(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第4828411号
(P4828411)

(45)発行日　平成23年11月30日(2011.11.30)　(24)登録日　平成23年9月22日(2011.9.22)

(51)Int.Cl.		F I		
G06F 13/00	(2006.01)	G06F	13/00	610Q
H04L 12/58	(2006.01)	G06F	13/00	540E
H04L 12/66	(2006.01)	H04L	12/58	100Z
		H04L	12/66	B

請求項の数 30　(全 32 頁)

(21)出願番号	特願2006-508818 (P2006-508818)
(86)(22)出願日	平成16年2月25日 (2004.2.25)
(65)公表番号	特表2006-521635 (P2006-521635A)
(43)公表日	平成18年9月21日 (2006.9.21)
(86)国際出願番号	PCT/US2004/005501
(87)国際公開番号	WO2004/079514
(87)国際公開日	平成16年9月16日 (2004.9.16)
審査請求日	平成19年2月16日 (2007.2.16)
審判番号	不服2010-19005 (P2010-19005/J1)
審判請求日	平成22年8月23日 (2010.8.23)
(31)優先権主張番号	10/378,463
(32)優先日	平成15年3月3日 (2003.3.3)
(33)優先権主張国	米国(US)

(73)特許権者　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
(74)代理人　100077481
弁理士　谷　義一
(74)代理人　100088915
弁理士　阿部　和夫
(74)復代理人　100115624
弁理士　濱中　淳宏
(74)復代理人　100136490
弁理士　中西　英一

最終頁に続く

(54)【発明の名称】スパム防止のためのフィードバックループ

(57)【特許請求の範囲】

【請求項１】

スパム防止に関連して、電子メール（ｅメール）およびメッセージの内の少なくとも１つから成るアイテムを分類することを容易にするシステムであって、

１組の前記アイテムを受け取る構成要素と、

前記アイテムの対象とする受信者を識別し、およびポーリングされる前記アイテムの部分集合にタグ付けをする構成要素であって、前記アイテムの部分集合は、スパムファイティングユーザと知られかつランダムに選択された、前記受信者の部分集合に対応し、ポーリングされる前記アイテムの部分集合は、前記アイテムがスパムまたは非スパムと分類される前に決定され、現在使用されているスパムフィルタによってスパムと指定されるアイテムを含んだ全てのアイテムに対してポーリングが考慮されるタグ付けをする構成要素と、

前記ポーリングされたアイテムの前記スパムファイティングユーザの分類に関する情報を受信し、スパムフィルタをトレーニングすることおよびスパムリストをポピュレートすることに関連して前記情報を使用する、機械学習法を採用したフィードバック構成要素と、

ポーリングアイテムとして識別するため、ポーリングのためにタグ付けされたアイテムを修正する構成要素であって、前記修正されたアイテムは、投票指示ならびに少なくとも２つの投票ボタンおよび前記ユーザによって前記アイテムの分類を容易にする少なくとも２つのアイテムクラスにそれぞれ対応するリンクの内のいずれか１つを含み、前記投票ボ

タンは前記それぞれのリンクに対応しており、ユーザによって前記投票ボタンのいずれか一方が選択されるとき、前記投票ボタン、前記それぞれのユーザおよび割り当てられた固有のＩＤに関する情報が記憶データベースに送られる、修正する構成要素と
を備えたことを特徴とするシステム。

【請求項２】
前記１組の前記アイテムを受信する構成要素は、電子メールサーバ、メッセージサーバおよびクライアント電子メールソフトウェアのうちいずれか１つであることを特徴とする請求項１に記載のシステム。

【請求項３】
前記ポーリングされる前記アイテムの部分集合は、すべての受信された前記アイテム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４】
前記受信者の部分集合は、すべての受信者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５】
前記ポーリング用にタグ付けされた前記アイテムの部分集合は、
ユーザ当たりに選択された前記アイテムの数と、
ある時間当たりおよびユーザ当たりに選択される前記アイテムの数と、
既知のユーザに対応するアイテムにタグ付けする確率と
のうち少なくとも１つに基づいて限定さ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６】
前記ポーリングアイテムは、前記タグ付けされたアイテムの要約を含んでおり、前記要約は、件名、日付、メッセージのテキストおよびテキストの最初の数行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７】
前記データベースは、ユーザプロパティ、タグ付けされたアイテムに関連付けられたアイテムコンテンツおよびプロパティ、ユーザ分類および投票統計データ、ユーザ当たりのポーリングおよびある時間当たりおよびユーザ当たりのポーリングの頻度分析データ、スパムリスト、正当なメールリスト、並びに、ブラックホールリストに関する情報並びにデータを記憶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８】
複数のスパムファイティング集団全体にわたって分布され、その結果、各集団からのフィードバックは、各集団と動作可能にインターフェースされた中央データベースに送信され、前記フィードバックの何らかの部分は、プライバシーの理由によって除去さ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９】
ユーザ信頼度および信頼性をテストするユーザ分類妥当性の検査構成要素をさらに備え、前記ユーザ分類妥当性の検査構成要素は、１つまたは複数の疑わしいユーザに適用できることを特徴とする請求項１に記載のシステム。

【請求項１０】
前記フィードバック構成要素は、ユーザフィードバック、ハニーポットフィードバックおよび、任意選択で、受信されたアイテムのユーザ受信者フィードバックに関する情報を受信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１１】
前記アイテムの対象とする受信者を識別し、およびポーリングされる前記アイテムの部分集合にタグ付けをする構成要素は、それぞれのユーザがその元の形態による前記メッセージの第１のコピーおよびポーリングのための形態による前記メッセージの第２のコピーを受信するように、元々受信された各タグ付けされたメッセージのコピーを作成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１２】
電子メールサーバ、メッセージサーバおよびクライアント電子メールソフトウエアを実行するクライアントコンピュータを含む、着信メッセージを受信するコンピュータによって実行され、スパム防止に関連してメッセージを分類することを容易にする方法であって、前記コンピュータが、
１組の前記メッセージを受け取るステップと、
前記メッセージの対象の受信者を識別するステップと、
既知のスパムファイティングユーザと知られかつランダムに選択された前記受信者の部分集合に対応する、ポーリングされる前記メッセージの部分集合にタグ付けをするステップであって、ポーリングされる前記メッセージの部分集合は、前記メッセージがスパムまたは非スパムと分類される前に決定され、現在使用されているスパムフィルタによってスパムと指定されるメッセージを含んだ全てのメッセージに対してポーリングが考慮されるステップと、
前記ポーリングメッセージの前記スパムファイティングユーザの分類に関する情報を受信するステップと、
スパムフィルタをトレーニングすることおよびスパムリストをポピュレートすることに関連して、前記情報を使用するステップであって、前記スパムフィルタをトレーニングすることは機械学習法を経由して使用されるステップと、
ポーリングメッセージとして識別するため、ポーリングのためにタグ付けされたメッセージを修正するステップであって、前記修正されたメッセージは、投票指示ならびに少なくとも２つの投票ボタンおよび前記ユーザによって前記メッセージの分類を容易にする少なくとも２つのメッセージクラスにそれぞれ対応するリンクの内のいずれか１つを含み、前記投票ボタンは前記それぞれのリンクに対応しており、ユーザによって前記投票ボタンのいずれか一方が選択されるとき、前記投票ボタン、前記それぞれのユーザおよび割り当てられた固有のＩＤに関する情報が、前記コンピュータ内にまたは外部に備えられた記憶データベースに送られるステップと
を備えることを特徴とする方法。
【請求項１３】
既知のスパムファイティングユーザである前記受信者の部分集合は、
新しいスパムフィルタをトレーニングするのを容易にするように、メッセージに対してフィードバックを提供するためにオプトインすることと、
オプトアウトしないことによって、メッセージに対してフィードバックを提供するために消極的にオプトインすることと、
参加しているメッセージサーバによって提供される電子メール／メッセージサービスに対して料金を支払うことと、
参加しているメッセージサーバで電子メールアカウントを開くことと
のうち少なくとも１つを実施する各受信者によって決定されることを特徴とする請求項１２に記載の方法。
【請求項１４】
タグ付けされたメッセージを修正する前記ステップは、
前記タグ付けされたメッセージを、ポーリングメッセージ用の別個のフォルダに移動することと、
前記タグ付けされたメッセージの「ｆｒｏｍ」アドレスを修正することと、
前記タグ付けされたメッセージの件名行を修正することと、
前記タグ付けされたメッセージ上のポーリングアイコンを使用し、それをポーリングメッセージとして識別することと、
固有の色を使用し、前記タグ付けされたメッセージをポーリングメッセージとして識別することと
のうち少なくとも１つを実施することを含むことを特徴とする請求項１２に記載の方法。

【請求項１５】
それぞれのユーザがその元の形態による前記メッセージの第１のコピーおよびポーリングのための形態による前記メッセージの第２のコピーを受信するように、元々受信された各タグ付けされたメッセージのコピーを作成するステップをさらに備えることを特徴とする請求項１２に記載の方法。
【請求項１６】
前記トレーニングされたスパムフィルタを１つまたは複数のサーバに配布するステップをさらに備え、前記配布は、自動的に、および／または、電子メールメッセージおよびダウンロードするためのウェブサイト上のポスティングのうちの少なくとも１つによる要求によって行われることを特徴とする請求項１２に記載の方法。
【請求項１７】
前記スパムフィルタをトレーニングすることおよび前記スパムリストをポピュレートすることは、ユーザ分類フィードバックに基づくデータ、ならびに、任意選択で１つまたは複数の追加ソースによって生成されたデータを使用して機械学習法によって実施され、前記１つまたは複数のソースは、ハニーポット、受信者非ユーザ分類フィードバックおよび能動学習法を含むことを特徴とする請求項１２に記載の方法。
【請求項１８】
偏らないデータのサンプリングを得るのを容易にするために、データの前記１つまたは複数のソースによって生成されたデータは、前記ソースによって生成されたデータのタイプに対しておよび前記ユーザ分類データに対して比例して再重み付けされることを特徴とする請求項１７に記載の方法。
【請求項１９】
着信メッセージのそれぞれの１つまたは複数の肯定的な特徴について前記着信メッセージを監視するステップと、
受信された肯定的な特徴の頻度を決定するステップと、
受信された１つまたは複数の肯定的な特徴が、少なくとも一部には履歴データに基づいて閾値頻度を超えているかどうか判定するステップと、
疑わしいメッセージがスパムであるかどうか判定するために他の分類データが使用可能になるまで、前記閾値頻度を超える前記１つまたは複数の肯定的な特徴に対応する疑わしいメッセージを隔離するステップと
をさらに備えることを特徴とする請求項１２に記載の方法。
【請求項２０】
使用された前記特徴は、前記送信側のＩＰアドレスおよびドメインのうち少なくとも１つを含む前記送信側についての情報であることを特徴とする請求項１９に記載の方法。
【請求項２１】
疑わしいメッセージを隔離する前記ステップは、
暫定的にスパムとして前記疑わしいメッセージに標識を付け、それらをスパムフォルダに移動する動作と、
他の分類データが使用可能になるまで、前記疑わしいメッセージを前記ユーザに送達することを遅らせる動作と、
前記疑わしいメッセージを、前記ユーザには見えないフォルダに保存する動作と
のうち少なくとも１つによって実施されることを特徴とする請求項１９に記載の方法。
【請求項２２】
前記スパムフィルタの最適化を容易にするために、前記スパムフィルタのフォールスポジティブ／捕捉率を決定するステップをさらに備え、前記フォールスポジティブ／捕捉率を決定するステップは、
ポーリング結果の第１の集合を含むトレーニングデータ集合を使用して、前記スパムフィルタをトレーニングすることと、
ユーザフィードバックを使用してポーリングメッセージの第２の集合を分類し、ポーリング結果の第２の集合を生み出すことと、

前記ポーリングメッセージの第２の集合を、前記トレーニングされたスパムフィルタに通すことと、
前記ポーリング結果の第２の集合を前記トレーニングされたスパムフィルタ結果に比較して、前記フィルタのフォールスポジティブ／捕捉率を決定し、それによって、最適なフィルタ性能に従ってフィルタパラメータを評価および調整することと
を含むことを特徴とする請求項１２に記載の方法。
【請求項２３】
各々が様々なパラメータを有し、各々が前記同じトレーニングデータセット上でトレーニングされる複数のスパムフィルタが構築され、その結果、スパムフィルタリング用の最適なパラメータを決定するために、各スパムフィルタの前記フォールスポジティブ／捕捉率が、少なくとも１つの他のスパムフィルタと比較されることを特徴とする請求項２２に記載の方法。
【請求項２４】
着信メッセージの追加の組を使用して、改善されたスパムフィルタを構築するステップであって、前記着信メッセージの部分集合は、前記改善されたスパムフィルタをトレーニングすることに関連して新しい情報を生み出すためにポーリングを受け、少なくとも一部には先に得られた情報がどれだけ過去に得られたかに基づいて、前記先に得られた情報が再重み付け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１２に記載の方法。
【請求項２５】
正当な送信側リストを構築するために、前記情報を使用するステップをさらに備えることを特徴とする請求項１２に記載の方法。
【請求項２６】
スパム送信者のアカウントを終了させるのを容易にするために、前記情報を使用するステップをさらに備えることを特徴とする請求項１２に記載の方法。
【請求項２７】
ＩＳＰを使用しているスパム送信者を識別するステップ、および、前記スパミングについて前記ＩＳＰに自動的に通知するステップをさらに備えることを特徴とする請求項２６に記載の方法。
【請求項２８】
スパムを送信する責任を負うドメインを識別するステップ、および、前記スパミングについて、前記ドメインの電子メールプロバイダの少なくとも１つに自動的に通知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２６に記載の方法。
【請求項２９】
前記スパムフィルタおよび前記スパムリストのうち少なくとも１つを、メールサーバ、電子メールサーバ、および、クライアント電子メールソフトウェアのうちいずれか１つに配布するステップをさらに備え、前記配布するステップは、
ウェブサイト上で通知を掲示し、前記スパムフィルタおよびスパムリストがダウンロードのために使用可能であることを通知することと、
前記スパムフィルタおよびスパムリストを、メールサーバ、電子メールサーバ、および、クライアント電子メールソフトウェアに自動的に送り出すことと、
前記スパムフィルタおよびスパムリストを、メールサーバ、電子メールサーバ、および、クライアント電子メールソフトウェアに手動で送り出すことと
のうち少なくとも１つを含むことを特徴とする請求項１２に記載の方法。
【請求項３０】
請求項１２乃至２９いずれかに記載の方法をコンピュータに実行させるためのコンピュータ実行命令を記憶したコンピュータ読取り可能記憶媒体。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、正当なメール（たとえば、善良なメール）と望ましくない情報（たとえば、

ジャンクメール）を共に識別するためのシステムおよび方法に関し、より詳細には、スパム防止のために電子メール通信文を分類することに関する。

【背景技術】

【０００２】

インターネットなど地球規模の通信ネットワークの出現により、膨大な数の潜在顧客に到達する商業機会が提供された。電子メッセージング、特に電子メール（ｅメール）は、（「スパム」とも呼ばれる）望ましくない広告およびプロモーションをネットワークユーザにまき散らすための手段として、ますます広まりつつある。

