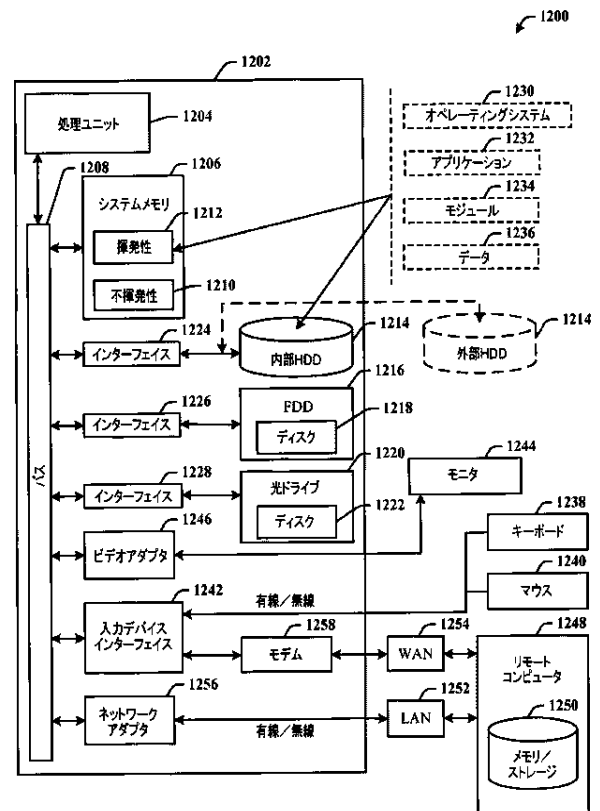
【図10】





【図11】



【図12】



【國際調查報告】

INTERNATIONAL SEARCH REPORT		International application No. PCT/US2009/036597
A. CLASSIFICATION OF SUBJECT MATTER		
<i>G06F 17/30(2006.01)</i> ; <i>G06F 17/27(2006.01)</i> ; <i>G06F 17/21(2006.01)</i>		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) IPC G06F		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Korean Utility models and applications for Utility models since 1975 Japanese Utility models and applications for Utility models since 1975		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) eKOMPASS (The internal search database of KIPO)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2004/0141354 A1 (CARNAHAN, J. M.) 22 JULY 2004 See abstract, figures 1-2, paragraphs 0002-0003, claims 1-4, 10-16, 20.	1, 2, 8
A	KR 2003-0080826 A (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 17 OCTOBER 2003 See abstract, figures 2-4, 6-7, page 3, claims 1-2.	1, 8, 17
A	US 2008/0005068 A1 (DUMAIS, S. T. et al.) 3 JANUARY 2008 See abstract, figures 1-4, paragraphs 0010-0012, 0039, claims 1-8.	1, 8, 17
A	US 2006/0149723 A1 (FINGER II, J. C.) 6 JULY 2006 See abstract, figures 2-5, 9, paragraphs 0010-0011, 0042-0045, 0047, claims 1-7	1, 8, 17
A	US 6738764 B2 (MAO, J. et al.) 18 MAY 2004 See abstract, figure 2, column 1, line.63-column 2, line.22, claims 1-6	1, 8, 17
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed		"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family
Date of the actual completion of the international search 28 AUGUST 2009 (28.08.2009)		Date of mailing of the international search report 28 AUGUST 2009 (28.08.2009)
Name and mailing address of the ISA/KR  Korean Intellectual Property Office Government Complex-Daejeon, 139 Seonsa-ro, Seo-gu, Daejeon 302-701, Republic of Korea Facsimile No. 82-42-472-7140		Authorized officer Tae-Geun KIM Telephone No. 82-42-481-8120 

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No. PCT/US2009/036597

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2004-0141354 A1	22.07.2004	WO 2004-066090 A2 WO 2004-066090 A3	05.08.2004 05.08.2004
KR 2003-0080826 A	17.10.2003	KR 10-0490748 B1 US 2003-0195882 A1	24.05.2005 16.10.2003
US 2008-0005068 A1	03.01.2008	None	
US 2006-0149723 A1	06.07.2006	US 07505961 B2 US 07039631 B1	17.03.2009 02.05.2006
US 6738764 B2	18.05.2004	None	

フロントページの続き

(81)指定国 AP(BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), EA(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), EP(AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TC), AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BC, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW

(72)発明者 ハン イ

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 エルシーエーインターナショナル パテント内

(72)発明者 ドミトリー マイヤーゾン

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 エルシーエーインターナショナル パテント内

(72)発明者 ジュン シュウ

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 エルシーエーインターナショナル パテント内

Fターム(参考) 5B075 ND36 PQ02 PQ46 PR06

5B084 AA26 AB01 BB01 CA12 CC05 CC19 CD22 CF12 DB02 DC02