

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2005-339545

(P2005-339545A)

(43) 公開日 平成17年12月8日(2005.12.8)

(51) Int. Cl.⁷

G06F 17/30

F I

G06F 17/30

220Z

G06F 17/30

180Z

テーマコード(参考)

5B075

審査請求 未請求 請求項の数 40 O L 外国語出願 (全 33 頁)

(21) 出願番号 特願2005-147965 (P2005-147965)
 (22) 出願日 平成17年5月20日(2005.5.20)
 (31) 優先権主張番号 10/850,623
 (32) 優先日 平成16年5月21日(2004.5.21)
 (33) 優先権主張国 米国(US)

(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 9805
 2-6399 レッドモンド ワン マイ
 クロソフト ウェイ
 (74) 代理人 100077481
 弁理士 谷 義一
 (74) 代理人 100088915
 弁理士 阿部 和夫
 (72) 発明者 バマ ラマラスナム
 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
 ョン内

最終頁に続く

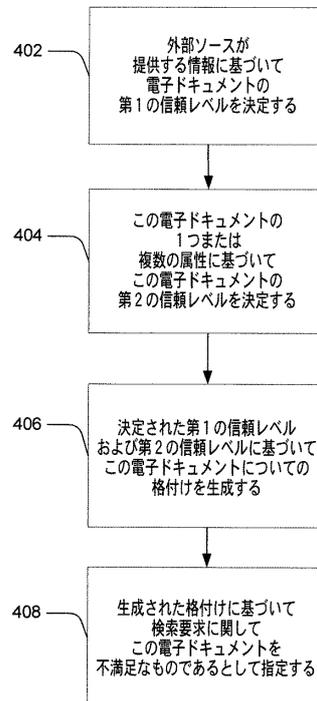
(54) 【発明の名称】 外部データを使用した検索エンジンスパムの検出

(57) 【要約】

【課題】 検索に関連して電子ドキュメントを評価する。

【解決手段】 外部ソースは、検索エンジンが検索する電子ドキュメントを評価するために使用するためのデータを提供する。この電子ドキュメントの第1の信頼レベルは、この外部提供データに基づいて決定される。第1の信頼レベルは、この電子ドキュメントが望ましくないものである可能性を示す。この電子ドキュメントの第2の信頼レベルは、この電子ドキュメントの属性に基づいて決定される。第2の信頼レベルは、この電子ドキュメントが検索に関連して不満足なものであるという可能性を示す。決定された第1の信頼レベル、および、決定された第2の信頼レベルの関数として生成された電子ドキュメントについての格付けを使用し、受信された検索要求に関してこの電子ドキュメントを不満足なものであるとして分類する。

【選択図】 図4



【特許請求の範囲】**【請求項 1】**

検索に関連して電子ドキュメントを評価する方法であって、

ユーザからの検索要求に回答して検索エンジンによって検索可能な電子ドキュメントの第 1 の信頼レベルを決定するステップであって、前記第 1 の信頼レベルは、前記検索エンジンの外部のソースが提供する情報に基づいて前記電子ドキュメントが望ましくないものである可能性を示していることと、

前記電子ドキュメントの第 2 の信頼レベルを決定するステップであって、前記第 2 の信頼レベルは、前記電子ドキュメントの 1 つまたは複数の属性に基づいて前記検索要求に関して前記電子ドキュメントが不満足なものである可能性を示していることと、

10

決定された前記第 1 の信頼レベルおよび決定された前記第 2 の信頼レベルの関数として、前記電子ドキュメントについての格付けを生成するステップと、

前記電子ドキュメントの前記生成された格付けに基づいて、前記検索要求に関して前記電子ドキュメントを不満足なものであるとして指定するステップと、

を備えることを特徴とする方法。

【請求項 2】

前記外部ソースは、電子メールスパム検出システムを備えることを特徴とする請求項 1 に記載の方法。

【請求項 3】

前記電子ドキュメントは、1 つまたは複数のウェブページおよびマルチメディアファイルを含むことを特徴とする請求項 1 に記載の方法。

20

【請求項 4】

前記第 1 の信頼レベルを決定する前記ステップは、

前記外部ソースから 1 つまたは複数のホスト名を受信するステップであって、前記受信されたホスト名が提供する情報は、望ましくないものであるという所定の可能性を有していることと、

前記受信されたホスト名のうちの 1 つによって提供されているものとして、前記電子ドキュメントを識別するステップと、

前記受信されたホスト名のうちの 1 つによって提供されるものとして前記電子ドキュメントを識別する前記ステップに回答して、前記電子ドキュメントについての、前記所定の可能性に基づいた前記第 1 の信頼レベルを指定するステップと、

30

を含むことを特徴とする請求項 1 に記載の方法。

【請求項 5】

前記電子ドキュメントからリンクされた 1 つまたは複数の他の電子ドキュメントについての前記第 1 の信頼レベルを指定するステップをさらに含むことを特徴とする請求項 4 に記載の方法。

【請求項 6】

前記第 1 の信頼レベルを決定する前記ステップは、

前記外部ソースから 1 つまたは複数のネットワークアドレスを受信するステップであって、前記外部ソースは、前記受信されたネットワークアドレスのうちの 1 つに配置された 1 つまたは複数の電子ドキュメントを、望ましくないものであるという所定の可能性を有するものとして識別することと、

40

前記受信されたネットワークアドレスのうちの 1 つに配置されているものとして、前記電子ドキュメントを識別するステップと、

前記受信されたネットワークアドレスのうちの 1 つに配置されているものとして前記電子ドキュメントを識別する前記ステップに回答して、前記電子ドキュメントについての前記所定の可能性に基づいた前記第 1 の信頼レベルを指定するステップと、

を含むことを特徴とする請求項 1 に記載の方法。

【請求項 7】

前記第 1 の信頼レベルを決定する前記ステップは、

50

前記外部ソースからの、前記電子ドキュメントを望ましくないものとして指定するユーザ提供情報を受信するステップと、

前記受信されたユーザ提供情報に応答して、前記電子ドキュメントの望ましくないパターンを特徴づける前記電子ドキュメントの1つまたは複数の属性を識別するステップと、

前記電子ドキュメントの前記識別された属性に基づいて、前記電子ドキュメントについての前記第1の信頼レベルを指定するステップと、

を含むことを特徴とする請求項1に記載の方法。

【請求項8】

前記第1の信頼レベルを決定する前記ステップは、

前記外部ソースからの1つまたは複数の用語を受信するステップであって、前記受信された用語のうち少なくとも1つが出現する1つまたは複数の電子ドキュメントは望ましくないものであるという所定の可能性を有することを、前記外部ソースが決定するステップと、

前記電子ドキュメント中の前記受信された用語のうち少なくとも1つの出現を検出するステップと、

前記電子ドキュメント中の前記受信された用語のうち少なくとも1つの出現を検出する前記ステップに応答して、前記電子ドキュメントについての、前記所定の可能性に基づいた前記第1の信頼レベルを指定するステップと、

を含むことを特徴とする請求項1に記載の方法。

【請求項9】

前記第2の信頼レベルを決定するステップは、

前記電子ドキュメントの前記属性を識別するために、前記電子ドキュメントを解析するステップであって、検索に関連して、前記属性は前記電子ドキュメントの望ましくないパターンを特徴づけていることと、

前記電子ドキュメントの前記識別された属性に基づいて、前記第2の信頼レベルを決定するステップと、

を含むことを特徴とする請求項1に記載の方法。

【請求項10】

前記第2の信頼レベルを決定する前記ステップは、

前記電子ドキュメントに関してユーザ提供情報を受信するステップであって、前記電子ドキュメントを検索結果中において、前記受信されたユーザ提供情報は、前記電子ドキュメントを望ましくないものとして指定することと、

前記受信されたユーザ提供情報に応答して、検索に関して前記電子ドキュメントの望ましくないパターンを特徴づける前記電子ドキュメントの1つまたは複数の属性を識別するステップと、

前記電子ドキュメントの前記識別された属性に基づいて、前記第2の信頼レベルを決定するステップと、

を含むことを特徴とする請求項1に記載の方法。

【請求項11】

前記受信された検索要求に応答して、検索結果を前記ユーザに提供するステップと、

以下の、前記提供された検索結果中において不満足なものであるとして指定された前記電子ドキュメントを格下げするステップ、前記提供された検索結果から不満足なものであるとして指定された前記電子ドキュメントを除外するステップ、および、前記電子ドキュメントのランキングが前記提供された検索結果中において所定のランクを超えるとときに、前記提供された検索結果中の前記電子ドキュメントのランキングを保持するステップ、

のうちの1つまたは複数のステップを実施するステップと、

をさらに備えることを特徴とする請求項1に記載の方法。

【請求項12】

1つまたは複数のコンピュータ読取り可能媒体が、請求項1に記載の方法を実施するためのコンピュータ実行可能命令を有することを特徴とする請求項1に記載の方法。

10

20

30

40

50

【請求項 13】

検索に関連して電子ドキュメントを評価する方法であって、

ユーザからの検索要求に応答して検索エンジンによって検索可能である電子ドキュメントに関して、ユーザ提供情報を受信するステップであって、前記ユーザ提供情報は前記電子ドキュメントを望ましくないものとして特徴づけていることと、

前記受信されたユーザ提供情報の関数として、前記電子ドキュメントについての格付けを生成するステップと、

前記電子ドキュメントの前記生成された格付けに従って、前記検索要求に関して前記電子ドキュメントを不満足なものであるとして指定するステップと、

を備えることを特徴とする方法。

10

【請求項 14】

前記電子ドキュメントは、1つまたは複数のウェブページおよびマルチメディアファイルを含むことを特徴とする請求項 13 に記載の方法。

【請求項 15】

前記受信されたユーザ提供情報は、前記電子ドキュメントを望ましくない電子メールに関連付けられているものとして指定することを特徴とする請求項 13 に記載の方法。

【請求項 16】

前記受信されたユーザ提供情報は、前記電子ドキュメントを検索結果中において望ましくないものとして指定することを特徴とする請求項 13 に記載の方法。

【請求項 17】

前記電子ドキュメントについての前記格付けを生成する前記ステップは、

前記電子ドキュメントの1つまたは複数の属性を識別するために、前記電子ドキュメントを解析するステップと、

前記電子ドキュメントについての前記格付けを生成するために、確率的分類機構に前記識別された属性を適用するステップであって、前記確率的分類機構は前記識別された属性が望ましくないものであるかどうかを認識するようにトレーニングされていることと、

を含むことを特徴とする請求項 13 に記載の方法。

20

【請求項 18】

前記確率的分類機構は、ナイーブベイジアン分類機構、制限依存性ベイジアン分類機構、ベイジアンネットワーク分類機構、判定ツリー、サポートベクトルマシン、コンテンツマッチング分類機構、最大エントロピー分類機構、および、これらの組合せを含む群から選択される1つまたは複数の分類機構を含むことを特徴とする請求項 13 に記載の方法。

30

【請求項 19】

前記電子ドキュメントについての前記格付けを生成する前記ステップは、

前記受信されたユーザ提供情報の信頼性を決定するステップと、

前記決定された信頼性の関数として、前記電子ドキュメントについての前記格付けを生成するステップと、

を含むことを特徴とする請求項 13 に記載の方法。

【請求項 20】

前記受信されたユーザ提供情報の前記信頼性を決定する前記ステップは、

前記電子ドキュメントに関する他のユーザ提供情報を受信するステップと、

前記他のユーザ提供情報が前記受信されたユーザ提供情報と一致するかどうかを判定するステップと、

を含むことを特徴とする請求項 19 に記載の方法。

40

【請求項 21】

1つまたは複数のコンピュータ読取り可能媒体は、請求項 13 に記載の方法を実施するためのコンピュータ実行可能命令を有することを特徴とする請求項 13 に記載の方法。

【請求項 22】

検索に関連して電子ドキュメントを評価するシステムであって、

ユーザからの検索要求を受信し、前記受信された検索要求に基づいて電子ドキュメント

50

を識別するプロセッサと、

前記電子ドキュメントが望ましくないものであるかどうかを評価するために使用するための、外部のソースにより前記プロセッサに提供されるデータを記憶するメモリ域と、
を備え、

前記プロセッサは、前記外部ソースにより提供された前記データに基づいて前記電子ドキュメントが望ましくないものである可能性を示す、前記電子ドキュメントの第1の信頼レベルを決定するように構成されており、

前記プロセッサは、前記電子ドキュメントの1つまたは複数の属性に基づいて前記電子ドキュメントが検索に関して不満足なものである可能性を示す、前記電子ドキュメントの第2の信頼レベルを設定するようにさらに構成されており、

前記プロセッサはさらに、前記決定された第1の信頼レベルおよび前記設定された第2の信頼レベルの関数として前記電子ドキュメントについての格付けを生成し、前記電子ドキュメントの前記生成された格付けに基づいて前記電子ドキュメントを前記受信された検索要求に関して不満足なものであるとして分類するように、さらに構成されていることを特徴とするシステム。

10

【請求項23】

前記外部ソースは、電子メールスパム検出システムを備えることを特徴とする請求項22に記載のシステム。

【請求項24】

前記外部ソースが提供する前記データは、1つまたは複数のホスト名を識別し、前記ホスト名のそれぞれは、望ましくないものであるという所定の可能性を有する情報を提供し、前記プロセッサは、前記ホスト名のうちの1つにより提供されているものとして前記電子ドキュメントを識別するように構成され、前記プロセッサはさらに、前記ホスト名のうちの1つにより提供されているものとして前記電子ドキュメントを識別することに応答して、前記所定の可能性に基づいている、前記電子ドキュメントについての前記第1の信頼レベルを指定するように構成されていることを特徴とする請求項22に記載のシステム。