【０００３】

コンサルティング／市場調査会社であるＲａｄｉｃａｔｉ　Ｇｒｏｕｐ，Ｉｎｃ．は、２００２年８月現在、毎日２０億通のジャンク電子メールメッセージが送られており、この数は２年ごとに３倍になることが予想されると見積もっている。個人および団体（たとえば、会社、政府機関）はますます、ジャンクメッセージによって迷惑を受け、しばしば不快な思いをさせられつつある。したがって、ジャンク電子メールは、今まさに、信頼できるコンピューティングに対する大きな脅威になるであろう。

【０００４】

ジャンク電子メールを阻止するために使用される主な技法は、フィルタリングシステム／方法を使用することである。１つの実証済みのフィルタリング技法は、機械学習手法に基づくものであり、機械学習フィルタが、着信メッセージに対し、そのメッセージがジャンクである確率を割り当てる。この手法では、一般に２種類の事例メッセージ（たとえば、ジャンクメッセージと非ジャンクメッセージ）から特徴が抽出され、この２つの種類間を確率的に弁別するために学習フィルタが適用される。多数のメッセージの特徴は内容（たとえば、メッセージの件名および／または本文内の単語および句）に関係するため、そのようなタイプのフィルタは、一般に「コンテンツベースのフィルタ」と呼ばれる。

【０００５】

いくつかのジャンク／スパムフィルタは適応型であり、これは、多言語ユーザや稀少言語を話すユーザが、自分たちの特定のニーズに適応することができるフィルタを必要とする点において重要である。さらに、何がジャンク／スパムであり何がジャンク／スパムでないかについて、すべてのユーザが一致しているわけではない。したがって、（たとえば、ユーザ挙動を観察することを介して）暗黙のうちにトレーニングすることができるフィルタを使用することにより、それぞれのフィルタを動的に調整し、ユーザ特定のメッセージ識別のニーズを満たすことができる。

【発明の開示】

【発明が解決しようとする課題】

【０００６】

フィルタリング適応のための１つの手法は、ジャンクまたは非ジャンクとしてメッセージに標識を付けるようにユーザに要求することである。残念ながら、そのような手動の集中トレーニング技法は、そのようなトレーニングを適正に行うために必要とされる時間の量は言うまでもなく、そのようなトレーニングに関連する複雑さのため、多数のユーザにとって望ましくない。さらに、そのような手動トレーニング技法は、しばしば個々のユーザによって損なわれる。たとえば、無料メーリングリストに対して加入していることは、ユーザによってしばしば忘れさられ、したがって、誤ってジャンクメールとして標識が付けられる。その結果、正当なメールがユーザのメールボックスから無期限に遮断される。別の適応型フィルタトレーニング手法は、暗黙のトレーニングキューを使用することである。たとえば、ユーザがメッセージに回答し、またはそれを転送した場合、この手法では、メッセージが非ジャンクであると仮定される。しかし、この種のメッセージキューだけを使用することは、統計上の偏りをトレーニングプロセス内に導入し、フィルタの各々の精度が低くなる。

【０００７】

さらに別の手法は、全ユーザ電子メールをトレーニングのために使用することであり、

この場合、初期標識は、既存のフィルタによって割り当てられ、ユーザが時々、明示的なキュー（たとえば、「ユーザ訂正（ｕｓｅｒ－ｃｏｒｒｅｃｔｉｏｎ）」メソッド）－たとえば、「ジャンクとして削除する（ｄｅｌｅｔｅ　ａｓ　ｊｕｎｋ）」や「ジャンクでない（ｎｏｔ　ｊｕｎｋ）」などオプションを選択すること－および／または暗黙のキューを用いてこれらの割り当てを無効にする。そのような手法は、先に論じた技法より良いが、依然として、以下で述べられている、また特許請求の範囲に記載された本発明に比べて不十分である。

【課題を解決するための手段】

【０００８】

以下、本発明のいくつかの態様を基本的に理解するために、本発明について簡単にまとめる。この概要は、本発明の広範な全体像ではない。本発明の主な／重大な要素を特定すること、あるいは本発明の範囲を述べることは意図されていない。後から提供されるより詳しい説明の序文として、本発明のいくつかの概念を簡単な形態で示すことを目的とするにすぎない。

【０００９】

本発明は、スパム防止に関連してアイテムを分類することを容易にするフィードバックループシステムおよび方法を提供する。本発明は、スパムフィルタに適用される機械学習手法を利用し、特に、トレーニングデータの集合を生成するために正当なメールとジャンク／スパムメール双方の例が得られるように、着信電子メールメッセージをランダムにサンプリングする。予め選択された個人がスパムファイタ（ｓｐａｍ　ｆｉｇｈｔｅｒｓ）として働き、そのサンプルの（任意選択でわずかに修正することができる）それぞれの複製を類別する際に参加する。

【００１０】

一般に、ポーリング用に選択されたメッセージは、ポーリングメッセージとして見えるように様々な態様で修正される。本発明の独自の態様は、ポーリング用に選択された着信メッセージのコピーが作られ、その結果、何人かのユーザ（たとえば、スパムファイタ）が（たとえば、メッセージの内容の点において）同じメッセージを２回、すなわち１回はポーリングメッセージの形態で、また再度その元の形態で受信することになる。本発明の他の独自の態様は、既存のフィルタによってスパムとして標識が付けられているものを含めて、メッセージすべてがポーリングすることを考慮される。スパムと標識が付けられたメッセージは、ポーリングすることを考慮され、選択された場合、既存のフィルタの指定に従ってスパムとして処理される（たとえば、ジャンクフォルダに移動、削除など）ものでなくなる。

【００１１】

従来のスパムフィルタと異なり、善良なメールとスパムとを区別することを学習し、それにより、偏った不正確なフィルタリングを軽減するように、本発明のフィードバック技法によるトレーニングスパムフィルタによって、より正確なスパムフィルタを作成することができる。フィードバックは、少なくとも一部には、任意の適切な数のユーザにポーリングし、それらの着信電子メールについてフィードバックを得ることによって行われる。スパムファイタとして識別されたユーザは、着信メッセージの選択されたものが正当なメールか、それともジャンクメールかについて投票するというタスクを負う。着信電子メールの肯定的分類と否定的分類は共に、ユーザ向けの善良な（たとえば、スパムでない）ものであるメールをスパムメールとして不適切にフィルタして除去することを軽減するために望ましいものである。それぞれの分類は、各メールトランザクションに関連する任意の他の情報と共に、スパムフィルタのトレーニングを容易にするためにデータベースに移動される。データベースおよび関連構成要素は、機械学習システム用トレーニングデータの集合を生成するために、ユーザプロパティ、ユーザ投票情報／履歴、またはメッセージプロパティ、たとえば各選択されたメッセージに割り当てられた一意の識別番号、メッセージ分類、メッセージ内容の概要、または上記のいずれかに関連する統計データを含む、選択されたメッセージ（または選択されたメールトランザクション）のためのプロパティを

編集および記憶することができる。機械学習システム（たとえば、ニューラルネットワーク、サポートベクタマシン（ＳＶＭ）、ベイジアン信念ネットワーク（Ｂａｙｅｓｉａｎ Ｂｅｌｉｅｆ Ｎｅｔｗｏｒｋｓ））は、正当なメールとスパムメールを共に認識し、さらにそれらを区別するようにトレーニングされる改良型スパムフィルタを作成することを容易にする。新しいスパムフィルタは、本発明に従ってトレーニングされた後に、メールサーバ、およびクライアント電子メールソフトウェアプログラムに配布することができる。さらに、この新しいスパムフィルタは、個別化されたフィルタの性能を改善するために、特定のユーザに関してトレーニングすることができる。新しいトレーニングデータ集合が構築されたとき、スパムフィルタは、その性能と精度を最適化するために、機械学習を介してさらにトレーニングを受けることができる。また、メッセージ分類によるユーザフィードバックを使用し、スパムフィルタ性能をテストするために、かつ／またはスパム発信元を識別するために、スパムフィルタおよび保護者による規制（ｐａｒｅｎａｔａｌ ｃｏｎｔｒｏｌ）のリストを生成することができる。

【００１２】

本発明の他の態様は、交差検定法を介して、かつ／または結果が既知のテストメッセージによって、信頼できないユーザを検出する方法を提供する。交差検定は、何人かのユーザのポーリング結果が除外されたフィルタをトレーニングすることを必要とする。すなわち、このフィルタは、ユーザの部分集合からのポーリング結果を使用してトレーニングされる。概して、このユーザの部分集合は、いくつかの間違いがある場合でさえも、そのユーザの部分集合とほとんどの場合に一致しているないユーザを検出するのに十分良好に機能することになる。除外されたユーザからのポーリング結果は、トレーニング済みフィルタのものと比較される。この比較により、本質的に、トレーニング用部分集合からのユーザが、除外されたユーザに属するメッセージに対してどのように投票したかが判定される。除外されたユーザの投票とフィルタとの一致が低い場合には、そのユーザからのポーリング結果を廃棄する、または手動検査のためにマークをすることができる。この技法は、望むなら、毎回異なるユーザからのデータを除外して繰り返すことができる。

【００１３】

フィルタとユーザ投票がひどく一致しないメッセージなど、個々のメッセージに対する誤りもまた検出することができる。これらのメッセージは、自動除去および／または手動検査のためにフラグ付けをすることができる。交差検定の代替として、すべての、または実質的にすべてのユーザについてフィルタをトレーニングすることができる。フィルタと一致しないユーザ投票および／またはメッセージは、廃棄することができる。交差検定の別の代替は、その結果が既知である場合にユーザがメッセージに対して投票するように依頼される、結果が既知のテストメッセージを必要とする。ユーザによるメッセージの正確な分類（たとえば、ユーザ投票がフィルタアクションに一致する）により、そのユーザの信頼性が検証され、そのユーザの分類をトレーニングから除去するかどうか、また、そのユーザを今後のポーリングから除去するかどうか判定される。

【００１４】

本発明の他の態様は、着信メールをスパムとして識別するために、かつ／または特定の商用電子メールアドレス処理を追跡するために、既知のスパムターゲット（たとえば、ハニーポット）を作成することを可能にする。既知のスパムターゲットまたはハニーポットは、正当なメールの集合を決定することができ、他のメールすべてをスパムと考えることができる電子メールアドレスである。たとえば、人に見つけられそうにない制限的な形で、その電子メールアドレスをウェブサイト上で開示することができる。したがって、このアドレスに送信されるどのメールも、スパムと考えることができる。別法として、その電子メールアドレスは、そこから正当なメールを受信することが予想されるマーチャントだけに開示しておくことができる。したがって、そのマーチャントから受信されたメールは正当なメールであるが、受信された他のメールはすべて、安全にスパムであると考えることができる。ハニーポットおよび／または他のソース（たとえば、ユーザ）から引き出されたスパムデータは、フィードバックループシステム内に一体化することができるが、ハ

ニーポットを用いたスパム分類の実質的な増大により、以下でより詳しく述べるように、そのようなデータを減量し、偏ったポーリング結果を得ることを軽減すべきである。

【００１５】

本発明の他の態様は、フィードバックループシステムによって、またはフィルタによって不確実とみなされるメッセージを隔離することを可能にする。そのようなメッセージは、廃棄または分類されるのではなく、任意の適切な時間の間、保持される。この時間は予め設定することができ、または、たとえば同じＩＰアドレスからの、または類似の内容を有するそのメッセージに似た、所定の数の得票結果を受信するまでそのメッセージを保持することができる。

【００１６】

前述のおよび関連の目的を達成するために、本明細書では、本発明のある種の例示的な態様が、以下の説明および添付の図面と共に述べられている。しかし、これらの態様は、本発明の原理を使用することができる様々な方法のいくつかを示すものにすぎず、本発明は、そのような態様とそれらの均等物すべてを含むものとする。図面と共に考察し、以下の本発明の詳細な説明により、本発明の他の利点および新規の特徴は明らかになるだろう。

【発明を実施するための最良の形態】

【００１７】

次に、本発明について図面を参照しながら述べる。図面では、全体を通して同じ要素を参照するために同じ符号が使用される。以下の説明では、説明する目的で、本発明を十分理解するために、多数の特定の詳細が述べられている。しかし、これらの特定の詳細なしに本発明を実施することができることは自明である。場合によっては、本発明について説明するのを容易にするために、周知の構造およびデバイスがブロック図の形態で示される。

【００１８】

本願では、「構成要素」および「システム」という用語は、ハードウェア、ハードウェアとソフトウェアの組合せ、ソフトウェア、または、実行中のソフトウェアであるコンピュータ関連のエンティティを指すものとする。たとえば、構成要素は、それだけには限らないが、プロセッサ上で動作するプロセス、プロセッサ、オブジェクト、実行可能物、実行のスレッド、プログラム、および／またはコンピュータとすることができる。例示のために、サーバ上で動作するアプリケーションもそのサーバも共に構成要素とすることができる。プロセスおよび／または実行のスレッド内に１つまたは複数の構成要素が常駐することができ、構成要素は、１つのコンピュータ上に存在し、および／または２つ以上のコンピュータ間で分散することができる。

【００１９】

本発明は、機械学習式スパムフィルタリング用のトレーニングデータを生成することに関連して、様々な推論スキームおよび／または技法を組み込むことができる。本明細書では、「推論」という用語は、概して、事象および／またはデータを介して取り込まれた観察結果の集合から、システムの状態、環境、および／またはユーザについて推論し、または推理するプロセスを指す。推論は、たとえば、特定の内容または動作を識別するために使用することができ、あるいは、状態全体にわたって確率分布を生成することができる。推論は、確率的なもの、すなわち、データおよび事象の考察に基づいた当該の状態全体にわたる確率分布の計算とすることができる。推論はまた、事象および／またはデータの集合から、より高いレベルの事象を構成するために使用される技法を指す場合がある。そのような推論により、事象群が時間の近接性で相関されていようとなかろうと、また、事象やデータが１つの、あるいは複数の事象源やデータ源に由来していようと、観察された事象および／または記憶された事象データの集合から新しい事象またはアクションが構築される。

【００２０】

本明細書全体にわたって「メッセージ」という用語が広く使用されるが、そのような用

語は、電子メールそれ自体に限定されず、任意の好適な通信アーキテクチャを介して配布することができるどの形態の電子メッセージングをも含むように適切になすことができることを理解されたい。たとえば、２名以上の人同士の会議を容易にする会議アプリケーション（たとえば、対話型チャットプログラムおよびインスタントメッセージングプログラム）もまた、本明細書に開示されているフィルタリングの利益を利用することができる。というのは、望ましくないテキストは、ユーザがメッセージを交換しているとき、通常のチャットメッセージ内に電子的にばらまかれ、および／または、リードオフメッセージ、クロージングメッセージ、もしくは上記のすべてとして挿入される可能性があるからである。この特定のアプリケーションでは、望ましくない内容（たとえば、コマーシャル、プロモーション、または広告）を取り込み、ジャンクとしてタグ付けするために、特定のメッセージ内容（テキストおよび画像）を自動的にフィルタするようにフィルタをトレーニングすることができる。

【００２１】

本発明では、「受信者」という用語は、着信メッセージまたはアイテムのアドレスを指す。「ユーザ」という用語は、本明細書に述べられているフィードバックループシステムおよびプロセスに、受動的に、または積極的に参加するように選択された受信者を指す。

【００２２】

次に図１Ａを参照すると、本発明の一態様によるフィードバックトレーニングシステム１０の全体的なブロック図が示されている。メッセージ受信構成要素１２は、（ＩＭと呼ばれる）着信メッセージを受信し、対象となる受信者１４に送達する。メッセージ受信構成要素１２は、望ましくないメッセージ（たとえば、スパム）の送達を軽減するために、多数のメッセージ受信構成要素の場合に通例であるように、少なくとも１つのフィルタ１６（たとえば、ジャンクメールフィルタ）を含むことができる。メッセージ受信構成要素１２は、フィルタ１６と共に、メッセージ（ＩＭ）を処理し、メッセージのフィルタ済み部分集合（ＩＭ’）を対象となる受信者１４に送る。

【００２３】

本発明のフィードバック態様の一部として、ポーリング構成要素１８は、着信メッセージ（ＩＭ）をすべて受信し、それぞれの対象となる受信者１４を識別する。ポーリング構成要素１８は、たとえば（ＩＭ’’と呼ばれる）着信メッセージの部分集合をスパム、またはスパムでないものとして分類するために、（スパムファイタ２０と呼ばれる）対象となる受信者１４の部分集合を選択する。（ＶＯＴＩＮＧ　ＩＮＦＯと呼ばれる）分類関連情報がメッセージストア／投票ストア２２に送信され、メッセージストア／投票ストア２２において、投票情報、並びにそれぞれのＩＭ’’のコピーが、フィードバック構成要素２４によってなど、後で使用するために記憶される。具体的には、フィードバック構成要素２４は、投票情報を利用する機械学習技法（たとえば、ニューラルネットワーク、ＳＶＭ、ベイジアンネットワーク、または、本発明と共に使用するのに適した任意の機械学習システム）を使用し、たとえばスパムメールを識別することに関してフィルタ１６をトレーニングおよび／または改善する（および／または、新しいフィルタを構築する）。着信メッセージの新しいストリームが、新たにトレーニングされたフィルタ１６を介して処理されるにつれて、スパムがより少なく、より多くの（ＩＭ’と呼ばれる）正当なメッセージが対象となる受信者１４に送達される。したがって、システム１０は、スパムファイタ２０によって生成されるフィードバックを使用することによって、スパムの識別と、改善されたスパムフィルタのトレーニングとを容易にする。本発明のそのようなフィードバック態様は、スパム検出システムを洗練するための豊かな、非常に動的なスキームを提供する。本発明のより細かい態様に関する様々な詳細について、下記で論じる。

【００２４】

次に図１Ｂを参照すると、本発明による、スパムファイティングおよびスパム防止に関連するフィードバックループトレーニングの流れ図１００が示されている。トレーニングプロセスの準備の際に、および／またはトレーニングプロセスの前に、（たとえば、すべての電子メールユーザを含むマスタ集合から）スパムファイタとすべきユーザが選択され

、この選択は、ランダムサンプリング、または信頼レベル、または本発明による任意の好適な選択スキーム／基準に基づくことができる。たとえば、選択されたユーザの部分集合は、すべてのユーザ、ランダムに選択されたユーザの集合、スパムファイタとしてオプトインしているユーザ、またはオプトアウトしなかったユーザ、および／またはそれらの任意の組合せを含み、および／または、一部にはそれらの人口学的場所および関連情報に基づくことができる。

【００２５】

別法として、選択された電子メールユーザのマスタ集合は、スパム送信者が本発明を無力化することがよりコストの掛かるものとなるよう、有料ユーザに制限することができる。したがって、スパムファイティングに参加するように選択されたユーザの部分集合は、有料ユーザだけを含むことができることになる。次いで、選択されたユーザ（たとえば、スパムファイタ）の名前およびプロパティを含むリストまたは顧客表を作成することができる。

【００２６】

メッセージの着信ストリーム１０２が受信されたとき、各メッセージの受信者は、１０４で、すべてのスパムファイタのリストに突き合わせてチェックされる。受信者がリスト上にある場合には、そのメッセージはポーリングのために考慮される。次に、ポーリング用のメッセージを選択すべきかどうか判定される。従来のスパムフィルタと異なり、本発明は、少なくとも着信メールすべてがポーリングのために考慮されるまで、どのメッセージ（たとえば、スパム）をも削除しない。すなわち、メールは、任意の標識（たとえば、スパム、非スパム）の対象となる前に分類され、これは、ユーザポーリングに使用可能な、偏らないメッセージのサンプルを得るのを容易にする。

【００２７】

メッセージ選択用の構成要素（図示せず）を使用し、データの偏りを軽減するために、何らかのランダムな確率を用いてメッセージを選択することができる。別の手法は、人口学的情報、並びに他のユーザ／受信者属性およびプロパティを使用することを必要とする。したがって、少なくとも一部にはユーザ／受信者に基づいて、メッセージを選択することができる。メッセージを選択するための他の代替アルゴリズムも存在する。しかし、ユーザ当たり、もしくはある時間当たりのユーザ当たり選択されるメッセージの数に対して、もしくは、任意の所与のユーザからのメッセージを選択する確率に対して制限がある可能性がある。そのような制限がないと、スパム送信者は、アカウントを作成し、そこに数百万のスパムメッセージを送信し、そのようなメッセージすべてを善良として分類することができることになり、これによりスパム送信者は、誤って標識が付けられたメッセージによってトレーニングデータベースを改悪することが可能になるであろう。

【００２８】

いくつかの形態のスパムフィルタリング、特にブラックホールリストと呼ばれるものは、スキップ可能でない可能性がある。ブラックホールリストは、サーバが、インターネットプロトコル（ＩＰ）アドレスのあるリストからメールを受信するのを防止する。したがって、メッセージの選択は、ブラックホールリストからのものでないメールの集合から選ぶことができる。

【００２９】

本発明の独自の態様は、現在実施されているフィルタによってスパムとしてマークされた、ポーリング用に選択されたメッセージが、削除されることも、ジャンクメールフォルダに移動されることもないことである。その代わりに、それらは、他のメッセージすべてが受信されてポーリングを考慮される通常の受信箱またはメールボックス内に配置される。しかし、そのメッセージのコピーが２つあり、そのメッセージがフィルタによってスパムと考えられた場合には、一方のコピーがスパムフォルダに送達され、そうでない場合には、設定されたパラメータに従って処理される（たとえば、削除され、または特別にマークされ、またはジャンクフォルダに移動される）。

【００３０】

メッセージは、選択されたとき、ユーザに転送され、それがポーリングメッセージであることを示すように何らかの特別な方法でマーク付けされる。具体的には、選択されたメッセージは、メッセージ修正構成要素１０６によって修正することができる。メッセージ修正の例には、それだけには限らないが、別個のフォルダ内にポーリングメッセージを置くこと、「発信元（ｆｒｏｍ）」アドレスまたは件名行を変更すること、および／またはそのメッセージをそのユーザに対するポーリングメッセージとして識別する特別なアイコンまたは特別な色を使用することが含まれる。選択されたメッセージはまた、別のメッセージ内でカプセル化することができ、その別のメッセージは、カプセル化されたメッセージに対してどのように投票し、および／または分類するかについてユーザに使用説明（または指示：instruction）を送ることになる。これらの使用説明は、たとえば、そのメッセージをスパムとして投票するためのものと、そのメッセージをスパムでないものとして投票するためのものという少なくとも２つのボタンまたはリンクを含むことができる。

【００３１】

投票用ボタンは、ポーリングメッセージのコピーをユーザに送信する前に、メッセージの内容を修正することによって実装することができる。本発明が（メールサーバではなく）クライアント電子メールソフトウェアに関連して使用されるとき、ユーザインターフェースは、投票用ボタンを含むように修正することができる。

【００３２】

さらに、ポーリングメッセージは、使用説明と投票用ボタン、並びに、そこに添付された、選択されたメッセージを含むことができる。また、ポーリングメッセージは、件名行、発信元（ｆｒｏｍ）アドレス、送信および／または受信された日付、テキストまたはテキストの少なくとも最初の数行など、選択されたメッセージの概要を含むことができる。別の手法は、投票用使用説明および投票用ボタンがその先頭に追加された状態でメッセージを送信することを必要とする。実際には、ユーザがポーリングメッセージのコピーを開いた、および／またはダウンロードしたとき、それだけには限らないが「スパム」ボタンおよび「スパムでない」ボタンを含むボタン（またはリンク）を、ユーザインターフェース上でポップアップさせることができ、もしくは、ポーリングメッセージ内に組み込むことができる。したがって、各ポーリングメッセージが１組の使用説明および好適な投票用ボタンを含むことが可能である。おそらくは（使用説明またはボタンのテキストを不明瞭にする可能性がある）ＨＴＭＬのｂａｃｋｇｒｏｕｎｄ命令を削除することを含めて、他の修正が必要となる場合がある。

【００３３】

望ましい情報のタイプに応じて、「送信請求型商業電子メール」ボタンなど別のボタンを設けることもできる。このメッセージはまた、今後のポーリングからオプトアウトするためのボタン／リンクを含むことができる。使用説明は、ユーザの好ましい言語にローカライズされ、また、ポーリングメッセージ内に埋め込むことができる。

【００３４】

さらに、ポーリング用に選択されたメッセージは、メッセージ修正構成要素１０６によって、または何らかの他の好適なウィルススキャン構成要素（図示せず）によって、ウィルスがあるかどうかスキャンすることができる。ウィルスが見つかった場合、そのウィルスを除去することも、そのメッセージを廃棄することもできる。ウィルス除去は、そのメッセージが選択されたとき、また、ユーザがそのメッセージをダウンロードする直前を含めて、システム１００の任意の時点で行うことができる。

【００３５】

メッセージの修正の後で、メッセージ送達構成要素１０８は、投票のために、ポーリングメッセージをユーザに送達する。ユーザフィードバック（たとえば、ポーリングメッセージ、ユーザの投票とそれに関連付けられた任意のユーザプロパティ）には、固有の識別子（ＩＤ）１１０（たとえば、メタデータ）が割り当てられる。ＩＤ１１０および／またはそれに対応する情報は、ユーザ分類／投票が編集および記憶されるメッセージストア／投票ストア１１２（たとえば、中央データベース）に送信される。

【００３６】
データベースレベルでは、ポーリングに使用可能な選択されたメッセージを、後でポーリングしまたは使用するために、保持することができる。さらに、データベースは、特定のユーザが過剰にサンプリングされていないように、もしくは、そのユーザによって指定された制限内である量のデータがそのユーザから確実に収集されつつあるようにするために、指定時刻ごとに頻度分析を実行することができる。具体的には、フィードバックシステム１００は、サンプリングとデータ双方の偏りを軽減するために、ユーザのメールの割合制限、並びにサンプリング期間を監視する。これは、使用量の少ないユーザと使用量の多いユーザを共に含めて、ユーザが、利用可能なユーザすべてから選択される場合、特に重要である。たとえば、使用量の少ないユーザは、一般に、使用量の多いユーザに比べて、非常に少ないボリュームのメールを送受信する。したがって、システム１００は、選択されたメッセージが、確実に、そのユーザによって受信されたメッセージのＴ個ごとに約１つとなるように、並びに、そのユーザによってＺ時間ごとに１つ以下のメッセージが受信されるように、メッセージ選択プロセスを監視する。したがって、例えば、システムは、サンプリングすべき（たとえば、ポーリングするかどうかが考察される）着信メッセージ１０個ごとに１つ、ただし、２時間ごとに１つ以下をポーリングすることができる。頻度制限、または割合制限は、使用量の多いユーザに比べて、使用量の少ないユーザに対して、不釣り合いな量のメッセージをサンプリングすることを軽減し、また、ユーザを過剰に煩わせるのを緩和する。

【００３７】
中央データベース１１２は、ある頻度ごとに、ポーリングのためにシステム１００によってサンプリングされた、しかし分類されていないメッセージがあるかどうかスキャンする。データベースは、これらのメッセージを引き出し、それぞれのユーザの人口学的プロパティに対してそれらをローカライズし、ユーザに投票するように、またメッセージを分類するように要求するためにポーリングメッセージを作成する。しかし、スパムフィルタは、あらゆる新しい着信分類を受信した直後に、修正またはトレーニングを受けることができない。逆に、オフライントレーニングでは、指定スケジュールごとに、または継続的に、または日ごとに、データベース１１２内に受信されるデータをトレーナが連続的に調べることが可能になる。すなわち、トレーナは、規定された開始点から、または、過去における設定された量の時間で開始し、フィルタをトレーニングするために、その時点以降、すべてのデータを調べる。たとえば、規定された時間は、午前０時から午前６時とすることができる。

【００３８】
新しいスパムフィルタは、機械学習技法１１４（たとえば、ニューラルネットワーク、サポートベクタマシン（ＳＶＭ））により、データベース１１２内で維持されているメッセージ分類を分析することによって、継続的にトレーニングすることができる。機械学習技法は、学習のために善良なメールとスパムの両方の例を必要とし、これにより、それらを区別するように学習することができる。スパムの既知のサンプルをマッチングすることに基づく技法であっても、誤って善良なメールが捕捉されないようにすることができるよう、善良なメールの例を有することは役立つ。

【００３９】
したがって、単なる苦情だけではなく、スパムの肯定的例と否定的例を共に有することが重要である。フリーメーリングリストなど、スパムと正当なメールを共に大量に送信するいくつかのドメインがある。苦情だけに基づいてシステムを構築した場合、これらのドメインからのメールすべてがフィルタされる可能性があり、間違いが多数に上る。したがって、そのドメインは大量の善良なメールをも送信することを認識していることは重要である。さらに、ユーザは、しばしば、フリーメーリングリストに対してサインアップしたことを忘れることなど、間違いを犯す。たとえば、Ｎｅｗ　Ｙｏｒｋ　Ｔｉｍｅｓなど大規模な合法プロバイダは、定期的に正当なメールを送信する。数人のユーザは、サインアップしていたことを忘れて苦情を言い、これらのメッセージをスパムとして分類する。こ

のメールが本物であることを大抵のユーザが理解しているというデータがない場合、このサイトからのメールは、他の方法で阻止することができる。

【００４０】

新しいフィルタ１１６は、参加しているインターネットサービスプロバイダ（ＩＳＰ）全体にわたって、および／または電子メールもしくはメッセージサーバに対して、および／または個々の電子メールクライアントに対して、および／または更新サーバに対して、および／または個々の企業の中央データベースに対して、配布構成要素１１８によって継続的に配布されることができる。さらに、フィードバックシステム１００は、考慮されてポーリングするのに使用されたメッセージのサンプルが、システム１００によって受信された電子メールの実際の配布に従うことができるように、継続的に機能する。その結果、新しいスパムフィルタをトレーニングするために使用されたトレーニングデータ集合は、適応スパム送信者に対して最新に保たれる。新しいフィルタが構築されたとき、ポーリングデータは、それがどれだけ過去に得られたかに基づいて、廃棄するか、もしくは少なく重み付けする（たとえば、軽視する）ことができる。