20

【請求項25】

前記外部ソースが提供する前記データは、1つまたは複数のネットワークアドレスを識別し、前記外部ソースは、前記ネットワークアドレスのうちの1つに配置される1つまたは複数の電子ドキュメントを望ましくない所定の可能性を有するものとして識別し、前記プロセッサは、前記電子ドキュメントを前記ネットワークアドレスのうちの1つに配置されているものとして識別するように構成され、前記プロセッサは、前記電子ドキュメントを前記ネットワークアドレスのうちの1つに配置されているものとして識別することに応答して、前記所定の可能性に基づいている前記電子ドキュメントについての前記第1の信頼レベルを指定するように構成されることを特徴とする請求項22に記載のシステム。

30

【請求項26】

前記外部ソースが提供する前記データは、1つまたは複数の用語を識別し、前記用語のうちの少なくとも1つが出現する1つまたは複数の電子ドキュメントは、望ましくないものであるという所定の可能性を有することを、前記外部ソースが決定し、前記プロセッサは、いつ前記用語のうちの少なくとも1つが前記電子ドキュメント中に出現するかを決定するように構成され、前記プロセッサは、前記用語のうちの少なくとも1つが前記電子ドキュメント中に出現することを決定することに応答して、前記所定の可能性に基づいている前記電子ドキュメントについての前記第1の信頼レベルを指定するように構成されることを特徴とする請求項22に記載のシステム。

40

【請求項27】

前記プロセッサは、前記電子ドキュメントの前記属性を識別するために、前記電子ドキュメントを解析し、前記検索に関して前記電子ドキュメントの望ましくないパターンを特徴づける前記電子ドキュメントの前記識別された属性に基づいて、前記第2の信頼レベルを設定するように構成されることを特徴とする請求項22に記載のシステム。

【請求項28】

50

前記プロセッサは、前記受信された検索要求に応答して検索結果を前記ユーザに提供し

、
前記提供された検索結果中において不満足なものであるとして分類される前記電子ドキュメントを格下げするステップ、前記提供された検索結果から不満足なものであるとして分類される前記電子ドキュメントを除外するステップ、および、前記電子ドキュメントのランキングが前記提供された検索結果中において所定のランクを超えるときに、前記提供された検索結果中の前記電子ドキュメントの前記ランキングを保持するステップ、のうちの1つまたは複数のステップを実行するように構成されることを特徴とする請求項22に記載のシステム。

【請求項29】

10

検索に関連して電子ドキュメントを評価するコンピュータ実行可能コンポーネントを有する1つまたは複数のコンピュータ読取り可能媒体であって、

ユーザからの検索要求に応答して検索可能な電子ドキュメントに関するユーザ提供情報を受信するインターフェースコンポーネントであって、前記ユーザ提供情報は前記電子ドキュメントを望ましくないものであるとして特徴付けていることと、

前記受信されたユーザ提供情報の関数として、前記電子ドキュメントについての格付けを生成する解析コンポーネントと、

前記電子ドキュメントの前記生成された格付けに従って、前記電子ドキュメントを前記検索要求に関して不満足なものであるとして分類するクエリコンポーネントと、

を備えることを特徴とするコンピュータ読取り可能媒体。

20

【請求項30】

前記受信されたユーザ提供情報は、前記電子ドキュメントが、望ましくない電子メールのソースに関連付けられていることを指定することを特徴とする請求項29に記載のコンピュータ読取り可能媒体。

【請求項31】

前記受信されたユーザ提供情報は、前記電子ドキュメントが検索結果中で望ましくないものであることを指定することを特徴とする請求項29に記載のコンピュータ読取り可能媒体。

【請求項32】

前記解析コンポーネントは、前記電子ドキュメントの1つまたは複数の属性を識別するために前記電子ドキュメントを解析するよう構成され、前記解析コンポーネントは、前記電子ドキュメントについての前記格付けを生成するために、前記識別された属性が望ましくないものであるかどうかを認識するようにトレーニングされる確率的分類機構に、前記識別された属性を適用するようにさらに構成されることを特徴とする請求項29に記載のコンピュータ読取り可能媒体。

30

【請求項33】

前記確率的分類機構は、ナイーブベイジアン分類機構、制限依存性ベイジアン分類機構、ベイジアンネットワーク分類機構、判定ツリー、サポートベクトルマシン、コンテンツマッチング分類機構、最大エントロピー分類機構、および、これらの組合せを含む群から選択される1つまたは複数の分類機構を含むことを特徴とする請求項32に記載のコンピュータ読取り可能媒体。

40

【請求項34】

前記解析コンポーネントは、前記受信されたユーザ提供情報の信頼性を決定し、前記決定された信頼性の関数として前記電子ドキュメントについての前記格付けを生成するように構成されることを特徴とする請求項29に記載のコンピュータ読取り可能媒体。

【請求項35】

前記インターフェースコンポーネントは、前記電子ドキュメントに関する他のユーザ提供情報を受信するように構成され、前記解析コンポーネントは、前記受信電子ドキュメントの前記信頼性を決定するために前記他のユーザ提供情報が前記受信されたユーザ提供情報に対応するかどうかを検査するように構成されることを特徴とする請求項34に記載の

50

コンピュータ読取り可能媒体。

【請求項 36】

検索に関連して電子ドキュメントを評価するコンピュータ実行可能コンポーネントを有する1つまたは複数のコンピュータ読取り可能媒体であって、

ユーザからの検索要求を受信し、前記受信された検索要求に基づいて電子ドキュメントを識別するクエリコンポーネントと、

前記電子ドキュメントが望ましくないものであるかどうかを評価するために使用するためのデータを提供する外部コンポーネントと、

前記外部コンポーネントにより提供される前記データに基づいて、前記電子ドキュメントが望ましくないものであるという可能性を示す前記電子ドキュメントの第1の信頼レベルを決定する内部コンポーネントであって、前記内部コンポーネントは、前記電子ドキュメントの1つまたは複数の属性に基づいて、前記電子ドキュメントが検索に関して不満足なものである可能性を示す前記電子ドキュメントの第2の信頼レベルを設定するように構成されることと、

前記決定された第1の信頼レベルおよび前記設定された第2の信頼レベルの関数として、前記電子ドキュメントについての格付けを生成する解析コンポーネントと、

を備え、

前記クエリコンポーネントは、前記電子ドキュメントの前記生成された格付けに基づいて、前記受信された検索要求に関して前記電子ドキュメントを不満足なものであるとして分類するように構成されることを特徴とするコンピュータ読取り可能媒体。

【請求項 37】

前記外部コンポーネントにより提供された前記データは、1つまたは複数のホスト名を識別し、前記ホスト名のそれぞれが、望ましくないものであるという所定の可能性を有する情報を提供し、前記内部コンポーネントは、前記ホスト名のうちの1つにより提供されているものとして前記電子ドキュメントを識別するように構成され、前記内部コンポーネントは、前記ホスト名のうちの1つにより提供されているものとして前記電子ドキュメントを識別することに応答して、前記所定の可能性に基づいている前記電子ドキュメントについての前記第1の信頼レベルを指定するようにさらに構成されることを特徴とする請求項 36 に記載のコンピュータ読取り可能媒体。

【請求項 38】

前記外部コンポーネントが提供する前記データは、1つまたは複数のネットワークアドレスを識別し、前記外部コンポーネントは、前記ネットワークアドレスのうちの1つに配置された1つまたは複数の電子ドキュメントを望ましくないものであるという所定の可能性を有するものとして識別し、前記内部コンポーネントは、前記電子ドキュメントを前記ネットワークアドレスのうちの1つに配置されているものとして識別するように構成され、前記内部コンポーネントは、前記電子ドキュメントを前記ネットワークアドレスのうちの1つに配置されているものとして識別することに応答して、前記所定の可能性に基づいている前記電子ドキュメントについての前記第1の信頼レベルを指定するように構成されることを特徴とする請求項 36 に記載のコンピュータ読取り可能媒体。

【請求項 39】

前記外部コンポーネントが提供する前記データは、1つまたは複数の用語を識別し、前記用語のうちの少なくとも1つが出現する1つまたは複数の電子ドキュメントが、望ましくないものであるという所定の可能性を有することを、前記外部コンポーネントが決定し、前記内部コンポーネントは、いつ前記用語のうちの少なくとも1つが前記電子ドキュメント中に出現するかを決定するように構成され、前記内部コンポーネントは、前記用語のうちの少なくとも1つが前記電子ドキュメント中に出現することを決定することに応答して、前記所定の可能性に基づいている前記電子ドキュメントについての前記第1の信頼レベルを指定するように構成されることを特徴とする請求項 36 に記載のコンピュータ読取り可能媒体。

【請求項 40】

10

20

30

40

50

前記クエリコンポーネントは、前記受信された検索要求に応答して検索結果を前記ユーザに提供し、

前記提供された検索結果中において不満足なものであるとして分類される前記電子ドキュメントを格下げするステップ、前記提供された検索結果から不満足なものであるとして分類される前記電子ドキュメントを除外するステップ、および、前記電子ドキュメントのランキングが前記提供された検索結果中において所定のランクを超えるときに、前記提供された検索結果中の前記電子ドキュメントの前記ランキングを保持するステップ、のうちの1つまたは複数のステップを実施するように、構成されることを特徴とする請求項36に記載のコンピュータ読取り可能媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本発明の実施形態は、通信ネットワークを使用した関連性のあるデータエンティティについての検索の分野に関する。詳細には、本発明の実施形態は、電子ドキュメントの作成者が誤解をさせることにより、検索エンジンにその電子ドキュメントに対して不当に高いランクを付与させる作為的な操作を、外部データを使用して防止することに関する。

【背景技術】

【0002】

インターネットは、多数のコンピュータ上に分散された膨大な量の情報を有している。したがって、様々な話題に関する大量の情報をユーザに提供している。これは、イントラネットやエクストラネット (e x t r a n e t) など、他のいくつかの通信ネットワークについても言えることである。大量の情報がネットワーク上で利用可能ではあるが、所望の情報を見出すことは、通常簡単でもなく、また迅速にも行うこともできない。

【0003】

ネットワーク上で所望の情報を見出すという課題に対処するために、検索エンジンが、開発されてきている。一般的に、所望の情報のタイプについてのアイデアを持つユーザは、検索エンジンに対して1つまたは複数の検索用語を入力する。次いで、ユーザが指定した検索用語に関連した電子ドキュメントを含んでいるとこの検索エンジンが判断したネットワークロケーション (例えば、URL) のリストを、この検索エンジンは返す。多数の検索エンジンはまた、関連性ランキング (r e l e v a n c e r a n k i n g) を提供する。典型的な関連性ランキングは、所与のネットワークロケーションにある電子ドキュメントが、他の電子ドキュメントに対して、相対的にこのユーザが指定した検索用語に関連している可能性についての相対的評価のことである。例えば、従来の検索エンジンは、特定の検索用語が電子ドキュメント中で出現する回数、その電子ドキュメント中における配置 (例えば、しばしば、その電子ドキュメントの末尾に出現する場合に比べて、タイトル中に出現する用語は、より重要であると見なすことができる) に基づいて関連性ランキングを提供することもある。さらに、リンク解析が、ウェブページおよび他のハイパーリンクされたドキュメントをランク付けする際における強力な技法となってきた。アンカーテキスト解析 (a n c h o r - t e x t a n a l y s i s)、ウェブページ構造解析、キー用語リストの使用、およびURLテキストが、関連性ランキングを提供するのに使用される他の技法である。

【0004】

電子ドキュメントの作成者は、しばしば、彼らの電子ドキュメントをユーザに表示する作為的な努力を介して、関連性ランキングの問題を複雑にしてしまう。例えば、一部の作成者は、彼らの電子ドキュメントについて、別の方法で保証され得るよりも高いランクの統計を生成するように、検索エンジンを誘導しようと試みる。検索エンジンからの不当に高いランクを達成しようという試みにおける、作成者による電子ドキュメントの作為的な操作は、一般に検索エンジンスパミング (s e a r c h e n g i n e s p a m m i n g) と呼ばれる。検索エンジンスパム (s e a r c h e n g i n e s p a m) の目的は、ユーザを詐欺的に誘導して操作された電子ドキュメントを訪問させることである。操

10

20

30

40

50

作の一形態は、電子ドキュメント中に（例えば、電子ドキュメントのメタタグ中に）数百個のキー用語を挿入すること、または、1つもしくは複数の検索用語に関するその電子ドキュメントの関連性について検索エンジンを混乱させて過大評価させる（または誤って識別させることさえする）他の技法を利用することを含んでいる。例えば、自動車についてのクラシファイド広告（classified advertising）ウェブページの作成者は、用語「car（車）」の反復を用いてこの「キー用語」セクションを満たすこともある。この作成者がこれを行う結果、検索エンジンは、ユーザが用語「car」を検索するときにはいつでも、そのウェブページをより関連性の高いウェブページとして識別するようになる。しかし、このウェブページのトピックをより正確に表す「キー用語」セクションは、「automobile（自動車）」、「car」、「classified（クラシファイド）」、および「for sale（売り物）」という用語を含むこともあるだろう。

10

【0005】

検索エンジンスパムを作成する他の一部の技法には、実際のユーザに対するものと異なる電子ドキュメントを検索エンジンに対して返すこと（すなわち、クローキング（cloaking）技法）、電子ドキュメントと無関係のキー用語を対象とすること、キー用語カウントを増大させるためにユーザが見ることのない区域にキー用語を挿入すること、リンクの人気を増大させるためにユーザが見ることのない区域にリンクを挿入すること、低品質のドアウェイウェブページ（doorway web page）を生成すること、高ランク付け電子ドキュメントから無関係の電子ドキュメントへとユーザを詐欺的に転送して、無関係の電子ドキュメントをそのユーザに対して表示することなどが含まれる。その結果、検索エンジンは、クエリ（query）を実行するユーザに対して、実際には関連していないこともある、高ランク付けされた電子ドキュメントを提供することになる。したがって、この検索エンジンは、このような作為的なランキング操作からユーザを保護をしていない。