【００４１】

システム１００は、ゲートウェイサーバ、電子メールサーバ、および／またはメッセージサーバなど、サーバ部でメールが受信されたとき実装することができる。たとえば、メールが電子メールサーバ内に来たとき、サーバは、対象となる受信者のプロパティをルックアップし、その受信者がシステム１００にオプトインしているかどうか判定する。受信者のプロパティがそのように示している場合、その受信者のメールは、ポーリングに使用できる可能性がある。クライアントだけのアーキテクチャも存在する。たとえば、クライアント電子メールソフトウェアは、単一のユーザについてポーリング判断を下し、電子メールを中央データベースに送達し、または、個別化されたフィルタの性能を改善するためにポーリング情報を使用することができる。これら本明細書に述べられているものに加えて、このシステム１００のための他の代替的な構成が存在し、それらが本発明の範囲内に入ることが企図されている。

【００４２】

次に図２を参照すると、本発明の一態様による基本的なフィードバックループプロセス２００の流れ図が示されている。説明を簡単にするために、本方法は、一連の動作として示され、述べられているが、本発明によれば、いくつかの動作は、本明細書に示され述べられているものと異なる順序で、および／または他の動作と同時に行われる可能性があるため、本発明は動作の順序によって制限されないことを理解されたい。たとえば、方法は、別法として、状態図内でなどの一連の相互に関係のある状態または事象として表すことができることを、当業者なら理解するであろう。さらに、本発明による方法を実施するために図中の動作すべてを必要とするわけではない。

【００４３】

プロセス２００は、プロセス２０２において、サーバなど構成要素内にメールが到来し、受信されたことによって始まる。メールがサーバに到着したとき、サーバは、（プロセス２０４において）対象となる受信者のプロパティを識別し、ポーリングのために、対象となる受信者が先にスパムファイタとしてオプトインしているかどうか判定する。したがって、プロセス２００は、受信者がフィードバックシステムにオプトインしているかどうか示すことができるユーザプロパティフィールドを使用するか、もしくは、オプトインしているユーザのリストを調べる。ユーザがフィードバックシステムの参加者であると判定され、プロセス２０６においてポーリング用に選択されている場合、フィードバックシステムは、（プロセス２０８において）どのメッセージがポーリング用に選択されるか判定することによって、アクションを起こす。そうでない場合、プロセス２００は、着信メッセージの少なくとも１人の対象となる受信者がユーザ（たとえば、スパムファイタ）であると決定されるまで、プロセス２０２に戻る。

【００４４】

実際には、現在使用されているフィルタ（たとえば、パーソナライズ化されたフィルタ

、Ｂｒｉｇｈｔｍａｉｌフィルタ）によってスパムとして指定される（またはスパムであるはずの）メッセージを含めて、メッセージすべてがポーリングのために考慮される。したがって、メッセージは、ポーリングのために考慮される前に削除されることも、廃棄されることも、ジャンクフォルダに送られることもない。

【００４５】

サーバによって受信された各メッセージまたはメールアイテムは、そのメールトランザクションに対応する１組のプロパティを有する。サーバは、これらのプロパティを編集し、ポーリングメッセージと共に中央データベースに送る。プロパティの例には、（たとえば、「Ｔｏ：」「ｃｃ：」および／または「ｂｃｃ：」フィールド内にリストされる）受信者リスト、現在使用されているフィルタの判断（たとえば、フィルタがメッセージをスパムとして識別したかどうか）、別の任意選択のスパムフィルタ（たとえば、Ｂｒｉｇｈｔｍａｉｌフィルタ）の判断、およびユーザ情報（たとえば、ユーザ名、パスワード、実名、ポーリングされるメッセージの頻度、使用量データなど）が含まれる。ポーリングメッセージおよび／またはその内容、並びに対応するユーザ／受信者には、それぞれ固有の識別子が割り当てられる。識別子は、データベースに送り、必要に応じて後で更新することもできる。

【００４６】

プロセス２１４において、ポーリング用に選択されたメッセージ（たとえば、元のメッセージ$_{1\text{-}M}$、ただし、Ｍは１以上の整数）は、メッセージ$_{1\text{-}M}$がポーリングメッセージ$_{P1\text{-}PM}$であることをユーザに示すように修正され、次いで、（プロセス２１６において）ポーリングのためにユーザに送達される。たとえば、ポーリングメッセージは、添付ファイルとして投票を受ける元のメッセージと、そのメッセージに対してどのように投票するかについての１組の使用説明とを含むことができる。その１組の使用説明は、たとえば「善良なメール」ボタンと「スパム」ボタンなど、少なくとも２つのボタンを含む。ユーザが、メッセージを善良なメールまたはスパムとして分類するために（プロセス２１８において）１つのボタン上でクリックしたとき、ユーザは、そのユーザが送信している分類のための固有の識別子に対応するユニフォームリソースロケータ（ＵＲＬ）に導かれる。この情報は掲示され、その元のメッセージ$_{1\text{-}M}$に対して中央データベース内の関連レコードが更新される。

【００４７】

プロセス２１６において、またはプロセス２００中における任意の他の適切な時間に、元のメッセージをオプションでユーザに送達することができる。したがって、ユーザは、メッセージを２回、すなわち１回はその元の形態で、また再度その修正されたポーリング形態で受信する。

【００４８】

さらに後のある時間には、新しいスパムフィルタが、少なくとも一部にはユーザフィードバックに基づいて、プロセス２２０において作成およびトレーニングされる。新しいスパムフィルタが作成およびトレーニングをされた後で、そのフィルタは、（２２２において）直ちに電子メールサーバ上で使用することができ、および／またはクライアントサーバ、クライアント電子メールソフトウェアなどに配布することができる。新しい、または更新されたスパムフィルタをトレーニングおよび配布することは、継続的な活動である。したがって、プロセス２００は、着信メッセージの新しいストリームが受信されたとき２０４において継続される。新しいフィルタが構築されたとき、より古いデータは、それらがどれだけ過去に得られたかに基づいて、廃棄され、または、少なく重み付けされる。

【００４９】

フィードバックシステム１００およびプロセス２００は、その参加しているユーザのフィードバックに依拠する。残念ながら、何人かのユーザは信頼することができないか、あるいは単に怠惰であり、一貫して正確な分類を提供することができない。中央データベース１１２（図１ａ）は、ユーザ分類の履歴を保持する。したがって、フィードバックシステム１００は、矛盾の数、そのユーザの気が変わった回数、既知の善良なメールまたは既

知のスパムに対するそのユーザの応答、並びに、ポーリングメッセージに対するユーザ回答の数もしくは頻度を追跡することができる。

【００５０】

これらの数のいずれか１つが、規定された閾値を超えたとき、または単にシステムのあらゆるユーザについて、フィードバックシステム１００は、１つの、またはいくつかの妥当性検査技法を呼び出し、特定の１人または複数のユーザの信頼性を査定することができる。１つの手法は、本発明の他の態様による、図３に示されている交差検定法３００である。

【００５１】

交差検定技法は、３０２において、中央データベースが、ポーリング結果およびそれぞれのユーザ情報など着信データを受信して始まる。次に、３０４において、適切な数のユーザをテストするために交差検定をするのが望ましいかどうかを判定しなければならない。望ましい場合には、３０６において、着信データの何らかの部分を使用して、新しいスパムフィルタがトレーニングされる。すなわち、テストされているユーザからのデータは、トレーニングから除外される。たとえば、フィルタは、（９０％フィルタと呼ばれる）ポーリングされたユーザデータの約９０％を用いてトレーニングされ、それによって、テストされているユーザによって送信されたデータに対応する（１０％テストユーザと呼ばれる）データの約１０％を除外する。

【００５２】

３０８において、９０％ユーザが、テストユーザのメッセージに対してどのように投票したかどうかを判定するために、９０％フィルタは、残りの１０％テストユーザデータに対して実行される。（３１０において）９０％フィルタと１０％テストユーザデータとの不一致の量が、規定された閾値を超えた場合には、３１２において、ユーザの分類を手動で検査することができる。別法で、またはそれに加えて、テストメッセージを疑わしいまたは信頼できないユーザに送信することができ、および／または、これらの特定のユーザを今後のポーリングから除外することができ、あるいは／または、これらの過去のデータを廃棄することができる。しかし、閾値を超えない場合には、プロセスは３０６に戻る。交差検定技法３００は、投票／分類データの信頼性を判定および維持するために、必要に応じて様々なユーザを除外して、テストユーザの任意の適切な集合と共に使用することができる。

【００５３】

ユーザ忠実度および信頼度を査定するための第２の手法は、所与の期間内で収集されたデータすべてに対してフィルタをトレーニングし、次いで、そのフィルタを使用して、トレーニングデータに対してテストすることを含む。この技法は、テスト・オン・トレーニング（ｔｅｓｔ－ｏｎ－ｔｒａｉｎｉｎｇ）として知られている。あるメッセージがトレーニングに含まれていた場合、フィルタは、その評価を学習しているべきであった。たとえば、学習済みフィルタは、ユーザが行ったのと同じ方法でそのメッセージを分類すべきである。しかし、フィルタは、ユーザが「スパムでない」として標識を付けたときスパムとして標識を付けることによって、引き続き間違いを犯す可能性があり、逆も同様である。フィルタがそのトレーニングデータと一致しないためには、メッセージは、他のメッセージとひどく一致しないことを必要とする。そうでない場合、トレーニング済みフィルタは、ほぼ確実に、何らかの方法で正しく分類しているはずである。したがって、そのメッセージは、信頼できない標識を有するものとして廃棄することができる。この技法または交差検定のどちらかを使用することができる。すなわち、交差検定は、分類において、より多くの誤りを生み出し、信頼度があまり高くない可能性があり、逆に、テスト・オン・トレーニングは、誤りがより少なく、より信頼度が高い。

【００５４】

テスト・オン・トレーニングと交差検定技法３００は共に、個々のメッセージに適用することができ、個々のユーザの分類またはメッセージの評価は、（たとえば、多数評価（ｍａｊｏｒｉｔｙ　ｒａｔｉｎｇ）に従って）一般的な一致によって除外される。別法と

して、両技法を使用し、潜在的に信頼可能でないユーザを識別することができる。

【００５５】

さらに、または交差検定技法および／もしくはテスト・オン・トレーニング技法の代わりに、「結果が既知の」技法を使用し、ユーザ信頼性を検証することができる（図４への３１４に引き続く）。図３および図４の技法は別々に示されているが、両手法を同時に使用することができることを理解されたい。すなわち、既知の善良なメッセージと既知のスパムメッセージからの情報を交差検定またはテスト・オン・トレーニングの結果と組合せ、どのユーザを廃棄すべきか判定することができる。

【００５６】

図４を参照すると、本発明の一態様による、投票するユーザの忠実度を妥当性検査するためのプロセス４００の流れ図が示されている。プロセス４００は、図３に示されている３１４から参照される。４０２において、結果が既知のテストメッセージが、疑わしいユーザ（またはすべてのユーザ）に送信される。たとえば、テストメッセージを着信メールに注入し、次いで、データベースが「既知の」結果を受信するように手分類することができる。そうでない場合、プロセス４００は、既知の結果メッセージが第三者によって送信されるまで待つことができる。ユーザは、同じテストメッセージに対して投票することが許される。投票結果は、４０４において、既知の結果と比較される。４０６において、ユーザの投票が一致しない場合には、一貫性と信頼度を実証するまで、それらの現在および／または今後および／または過去の分類を（４０８において）適切な時間の間、手検査することができる。別法として、それらの現在または今後または過去の分類を、廃棄または除去することができる。最後に、そのユーザを今後のポーリングから除去することができる。しかし、それらの投票結果がテストメッセージ結果と一致する場合には、４１０においてそのユーザを信頼できるものと考えることができる。プロセス４００は、４１２で図３に戻り、疑わしいユーザの次のグループについてどのタイプの妥当性の検査技法が望ましいか判定する。

【００５７】

ユーザ信頼度を査定するための第４の手法（図示せず）は、能動学習（ａｃｔｉｖｅ　ｌｅａｒｎｉｎｇ）である。能動学習の場合、メッセージはランダムに拾い上げられない。その代わりに、フィードバックシステムは、そのメッセージがシステムにとってどれだけ有用になるか推定することができる。たとえば、フィルタがスパムの確率を返す場合、ポーリングのために、現在のフィルタによって、もっとも不確実に分類されるメッセージ、すなわち、そのスパムの確率が５０％に最も近いものを優先的に選択することができる。メッセージを選択するための別の方法は、メッセージがどれだけ一般的であるか判定することである。メッセージが一般的であるほど、ポーリングするためにより有用である。独特の各メッセージは、あまり一般的でないため、あまり有用でない。既存のフィルタの信頼レベル（ｃｏｎｆｉｄｅｎｃｅ　ｌｅｖｅｌ）を使用すること、メッセージの特徴がどれだけ一般的であるかを使用すること、並びに、既存のフィルタのその設定または内容の信頼レベル（たとえば、メタコンフィデンス（ｍｅｔａｃｏｎｆｉｄｅｎｃｅ））を使用することによって、能動学習を用いることができる。機械学習の当業者に周知のＱＢＣ（ｑｕｅｒｙ－ｂｙ－ｃｏｍｍｉｔｔｅｅ）など、多数の他の能動学習技法があり、これらの技法のいずれも使用することもできる。

【００５８】

次に、図５を参照すると、本発明の一態様による、ユーザフィードバックに加えてハニーポットフィードバックをスパムフィルタトレーニングに組み込むためのプロセス５００の流れ図が示されている。ハニーポットは、それらに誰が電子メールを送信しつつあるべきかが知られている電子メールアドレスである。たとえば、（５０２において）新たに作成される電子メールアドレスを私的なものに保ち、選択された個人だけに開示することができる。また、公に、しかし（たとえば、メールインクとして白い背景上に白い書体を置いて、）人に見られない制限された方法で開示することもできる。ハニーポットは、スパム送信者による辞書攻撃において特に有用である。辞書攻撃は、おそらくは辞書内のアド

レスすべて、あるいは、辞書内の単語の対、または有効なアドレスを見出すための類似の技法で作成された非常に大量のアドレスに、スパム送信者が電子メールを送信しようと試みるものである。（５０４において）ハニーポットに送信される電子メール、または（５０６において）何人かの選択された個人からのものでない電子メールは、（５０８において）スパムと考えられる。また、電子メールアドレスは、疑わしいマーチャントとサインアップさせることができる。したがって、そのマーチャントから受信するどの電子メールも（５１０において）善良なメールと考えられるが、他のメールすべてはスパムと考えられる。スパムフィルタは、（５１２において）それに応じてトレーニングにすることができる。さらに、疑わしいマーチャントがユーザの情報（たとえば、少なくとも電子メールアドレス）を第三者に販売する、または他の方法で開示することが決定される。これは、他の疑わしいマーチャントを用いて繰り返すことができ、ユーザの情報がスパム送信者に配布される可能性があることをユーザに警告するために、リストを生成することができる。これらは、安全にスパムと考えることができるハニーポットに電子メールを送信させるいくつかの技法にすぎない。実際には、安全にスパムと考えることができるハニーポットに電子メールを送信させるための他の代替的な方法がある。

【００５９】

ハニーポットはスパムの良いソースであり、しかし正当なメールの非常に不十分なソースであるため、ハニーポットからのデータは、フィードバックループシステム（図１）からのデータと組み合わせて、新しいスパムフィルタをトレーニングすることができる。異なるソースまたは異なる分類からのメールは、異なるように重み付けすることができる。たとえば、それらのメールの１０％に対してポーリングされる１０個のハニーポットおよび１０人のユーザがある場合、ポーリングからのものの約１０倍と同程度のものをハニーポットから予想すべきである。したがって、この差を埋め合わせるために、ポーリングからの正当なメールをスパムの１０倍または１１倍と同程度で重み付けすることができる。別法として、ハニーポットデータは、選択的に減じて重み付けすることができる。たとえば、ユーザのメールの約５０％が善良なメールであり、約５０％がスパムである。同じボリュームのスパムがハニーポットに届く。したがって、ハニーポットは１００％のスパムを有するように思われ、１０％だけでなくそのすべてがサンプリングされる。組み合わされたシステム内で正しい比率のスパムと善良なメールでトレーニングするために、ハニーポットデータは９５％だけ減じて重み付けされ、ユーザスパムは５０％だけ減じて重み付けされ、１：１の全体的な比率を得る。

【００６０】

スパムレポートの他のソースには、フィードバックループシステムの参加者として含まれないユーザが含まれる。たとえば、フィルタを通過したスパムをレポートするために、全メールのすべてのユーザに対して「スパムをレポート（Ｒｅｐｏｒｔ　Ｓｐａｍ）」ボタンを使用可能にすることができる。このデータは、フィードバックループシステムからのデータを組み合わせることができる。この場合も、スパムのこのソースは、様々な側面で偏っているか、または、信頼できない可能性があるため、様々に減じて重み付け、または重み付けすべきである。また、再重み付けを行い、フィルタされなかったメールだけ「スパムとしてレポート（Ｒｅｐｏｒｔ－ａｓ－ｓｐａｍ）」ボタンによるレポートの対象となることを反映すべきである。

【００６１】

スパムフィルタに加えて、フィードバックループシステムによって隔離フィルタを作成および使用することができる。隔離フィルタは、肯定的なメール特徴と否定的なメール特徴を共に利用する。たとえば、人気のあるオンラインマーチャントからのメールは、ほぼ常に善良である。スパム送信者は、自分のスパム内で善良なマーチャントのメールの側面を真似ることによってシステムを利用する。別の例は、スパム送信者が、意図的に、あるＩＰアドレスを介して少量の善良なメールを送信することによってフィードバックシステムをだますことである。フィードバックループシステムは、このメールを善良なメールとして分類することを学習し、そのようなとき、スパム送信者は、同じＩＰアドレスからス

パムの送信を開始する。

【００６２】

したがって、隔離フィルタは、履歴データに基づいて、システムがそのために使用されるよりはるかに大量に、特定の肯定的な特徴が受信されつつあることに気付く。これにより、システムは、そのメッセージに疑問を抱き、したがって、そのメールをスパムとして送達またはマークするために選ぶ前に、十分な得票結果が得られるまでそのメッセージを隔離する。また、隔離フィルタは、メールがスパムか否か知られておらず、または確実でなく、そのことがしばらくの間わからない新しいＩＰアドレスからメールが受信されたときに使用することができる。隔離は、暫定的にそのメールをスパムとしてマークし、スパムフォルダに移動することを含めて、あるいは、ユーザに送達しないで、またはメールが見られなくなるようどこかに保存することによって、いくつかの方法で実行することができる。隔離は、スパムフィルタ閾値に近いメッセージについて行うことができる。すなわち、ポーリングからの追加の情報が正しい判断をする助けとなると想定することができる。また、隔離は、多数の同様なメッセージが受信されたときに行うことができる。すなわち、そのメッセージのいくつかを、フィードバックループを用いてポーリングのために送信することができ、トレーニング済みフィルタを使用し、そのメッセージを正しく分類することができる。

【００６３】

フィルタを構築することに加えて、本明細書で述べられているフィードバックループシステムは、フィルタを評価するためにも使用することができる。すなわち、スパムフィルタのパラメータを必要に応じて調整することができる。たとえば、フィルタは、昨夜の午前０時を介してトレーニングされる。午前０時の後に、データベース内に入るデータを取り、ユーザの分類に比べて、スパムフィルタの誤り率を決定する。さらに、フィードバックループを使用し、スパムフィルタのフォールスポジティブ／捕捉率（ｆａｌｓｅ　ｐｏｓｉｔｉｖｅ　ａｎｄ　ｃａｔｃｈ　ｒａｔｅｓ）を決定する。たとえば、ユーザ投票を取ることができ、メールを潜在的フィルタに通し、フォールスポジティブ／捕捉率を決定することができる。次いで、この情報を使用し、フィルタを調整および最適化することができる。最も低いフォールスポジティブ／捕捉率を得るために、それぞれが異なる設定またはアルゴリズムを使用するいくつかのフィルタを構築することによって、様々なパラメータ設定または様々なアルゴリズムを手動または自動で試みることができる。したがって、結果同士を比較し、最良の、または最適なフィルタパラメータを選択することができる。

【００６４】

フィードバックループは、常にスパムとして投票される、または、常に善良として投票される、または、少なくとも９０％善良と投票される、などのＩＰアドレスもしくはドメインもしくはＵＲＬのリストを構築およびポピュレートするために使用することができる。これらのリストは、他の方法でのスパムフィルタリングのために使用することができる。たとえば、少なくとも９０％スパムと投票されたＩＰアドレスのリストは、メールを受け入れないアドレスのブラックホールリストを構築するために使用することができる。フィードバックループはまた、スパム送信者のアカウントを打ち切るために使用することもできる。たとえば、あるＩＳＰの特定のユーザがスパムを送信していると思われる場合、そのＩＳＰに自動的に通知することができる。同様に、特定のドメインが大量のスパムに責任があると思われる場合、そのドメインの電子メールプロバイダに自動的に通知することができる。

【００６５】

フィードバックループシステムを実装するために使用することができるいくつかのアーキテクチャがある。１つの例示的なアーキテクチャは、図７で述べることになるように、サーブドベースであり、選択プロセスは、メールが電子メールサーバに到達したとき発生する。代替のアーキテクチャは、図６で述べられているように、クライアントをベースとするものである。クライアントをベースとするフィードバックループでは、ポーリング情

報を使用し、個別化されたフィルタの性能を改善することができ、あるいは、ここで示されている例示的な実装では、その情報を（たとえば、全社的、またはグローバルな）共用フィルタ用のトレーニングデータとして共用リポジトリに送信することができる。下記に述べられている以下のアーキテクチャは単に例示的なものであり、示されていない追加の構成要素および特徴を含むことができることを理解されたい。

【００６６】

次に図６を参照すると、クライアントをベースとするアーキテクチャにおけるフィードバックループ技法の例示的、一般的なブロック図が示されている。ネットワーク６００は、（クライアント$_1$、クライアント$_2$．．．クライアント$_N$とも呼ばれ、ただし、Ｎは１以上の整数の）１つまたは複数のクライアント６０２、６０４、６０６との間で電子メールの送信を容易にするために設けられる。ネットワークは、インターネットなど地球規模の通信ネットワーク（ＧＣＮ）、またはＷＡＮ（広域ネットワーク）、ＬＡＮ（ローカルエリアネットワーク）、あるいは任意の他のネットワーク構成とすることができる。この特定の実装では、ＳＭＴＰ（簡易メール転送プロトコル）ゲートウェイサーバ６０８がネットワーク６００とインターフェースし、ＬＡＮ６１０にＳＭＴＰサービスを提供する。ＬＡＮ６１０上で動作可能に配置された電子メールサーバ６１２は、ゲートウェイ６０８とインターフェースし、クライアント６０２、６０４、６０６の着信電子メールおよび発信電子メールを制御および処理する。そのようなクライアント６０２、６０４、６０６もまたＬＡＮ６１０上に配置され、少なくともそこで提供されるメールサービスにアクセスする。

【００６７】

クライアント$_1$６０２は、クライアントプロセスを制御する中央処理装置（ＣＰＵ）６１４を含む。ＣＰＵ６１４は、複数のプロセッサを備えることができる。ＣＰＵ６１４は、上述の１つまたは複数のデータ収集／フィードバック機能のいずれかを提供することと関連して、命令を実行する。命令には、それだけには限らないが、少なくとも上述の基本的なフィードバックループ方法、クライアント／メッセージ選択に対処するためにそれと組み合わせて使用することができる手法の少なくともいくつかもしくは全部、ポーリングメッセージ修正、データ保持、クライアント信頼度／分類の妥当性検査、フィードバックループシステムを含む複数のソースからのデータの再重み付け、スパムフィルタ最適化／調整、隔離フィルタ、スパムリストの作成、並びに、それぞれのＩＳＰや電子メールプロバイダに対するスパム送信者についての自動通知を実行する符号化命令が含まれる。ユーザインターフェース６１６は、ＣＰＵ６１４およびクライアントオペレーティングシステムとの通信を容易にするために設けられる。クライアント$_1$が対話し、電子メールにアクセスし、および、ポーリングメッセージに対して投票することができるようにする。

【００６８】

サーバ６１２から取り出されたクライアントメッセージのサンプリングは、メッセージセレクタ６２０によってポーリング用に選択することができる。対象となる受信者（クライアント）が先に参加することに合意している場合、メッセージがポーリング用に選択および修正される。メッセージ修正器６２２は、ポーリングメッセージになるようにメッセージを修正する。たとえば、メッセージは、上記のメッセージ修正説明に従って、投票用の使用説明と、投票用ボタンおよび／またはリンクとを含むように修正することができる。投票用ボタンおよび／またはリンクは、クライアント電子メールソフトウェアのユーザインターフェース６１６を修正することによって実装される。さらに、メッセージ修正器６２２は、クライアント６０２によって、閲覧するのに開かれ、または、ダウンロードされる前に、メッセージ（ポーリングメッセージおよび非ポーリングメッセージ）内のどのウィルスをも除去することができる。

【００６９】

一実装においては、スパムファイティングクライアント６０２のユーザは、各メッセージを１回だけ見ており、いくつかのメッセージはポーリングメッセージとして特別にマークされ、投票用ボタンなどを含む。本実装では、スパムファイティングクライアント６０

２のユーザは、いくつかのメッセージを２回見ることもあり、一方は通常のメッセージであり、他方はポーリングメッセージである。これは、いくつかの方法で実施することができる。たとえば、ポーリングメッセージは、サーバ６１２に返され、ポーリング済みメッセージストアに記憶することができる。別法として、クライアント６０２は、追加のメッセージを電子メール（Ｅ－Ｍａｉｌ）サーバ６１２に記憶することができる。別法として、クライアント６０２は、ユーザに各メッセージを２回、すなわち１回は通常のメッセージとして、１回は修正された形態で示すことができる。

【００７０】

ポーリング結果６２６は、ＣＰＵ６１４に次いでデータベース６３０に送信することができ、データベース６３０は、クライアントフィードバックアーキテクチャの特定の構成に応じて１つのクライアントからの、または、複数のクライアントからのデータを記憶するように構成することができる。中央データベース６３０は、ポーリングメッセージ、ポーリング結果、並びにそれぞれのクライアントユーザ情報を記憶する。関連構成要素を使用し、ポーリング頻度、クライアントユーザ信頼性（たとえば、ユーザの妥当性検査６３２）、並びに、他のクライアント統計を決定するためなど、そのような情報を解析することができる。妥当性の検査技法は、クライアントの投票の信頼度が疑わしいとき特に使用することができる。疑いは、矛盾の数、心変わりの数、並びに、特定の１人もしくは複数のユーザについてポーリングされた数を解析することにより発生する可能性があり、別法として、妥当性の検査技法は、あらゆるユーザについて使用することができる。中央データベース６３０内に記憶された任意の適切な量のデータを機械学習技法６３４において使用し、新しい、かつ／または改善されたスパムフィルタのトレーニングを容易にすることができる。

【００７１】

クライアント６０４および６０６は、特定のクライアントに対してパーソナライズ化されたフィルタを得て、トレーニングするために、上述のような同様の構成要素を含む。記載されているものに加えて、ポーリング済みメッセージスクラバ６２８は、データ集約、データ圧縮など様々な理由のためにポーリング済みメッセージの諸側面を除去することができるように、ＣＰＵ６１４と中央データベース６３０の間でインターフェースすることができる。ポーリング済みメッセージスクラバ６２８は、ポーリング済みメッセージの関係のない部分、並びに、それに関連する任意の望ましくないユーザ情報を一掃することができる。

【００７２】

次に、図７を参照すると、マルチユーザログインを容易にし、ポーリングデータを得る、本発明のフィードバックループ技法による例示的なサーバベースのフィードバックループシステム７００が示されている。ネットワーク７０２は、（ユーザ$_1$７０４$_1$、ユーザ$_2$７０４$_2$．．．ユーザ$_N$７０４$_N$とも呼ばれ、ただし、Ｎは１以上の整数の）１つまたは複数のユーザ７０４との間で電子メールの送信を容易にするために設けられる。ネットワーク７０２は、インターネットなど地球規模の通信ネットワーク（ＧＣＮ）、またはＷＡＮ（広域ネットワーク）、ＬＡＮ（ローカルエリアネットワーク）、あるいは任意の他のネットワーク構成とすることができる。この特定の実装では、ＳＭＴＰ（簡易メール転送プロトコル）ゲートウェイサーバ７１０がネットワーク７０２とインターフェースし、ＬＡＮ７１２にＳＭＴＰサービスを提供する。ＬＡＮ７１２上で動作できるように配置された電子メールサーバ７１４は、ゲートウェイ７１０とインターフェースし、ユーザ７０４の着信電子メール並びに発信電子メールを制御し、処理する。

【００７３】

システム７００は、メッセージ選択７１６、メッセージ修正７１８、メッセージポーリング（７２０、７２２、７２４）が、システム７００にログインする各異なるユーザについて行われるように、マルチログイン機能を提供する。したがって、コンピュータオペレーティングシステムのブートアッププロセスの一部として、ログイン画面を提示する、または、必要に応じて、ユーザ７０４が自分の着信メッセージにアクセスできるようになる

前に、関連ユーザプロフィルを保証するために、ユーザインターフェース７２６が提供される。したがって、第１のユーザ７０４$_1$（ユーザ$_1$）がメッセージにアクセスすることを選んだとき、第１のユーザ７０４$_1$は、典型的にはユーザ名およびパスワードの形態でアクセス情報を入力することによって、ログイン画面７２８を介してシステムにログインする。ＣＰＵ７３０は、アクセス情報を処理し、メッセージ通信アプリケーション（たとえば、電子メールクライアント）を介して第１のユーザ受信箱ロケーション７３２だけにユーザアクセスを許す。

【００７４】

着信メールがメッセージサーバ７１４上で受信されたとき、メッセージはランダムにポーリング用に選択され、これは、メッセージのうち少なくとも１つがポーリング用にタグ付けされることを意味する。タグ付けされたメッセージの対象となる受信者は、その受信者の誰か１人もまた、指定されたスパムファイティングユーザであるかどうか判定するために調べられる。そのような情報を示す受信者プロパティは、必要に応じて、メッセージサーバ７１４上で、またはシステム７００における任意の他の構成要素上で保持することができる。対象となる受信者の誰がスパムファイタでもあるか判定された後で、そのそれぞれのメールのコピーは、メールトランザクションに関する任意の他の情報とともに、記憶するために中央データベース７３４に送信することができる。ポーリング用にタグ付けされたメッセージは、メッセージ修正器７１８によって、上述した任意の数の方法によって修正される。ポーリング用に選択されたメッセージもまた、ユーザ７０４特有のものとすることができる。たとえば、ユーザ７０４は、いくつかのタイプのメッセージだけポーリングのために利用可能であることを示すことができる。これによりデータのサンプリングが偏る可能性があるため、そのようなデータは、他のクライアントデータに対して再重み付けし、不適切なトレーニングデータセットを構築するのを緩和することができる。