20

【0006】

既存の検索エンジンは、各スパム技法を別々に解析して操作された電子ドキュメントのパターンを識別することによって、検索エンジンスパムを防止しようと試みる。このような検索エンジンが識別されたパターンを有する電子ドキュメントを検出するとき、この検索エンジンは、この電子ドキュメントにスパムとしてラベルを付けて、この電子ドキュメントを検索結果の形でユーザに表示させず、または、その結果を格下げする。例えば、特定の検索エンジンでは、エンドユーザのためではなくその検索エンジンのために主として構築された電子ドキュメントに、検索エンジンスパムとしてラベルを付けることができる。同様に、ある検索エンジンでは、電子ドキュメント中の隠しテキストおよび/または隠しリンクを検出し、この電子ドキュメントに検索エンジンスパムとしてラベルを付けることができる。一部の検索エンジンではまた、不必要な多数のホスト名（例えば、poker.foo.com、blackjack.foo.comなど）を有し、過剰なクロスリンクを使用してそのウェブサイトの外見上の人気を人為的に上げたウェブサイトを検出し、このウェブサイトに検索エンジンスパムとしてラベルを付けることができる。さらに、既存の検索エンジンでは、クローキング技法、またはウェブサイトが別のウェブサイトと相互リンクを交換して検索エンジンの最適化を増大させるリンクファーム（link farming）を使用したウェブサイトを検出することもできる。

30

40

【0007】

検索エンジンスパムとは著しく違って、電子メール（またはEメール）スパムは、通常、同時に多数の受信者に送信される迷惑電子メールメッセージである。電子メールスパムは、ジャンクメール（junk mail）の電子的な等価物である。ほとんどの場合に、電子メールスパムメッセージの内容は、その受信者の関心とは無関係である。したがって、電子メールスパムを作成することは、最小のコストで非常に多数の人々にメッセージを配信するインターネットの不正な使用である。

【発明の開示】

50

【発明が解決しようとする課題】

【0008】

電子メールスパムは、いくつかの点で検索エンジンスパムから区別される。例えば、プログラムは、電子メールメッセージを自動的に生成して、電子メールスパムを多数の受信者に送信することができる。対照的に、検索エンジンスパムは、電子メールアドレス、送信者、または受信者を含んではいない。しかし、それにもかかわらず、検索エンジンスパムは、ある種の特徴を電子メールスパムと共有している。例えば、検索エンジンスパムも電子メールスパムも共に、それらが詐欺的にユーザを誘導して特定の製品またはサービスへと訪問させるために作成されているという点において、望ましいものではない。したがって、少なからず電子メールスパムの作成者は、検索エンジンスパムを生成して、製品もしくはサービスに関連した1つまたは複数の電子ドキュメントの露出を増大させようとすることもある。すなわち、スパマー (s p a m m e r) はしばしば、電子メールスパムも検索エンジンスパムも共に利用して製品またはサービスを売り込むことがある。したがって、一般的に、電子メールスパムと検索エンジンスパムの間には強い相関がある。それにもかかわらず、従来技術のシステムおよび方法では、電子メールスパムおよび検索エンジンスパムの可能性のあるソースの間におけるこのような相関が見逃されてきている。特に、従来技術においては、電子メールスパムおよび検索エンジンスパムを、まったく異なる解決方法を必要とする別々の問題として取り扱っている。

10

【0009】

したがって、検索エンジンスパムを効果的に識別し防止する解決方法が、強く望まれている。

20

【課題を解決するための手段】

【0010】

本発明の実施形態では、とりわけ外部ソースを使用して、検索に関連して望ましくないものである得る電子ドキュメントを検出し、それによってより良好な検索エンジンの結果を実現することにより、従来技術における1つまたは複数の欠陥を克服している。本発明の一実施形態によれば、電子メールスパム検出システムは、電子メールメッセージを電子メールスパムの可能性の高いものとして識別する。次いで、データベースなどのメモリ領域は、この電子メールメッセージ中に含まれるリンクのリストを記憶する。本発明の一実施形態は、このデータベースにアクセスし、このデータベースに記憶されたリンクが提供する電子ドキュメントについての信頼レベル (c o n f i d e n c e l e v e l) を決定する。この電子ドキュメントの信頼レベルは、この電子ドキュメントが検索エンジンスパムであるという可能性 (l i k e l i h o o d) を示す。別の実施形態においては、本発明は、電子メールスパムの可能性が高い電子メールの起源となるネットワークアドレスを識別する。次いで、このデータベースは、このネットワークアドレスを記憶する。このデータベースにアクセスすることにより、本発明の実施形態は、このネットワークアドレスに位置する電子ドキュメントについての信頼レベルを決定することができる。それによって検索エンジンスパムを、より良好に識別することができる。さらに、この電子メールスパム検出システムは、電子メールスパム中に頻繁に出現する用語 (例えば、単語、単語の組合せ、フレーズ、ストリング、n-グラム、バイナリデータなど) のリストを識別することができる。次いで、データベースは、この用語リストを記憶する。したがって、本発明の一実施形態は、この電子ドキュメントが検索エンジンスパムである可能性を示す1つまたは複数のこの記憶された電子メールスパム用語を含んでいる検索エンジンスパムに関して、電子ドキュメントについての信頼レベルを生成する。電子ドキュメントが、検索エンジンスパムであるという高い信頼レベルを有する場合には、本発明の実施形態は、ユーザに提供される検索結果中において、この電子ドキュメントを格下げすることができる。代わりに、本発明の実施形態は、この提供された検索結果からこの電子ドキュメントを除去することもできる。

30

40

【0011】

1つまたは複数の他の実施形態によれば、本発明により、ユーザは電子ドキュメントの

50

望ましさについての情報を提供できるようになる。ユーザは、電子メールスパムまたは検索エンジンスパムに回答して、この情報を提供することができる。このユーザ提供情報がこの電子ドキュメントを望ましくないとして特徴づける場合には、本発明の実施形態は、この電子ドキュメントの1つまたは複数の属性 (a t t r i b u t e) を識別して、この電子ドキュメントについて格付けを生成する。この電子ドキュメントが高い格付けを有する場合には、この電子ドキュメントは、検索エンジンスパムであるという高い可能性を有する。したがって、本発明の実施形態は、検索結果におけるこの電子ドキュメントのランキングを調整して、正確な関連性ランキングをユーザに対して提供することができる。さらに、本明細書中で説明している本発明の実施形態の特徴は、経済的に実現可能であり、商業的に実用可能であり、また現在使用可能な技法に比べて実装するのがより簡単である。

10

【0012】

簡単に説明すれば、検索に関連して、本発明の態様を使用した方法は、電子ドキュメントを評価する。この方法は、電子ドキュメントの第1の信頼レベルを決定するステップを含んでいる。この電子ドキュメントは、ユーザからの検索要求に回答して、検索エンジンによって検索が可能である。この第1の信頼レベルは、この検索エンジンの外部のソースが提供する情報に基づいて、この電子ドキュメントが望ましくないものである可能性を示す。この方法は、この電子ドキュメントの第2の信頼レベルを決定するステップも含んでいる。この第2の信頼レベルは、この電子ドキュメントの1つまたは複数の属性に基づいて、この電子ドキュメントがこの検索要求に関して不満足なものである可能性を示す。この方法は、この決定された第1の信頼レベルと決定された第2の信頼レベルの関数としてこの電子ドキュメントについての格付けを生成するステップをさらに含んでいる。この方法は、この電子ドキュメントの生成された格付けに基づいて、この電子ドキュメントを、この検索要求に関して不満足なものであるとして指定するステップも含んでいる。

20

【0013】

本発明の他の実施形態においては、検索に関連して、本発明の態様を使用した方法は、電子ドキュメントを評価する。この方法は、電子ドキュメントに関してユーザ提供情報を受信するステップを含んでいる。この電子ドキュメントは、ユーザからの検索要求に回答して、検索エンジンによって検索が可能である。このユーザ提供情報は、この電子ドキュメントを望ましくないものとして特徴づける。この方法はまた、この受信されたユーザ提供情報の関数としてこの電子ドキュメントに関する格付けを生成するステップを含んでいる。この方法は、この電子ドキュメントの生成済みの格付けに従って、この検索要求に関して不満足であるものとして、この電子ドキュメントを指定するステップをさらに含んでいる。

30

【0014】

本発明のさらに他の実施形態においては、本発明の態様を使用したシステムは、検索に関連して、電子ドキュメントを評価する。このシステムは、ユーザからの検索要求を受信し、この受信された検索要求に基づいて電子ドキュメントを識別するためのプロセッサを含んでいる。このシステムはまた、この電子ドキュメントが望ましくないかどうかを評価する際に使用するための、このプロセッサの外部のソースが提供するデータを記憶するメモリ領域を含んでいる。このプロセッサは、この電子ドキュメントの第1の信頼レベルを決定するように構成されている。この第1の信頼レベルは、この外部のソースが提供するデータに基づいて、この電子ドキュメントが望ましくないものである可能性を示す。このプロセッサは、この電子ドキュメントの第2の信頼レベルを設定するようにも構成されている。この第2の信頼レベルは、この電子ドキュメントの1つまたは複数の属性に基づいて、この電子ドキュメントが検索に関連して不満足である可能性を示す。このプロセッサはさらに、この決定された第1の信頼レベルとこの設定された第2の信頼レベルの関数としてこの電子ドキュメントについての格付けを生成し、この電子ドキュメントの生成済みの格付けに基づいて、この受信された検索要求に関して不満足なものであるとして、この電子ドキュメントを分類するように構成される。

40

50

【0015】

本発明のさらに他の実施形態においては、本発明の態様を使用したコンピュータ読取り可能媒体は、検索に関連して、電子ドキュメントを評価するためのコンピュータ実行可能コンポーネントを有している。このコンピュータ読取り可能媒体は、電子ドキュメントに関してユーザ提供情報を受信するためのインターフェースコンポーネント(interface component)を含んでいる。この電子ドキュメントは、ユーザからの検索要求に应答して、検索可能である。このユーザ提供情報は、この電子ドキュメントを望ましくないとして特徴づける。このコンピュータ読取り可能媒体は、この受信されたユーザ提供情報の関数としてこの電子ドキュメントについての格付けを生成するための解析コンポーネント(analyzing component)も含んでいる。このコンピュータ読取り可能媒体は、この電子ドキュメントの生成済みの格付けに従って、この検索要求に関して不満足なものであるとして、この電子ドキュメントを分類するためのクエリコンポーネント(query component)をさらに含んでいる。

10

【0016】

本発明のさらに他の実施形態においては、本発明の態様を使用したコンピュータ読取り可能媒体は、検索に関連して、電子ドキュメントを評価するためのコンピュータ実行可能コンポーネントを有している。このコンピュータ読取り可能媒体は、ユーザからの検索要求を受信し、この受信された検索要求に基づいて電子ドキュメントを識別するクエリコンポーネントを含んでいる。このコンピュータ読取り可能媒体はまた、この電子ドキュメントが望ましくないかどうかを評価するためのデータを提供するための外部コンポーネントも含んでいる。このコンピュータ読取り可能媒体は、この電子ドキュメントの第1の信頼レベルを決定するための内部コンポーネントをさらに含んでいる。この第1の信頼レベルは、この外部コンポーネントが提供するデータに基づいてこの電子ドキュメントが望ましくないものである可能性を示す。この内部コンポーネントは、さらにこの電子ドキュメントの第2の信頼レベルを設定するように構成されている。この第2の信頼レベルは、この電子ドキュメントの1つまたは複数の属性に基づいて、この電子ドキュメントが検索に関連して不満足なものである可能性を示す。コンピュータ読取り可能媒体はまた、この決定された第1の信頼レベルとこの設定された第2の信頼レベルの関数としてこの電子ドキュメントについての格付けを生成する解析コンポーネントも含んでいる。クエリコンポーネントは、この電子ドキュメントの生成済みの格付けに基づいて、この受信された検索要求に関して不満足なものであるとして、この電子ドキュメントを分類するように構成されている。

20

30

【0017】

検索に関連して不満足な電子ドキュメントを検出する方法を実施するためのコンピュータ実行可能命令を有するコンピュータ読取り可能媒体は、本発明のさらなる態様を実施している。

【0018】

代わりに、本発明の実施形態は、他の様々な方法および装置を含むこともできる。

【0019】

他の特徴については、本明細書中で以下において部分的には明白となり、また部分的には指摘もしている。以下、同一の参照文字は、図面全体を通して同一の要素を示している。

40

【発明を実施するための最良の形態】

【0020】

望ましくない電子ドキュメントを検出するための例示的なネットワーク環境

最初に図1を参照すると、ブロック図が、本発明の実施形態を利用することができる適切なネットワーク環境の一実施例を示している。サーバコンピュータ102は、検索エンジン104などのプロセッサを含んでいる。この検索エンジン104は、クローラ(crawler)106をさらに含んでいる。クローラ106は、図1に示すリモートサーバコンピュータ110やリモートサーバコンピュータ112など、通信ネットワーク108

50

に接続された1つまたは複数のコンピュータ上に分散された電子ドキュメントを検索する。通信ネットワーク108は、イントラネットなどのローカルエリアネットワーク、インターネットなどのワイドエリアネットワーク、またはサーバコンピュータ102が、それによってサーバコンピュータ110および112などのリモートコンピュータと直接的または間接的に情報をやりとりすることができるネットワークの組合せであってもよい。