【００７５】

ポーリングメッセージのウィルススキャンもまた、この時点で、または、ポーリングメッセージがユーザ７０４によってダウンロードされ、かつ／または開かれる前の任意の他の時点で行うことができる。メッセージは、適切な形で修正された後で、受信箱$_1$７３２、受信箱$_2$７３６、受信箱$_N$７３８と呼ばれるそれぞれのユーザの受信箱に送達され、そこでポーリングのためにメッセージを開くことができる。ポーリングプロセスを容易にするために、各ポーリングメッセージは、ユーザによって選択されたときポーリングメッセージおよびポーリング結果に関する情報を生成する２つ以上の投票用ボタンまたはリンクを含む。各ポーリングメッセージのテキストは、投票用ボタンまたはリンクを組み込むように修正することができる。

【００７６】

分類に起因する任意の情報（たとえば、ポーリングメッセージまたはそれに関連付けられたＩＤ、ユーザプロパティ）を含む（メッセージ得票$_1$７２０、メッセージ得票$_2$７２２、メッセージ得票$_N$７２４と呼ばれる）メッセージ得票結果は、ＬＡＮ７１２上のネットワークインターフェース７４０を介して、中央データベース７３４に送信される。中央データベース７３４は、機械学習技法に適用し、新しく、かつ／または改善されたスパムフィルタ７４２を構築もしくは最適化するために、それぞれのユーザからのポーリング／ユーザ情報（７２０、７２２、７２４）を記憶することができる。しかし、プライバシーおよび／またはセキュリティの理由で、秘密情報は、中央データベース７１４に送信される前に、情報から除去し、または、取り去ることができる。ポーリングを介してユーザ７０４によって生成された情報もまた、統計データ内に集約することができる。したがって、情報を送信するためにあまり帯域幅は使用されない。

【００７７】

次いで、新たにトレーニングされたスパムフィルタ７４２は、新しいフィルタが使用可能であるときなど継続的に、特定の要求によってまたは自動的に、他のサーバ（図示せず）、並びに、ＬＡＮ７１２とインターフェースするクライアント電子メールソフトウェア（図示せず）に、配布することができる。たとえば、最も新しいスパムフィルタは、自動

的にサーバなどに送り出し、かつ／またはウェブサイトを介してダウンロードするために、使用可能にすることができる。より新しいスパムフィルタを構築するために新しいトレーニングデータセットが生成されたとき、より古いデータセット（たとえば、先に得られ、および／またはフィルタをトレーニングするために使用された情報）は、データの寿命に応じて、廃棄または無視することができる。

【００７８】

次に、スパムファイティングを投入している組織が、多数の異なるフィルタの使用組織によって共用されるフィルタを使用できるようにする、代替的なシナリオを考えてみる。本発明の一態様では、フィルタプロバイダはまた、非常に大規模な電子メールサービス（たとえば、有料および／または無料電子メールアカウント）のプロバイダである。それ自体の組織からの電子メールに排他的に頼るのではなく、フィルタプロバイダは、善良なメールとスパムの範囲をより良く取り込むように、いくつかのフィルタ使用組織からのいくつかのデータをも使用することを選ぶ。上述のようなフィードバックループシステムもまた、サーバまたはクライアントをベースとするアーキテクチャにおいて、そのような組織横断シナリオにおいて使用することができる。データをそれ自体のユーザからおよび様々なフィルタ使用組織から集約するフィルタプロバイダを、「内部」組織と呼び、参加しているフィルタ使用組織の１つに常駐する構成要素を「外部」と呼ぶことにする。一般に、組織横断システムは、それだけには限らないがＨｏｔｍａｉｌなど、フィルタプロバイダ部のメールデータベースサーバ（内部）並びに、１つもしくは複数の個々の企業内に常駐することができるものなどの１つもしくは複数のメッセージサーバ（外部）とを含む。この場合には、内部メールデータベースサーバはまた、それ自体の顧客からの実質的な電子メールフィードバックを記憶する。本発明のこの態様によれば、トレーニングデータセットは、内部データベース（たとえば、ＨｏｔｍａｉｌまたはＭＳＮサーバ上の無料電子メール／メッセージング）上で記憶された情報、並びに、それぞれの外部サーバに関連付けられた１つもしくは複数の外部データベース上で記憶された情報に基づいて生成することができる。外部データベース上で保持されている情報は、たとえば、機械学習技法において使用するために、インターネットなどネットワークを介して内部サーバに通信することができる。最終的には、外部データベースからのデータを使用することによって、新しいスパムフィルタをトレーニングし、および／または、外部に位置し（たとえば、それぞれの企業内の）、もしくは、内部メールサーバに関連付けられた既存のスパムフィルタを改善することができる。

【００７９】

外部データベースの１つまたは複数からのデータは、ポーリングメッセージ、ポーリング結果（分類）、ユーザ情報／プロパティ、並びに、ユーザ当たりの、もしくはユーザのグループ当たりの、もしくは各企業について平均した投票統計データのうち少なくとも１つを含むべきである。投票統計データは、それぞれの企業によって生成された情報の信頼度を決定し、外部データの偏りを軽減するのを容易にする。したがって、１つまたは複数の外部データベース（企業）からのデータは、再重み付けされ、または、他の外部データベースの１つもしくは複数と異なるように重み付けするされ得る。さらに、外部エンティティは、上述のように同様な妥当性の検査技法を使用して、信頼度および信頼性に関してテストをすることができる。

【００８０】

企業セキュリティ、プライバシー、および機密性のために、たとえば各企業から電子メールサーバに、インターネットを渡って通信される情報またはデータは、その元の形態からスクラブ（ｓｃｒｕｂ）し、および／または短縮し、および／または圧縮することができる。元の形態は、それぞれの外部データベース上で維持され、かつ／または他の方法によって、各企業の嗜好に従って処理される。したがって、電子メールサーバまたは任意の他の内部メールサーバは、スパム分類、送信側ドメイン、送信側名、スパムに分類されたメッセージの内容など、トレーニングデータを生成するために必要な関連情報だけを受信する。

【００８１】
次に図８を参照すると、例示的な組織横断フィードバックシステム８００が示されている。このシステム８００は、内部データベースサーバおよび外部メールサーバがネットワークを介してデータベース情報を通信および交換し、改善されたスパムフィルタを構築するために機械学習技法で使用されるトレーニングデータセットの生成を容易にすることができる。システム８００は、（たとえば、少なくとも１つの企業に関連する）少なくとも１つの外部メッセージサーバ８０２、並びに、内部データベースサーバ８０４を含む。組織横断システムの性質により、外部サーバ８０２および内部電子メールサーバ８０４は、それぞれ、それ自体のデータベースを維持する。すなわち、電子メールサーバ８０４は、やはり新しいスパムフィルタ８０８をトレーニングするために使用することができる内部データベース８０６に関連付けられる。同様に、外部サーバ８０２は、少なくとも１つの新しいスパムフィルタ８１２、並びに、電子メールサーバ８０４に対して内部に位置するスパムフィルタ８０８をトレーニングするために使用することができる外部データベース８１０に関連付けられる。したがって、外部データベース８１０上において記憶された情報を使用し、電子メールサーバ上に位置するスパムフィルタ８０８をトレーニングすることができる。

【００８２】
ＧＣＮ８１４は、内部電子メールサーバ８０４および１つもしくは複数の外部メッセージサーバ８０２の間で、情報の通信を容易にするために設けられる。組織横断システムの外部サーバ構成要素は、サーバをベースとするフィードバックループシステム（たとえば、上記図７）と同様な形で動作する。たとえば、メッセージサーバ８０２、外部データベース８１０、フィルタ８１２は、ＬＡＮ８１５上に位置することができる。さらに、コンピュータオペレーティングシステムのブートアッププロセスの一部としてログイン画面８１８を提示する、または、必要に応じて、ユーザが自分の着信メッセージにアクセスできるようになる前に関連ユーザプロフィルを保証するために、ユーザインターフェース８１６が提供される。

【００８３】
このサーバをベースとするシステムにおいては、（ユーザ$_1$８２０、ユーザ$_2$８２２、ユーザ$_N$８２４と呼ばれる）１人または複数のユーザが、利用可能なメールサービスを使用するために、同時にシステムにログインすることができる。実際には、第１のユーザ８２０（ユーザ$_1$）がメッセージにアクセスすることを選んだとき、第１のユーザ８２０は、典型的にはユーザ名およびパスワードの形態でアクセス情報を入力することによって、ログイン画面８１８を介してシステムにログインする。ＣＰＵ８２６は、アクセス情報を処理し、メッセージ通信アプリケーション（たとえば、電子メールクライアント）を介して第１のユーザ受信箱ロケーション８２８だけにユーザアクセスを許す。

【００８４】
着信メールがメッセージサーバ８０２上で受信されたとき、メッセージは、ランダムに、または、具体的に、ポーリングのための対象とされる。メッセージをポーリング用に選択することができるようになる前に、そのような目標をしぼったメッセージの対象となる受信者はスパムファイタユーザリストと比較され、その受信者の誰かもまた、指定されたスパムファイティングユーザであるかどうかを判定する。そのような情報を示す受信者プロパティは、メッセージサーバ８０２、データベース８１０上で、または、必要に応じて、システム８００における任意の他の構成要素上で保持することができる。対象となる受信者の誰がやはりスパムファイタであるかを判定された後に、メッセージはポーリング用に選択され、ポーリングメッセージのコピー、並びに、メールトランザクションに関連する任意の他の情報を、データベース８１０に送信することができる。

【００８５】
ポーリング用に選択されたメッセージは、メッセージ修正器８３０によって、上述した任意の数の方法によって修正される。実際には、固有の識別（ＩＤ）を、各ポーリングメッセージに、および／またはスパムファイタに、および／または各ポーリング結果に割り

当て、データベース８１０内に記憶することができる。先に述べたように、ポーリング用に選択されるメッセージは、ランダムに選ぶことができ、または、それぞれのユーザ（８２０、８２２、８２４）特有のものとすることができる。たとえば、ユーザ$_1$８２０は、あるタイプのメッセージ（たとえば、企業の外から送信されたメッセージ）だけがポーリング用に使用可能であることを示すことができる。そのような特定のメッセージから生成されたデータは、データのサンプリングが偏るのを緩和するために、再重み付けおよび／または軽視される。

【００８６】

ポーリングメッセージのウィルススキャンもまた、この時点において、または、ポーリングメッセージがユーザによってダウンロードされ、および／または開かれる前の、任意の他の時点において行うことができる。メッセージは、適切な形で修正された後に、受信箱$_1$８２８、受信箱$_2$８３２、受信箱$_N$８３４と呼ばれるそれぞれのユーザの受信箱に送達され、そこでポーリングのためにメッセージを開くことができる。ポーリングプロセスを容易にするために、各ポーリングメッセージは、ユーザによって選択されたときポーリングメッセージおよびポーリング結果に関する情報を生成する、２つ以上の投票用ボタンまたはリンクを含む。各ポーリングメッセージのテキストは、投票用ボタンまたはリンクを組み込むように修正することができる。

【００８７】

分類に起因する任意の情報（たとえば、ポーリングメッセージまたはそれに関連付けられたＩＤ、ユーザプロパティ）を含む（メッセージ得票$_1$８３６、メッセージ得票$_2$８３８、メッセージ得票$_N$８４０と呼ばれる）メッセージ得票結果は、ＬＡＮ８１５上に位置するネットワークインターフェース８４２を介して、データベース８１０に送信される。データベース８１０は、後に、新しい、および／または改善されたスパムフィルタ８１２、８０８を構築および／または最適化するために使用される機械学習技法で使用するために、それぞれのユーザからのポーリング／ユーザ情報を記憶する。

【００８８】

プライバシーの理由で、各企業は、ポーリング済みメッセージおよび／またはユーザ情報を、それ自体のデータベース８１０に、および／または、たとえばＧＣＮ８１４を介して電子メールデータベース８０６に送信する前に、重要な情報を取り去りたいと望む場合がある。１つの手法は、スパムメッセージについてだけデータベース（８０６および／または８１０）にフィードバックを送り、それによって正当なメールについてのフィードバックを除外することである。別の手法は、送信側および送信側のＩＰアドレスなど、正当なメールに関する情報の一部の部分集合だけ送ることである。別の手法は、フィルタによって悪質とマークされるはずのユーザによって善良とマークされたもの、またはその逆など、選択されたメッセージについて、それらをフィルタに送信する前に、明示的にユーザ許可を依頼することである。これらの手法のいずれか、または、それらの組合せは、参加しているクライアントについて秘密情報のプライバシーを維持し、一方、スパムフィルタ（８０８および／または８１２）をトレーニングするためにデータを連続的に提供することを容易にする。

【００８９】

上述のものなどユーザ妥当性の検査スキームもまた、各企業に、並びに、企業内の各ユーザに適用することができる。たとえば、ユーザは、個々に疑わしいユーザの分類がフィルタトレーニングから除外される交差検定技法にかけることができる。フィルタは、残りのユーザからのデータを使用してトレーニングすることができる。次いで、トレーニングされたフィルタが、除外されたユーザからのメッセージを調べ、そのメッセージをどのように分類しているか判定する。不一致の数が閾値レベルを超えた場合には、その疑わしいユーザは信頼できないものと考えられる。さらに、信頼できないユーザからのメッセージ分類は、データベースおよび／またはフィルタによって受け入れられる前に、手動で検査することができる。そうでない場合は、そのユーザを今後のポーリングから除去することができる。

【００９０】
次に、図９を参照すると、本発明の様々な態様を実施するための例示的な環境９１０は、コンピュータ９１２を含んでいる。コンピュータ９１２は、処理装置９１４、システムメモリ９１６、システムバス９１８を含む。システムバス９１８は、それだけには限らないが、システムメモリ９１６を含むシステム構成要素を処理装置９１４に結合する。処理装置９１４は、様々な使用可能なプロセッサのいずれかとすることができる。デュアルマイクロプロセッサおよび他のマルチプロセッサアーキテクチャもまた、処理装置９１４として使用することができる。
【００９１】
システムバス９１８は、メモリバスもしくはメモリコントローラ、周辺機器バスもしくは外部バス、および／または任意の様々な使用可能なバスアーキテクチャを使用する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、いずれかとすることができ、バスアーキテクチャには、それだけには限らないが、１１ビットバス、ＩＳＡ、ＭＳＡ、ＥＩＳＡ（Ｅｘｔｅｎｄｅｄ　ＩＳＡ）、ＩＤＥ、ＶＥＳＡローカルバス（ＶＬＢ）、ＰＣＩ、ＵＳＢ、ＡＧＰ、ＰＣＭＣＩＡバス、ＳＣＳＩが含まれる。
【００９２】
システムメモリ９１６には、揮発性メモリ９２０および不揮発性メモリ９２２が含まれる。起動中などにコンピュータ９１２内の要素間で情報を転送するための基本ルーチンを含む基本入出力システム（ＢＩＯＳ）は、不揮発性メモリ９２２内に記憶される。限定ではなく例を挙げると、不揮発性メモリ９２２には、読出し専用メモリ（ＲＯＭ）、プログラム可能なＲＯＭ（ＰＲＯＭ）、電気的プログラム可能なＲＯＭ（ＥＰＲＯＭ）、電気的消去可能なＲＯＭ（ＥＥＰＲＯＭ）、またはフラッシュメモリが含まれる。揮発性メモリ９２０には、外部キャッシュメモリとして動作するランダムアクセスメモリ（ＲＡＭ）が含まれる。限定ではなく例を挙げると、ＲＡＭは、シンクロナスＲＡＭ（ＳＲＡＭ）、ダイナミックＲＡＭ（ＤＲＡＭ）、シンクロナスＤＲＡＭ（ＳＤＲＡＭ）、ダブルデータレートＳＤＲＡＭ（ＤＤＲ　ＳＤＲＡＭ）、ＥＳＤＲＡＭ（ｅｎｈａｎｃｅｄ　ＳＤＲＡＭ）、ＳＬＤＲＡＭ（Ｓｙｎｃｈｌｉｎｋ　ＤＲＡＭ）、ダイレクトラムバスＲＡＭ（ＤＲＲＡＭ）など、多数の形態で使用可能である。
【００９３】
コンピュータ９１２はまた、取外し式／非取外し式、揮発性／不揮発性コンピュータ記憶媒体を含む。図９は、たとえば、ディスクストレージ９２４を示す。ディスクストレージ９２４には、それだけには限らないが、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－１００ドライブ、フラッシュメモリカード、メモリースティックのようなデバイスが含まれる。さらに、ディスクストレージ９２４には、それだけには限らないが、コンパクトディスクＲＯＭデバイス（ＣＤ－ＲＯＭ）、記録可能なＣＤドライブ（ＣＤ－Ｒドライブ）、再書込み可能なＣＤドライブ（ＣＤ－ＲＷドライブ）、またはデジタル多用途ディスクＲＯＭドライブ（ＤＶＤ－ＲＯＭ）など光ディスクドライブを含めて、記憶媒体が別個に、または他の記憶媒体との組合せで含まれる可能性がある。ディスク記憶装置９２４の、システムバス９１８に対する接続を容易にするために、インターフェース９２６など取外し式または非取外し式インターフェースが一般に使用される。
【００９４】
図９は、ユーザと、好適な動作環境９１０に述べられている基本的なコンピュータ資源との間の媒介物として動作するソフトウェアについて述べていることを理解されたい。そのようなソフトウェアには、オペレーティングシステム９２８が含まれる。オペレーティングシステム９２８は、ディスクストレージ９２４に記憶することができ、コンピュータシステム９１２の資源を制御し、割り当てるように動作する。システムアプリケーション９３０は、システムメモリ９１６内またはディスクストレージ９２４に記憶されたプログラムモジュール９３２およびプログラムデータ９３４を介して、オペレーティングシステム９２８による資源の管理を利用する。本発明は、様々な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合せと共に実施することができることを理解されたい。
【００９５】
ユーザは、入力デバイス９３６を介してコンピュータ９１２にコマンドまたは情報を入力する。入力デバイス９３６には、それだけには限らないが、マウスなどポインティングデバイス、トラックボール、スタイラス、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、衛星パラボラアンテナ、スキャナ、ＴＶ同調器カード、デジタルカメラ、デジタルビデオカメラ、ウェブカメラなどが含まれる。これら、および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を通じて処理装置９１４に接続する。インターフェースポート９３８には、たとえば、シリアルポート、パラレルポート、ゲームポート、ユニバーサルシリアルバス（ＵＳＢ）が含まれる。出力デバイス９４０は、入力デバイス９３６と同じタイプのポートのいくつかを使用する。したがって、たとえばＵＳＢポートは、コンピュータ９１２に入力を送るために、また、コンピュータ９１２から出力デバイス９４０に情報を出力するために使用することができる。出力アダプタ９４２は、出力デバイス９４０の中でも、特別なアダプタを必要とするモニタ、スピーカ、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ている。限定ではなく例を挙げると、出力アダプタ９４２には、出力デバイス９４０とシステムバス９１８の間で接続手段を提供するビデオカードおよびサウンドカードが含まれる。他のデバイスおよび／またはデバイスのシステムは、リモートコンピュータ９４４など、入力機能と出力機能を共に提供することに留意されたい。
【００９６】
コンピュータ９１２は、リモートコンピュータ９４４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内で動作することができる。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、マイクロプロセッサをベースとする家電、ピアデバイスまたは他の共通ネットワークノードなどとすることができ、一般に、コンピュータ９１２に関して述べられている要素の多数または全部を含む。簡潔にするために、メモリ記憶装置９４６だけがリモートコンピュータ９４４と共に示されている。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を介してコンピュータ９１２に論理的に接続され、次いで、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、ローカルエリアネットワーク（ＬＡＮ）および広域ネットワーク（ＷＡＮ）など、通信ネットワークを含む。ＬＡＮ技術には、光ファイバ分散データインターフェース（ＦＤＤＩ）、より線ＦＤＤＩ（ＣＤＤＩ）、イーサネット（登録商標）／ＩＥＥＥ１１０２．３、トークンリング／ＩＥＥＥ１１０２．５などが含まれる。ＷＡＮ技術には、それだけには限らないが、ポイント・トゥ・ポイント・リンク、ＩＳＤＮとその変形形態のような回路交換ネットワーク、パケット交換ネットワーク、およびデジタル加入者回線（ＤＳＬ）が含まれる。
【００９７】
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に接続するために使用されるハードウェア／ソフトウェアを指す。通信接続９５０は、図が見やすいようにコンピュータ９１２の内側で示されているが、コンピュータ９１２の外側とすることもできる。例示する目的にすぎないが、ネットワークインターフェース９４８に対する接続に必要なハードウェア／ソフトウェアには、通常の電話級モデム、ケーブルモデムおよびＤＳＬモデムを含むモデム、ＩＳＤＮアダプタ、並びにイーサネット（登録商標）カードなど、内部技術および外部技術が含まれる。
【００９８】
図１０は、本発明が相互作用することができるコンピューティング環境例１０００の概略ブロック図である。システム１０００は、１つまたは複数のクライアント１０１０を含む。クライアント１０１０は、ハードウェアおよび／またはソフトウェアとすることができる（たとえば、スレッド、プロセス、コンピューティングデバイス）。また、システム１０００は、１つまたは複数のサーバ１０３０を含む。サーバ１０３０もまた、ハードウ

ェアおよび／またはソフトウェアとすることができる（たとえば、スレッド、プロセス、コンピューティングデバイス）。サーバ１０３０は、たとえば、本発明を使用することによって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１０１０とサーバ１０３０の間の、１つの可能な通信は、２つ以上のコンピュータプロセス間で伝送されるように適合されたデータパケットの形態にあるものとすることができる。システム１０００は、クライアント１０１０とサーバ１０３０の間の通信を容易にするために使用することができる通信フレームワーク１０５０を含む。クライアント１０１０は、クライアント１０１０のローカルな情報を記憶するために使用することができる１つまたは複数のクライアントデータストア１０６０に動作可能に接続される。同様に、サーバ１０３０は、サーバ１０３０のローカルな情報を記憶するために使用することができる１つまたは複数のサーバデータストア１０４０に動作可能に接続される。

【００９９】

上述したものには、本発明の諸例が含まれる。当然ながら、本発明について述べるために構成要素または方法の考えられるあらゆる組合せについて述べることは可能でなく、本発明に関する多数の他の組合せおよび変形が可能であることを、当業者なら理解することができる。したがって、本発明は、添付した特許請求の範囲の精神および範囲内に入るそのような変更、修正、並びに変形形態をすべて包含するものとする。さらに「ｉｎｃｌｕｄｅｓ（含む）」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ている限り、そのような用語は、「ｃｏｍｐｒｉｓｉｎｇ（含む、備える）」が特許請求の範囲内で移行句として使用されたとき解釈されるように「ｃｏｍｐｒｉｓｉｎｇ」という用語と同様に包括的であるものとする。

【図面の簡単な説明】

【０１００】

【図１Ａ】本発明の一態様によるフィードバックループトレーニングシステムのブロック図である。

【図１Ｂ】本発明の一態様による例示的なフィードバックループトレーニングプロセスの流れ図である。

【図２】本発明の一態様による、スパムフィルタを作成するためにユーザによるメール分類を容易にする例示的な方法の流れ図である。

【図３】本発明の一態様による、図２の方法に参加するユーザの交差検定を容易にする例示的な方法の流れ図である。

【図４】本発明の一態様による、ユーザが信頼できないかどうか判定することを容易にする例示的な方法の流れ図である。

【図５】本発明の一態様による、スパムを捕らえ、スパム発信元を決定するのを容易にする例示的な方法の流れ図である。

【図６】本発明の一態様による、クライアントをベースとするフィードバックループアーキテクチャのブロック図である。

【図７】本発明の一態様による、トレーニングデータを生成する１人または複数のユーザを有する、サーバをベースとするフィードバックループシステムのブロック図である。

【図８】本発明の一態様による、外部ユーザデータベースに記憶されたトレーニングデータを引き出すために、それ自体のデータベースを有する内部サーバをシステムが含む、組織横断サーバベースフィードバックループシステムのブロック図である。

【図９】本発明の様々な態様を実施するための例示的な環境を示す図である。

【図１０】本発明による例示的な通信環境の概略ブロック図である。

【符号の説明】

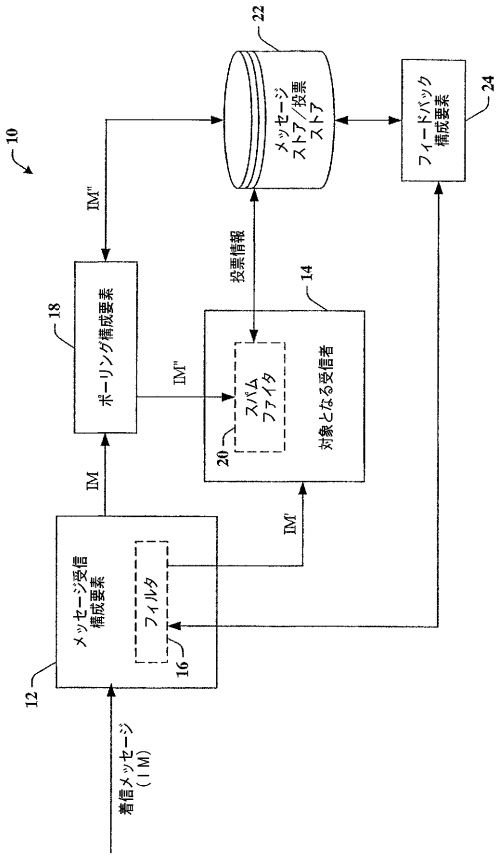
【０１０１】

１４　受信者

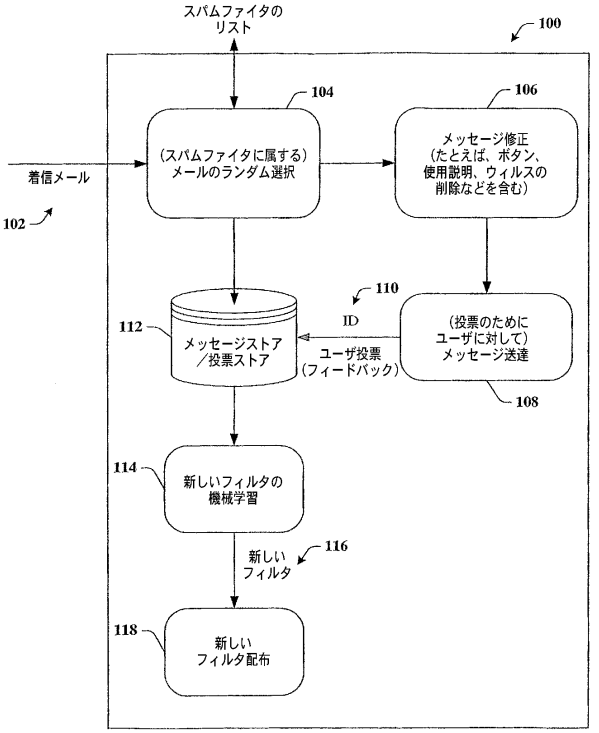
６１０、７１２、８１５　LAN

７００、８００　フィードバックシステム

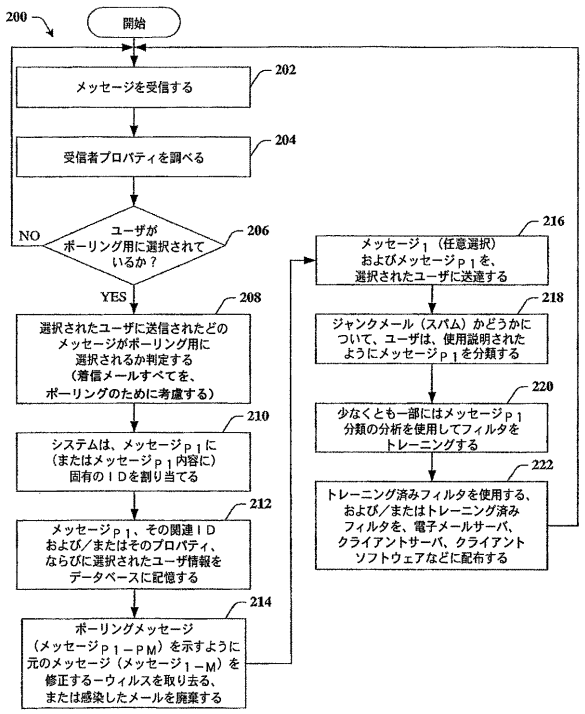
【図１Ａ】


10
22
メッセージストア／投票ストア
フィードバック構成要素
24
IM''
投票情報
ポーリング構成要素
18
IM''
14
スパムファイタ
20
対象となる受信者
IM
IM'
メッセージ受信構成要素
フィルタ
16
12
着信メッセージ（ＩＭ）

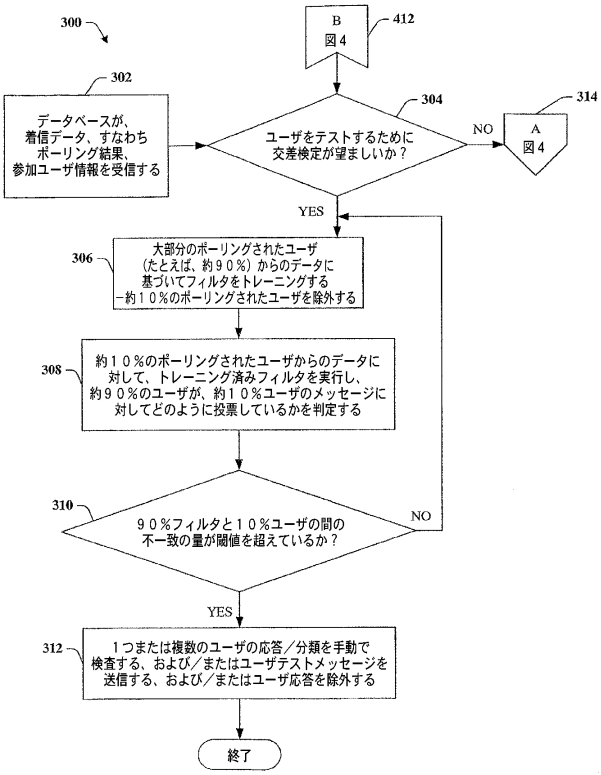
【図１Ｂ】


スパムファイタのリスト
100
104
106
（スパムファイタに属する）メールのランダム選択
メッセージ修正（たとえば、ボタン、使用説明、ウィルスの削除などを含む）
着信メール
102
110
112
ID
メッセージストア／投票ストア
ユーザ投票（フィードバック）
（投票のためにユーザに対して）メッセージ送達
108
114
新しいフィルタの機械学習
新しいフィルタ
116
118
新しいフィルタ配布