【0021】

クローラ106は、ネットワーク108に接続されたサーバコンピュータ110および112を検索し、サーバコンピュータ110上に記憶された電子ドキュメント114および116、並びに、サーバコンピュータ112上に記憶された電子ドキュメント118および120を見出す。これらのリモートサーバコンピュータ上に記憶された電子ドキュメントは、ウェブページ（例えば、HTMLページおよびXMLページ）、並びに、マルチメディアファイルを含むことができる。クローラ106は、これらの電子ドキュメントおよび関連付けられたデータを受信する。さらに、サーバコンピュータ102は、クローラ106によってアクセスされた電子ドキュメント122および124を含むことができる。

10

【0022】

図1に示すように、検索エンジン104の外部にあるソースを構成する電子メールスパム検出システム126もまた、ネットワーク108に接続される。電子メールスパム検出システム126は、システム126のユーザに配信される電子メールスパムを検出するシステムである。特に、サーバ110および/またはサーバ112など1つまたは複数のリモートコンピュータは、システム126のユーザに対して電子メールメッセージを生成し送信することができる。次いで電子メールスパム検出システム126は、特定の電子メールメッセージが電子メールスパムであり得ることを検出し、そのユーザを保護するアクションを実施する。例えば、システム126は、ユーザのメールボックスから検出された電子メールスパムを締め出すことができ、もしくは、ユーザに特定の電子メールメッセージが電子メールスパムであり得ることを警告することができる。あるいは、システム126は、その受信者にそのメッセージを配信する前に、信用のけるユーザに対してそのメッセージが電子メールスパムではないことを確認するための電子メールメッセージを配信することもできる。

20

【0023】

電子メールスパム検出システム126は、電子メールスパムを検出するためにいくつかの技法を利用することができる。1つの技法においては、システム126は、電子メールスパムのパターンを識別するようにトレーニングされた確率的分類機構（probabilistic classifier）を含んでいる。この確率的分類機構は、電子メールメッセージを分類するコンピュータ実行可能命令を含んでいる。一般に、この確率的分類機構は、電子メールスパムにおいて統計的に重要な属性（例えば、統計的に重要なキー用語および/またはコンテキスト情報）の組合せを識別する。迷惑電子メールメッセージは、しばしば一般的に共有されている一部の属性を含んでいる。このように一般的に共有され、したがって統計的に重要な属性の実施例には、製品またはサービスの非現実的な提案を記述するキー用語（例えば、無料の薬品、減量プログラム、またはクレジットカードの申し込み）が含まれている。さらに、このような属性は、電子メールスパムを送信していた判定される電子メールアドレスを含むこともある。特に、この確率的分類機構をトレーニングして、（例えば、電子メールスパムの「From:」行に基づいて）電子メールスパムの1人または複数の作成者のドメイン名を識別することができる。次いで、この確率的分類機構は、電子メールメッセージのこの「From:」行を解析することにより、この電子メールメッセージの送信者が既知の電子メールスパム作成者に対応するかどうかを決定することができる。

30

40

【0024】

同様にこの確率的分類機構をトレーニングして、電子メールスパムが起源とするネットワークアドレスを認識することができる。電子メールスパマーは、しばしば勝手に電子メ

50

ールスパムの「From:」行または他の情報を任意の値に設定することができる。しかし、この電子メールスパムの起源となるネットワークアドレス（例えば、IPアドレス）を隠すことは難しい。したがって、電子メールスパムを特徴づけるために、この着信SMTP接続のネットワークアドレスは、この確率的分類機構をトレーニングするのに価値のある属性である。さらに、この確率的分類機構をトレーニングして、電子メールスパムに関連付けられた1つもしくは複数のリンクまたはURLを識別することもできる。すなわち、電子メールスパムの可能性の高い電子メールに含まれるURLを特に解析して、電子メールスパムを特徴づける属性を生成する。多数の電子メールメッセージは、埋め込まれたURLを含んでいる。これらのURLの存在は、これらの電子メールメッセージが電子メールスパムであることを示すことができる。例えば、これらのURLは、迷惑な製品やサービスを提案する1つまたは複数のウェブページへと、電子メール受信者を誘導することもある。一実施形態においては、ホスト名（例えば、アルファベットの、小数点付き10進数の、16進数の、または、8進符号のホスト名）が、これらのURLから抽出されて、電子メールスパムを特徴づける助けになる。したがって、組み合わせられたURLが、`<URL1>@<URL2>@...@<URLn>`の形式である場合には、この最後の@記号の後のURL（すなわち、URL_n）が抽出すべきホスト名である。

10

【0025】

電子メールスパマーは、このスパマーに関連しているホスト名がこの確率的分類機構によって抽出されないようにするためのリダイレクタ（`redirector`）を、URL中に含めることもある。このリダイレクタがURL中に含まれることにより、この電子メールスパマーに関連するウェブサイトへこの電子メール受信者を転送する。このようなシナリオにおいては、この確率的分類機構は、リダイレクトURL中に隠された本物のホスト名を識別し、電子メールスパムを特徴づける属性としてこの本物のホスト名を使用するように構成される。

20

【0026】

説明したように、この確率的分類機構を電子メールスパムの可能性の高い電子メール上でトレーニングして、この電子メールスパムの1つまたは複数の属性を認識することができる。電子メールスパム検出システム126は、この確率的分類機構をトレーニングするために、いくつかの技法を使用して電子メールスパムの可能性のある電子メールを識別することができる。1つの技法においては、電子メール受信者は、特定の電子メールメッセージが電子メールスパムであるかどうかを示すことができる。別の技法では、システム126は、電子メールスパムを捕捉するためのハニーポット（`honeypot`）を保持する。ハニーポットは、決して存在したことがないか、または、所与の期間に終了してしまっている電子メールアカウントを表している。しかし、電子メールスパマーにとっては、ハニーポットは、通常の電子メールアカウントであるように見える。したがって、ハニーポットによって表される電子メールアカウントが、決して存在したことがないか、または、ある期間中に終了してしまっており、したがってこの電子メールアカウントが正当な電子メールを受信している理由はないと仮定すれば、ハニーポットに送信された電子メールメッセージは、電子メールスパムと考えることができる。

30

【0027】

この確率的分類機構をトレーニングするために電子メールスパムの可能性のある電子メールを識別するさらに他の技法においては、電子メールスパム検出システム126は、着信電子メールに対するチャレンジ応答機構（`challenge response`）を実装することができる。すなわち、システム126は、着信電子メールの送信者に、この電子メールがマシン生成されたものでないことを確認するチャレンジを解決するように要求することができる。この送信者がこのチャレンジを解決できない場合には、システム126は、この確率的分類機構がその属性を抽出する電子メールスパムの可能性のあるメールとして、この電子メールを識別することができる。

40

【0028】

広範な様々なトレーニング技法を利用して、この確率的分類機構をトレーニングするこ

50

とができる。スパムとして識別される電子メールおよび非スパムとして識別される電子メールが、コンピュータ実行可能トレーニング命令に供給される。次いで、これらのコンピュータ実行可能トレーニング命令は、スパムとして識別される電子メール中に存在するが、非スパムとして識別される電子メール中には存在しない属性を認識する。したがって、これらの認識された属性は、電子メールスパム中において統計的に意味のあるものとして分類される。これらのコンピュータ実行可能トレーニング命令は、さらに統計的に意味のあるものとして分類される属性ごとに、重みを決定することができる。これらのトレーニング命令は、どれほど頻繁にその属性が電子メールスパム中に現れるかを含めて、いくつかのファクタに基づいて所与の属性についての重みを決定する。これらのコンピュータ実行可能トレーニング命令は、いくつかの異なるアーキテクチャとして実装することができる。例えば、これらのコンピュータ実行可能トレーニング命令は、ナイーブベイジアン分類機構、制限依存性ベイジアン分類機構、ベイジアンネットワーク分類機構、判定ツリー、サポートベクトルマシン、コンテンツマッチング分類機構、最大エントロピー分類機構、およびこれらの組合せなどとして実装することができる。

10

【0029】

さらに、システム126の確率的分類機構は、パターン認識機構によってトレーニングして、キー用語マッチング技法では識別することができないこともある統計的に意味のある属性の組合せを識別することができる。特に、パターン認識機構がこの確率的分類機構をトレーニングするために使用するこれらの統計技法は、この確率的分類機構が所与の属性の変形を認識することができるように、トレーニングサンプルに基づいて属性を一般化

20

【0030】

電子メールメッセージから抽出された属性のその解析に基づいて、この確率的分類機構は、この電子メールメッセージについての格付け(rating)を生成する。例えば、確率的分類機構は、電子メールメッセージ中で識別される個々の属性(例えば、用語、ネットワークアドレス、ホスト名など)に対して絶対重みを割り当てることができる。前述のように、所与の属性についての重みは、この確率的分類機構のトレーニングプロセス中に決定される。次いで、この確率的分類機構は、この割り当てられた重みを(例えば、これらの重みの合計を取る)数学的関数に適用することにより、この電子メールメッセージについての格付けを生成する。一実施形態においては、電子メールメッセージの格付けは、パーセンテージ(例えば、60%)の形式とすることができる。電子メールメッセージの格付けが高くなればなるほど、この電子メールメッセージが電子メールスパムである可能性は高くなる。すなわち、電子メールメッセージの格付けは、この電子メールメッセージが、電子メールスパム中に現れる可能性の高い要素を含んでいる可能性を示す。別の実施形態においては、この確率的分類機構は、特定の属性がその電子メールメッセージ中に出現する頻度、並びに、この電子メールメッセージ中に存在する属性の組合せに基づいて、電子メールメッセージについての格付けを生成する。特に、それ自体では電子メールスパムであることを示すことができない属性は、電子メールメッセージが電子メールスパムを構成するコンテキスト情報または集約情報としての役割を果たすことができる。例えば、属性「クレジットカード」だけでは、電子メールメッセージが電子メールスパムであることを示唆しないこともある。しかし、属性「年会費無料」と組み合わせられた属性「クレジットカード」は、この電子メールメッセージが、おせっかいな提案、つまり電子メールスパムを構成することを示唆することができる。

30

40

【0031】

確率的分類機構は、さらに生成された格付けの関数としてこの電子メールメッセージを

50

分類する。すなわち、確率的分類機構が電子メールメッセージについての格付けを生成した後、確率的分類機構は、この電子メールメッセージがこの格付けに基づいて電子メールスパムを構成するかどうかを判定する。例えば、電子メールスパム検出システム126は、その上にしきい値レベル(例えば、70%)を記憶していることもあり、このしきい値レベルは、所定の電子メールメッセージが望ましくないものである可能性を表す。次いで、確率的分類機構は、この電子メールメッセージの格付けをこのしきい値レベルと比較する。一実施形態においては、この電子メールメッセージの格付けがこのしきい値レベルよりも大きい(またはしきい値レベル以上である)場合には、確率的分類機構は、この電子メールメッセージを電子メールスパムとして分類する。管理者が、このしきい値レベルを変更することにより、電子メールスパム検出システム126の感受性を変更することができることに留意されたい。例えば、管理者は、より高いしきい値レベルを設定し、その結果、より少ない電子メールメッセージが電子メールスパムとして分類されるようにすることができる。

【0032】

ある電子メールメッセージが電子メールスパムの可能性があるものとして分類される場合には、ネットワーク108に接続されたデータベース128などのメモリ領域に記憶するために、システム126はこの電子メールメッセージに関連付けられたある種の属性を抽出する。本発明の一実施形態によれば、システム126は、この電子メールメッセージに関連付けられた1つまたは複数のネットワークアドレス(例えば、IPアドレス)を識別する。例えば、システム126は、この電子メールメッセージの起源となるネットワークアドレスを識別することができる。したがって、この電子メールメッセージが、サーバ110に由来する場合、システム126は、サーバ110のネットワークアドレスをデータベース128に記憶する。本発明の別の実施形態によれば、システム126は、さらに電子メールスパムとして分類される電子メールメッセージ中に含まれる1つまたは複数のリンクを識別する。次いでシステム126は、データベース128中に、この識別されたリンクのホスト名を記憶する。したがって、電子メールスパムとして分類された電子メールメッセージが電子ドキュメント114のURLを含む場合には、システム126は、このURLのホスト名をデータベース128に記憶する。さらに、システム126は、電子メールスパムに関連付けられた用語リスト(例えば、単語、単語の組合せ、フレーズ、ストリング、n-グラム、2進データなど)を識別する。システム126は、この用語リストもまたデータベース128に記憶する。

【0033】

データベース128中に記憶されたネットワークアドレス、ホスト名、または用語ごとに、システム126は、さらにこのネットワークアドレス、ホスト名、または用語が電子メールスパムに関連付けられている信頼レベルを指定する。システム126は、このネットワークアドレス、ホスト名、または用語を含んでいる電子メールメッセージの格付けに基づいてこの信頼レベルを指定することができる。したがって、この確率的分類機構が特定の電子メールメッセージについて80%の格付けを生成する場合には、この確率的分類機構は、この電子メールメッセージから識別されるネットワークアドレス、ホスト名、および/または用語について80%の信頼レベルを指定する。このネットワークアドレス、ホスト名、および/または用語についての指定された信頼レベルは、同様にデータベース128に記憶される。

【0034】

検索エンジン104のクローラ106が、ネットワーク108をナビゲートして、ネットワーク108上に配置された1つまたは複数の電子ドキュメントを収集し、検索エンジン104のインデックスビルダ(index builder)129が、この収集された電子ドキュメントを解析して、インデックス付けたためのこれらの特徴を識別するときに、検索エンジン104は、収集された電子ドキュメントについての別の信頼レベルを設定して、この収集された電子ドキュメントが検索エンジンスパムである(すなわち、検索に関して是不満足なものである)可能性を示すことになる。詳細には、クローラ106は、