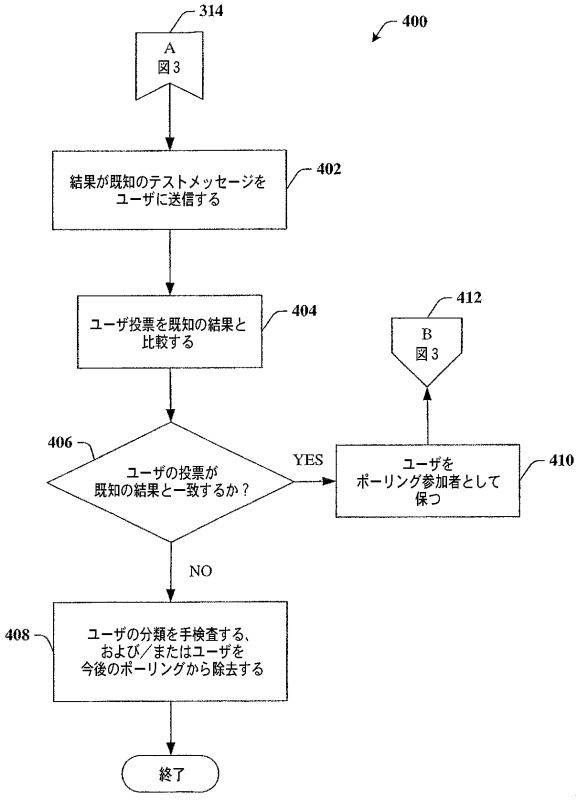
【図２】


200
開始
メッセージを受信する
202
受信者プロパティを調べる
204
NO
ユーザがポーリング用に選択されているか？
206
YES
選択されたユーザに送信されたどのメッセージがポーリング用に選択されるか判定する（着信メールすべてを、ポーリングのために考慮する）
208
システムは、メッセージＰ１に（またはメッセージＰ１内容に）固有のＩＤを割り当てる
210
メッセージＰ１、その関連ＩＤおよび／またはそのプロパティ、ならびに選択されたユーザ情報をデータベースに記憶する
212
ポーリングメッセージ（メッセージＰ１－ＰＭ）を示すように元のメッセージ（メッセージ１－Ｍ）を修正する－ウィルスを取り去る、または感染したメールを廃棄する
214
メッセージ１（任意選択）およびメッセージＰ１を、選択されたユーザに送達する
216
ジャンクメール（スパム）かどうかについて、ユーザは、使用説明されたようにメッセージＰ１を分類する
218
少なくとも一部にはメッセージＰ１分類の分析を使用してフィルタをトレーニングする
220
トレーニング済みフィルタを使用する、および／またはトレーニング済みフィルタを、電子メールサーバ、クライアントサーバ、クライアントソフトウェアなどに配布する
222

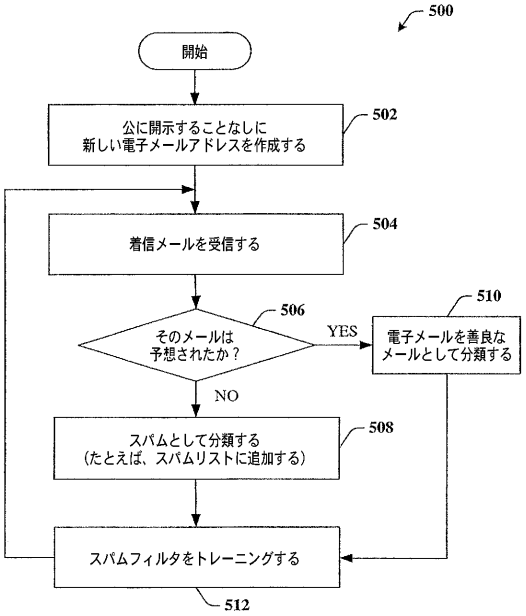
【図３】


300
B 図４
412
302
データベースが、着信データ、すなわちポーリング結果、参加ユーザ情報を受信する
304
ユーザをテストするために交差検定が望ましいか？
NO
A 図４
314
YES
306
大部分のポーリングされたユーザ（たとえば、約９０％）からのデータに基づいてフィルタをトレーニングする－約１０％のポーリングされたユーザを除外する
308
約１０％のポーリングされたユーザからのデータに対して、トレーニング済みフィルタを実行し、約９０％のユーザが、約１０％ユーザのメッセージに対してどのように投票しているかを判定する
310
９０％フィルタと１０％ユーザの間の不一致の量が閾値を超えているか？
NO
YES
312
１つまたは複数のユーザの応答／分類を手動で検査する、および／またはユーザテストメッセージを送信する、および／またはユーザ応答を除外する
終了

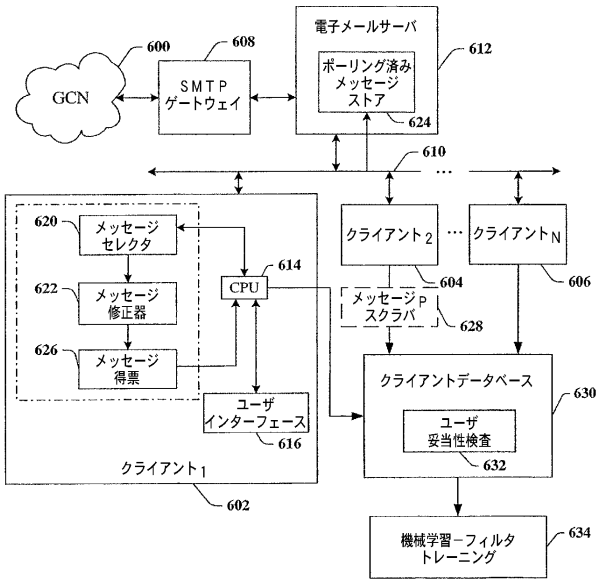
【図４】


400
314
A
図3
結果が既知のテストメッセージを
ユーザに送信する
402
ユーザ投票を既知の結果と
比較する
404
412
B
図3
406
ユーザの投票が
既知の結果と一致するか？
YES
ユーザを
ポーリング参加者として
保つ
410
NO
408
ユーザの分類を手検査する、
および／またはユーザを
今後のポーリングから除去する
終了

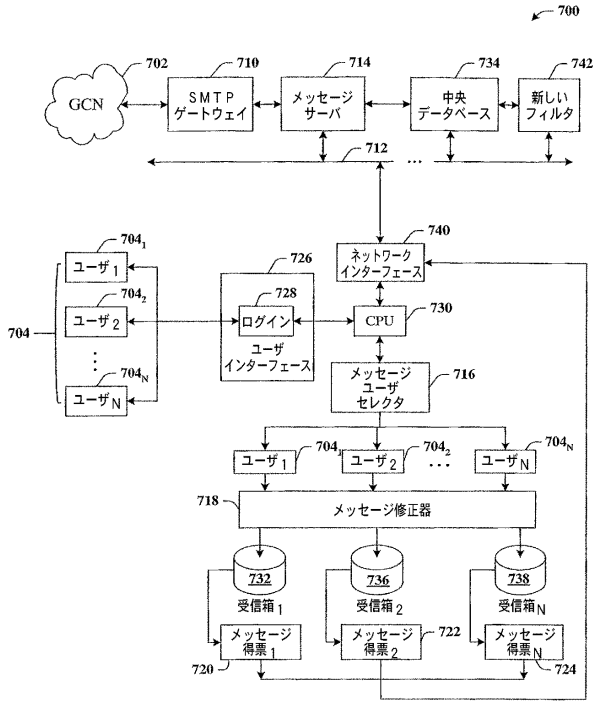
【図５】


500
開始
公に開示することなしに
新しい電子メールアドレスを作成する
502
着信メールを受信する
504
506
そのメールは
予想されたか？
YES
510
電子メールを善良な
メールとして分類する
NO
スパムとして分類する
（たとえば、スパムリストに追加する）
508
スパムフィルタをトレーニングする
512

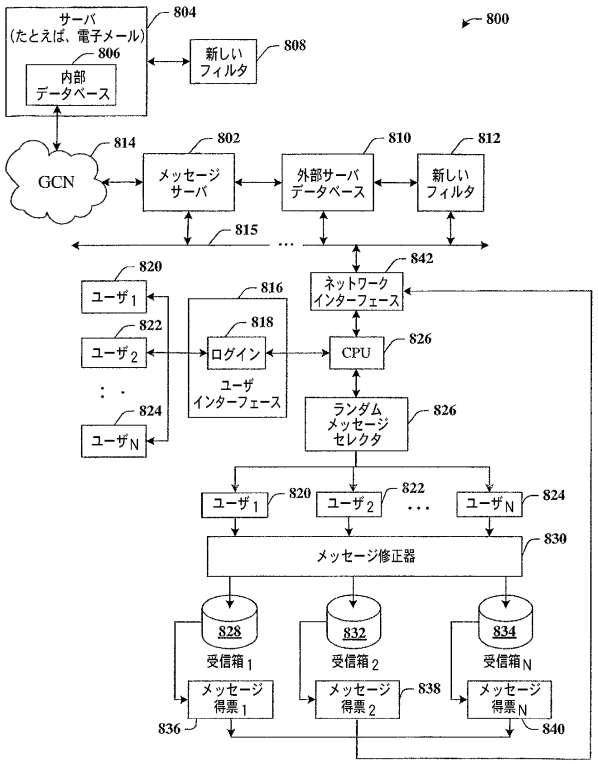
【図６】


GCN
600
SMTP
ゲートウェイ
608
電子メールサーバ
ポーリング済み
メッセージ
ストア
624
612
610
620
メッセージ
セレクタ
622
メッセージ
修正器
626
メッセージ
得票
CPU
614
クライアント2
604
クライアントN
606
メッセージP
スクラバ
628
ユーザ
インターフェース
616
クライアント1
602
クライアントデータベース
ユーザ
妥当性検査
632
630
機械学習－フィルタ
トレーニング
634

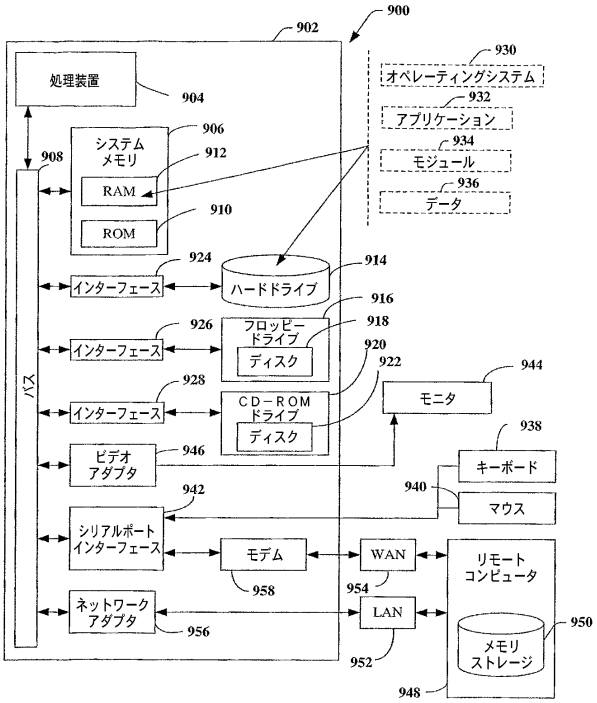
【図７】


700
GCN
702
SMTP
ゲートウェイ
710
メッセージ
サーバ
714
中央
データベース
734
新しい
フィルタ
742
712
704
ユーザ1
ユーザ2
ユーザN
726
ログイン
ユーザ
インターフェース
728
ネットワーク
インターフェース
740
CPU
730
メッセージ
ユーザ
セレクタ
716
718
メッセージ修正器
732
受信箱1
736
受信箱2
738
受信箱N
メッセージ
得票1
メッセージ
得票2
722
メッセージ
得票N
720
724

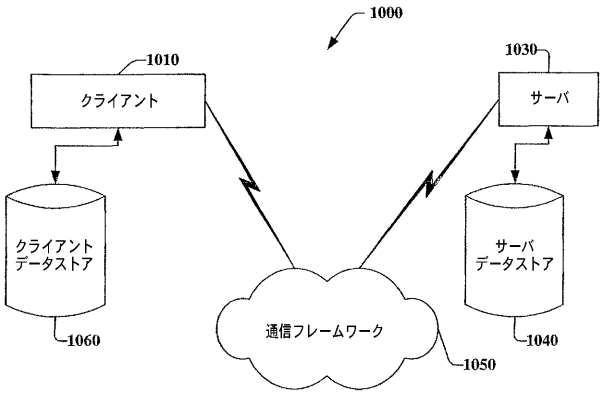
【図８】


800
サーバ
(たとえば、電子メール)
804
内部
データベース
806
新しい
フィルタ
808
GCN
814
メッセージ
サーバ
802
外部サーバ
データベース
810
新しい
フィルタ
812
815
ネットワーク
インターフェース
842
ユーザ1
820
ユーザ2
822
ユーザN
824
ログイン
818
ユーザ
インターフェース
816
CPU
826
ランダム
メッセージ
セレクタ
826
メッセージ修正器
830
828
受信箱1
832
受信箱2
834
受信箱N
メッセージ
得票1
メッセージ
得票2
838
メッセージ
得票N
836
840

【図９】


900
902
処理装置
904
システム
メモリ
906
RAM
912
ROM
910
908
バス
インターフェース
924
ハードドライブ
914
インターフェース
926
フロッピー
ドライブ
ディスク
916
918
インターフェース
928
CD-ROM
ドライブ
ディスク
920
922
ビデオ
アダプタ
946
シリアルポート
インターフェース
942
ネットワーク
アダプタ
956
モデム
958
WAN
954
LAN
952
リモート
コンピュータ
950
メモリ
ストレージ
948
オペレーティングシステム
930
アプリケーション
932
モジュール
934
データ
936
モニタ
944
キーボード
938
マウス
940

【図１０】


1000
1010
クライアント
1030
サーバ
クライアント
データストア
1060
通信フレームワーク
1050
サーバ
データストア
1040

フロントページの続き

(72)発明者 ロバート エル. ラウンスウェイト
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 デビッド イー. ヘッカーマン
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 ジョン ディー. メアー
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 ネーザン ディー. ハウエル
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 コーポレーション内
(72)発明者 ミカ シー. ルパーズバーグ
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 ディーン エー. スローソン
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 ジョシュア ティー. グッドマン
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

合議体
審判長 和田 志郎
審判官 中野 裕二
審判官 安久 司郎

(56)参考文献 米国特許第６４２１７０９（ＵＳ，Ｂ１）
特開平１０－７４１７２（ＪＰ，Ａ）

(58)調査した分野(Int.Cl., ＤＢ名)
G06F 13/00
H04L 12/58
H04L 12/66