この収集された電子ドキュメントの1つまたは複数のパターンを識別して、これらのパターンが検索エンジンスパムを特徴づけるパターンに対応するかどうかを決定する。例えば、クローラ106は、この収集された電子ドキュメントが、エンドユーザのためではなく検索エンジン104のために主として構築されているかどうかを識別することができる。クローラ106は、さらにこの収集された電子ドキュメントが検索エンジンスパムをしばしば特徴づける隠しテキストおよび/または隠しリンクを含んでいるかどうかを検出することができる。検索エンジンスパムを特徴づける他のパターンの一部には、非常に多数の不必要なホスト名、過剰なクロスリンク、リンクファームिंग(link farming)などを含んでいるものがある。収集された電子ドキュメントの識別済みのパターンに基づいて、検索エンジン104は、この収集された電子ドキュメントが検索エンジンスパムを構成する信頼レベルを生成することができる。

10

【0035】

検索エンジン104は、さらにデータベース128にアクセスしてクローラ106によって収集された1つまたは複数の電子ドキュメントに関連した情報を抽出するように構成されている。一実施形態においては、検索エンジン104は、データベース128に記憶されたネットワークアドレスリストを取得する。検索エンジン104が、取得されたネットワークアドレスが収集された電子ドキュメントのロケーションに対応していると判定する場合には、この検索エンジンは、このネットワークアドレスに関連付けられた信頼レベルをデータベース128から抽出する。同様に、検索エンジン104は、データベース128からホスト名リストを取得し、取得されたホスト名が収集された電子ドキュメントを提供するホスト名に対応するかどうかを判定することができる。対応している場合には、検索エンジン104は、この取得されたホスト名に関連付けられた信頼レベルをデータベース128から抽出する。さらに、検索エンジン104は、このホスト名が提供する電子ドキュメントからリンクされた1つまたは複数の電子ドキュメントを、この信頼レベルを有するものとして指定することができる。データベース128に記憶された用語について、検索エンジンは、この用語が収集された電子ドキュメント中に現れるかどうかを判定する。この記憶された用語がこの収集された電子ドキュメント中に現れる場合には、検索エンジンは、この記憶された用語に関連付けられた信頼レベルをデータベース128から抽出する。

20

【0036】

収集された電子ドキュメントが検索エンジンスパムを構成しているという可能性を示す検索エンジン104によって決定される信頼レベル、並びに、この収集された電子ドキュメントに関連付けられたネットワークアドレス、ホスト名、および/または用語の信頼レベルに基づいて、検索エンジン104は、この収集された電子ドキュメントについての重み付けされた格付けを計算する。特に、ネットワーク108のクローリング中に検索エンジン104が決定する信頼レベルは、この収集された電子ドキュメントが検索に関して望ましくないものである可能性を示している。また、データベース128から取得された1つ(または複数)の信頼レベルは、この収集された電子ドキュメントが望ましくない電子メールメッセージ(すなわち、電子メールスパム)に関連付けられている可能性を示している。電子メールスパムと検索エンジンスパムとの間の所有者の関連性(すなわち、電子メールスパムの作成者は、検索エンジンスパムを生成する可能性が高い)のために、検索エンジン104は、これらの2つのタイプの信頼レベルを結合させて、この収集された電子ドキュメントが検索エンジンスパムであるかどうかを、高い信頼度で示す重み付けされた格付けを生成することができる。

30

40

【0037】

電子ドキュメントが検索エンジンスパムを構成するというこの結合された可能性を、高い信頼度で決定する特定の一方法として、格付けを生成するため、様々なタイプの信頼レベルを重み付け平均する。例えば、この電子ドキュメントは検索エンジンスパムであるという60%の信頼レベルを有し、この電子ドキュメントのネットワークアドレスは電子メールスパムに関連付けられているという80%の信頼レベルを有し、この電子ドキュメント

50

中に出現する用語は電子メールスパムに関連付けられているという70%の信頼レベルを有する場合には、検索エンジン104は、これらの信頼レベルを平均してこの電子ドキュメントについて70%の格付けを生成することができる。代わりに、この電子ドキュメントの格付けを検索エンジンスパムである信頼レベル、および、電子メールスパムに関連した信頼レベルの重み付け平均とすることもできる。したがって、上記の実施例において、この電子ドキュメントのネットワークアドレスが電子メールスパムに関連付けられたという80%の信頼レベルは、この電子ドキュメント中に出現する用語が電子メールスパムに関連付けられたという70%の信頼レベルと共に重み付けされて、この電子ドキュメントは電子メールスパムに関連しているという75%の信頼レベルを生成することになる。次いで検索エンジン104は、この重み付けされた信頼レベルを検索エンジンスパムであるという60%の信頼レベルと共に平均して、67.5%の格付けを生成する。この67.5%の格付けは、この電子ドキュメントが電子メールスパムを構成しているという重み付けされた確率を示している。

10

【0038】

代わりに、これら2つの異なる信頼レベルは、電子ドキュメントがスパムに関連している可能性があるかどうかを判定する相異なるメカニズムを使用しているため、この電子ドキュメントが検索エンジンスパムを構成しているという結合された可能性は、これらのどちらのタイプの信頼レベルよりも高くなり得る。例えば、電子ドキュメントが検索エンジンスパムであることについて70%の信頼レベルを有し、この電子ドキュメントのネットワークアドレスが電子メールスパムに関連付けられていることについて80%の信頼レベルを有する場合には、この電子ドキュメントが検索エンジンスパムを構成しているこの結合された可能性は、90%になることもある。したがって、電子メールスパムの可能性のある電子メールに対する電子ドキュメントの関連付けを考慮することにより、検索エンジン104は、この電子ドキュメントが検索エンジンスパムであるかどうかを正確に判定することができる。

20

【0039】

検索エンジン104は、特定の電子ドキュメントが検索エンジンスパムを構成している可能性があるかと判定した後（例えば、この電子ドキュメントの格付けが、あるしきい値レベルよりも大きいときには）、検索エンジン104のクエリプロセッサは、様々なアクションを実施して、検索結果においてこの電子ドキュメントをユーザに対して表示しないようにすることができる。したがって、ユーザによって送られた検索要求に基づいて、このクエリプロセッサは、検索エンジンスパムを構成していると判定された電子ドキュメントをこの送信された検索要求の「ヒット(hit)」として、識別することができる。このようなシナリオにおいては、このクエリプロセッサは、このユーザに対して提供される検索結果中において、この電子ドキュメントを格下げすることができる。すなわち、この電子ドキュメントは検索エンジンスパムを構成している可能性があるため、検索エンジン104のクエリプロセッサは、この検索結果中においてこの電子ドキュメントのランキングを低下させる。代わりに、このクエリプロセッサは、このユーザに対して提供される検索結果からこの電子ドキュメントを取り除くこともできる。本発明の一実施形態においては、このクエリプロセッサによって取られるアクションを調整することが可能である。すなわち、電子ドキュメントが検索エンジンスパムであることがより確実な場合には、電子ドキュメントはさらに厳しいペナルティが課される。例えば、85%よりも大きな格付けの電子ドキュメントは、ユーザに対して提供される検索結果から取り除くことができる。一方、65%と85%の間の格付けをもつ電子ドキュメントは、検索結果中において50ランク分だけ格下げをすることができる。また、50%と65%の間の格付けをもつ電子ドキュメントは、25ランク分だけ格下げをすることができるが、一方で50%以下の格付けを有する電子ドキュメントは、ペナルティを受けないことになる。本発明の別の実施形態においては、電子ドキュメントの事前のランキングが、所定のランク（例えば、5番目のランク）よりも高い場合には、このクエリプロセッサは、検索結果中においてこの電子ドキュメントのランキングを保持する。すなわち、たとえその電子ドキュメントは検索

30

40

50

エンジンスパムであると判定されるとしても、高い関連性のある電子ドキュメントは、ペナルティを受けないこともある。

【0040】

次に図2を参照すると、ブロック図は、本発明の実施形態を利用することができる適切なネットワーク環境の別の実施例を示している。サーバコンピュータ202は、検索エンジン204を含んでいる。サーバコンピュータ202は、通信ネットワーク206に接続され、この通信ネットワークは、さらにリモートサーバコンピュータ208に接続される。通信ネットワーク206は、イントラネットなどのローカルエリアネットワーク、インターネットなどのワイドエリアネットワーク、または、このサーバコンピュータ202がリモートサーバコンピュータ208などのリモートコンピュータと直接または間接に通信を行うことができるネットワークの組合せであってもよい。リモートサーバコンピュータ208は、電子ドキュメント210および電子ドキュメント212を提供し、これらのドキュメントは、ウェブページまたはマルチメディアファイルとすることができる。さらに、リモートサーバコンピュータ208は、ネットワーク206に接続されるコンピュータを介して、1つまたは複数の電子メールメッセージをユーザ214に対して送信するように構成される。

10

【0041】

ユーザ214が、サーバコンピュータ208から電子メールメッセージを受信した後、このユーザは、この受信された電子メールメッセージを電子メールスパムまたは非スパムとして識別する。次いでユーザ214は、電子メールスパム検出システム216のインターフェースに対する入力（または一般的にユーザが提供する情報）として、この受信された電子メールメッセージのユーザの識別を送信する。この入力を受信するのに応答して、電子メールスパム検出システム216は、この電子メールメッセージが電子メールスパムであることについての信頼レベルを設定する。さらに、システム216はこの電子メールメッセージについての複数の入力を複数のユーザから受信し、これらの入力が互いに矛盾する場合には、システム216は、この電子メールメッセージについての信頼レベルを設定しないように決定することもできる。他方、これらの入力が互いに一致する場合には、システム216は、この電子メールメッセージが電子メールスパムを構成するという信頼レベルを設定することができる。本発明の他の実施形態においては、システム216は、ある規則を実装して1つまたは複数の入力を判定することができる。すなわち、ある種の入力には、これらの入力を送信したユーザはより信頼がおけるので、より高い重み付けが行われる。この他の実施形態においては、システム216は、特定の電子メールメッセージを電子メールスパムとして報告するユーザのパーセンテージを決定する。これらの大多数のユーザが、この電子メールメッセージは電子メールスパムであることに合意する場合には、少数のユーザからの入力は、あまり信頼することができない。すなわち、特定のユーザがある電子メールメッセージを電子メールスパムとして報告し、他の大多数のユーザがこの特定のユーザに同意する場合には、システム216は、このユーザは信頼がおけると判断することができる。他方、他の大多数のユーザが、この特定のユーザに同意しない場合には、システム216は、このユーザは信頼できないと判断することができる。したがって、システム216は、ユーザ提供からの入力の信頼性に少なくとも部分的に基づいて、電子メールメッセージについての信頼レベルを決定することができる。

20

30

40

【0042】

電子メールスパム検出システム216が、特定の電子メールメッセージが電子メールスパムを構成すると判定する場合には、この電子メールスパム検出システムはこの電子メールメッセージを解析して、この電子メールメッセージの1つまたは複数の属性を識別して、この電子メールスパムのパターンを決定する。この電子メールメッセージが、あるイメージを含む場合には、システム216は、このイメージ中のフレッシュトーン（fresh tone）のレベルを検出することによってこれらの属性を識別する。一実施形態においては、システム216は、この電子メールスパムに関連付けられた1つまたは複数の用語を識別することができる。さらに、システム216は、この電子メールスパムが起源

50

としているネットワークアドレス（例えば、サーバコンピュータ208のネットワークアドレス）を決定することができる。また、システム216は、この電子メールスパムに関連付けられたホスト名を識別することができる。例えば、電子ドキュメント210および/または電子ドキュメント212がこの電子メールスパムからリンクされる場合、システム216は、これらの電子ドキュメントのホスト名をこれらのリンクから抽出することができる。別の実施形態において、システム216は、ネットワーク206に接続されたデータベース216などのメモリ領域にこの電子メールスパムに関連したこの識別された属性を記憶する。

【0043】

サーバコンピュータ202の検索エンジン204は、データベース218にアクセスして、これらの記憶された属性を取得する。これらの記憶された属性に基づいて、検索エンジン204は、特定のネットワークアドレスに配置され、もしくは、特定のホスト名によって提供される1つまたは複数の電子ドキュメントについての格付けを生成する。さらに、検索エンジン204は、データベース218に記憶された用語がネットワーク206上に配置される特定の電子ドキュメント中において出現するかどうかを判定して、この電子ドキュメントについての格付けを生成する。電子ドキュメントの格付けは、この電子ドキュメントが検索エンジンスパムである可能性を示す。次いで、検索エンジン204は、この電子ドキュメントの格付けがしきい値レベルを超えている場合に、この電子ドキュメントを検索エンジンスパムとして分類する。検索エンジン204のクエリプロセッサは、さらにアクションを実施して、（例えば、この検索結果中の電子ドキュメントを格下げし、この検索結果からこの電子ドキュメントを取り除くなどを行って）ユーザに対して正確な検索結果を提供する。

【0044】

図3を参照すると、ブロック図は、本発明の実施形態が検索とともに電子ドキュメントを評価することができる適切なネットワーク環境のさらに他の実施例を示している。クライアントコンピュータ302は、ネットワーク306によってサーバコンピュータ304に接続される。この場合にも、ネットワーク306は、ローカルエリアネットワーク（例えば、イントラネット）、ワイドエリアネットワーク（例えば、インターネット）、または、複数のネットワークの組合せとすることができる。クライアントコンピュータ302は、電子ドキュメントの所在を突き止め、ユーザに対して表示する検索ユーザインターフェイス308（例えば、ブラウザ）または他のマシンアクセス可能プログラミングインターフェイス、あるいはプロトコルを含んでいる。

【0045】

クライアントコンピュータ302のユーザが、1つまたは複数の電子ドキュメントを検索したいと望むときに、このユーザはクエリストリング310を検索ユーザインターフェイス308に対して送信する。ユーザがクエリストリング310を送信した後に、クライアントコンピュータ302は、サーバコンピュータ304に配置された検索エンジン313のクエリプロセッサ312に対してクエリストリング310を送信して検索を要求する。この送信されたクエリストリング310に基づいて、クエリプロセッサ312は、リモートサーバコンピュータ316が提供する電子ドキュメント314をこの送信されたクエリストリング310の「ヒット」として識別する。リモートサーバコンピュータ316は、同様にネットワーク306に接続されている。次いで、クエリプロセッサ312は、クライアントコンピュータ302の検索ユーザインターフェイス308に対して電子ドキュメント314、または電子ドキュメント314のネットワークロケーションを返す。ユーザが、この返されたネットワークロケーションにアクセスして電子ドキュメント314を取得した後、ユーザは、電子ドキュメント314を検索エンジンスパムまたは非スパムとして識別することができる。次いでユーザは、このユーザの識別を検索エンジン313への入力として送信する。

【0046】

この入力を受信するのに応答して、検索エンジン313は、電子ドキュメント314が

検索エンジンシステムであることについての信頼レベルを設定する。さらに、検索エンジン 313 は電子ドキュメント 314 についての複数の入力を複数のユーザから受信し、これらの入力が互いに矛盾する場合には、検索エンジン 313 は、電子ドキュメント 314 についての信頼レベルを設定しないように決定することができる。他方、これらの入力が互いに一致する場合には、検索エンジン 313 は、電子ドキュメント 314 が検索エンジンシステムを構成することについての信頼レベルを設定することができる。本発明の他の実施形態においては、検索エンジン 313 は、ある規則を実装して 1 つまたは複数の入力を判定することができる。すなわち、ある種の入力は、これらの入力を送信したユーザはより信頼がおけるので、より高い重み付けが行われる。この他の実施形態においては、検索エンジン 313 は、電子ドキュメント 314 を検索エンジンシステムとして報告するユーザのパーセンテージを決定する。大多数のこれらのユーザが、この電子ドキュメント 314 は検索エンジンシステムであると同意する場合には、少数のユーザからの入力は、あまり信頼することができない。すなわち、特定のユーザが電子ドキュメント 314 を検索エンジンシステムとして報告し、他の大多数のユーザがこの特定のユーザに同意する場合には、検索エンジン 313 は、このユーザは信頼がおけると判断することができる。他方、他の大多数のユーザが、この特定のユーザに同意しない場合には、検索エンジン 313 は、このユーザは信頼できないと判断することができる。したがって、検索エンジン 313 は、ユーザ提供入力の信頼性に少なくとも部分的に基づいて、特定の電子ドキュメントについての信頼レベルを決定することができる。

10

20

30

40

50

【0047】

このユーザから提供される情報が電子ドキュメント 314 を検索エンジンシステムとして識別する場合には、検索エンジン 313 は電子ドキュメント 314 を解析して、検索エンジンシステムを特徴づける 1 つまたは複数の属性を検出する。電子ドキュメント 314 がイメージを含んでいる場合には、検索エンジン 313 は、そのイメージ中のフレッシュトーンのレベルを感知することによってこれらの属性を検出する。検索エンジン 313 は、電子ドキュメント 314 の 1 つまたは複数のパターンを識別して、これらのパターンが検索エンジンシステムを特徴づけるパターンに対応するかどうかを決定することになる。例えば、検索エンジン 313 は、電子ドキュメント 314 がエンドユーザのためにではなく検索エンジン 313 のために主として構築されているかどうかを、識別することができる。検索エンジン 313 は、さらに電子ドキュメント 314 が隠しテキストおよび/または隠しリンクを含んでいるかどうかを検出することができ、これらの隠しテキストおよび/または隠しリンクは、しばしば検索エンジンシステムを特徴づけている。検索エンジンシステムを特徴づける他の一部のパターンは、非常に多数の不必要なホスト名、過剰な相互リンク、リンクファームリングなどを含んでいる。

【0048】

この識別されたパターンまたは属性に基づいて、検索エンジン 313 は、電子ドキュメント 314 についての格付けを生成する。電子ドキュメント 314 の格付けは、電子ドキュメント 314 が検索エンジンシステムである可能性を示す。電子ドキュメント 314 の格付けがしきい値レベルを超える場合には、検索エンジン 313 は、電子ドキュメント 314 を検索エンジンシステムとして分類する。クエリプロセッサ 312 は、さらにアクションを実施して、(例えば、その検索結果中の電子ドキュメント 314 を格下げし、その検索結果から電子ドキュメント 314 を取り除くなどして) 正確な検索結果をユーザに対して提供する。

【0049】

望ましくない電子ドキュメントを検出する例示の方法

図 4 は、本発明の一実施形態による、検索に関する電子ドキュメントを評価するための例示的な方法を示している。402 において、電子ドキュメントの第 1 の信頼レベルが決定される。この第 1 の電子ドキュメントは、ユーザからの検索要求に応答して検索エンジンによって検索可能である。この第 1 の信頼レベルは、この検索エンジンの外部のソースが提供する情報に基づいて、この電子ドキュメントが望ましくないものである可能性を示

す。この外部ソースは、1つまたは複数の電子ドキュメントに関するデータを提供する電子メールスパム検出システムを含むこともある。例えば、この外部ソースは、望ましくないものであるという所定の可能性を有するものとしてこの外部ソースが識別する1つまたは複数の電子ドキュメントを表示するホスト名を、提供することができる。また、これらの電子ドキュメントからリンクされた電子ドキュメントについては、この第1の信頼レベルを指定することができる。この外部ソースは、望ましくないものであるという所定の可能性を有する1つまたは複数の電子ドキュメントが配置されたネットワークアドレスを、提供することもできる。この外部ソースは、さらに望ましくないものであるという所定の可能性を有する1つまたは複数の電子ドキュメント中出现する用語を、提供することもできる。この電子ドキュメントについての第1の信頼レベルは、この所定の可能性に基づいて決定される。 10

【0050】

404において、この電子ドキュメントの第2の信頼レベルが決定される。この第2の信頼レベルは、この電子ドキュメントの1つまたは複数の属性に基づいて、この電子ドキュメントがこの検索要求に関して不満足なものである可能性を示している。この電子ドキュメントの望ましくないパターンを特徴づけるこのような属性は、この電子ドキュメントを解析することによって識別される。代わりに、この電子ドキュメントに関連するユーザ提供情報を受信することもできる。このユーザ提供情報は、検索結果中においてこの電子ドキュメントを望ましくないものとして指定する。したがって、次いでこの電子ドキュメントの1つまたは複数の属性を識別して、望ましくないパターンを検出することができる。 20

【0051】

406において、この電子ドキュメントについての格付けが、決定された第1の信頼レベルおよび決定された第2の信頼レベルの関数として生成される。408において、この電子ドキュメントは、この電子ドキュメントの生成された格付けに基づいて、この検索要求に関して不満足なものであるとして指定される。さらに、このユーザからの検索要求に応答して、検索結果をこのユーザに対して提供することもできる。この電子ドキュメントが、不満足なものであるとして指定される場合、この電子ドキュメントは、提供された検索結果から除外することができる。代わりに、この電子ドキュメントを、ユーザに提供される検索結果中で格下げすることもできる。この電子ドキュメントのランキングが検索結果中において所定のランクを超える場合には、この電子ドキュメントのランキングを保持することもできる。 30

【0052】

図5は、本発明の一実施形態による、検索に関連して電子ドキュメントを評価するための別の例示的な方法を示している。502において、電子ドキュメントに関するユーザ提供情報が受信される。この電子ドキュメントは、ユーザからの検索要求に応答して検索エンジンによって検索可能である。このユーザ提供情報は、この電子ドキュメントを望ましくないものであるとして特徴づける。例えば、この受信されたユーザ提供情報は、この電子ドキュメントが、望ましくない電子メール(例えば、電子メールスパムの可能性のある電子メース)に関連していることを指定することができる。代わりに、この受信済されたユーザ提供情報は、この電子ドキュメントが、検索結果において望ましくないものである(例えば、検索エンジンスパムの可能性のあるドキュメントである)ことを指定することができる。504において、この電子ドキュメントについての格付けが、受信されたユーザ提供情報の関数として生成される。例えば、ユーザ提供情報によって望ましくないものであるとして特徴づけられた電子ドキュメントを解析して、この電子ドキュメントの1つまたは複数の属性を識別することができる。次いで、この識別された属性を確率的分類機構に適用して、この電子ドキュメントについての格付けを生成する。識別された属性が望ましいかどうかを認識するため、この確率的分類機構をトレーニングすることができる。また、確率的分類機構は、ナイーブベジアン分類機構、制限依存性ベジアン分類機構、ベジアンネットワーク分類機構、判定ツリー、サポートベクトルマシン、コンテンツ 40 50

マッチング分類機構、最大エントロピー分類機構、および、これらの組合せなどとして実装することができる。

【0053】

さらに、この受信されたユーザ提供情報の信頼性を決定することができる。また、この電子ドキュメントについての格付けをこの決定された信頼性の関数として生成することができる。一実施形態においては、この電子ドキュメントに関する他のユーザ提供情報を受信することができる。また、他のユーザ提供情報がこの受信されたユーザ提供情報に一致しているかどうかを判定することによって、この信頼性を決定することができる。506において、この電子ドキュメントは、この電子ドキュメントの生成された格付けに従って、この検索要求に関して不満足なものであるとして指定される。

10

【0054】

例示的なコンピュータ読取り可能媒体

図6は、本発明の一実施形態による例示的なコンピュータ読取り可能媒体600を示すブロック図である。図6に示すように、コンピュータ読取り可能媒体600は、クエリコンポーネント602、外部コンポーネント604、内部コンポーネント606、および解析コンポーネント608を含んでいる。しかし、コンピュータ読取り可能媒体600は、任意量のコンピュータ読取り可能媒体とすることができ、コンポーネント、および、各コンポーネントに関連付けられた機能の様々な組合せを含むことができることが企図されている。クエリコンポーネント602は、ユーザからの検索要求を受信し、この受信された検索要求に基づいて電子ドキュメントを識別する。外部コンポーネント604は、この電子ドキュメントが望ましくないものであるかどうかを評価するためのデータを提供する。内部コンポーネント606は、この電子ドキュメントの第1の信頼レベルを決定するように構成される。この第1の信頼レベルは、外部コンポーネント604が提供するデータに基づいて、この電子ドキュメントが望ましくないものである可能性を示す。例えば、外部コンポーネント604が提供するデータは、1つまたは複数のホスト名を識別する。各ホスト名は、望ましくないものであるという所定の可能性を有する情報を提供する。内部コンポーネント606は、これら提供された名前の中の1つが提供しているものとして、この電子ドキュメントを識別するように構成される。また、内部コンポーネント606はさらに、この電子ドキュメントを識別することに応答して、この電子ドキュメントについてのこの所定の可能性に基づいた第1の信頼レベルを、これらのホスト名の中の1つが提供しているものとして指定するようにも構成されている。

20

30

【0055】

同様に、外部コンポーネント604が提供するデータは、1つまたは複数のネットワークアドレスを識別することができる。外部コンポーネント604は、これらのネットワークアドレスの中の1つに配置された1つまたは複数の電子ドキュメントを、望ましくないものであるという所定の可能性を有するものとして識別する。内部コンポーネント606は、これらのネットワークアドレスの中の1つに配置されたものとして、これらの電子ドキュメントを識別するように構成される。また、内部コンポーネント606は、この電子ドキュメントを識別するのに応答して、この電子ドキュメントについてのこの所定の可能性に基づいた第1の信頼レベルを、これらのネットワークアドレスの中の1つに配置されたものとして指定するように構成される。

40

【0056】

さらに、外部コンポーネント604が提供するデータは、これらの用語の中の少なくとも1つが出現する1つまたは複数の電子ドキュメントが、望ましくないものであるという所定の可能性を有するように、1つまたは複数の用語を識別することができる。内部コンポーネント606は、いつこれらの用語の中の少なくとも1つがこの電子ドキュメント中に出現するかを決定するように、構成される。内部コンポーネント606はまた、これらの用語の中の少なくとも1つがこの電子ドキュメント中に出現することを決定するのに応答して、この電子ドキュメントについての第1の信頼レベルを指定するようにも構成される。この第1の信頼レベルは、この所定の可能性に基づいている。

50

【0057】

内部コンポーネント606はまた、この電子ドキュメントの第2の信頼レベルも設定する。この第2の信頼レベルは、この電子ドキュメントの1つまたは複数の属性に基づいて、この電子ドキュメントは検索に関連して不満足なものである可能性を示す。このような属性は、この検索に関するこの電子ドキュメントの望ましくないパターンを特徴づける。

【0058】

解析コンポーネント608は、決定された第1の信頼レベルおよび設定された第2の信頼レベルの関数として、この電子ドキュメントについての格付けを生成する。クエリコンポーネント602は、この電子ドキュメントの生成された格付けに基づいて、この受信された検索結果に関連してこの電子ドキュメントを不満足なものであるとして分類するように構成される。クエリコンポーネント602はまた、この受信された検索要求に回答して検索結果をそのユーザに対して提供する。また、クエリコンポーネント602は、この提供された検索結果中において不満足なものであるとして分類された電子ドキュメントを格下げし、または、不満足なものであるとして分類された電子ドキュメントをこの提供される検索結果から除外することができる。代わりに、この電子ドキュメントのランキングが提供された検索結果中の所定のランクを越えるときには、クエリコンポーネント602は、提供された検索結果中のこの電子ドキュメントのランキングを保持することもできる。

【0059】

図7は、本発明の一実施形態による例示的な別のコンピュータ読取り可能媒体700を示すブロック図である。図6に示すように、コンピュータ読取り可能媒体700は、インターフェースコンポーネント702、解析コンポーネント704、および、クエリコンポーネント706を含んでいる。しかし、コンピュータ読取り可能媒体700は、任意量のコンピュータ読取り可能媒体とすることができ、コンポーネント、および各コンポーネントに関連付けられた機能の様々な組合せを含むことができることが企図されている。インターフェースコンポーネント702は、電子ドキュメントに関するユーザ提供情報を受信する。この電子ドキュメントは、ユーザからの検索要求に回答して検索可能である。ユーザ提供情報は、この電子ドキュメントを望ましくないものであるとして特徴づける。例えば、この受信されたユーザ提供情報は、この電子ドキュメントが望ましくない電子メールのソースに関連付けられることを指定することができる。この受信されたユーザ提供情報は、この電子ドキュメントが検索結果中で望ましくないものであることを指定することもできる。

【0060】

解析コンポーネント704は、この受信されたユーザ提供情報の関数としてこの電子ドキュメントについての格付けを生成する。一実施形態においては、解析コンポーネント704は、この電子ドキュメントを解析してこの電子ドキュメントの1つまたは複数の属性を識別する。解析コンポーネント704はさらに、この識別された属性が望ましくないかどうかを認識するためにトレーニングされる確率的分類機構に対してこの識別された属性を適用して、この電子ドキュメントについての格付けを生成する。別の実施形態においては、解析コンポーネント704は、この受信されたユーザ提供情報の信頼性を決定し、この電子ドキュメントについての格付けをこの決定された信頼性の関数として生成する。例えば、インターフェースコンポーネント702は、この電子ドキュメントに関する他のユーザ提供情報を受信することができる。次いで、解析コンポーネント704は、他のユーザ提供情報が受信されたユーザ提供情報に対応するかどうか検査して、受信された電子ドキュメントの信頼性を決定する。解析コンポーネント704がこの電子ドキュメントについての格付けを生成した後に、クエリコンポーネント706は、この電子ドキュメントの生成された格付けに従って、この検索要求に関してこの電子ドキュメントを不満足なものであるとして分類する。

【0061】

例示の動作環境

図8は、コンピュータ130の形態における汎用コンピューティングデバイスの一実施

10

20

30

40

50

例を示している。本発明の一実施形態においては、コンピュータ130などのコンピュータは、本明細書中において例示し説明している他の図において使用するために適している。コンピュータ130は、1つまたは複数のプロセッサまたは処理装置132と、システムメモリ134を有している。この例示的な実施形態において、システムバス136は、システムメモリ134を含めて様々なシステムコンポーネントをプロセッサ132に結合している。バス136は、メモリバスまたはメモリコントローラ、ペリフェラルバス、アクセラレーテッドグラフィックスポート、および様々なバスアーキテクチャのうちのいずれかを使用したプロセッサバスまたはローカルバスを含めて、いくつかのタイプのバス構造のうちの1つまたは複数の任意のバス構造を表す。実施例として、限定するものではないが、かかるアーキテクチャには、ISAバス、MCAバス、EISAバス、VESALローカルバス、およびメザニンバスとしても知られているPCIバスが含まれる。

【0062】

コンピュータ130は、一般的に少なくとも一部の形態のコンピュータ読取り可能媒体を有する。コンピュータ読取り可能媒体は、コンピュータ130がアクセスすることができる使用可能な任意の媒体とすることができ、揮発性媒体も不揮発性媒体も、着脱可能媒体も着脱不能媒体も共に含んでいる。実施例として限定するものではないが、コンピュータ読取り可能媒体は、コンピュータストレージ媒体および通信媒体を含んでいる。コンピュータストレージ媒体は、コンピュータ読取り可能命令、データ構造、プログラムモジュール、他のデータなどの情報のストレージのための任意の方法または技術で実装される揮発性媒体および不揮発性媒体、着脱可能媒体および着脱不能媒体を含んでいる。例えば、コンピュータストレージ媒体には、RAM、ROM、EEPROM、フラッシュメモリまたは他のメモリ技術、CD-ROM、DVD(digital versatile disk デジタル多用途ディスク)または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他の磁気ストレージデバイス、あるいは所望の情報を記憶するために使用することができ、コンピュータ130がアクセスすることができる他の任意の媒体が含まれる。通信媒体は、一般的に搬送波や他の搬送メカニズムなどの変調データ信号の形のコンピュータ読取り可能命令、データ構造、プログラムモジュールまたは他のデータを実施し、任意の情報配信媒体を含んでいる。信号中の情報を符号化するようにして、その信号中の1つまたは複数の特性を設定または変更された被変調データ信号については、当業者ならよく理解されよう。有線ネットワークや直接配線接続などの有線媒体と、音響、RF、赤外線、他の無線媒体などの無線媒体は、通信媒体の実施例である。以上のうちの任意の組合せもまた、コンピュータ読取り可能媒体の範囲内に含まれる。

【0063】

システムメモリ134は、着脱可能および/または着脱不能な、揮発性および/または不揮発性のメモリの形態のコンピュータストレージ媒体を含んでいる。例示の実施形態においては、システムメモリ134は、ROM138およびRAM140を含んでいる。起動中などコンピュータ130内のエレメント間で情報を転送する助けをする基本ルーチンを含むBIOS(basic input/output system 基本入出力システム)142は、一般的にROM138に記憶される。RAM140は、処理装置132にとって直ちにアクセス可能な、もしくは、処理装置132によって現在動作させられている、または、その両方が行われるデータおよび/またはプログラムモジュールを一般的に含んでいる。実施例として、限定するものではないが、図8は、オペレーティングシステム144、アプリケーションプログラム146、他のプログラムモジュール148、およびプログラムデータ150を示している。

【0064】

コンピュータ130は、他の着脱可能/着脱不能な、揮発性/不揮発性のコンピュータストレージ媒体を含むこともできる。例えば、図8は、着脱不能な不揮発性磁気媒体から情報を読み取りまたはそれに情報を書き込むハードディスクドライブ154を示している。図8はまた、着脱可能な不揮発性磁気ディスク158から情報を読み取りまたはそれに

情報を書き込む磁気ディスクドライブ156、およびCD-ROMや他の光媒体など着脱可能な不揮発性の光ディスク162から情報を読み取りまたはそれに情報を書き込む光ディスクドライブ160を示している。この例示の動作環境中において使用することができる他の着脱可能/着脱不能な、揮発性/不揮発性のコンピュータストレージ媒体には、それだけには限定されないが、磁気テープカセット、フラッシュメモリカード、デジタル多用途ディスク、デジタルビデオテープ、ソリッドステートRAM、ソリッドステートROMなどが含まれる。ハードディスクドライブ154、ならびに磁気ディスクドライブ156および光ディスクドライブ160は、一般的にインターフェース166などの不揮発性メモリインターフェースによってシステムバス136に接続される。

【0065】

前述したように、図8に示されるこれらのドライブまたは他のマスストレージデバイスおよびその関連するコンピュータストレージ媒体は、コンピュータ130についてのコンピュータ読み取り可能命令、データ構造、プログラムモジュールおよび他のデータのストレージを提供する。図8において、例えばハードディスクドライブ154は、オペレーティングシステム170、アプリケーションプログラム172、他のプログラムモジュール174、およびプログラムデータ176を記憶するものとして示されている。これらのコンポーネントは、オペレーティングシステム144、アプリケーションプログラム146、他のプログラムモジュール148、およびプログラムデータ150と同じとすることもでき、また異なるものとすることもできることに留意されたい。オペレーティングシステム170、アプリケーションプログラム172、他のプログラムモジュール174、およびプログラムデータ176には、少なくともこれらが異なるコピーであることを示すために、ここでは異なる番号が付与されている。

【0066】

ユーザは、キーボード180やポインティングデバイス182（例えばマウス、トラックボール、ペン、もしくはタッチパッド）などの入力デバイスまたはユーザインターフェース選択デバイスを介してコンピュータ130にコマンドおよび情報を入力することができる。他の入力デバイス（図示せず）は、マイクロフォン、ジョイスティック、ゲームパッド、サテライトディッシュ、スキャナなどを含むことができる。これらおよび他の入力デバイスは、システムバス136に結合されるユーザ入力インターフェース184を介して処理装置132に接続されるが、これらは、パラレルポート、ゲームポート、USBなど他のインターフェースおよびバス構造によって接続することもできる。モニタ188または他のタイプのディスプレイデバイスもまた、ビデオインターフェース190などのインターフェースを介してシステムバス136に接続される。モニタ188に追加して、コンピュータは、しばしばプリンタやスピーカなど他のペリフェラル出力デバイス（図示せず）を含んでおり、これらは、出力ペリフェラルインターフェース（図示せず）を介して接続することができる。

【0067】

コンピュータ130は、リモートコンピュータ194など1つまたは複数のリモートコンピュータに対する論理接続を使用して、ネットワーク環境中で動作することができる。リモートコンピュータ194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、ピアデバイス、または他の共通ネットワークノードとすることができ、また一般的にコンピュータ130に関連した前述の要素の多くまたはすべてを含んでいる。図8に示す論理接続は、LAN196およびWAN198を含んでいるが、他のネットワークを含むこともできる。LAN136および/またはWAN138は、有線ネットワーク、無線ネットワーク、これらの組合せなどとすることができる。このようなネットワーキング環境は、オフィス、企業規模のコンピュータネットワーク、イントラネット、およびグローバルコンピュータネットワーク（例えば、インターネット）においては、一般的なものである。

【0068】

ローカルエリアネットワーキング環境中で使用する際には、コンピュータ130は、ネ

10

20

30

40

50

ットワークインターフェースまたはアダプタ186を介してLAN196に接続される。ワイドエリアネットワーキング環境中で使用する際には、コンピュータ130は、一般的にインターネットなどのWAN198上で通信を確立するためのモデム178または他の手段を含んでいる。モデム178は、内蔵でも外付けでもよいが、ユーザ入力インターフェース184または他の適切なメカニズムを介してシステムバス136に接続される。ネットワーク環境においては、コンピュータ130に関連して示されるプログラムモジュール、またはその一部分は、リモートメモリストレージデバイス(図示せず)に記憶することができる。実施例として、限定するものではないが、図8は、リモートアプリケーションプログラム192をこのメモリデバイス上に存在するものとして示している。図に示すこのネットワーク接続は、例示的なものであり、コンピュータ間で通信リンクを確立する他の手段を使用することもできる。

10

【0069】

一般に、コンピュータ130のデータプロセッサは、このコンピュータの様々なコンピュータ読取り可能ストレージ媒体に異なる時刻において記憶される命令の手段によってプログラムされる。プログラムおよびオペレーティングシステムは、一般的に例えばフロッピー(登録商標)ディスクまたはCD-ROM上に記憶されて配布される。そこから、これらは、コンピュータの二次メモリ中にインストールされ、またはロードされる。実行に際して、これらは、少なくとも部分的にこのコンピュータの一次電子メモリへとロードされる。このような媒体が、マイクロプロセッサまたは他のデータプロセッサに関連して以下で説明しているステップを実装するための命令またはプログラムを含んでいるときに、本明細書中で説明している本発明の実施形態は、これらおよび他の様々なタイプのコンピュータ読取り可能ストレージ媒体を含んでいる。本明細書中で説明している方法および技法に従ってプログラムされる際に、本発明の一実施形態はまた、コンピュータそれ自体も含んでいる。

20

【0070】

図8では、このオペレーティングシステムなどプログラムおよび他の実行可能なプログラムコンポーネントは、本明細書中において個別のブロックとして示されている。しかし、かかるプログラムおよびコンポーネントは、様々な時刻にこのコンピュータの異なるストレージコンポーネント中に存在し、このコンピュータの1つ(または複数)のデータプロセッサによって実行されることが理解されよう。

30

【0071】

コンピュータ130を含む例示的なコンピューティングシステム環境に関して説明しているが、本発明の一実施形態は、他の非常に多数の汎用または専用のコンピューティングシステム環境またはコンピューティングシステム構成と共に動作可能である。このコンピューティングシステム環境は、本発明の実施形態の使用または機能の範囲についてどのような限定を示唆することも意図してはいない。さらに、このコンピューティングシステム環境は、この例示的な動作環境中に示されるコンポーネントのどの1つもしくは組合せに関連したどのような依存性または必要要件を有するものとも解釈すべきではない。本発明の実施形態と共に使用するために適切とすることができるよく知られているコンピューティングシステム、コンピューティング環境、および/またはコンピューティングコンフィギュレーションの実施例には、それだけには限定されないが、パーソナルコンピュータ、サーバコンピュータ、ハンドヘルドデバイスまたはラップトップデバイス、マルチプロセッサシステム、マイクロプロセッサベースのシステム、セットトップボックス、プログラマブルな一般消費電子製品、携帯電話、ネットワークPC、ミニコンピュータ、メインフレームコンピュータ、以上のシステムもしくはデバイスのうちのいずれかを含んでいる分散コンピューティング環境などが含まれる。

40

【0072】

本発明の実施形態は、1台もしくは複数台のコンピュータまたは他のデバイスによって実行されるプログラムモジュールなどのコンピュータ実行可能命令の一般的な文脈で説明することができる。一般に、プログラムモジュールには、それだけには限定されないが、

50

特定のタスクを実施し、または特定の抽象データ型を実装するルーチン、プログラム、オブジェクト、コンポーネント、およびデータ構造が含まれる。本発明の実施形態は、タスクが、通信ネットワークを介してリンクされるリモート処理デバイスによって実施される分散コンピューティング環境中においても実行することができる。分散コンピューティング環境においては、プログラムモジュールは、メモリストレージデバイスを含めて、ローカルコンピュータストレージ媒体上にもリモートコンピュータストレージ媒体上にも配置することができる。

【0073】

動作中には、コンピュータ130は、本明細書中で説明している命令などのコンピュータ実行可能命令を実行して、検索に関連して電子ドキュメントを評価する。コンピュータ実行可能命令は、電子ドキュメントの第1の信頼レベルを決定するように構成される。この電子ドキュメントは、ユーザからの検索要求に応答して検索エンジンによって検索可能である。第1の信頼レベルは、この検索エンジンの外部にあるソースが提供する情報に基づいて、この電子ドキュメントが望ましくない可能性を示す。コンピュータ実行可能命令はまた、この電子ドキュメントの第2の信頼レベルを決定するようにも構成される。この第2の信頼レベルは、この電子ドキュメントの1つまたは複数の属性に基づいてこの検索要求に関してこの電子ドキュメントが不満足なものである可能性を示す。コンピュータ実行可能命令は、さらにこの決定された第1の信頼レベルおよびこの決定された第2の信頼レベルの関数として、この電子ドキュメントについての格付けを生成するように構成される。コンピュータ実行可能命令はまた、この電子ドキュメントのこの生成された格付けに基づいて、この検索要求に関してこの電子ドキュメントを不満足なものであるものとして指定するようにも構成される。

10

20

【0074】

コンピュータ130はまた、本明細書中で説明している命令などコンピュータ実行可能命令を実行して、検索に関連して電子ドキュメントを評価する。コンピュータ実行可能命令は、電子ドキュメントに関するユーザ提供情報を受信するように構成される。この電子ドキュメントは、ユーザからの検索要求に応答して検索エンジンによって検索可能である。このユーザ提供情報は、この電子ドキュメントを望ましくないものであるとして特徴づける。コンピュータ実行可能命令はまた、この受信されたユーザ提供情報の関数として、この電子ドキュメントについての格付けを生成するようにも構成される。コンピュータ実行可能命令は、さらにこの電子ドキュメントの生成された格付けに従って、この検索要求に関してこの電子ドキュメントを不満足なものであるとして指定するようにも構成される。

30

【0075】

本明細書中で示し、説明している方法の実行または実施の順序は、他に特に指定していない限り本質的なものではない。すなわち、これらの方法の要素は、他に特に指定していない限りどのような順序で実施することもでき、またこれらの方法は、本明細書中に開示されているよりも多い要素または少ない要素を含むこともできることを本発明者らは企図している。

【0076】

本発明または本発明の実施形態の要素を導入するときに、冠詞「1つの(a)」、「1つの(an)」、「その(the)」、および「前記(said)」は、1つまたは複数の要素が存在することを意味するように意図している。用語「含む(comprising)」、「含む(including)」、および「有する(having)」は、包含することを意図しており、これらのリストアップされた要素以外に追加の要素が存在し得ることを意味している。

40

【0077】

以上を考慮すれば、本発明のいくつかの目的が達成され、他の有利な結果が実現されることが理解されよう。

【0078】

50

本発明の実施形態の範囲を逸脱することなく、以上の構成および方法において様々な変更を行うことができる。以上の説明に含まれ、添付図面に示されるすべての事項は、例示的であり、限定的な意味ではないことが意図されている。

【図面の簡単な説明】

【0079】

【図1】本発明の実施形態を利用することができる例示のネットワーク環境を示すブロック図である。

【図2】本発明の実施形態を利用することができる他の例示のネットワーク環境を示すブロック図である。

【図3】本発明の実施形態を利用することができるさらに他の例示のネットワーク環境を示すブロック図である。 10

【図4】検索に関連して電子ドキュメントを評価するための、本発明の一実施形態によるプロセスフローを示す流れ図の一例である。

【図5】検索に関連して電子ドキュメントを評価するための、本発明の一実施形態によるプロセスフローを示す流れ図の一例である。

【図6】本発明の一実施形態によるコンピュータ読取り可能媒体の一例を示すブロック図である。

【図7】本発明の一実施形態によるコンピュータ読取り可能媒体の他の一例を示すブロック図である。

【図8】本発明の一実施形態を実装することができる適切なコンピューティングシステム環境の実施形態の一例を示すブロック図である。 20

【符号の説明】

【0080】

102、110、112、202、208、302、304、316 サーバ

130 コンピュータ

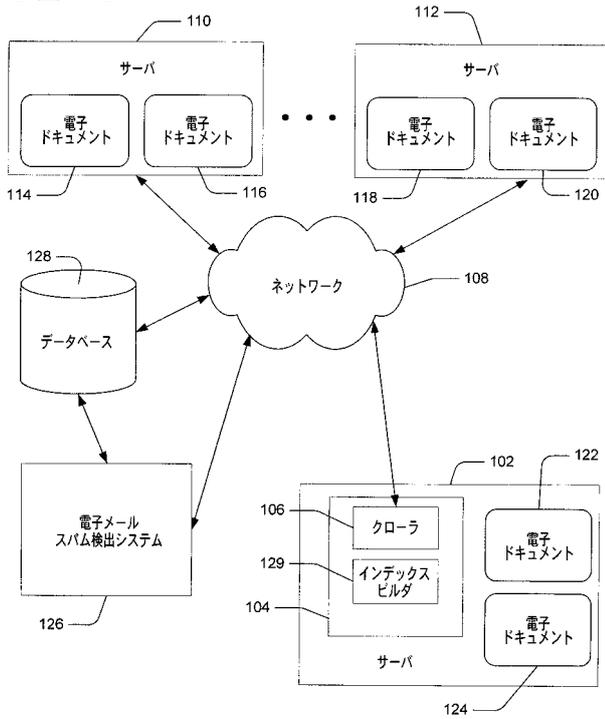
154 ハードディスクドライブ

156 磁気ディスクドライブ

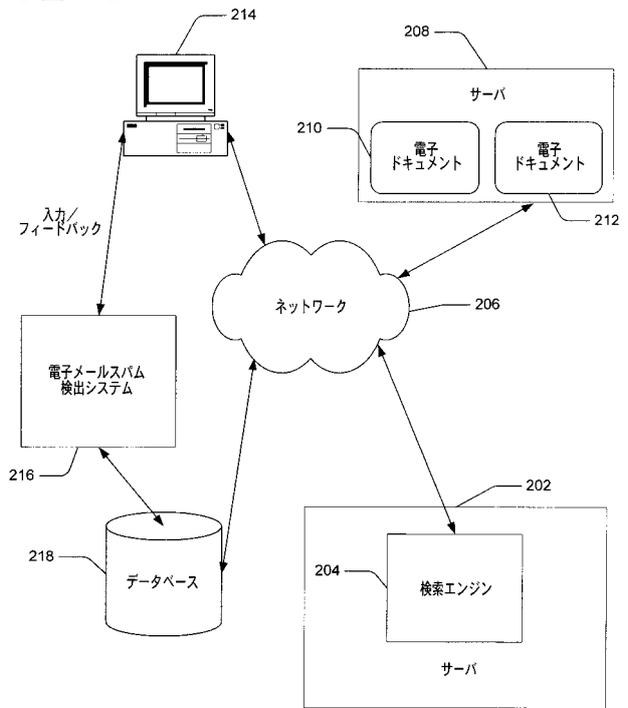
160 光ディスクドライブ

214 ユーザ

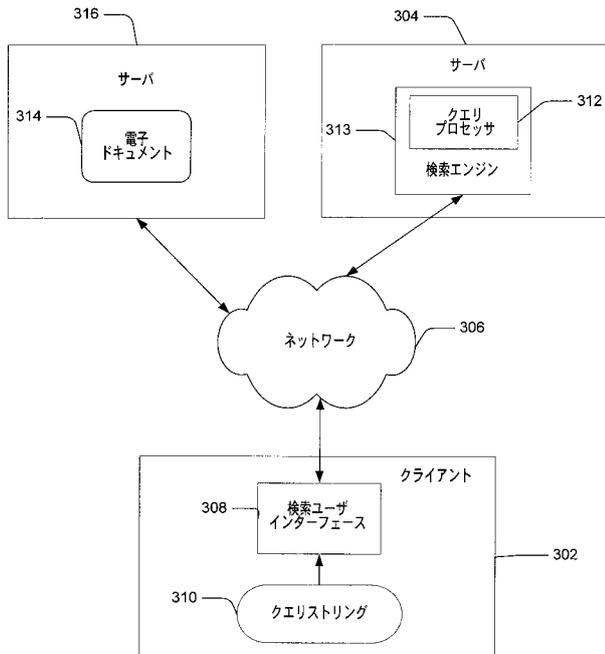
【 図 1 】



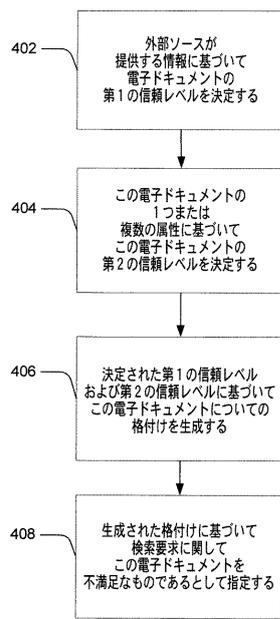
【 図 2 】



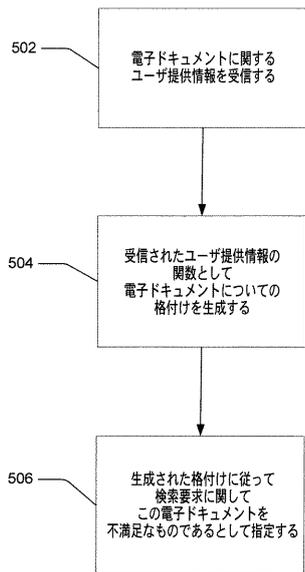
【 図 3 】



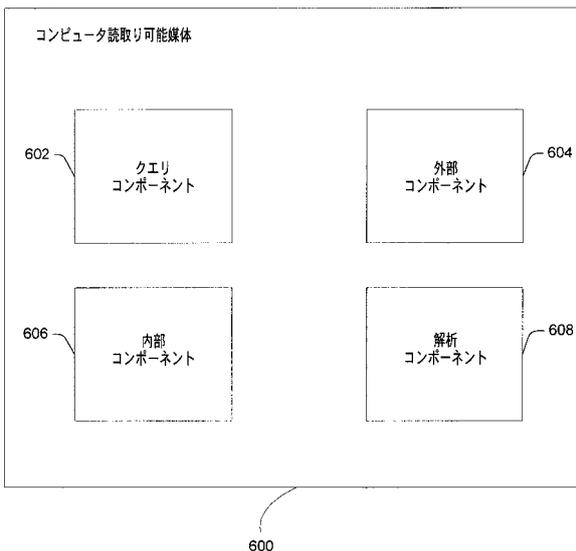
【 図 4 】



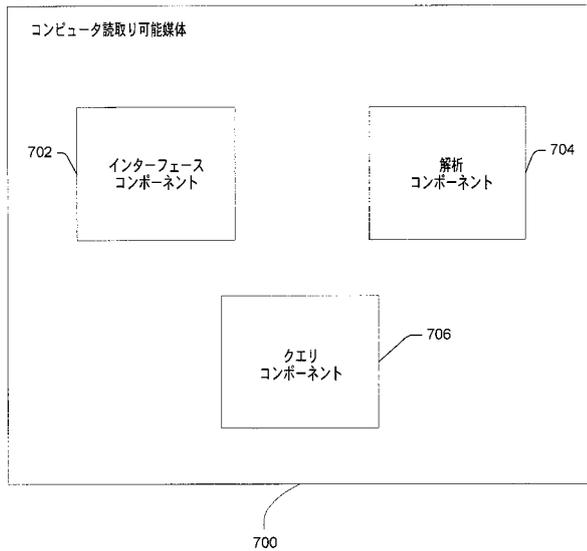
【 図 5 】



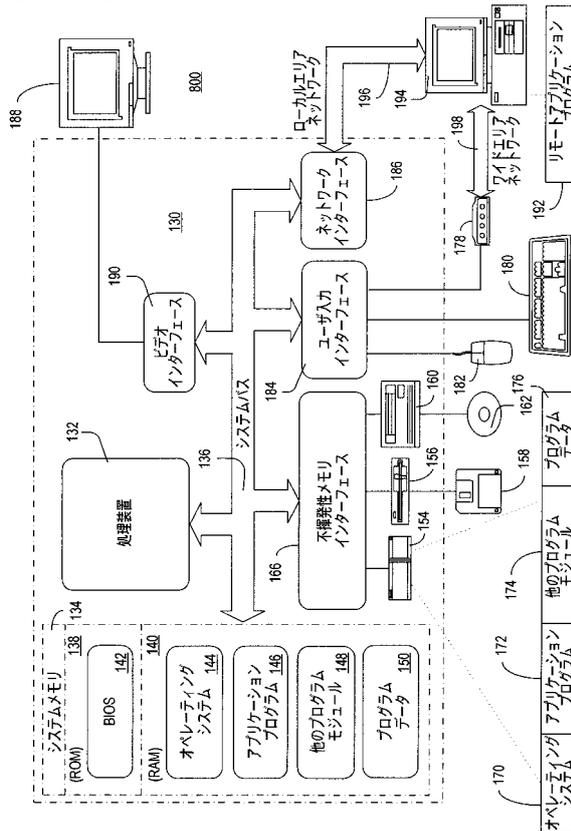
【 図 6 】



【 図 7 】



【 図 8 】



フロントページの続き

(72)発明者 エリック ビー . ワトソン

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 コーポレーション内

(72)発明者 ジャニーン ラス クラム

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 コーポレーション内

Fターム(参考) 5B075 KK02 KK07 ND03 ND40 NK02 NK32 NR12 NS10 PQ02

【外国語明細書】

2005339545000001.pdf