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

特許第5744228号
(P5744228)

(45)発行日　平成27年7月8日(2015.7.8)　(24)登録日　平成27年5月15日(2015.5.15)

(51)Int.Cl.　　F I
G06F 13/00 (2006.01)　　G06F 13/00 540E
G06F 17/27 (2006.01)　　G06F 17/27

請求項の数 16　(全 14 頁)

(21)出願番号　特願2013-545039(P2013-545039)
(86)(22)出願日　平成23年12月26日(2011.12.26)
(65)公表番号　特表2014-502754(P2014-502754A)
(43)公表日　平成26年2月3日(2014.2.3)
(86)国際出願番号　PCT/CN2011/084699
(87)国際公開番号　WO2012/083892
(87)国際公開日　平成24年6月28日(2012.6.28)
審査請求日　平成26年3月28日(2014.3.28)
(31)優先権主張番号　201010621142.1
(32)優先日　平成22年12月24日(2010.12.24)
(33)優先権主張国　中国(CN)

早期審査対象出願

(73)特許権者　507231932
北大方正集▲団▼有限公司
PEKING UNIVERSITY FOUNDER GROUP CO., LTD
中華人民共和国北京市▲海▼淀区成府路２９８号中▲関▼村方正大厦５▲層▼
5 Floor, Zhongguancun Founder Building, No. 298, Chengfu Road, Haidian District, Beijing 100871, China

最終頁に続く

(54)【発明の名称】インターネットにおける有害情報の遮断方法と装置

(57)【特許請求の範囲】
【請求項１】
インターネットにおける有害情報の遮断方法であって、
遮断待ちテキスト情報、システムプレリサーチモデル情報及びユーザーフィードバックモデル情報を取得するステップと、
前記遮断待ちテキスト情報を前処理するステップと、
前記前処理された遮断待ちテキスト情報と前記システムプレリサーチモデル情報とに対して特徴情報マッチングし、第一マッチング結果を取得するステップと、
前記前処理された遮断待ちテキスト情報と前記ユーザーフィードバックモデル情報とに対して特徴情報マッチングし、前記第一マッチング結果から独立した第二マッチング結果を取得するステップと、
第一マッチング結果と第二マッチング結果とが一致しているか否かに基いて、前記遮断待ちテキスト情報を遮断するステップと、
を備えることを特徴とする方法。
【請求項２】
更に、
前記システムプレリサーチモデル情報のコーパス及び前記ユーザーフィードバックモデル情報のコーパスを取得するステップ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】
前記ユーザーフィードバックモデル情報のコーパスには、ユーザーフィードバックコー

パス及び/または被遮断コーパスが含まれることを特徴とする請求項２に記載の方法。
【請求項４】
更に、
前記ユーザーフィードバックモデル情報のコーパス量及びそれに対応する閾値を取得するステップと、
前記ユーザーフィードバックモデル情報のコーパス量及びそれに対応する閾値に基づいて、前記ユーザーフィードバックモデル情報を更新するステップと、
を備えることを特徴とする請求項３に記載の方法。
【請求項５】
前記遮断待ちテキスト情報を前処理するステップにおいては、
前記遮断待ちテキスト情報に対してセグメント処理をし、
前記セグメント処理がされた特徴項候補量を統計することを特徴とする請求項２、３または４に記載の方法。
【請求項６】
前記前処理された遮断待ちテキスト情報と前記システムプレリサーチモデル情報とに対して特徴情報マッチングし、第一マッチング結果を取得するステップにおいては、
前記前処理された遮断待ちテキスト情報及び前記システムプレリサーチモデル情報を取得し、
前記前処理された遮断待ちテキスト情報と前記システムプレリサーチモデル情報とをマッチングし、特徴項を取得し、
前記特徴項のコーパス情報の得点を計算し、
コーパス情報の得点に基づいて、前記特徴項に対応する遮断待ちテキスト情報が有害情報であるかどうかを判断し、
判断結果に基づいて、第一マッチング結果を取得することを特徴とする請求項５に記載の方法。
【請求項７】
前記前処理された遮断待ちテキスト情報と前記ユーザーフィードバックモデル情報とに対して特徴情報マッチングし、第二マッチング結果を取得するステップにおいては、
前記前処理された遮断待ちテキスト情報及び前記ユーザーフィードバックモデル情報を取得し、
前記前処理された遮断待ちテキスト情報と前記ユーザーフィードバックモデル情報とをマッチングし、特徴項を取得し、
前記特徴項のコーパス情報の得点を計算し、
コーパス情報の得点に基づいて、前記特徴項に対応する遮断待ちテキスト情報が有害情報であるかどうかを判断し、
判断結果に基づいて、第二マッチング結果を取得することを特徴とする請求項６に記載の方法。
【請求項８】
前記システムプレリサーチモデル情報は規則索引データベースとシステムリサーチモデルの特徴項情報とを含んでおり、
前記ユーザーフィードバックモデル情報は規則索引データベースとユーザーフィードバックモデルの特徴項情報とを含むことを特徴とする請求項６または７に記載の方法。
【請求項９】
前記システムプレリサーチモデル情報の規則索引データベースは、システムプリセット規則を含んでおり、
前記ユーザーフィードバックモデル情報の規則索引データベースは、ユーザー設定規則を含むことを特徴とする請求項８に記載のインターネットにおける有害情報の遮断方法。
【請求項１０】
インターネットにおける有害情報の遮断装置であって
遮断待ちテキスト情報、システムプレリサーチモデル情報及びユーザーフィードバック

モデル情報を取得する情報取得モジュールと、
前記遮断待ちテキスト情報を前処理する前処理モジュールと、
前記前処理された遮断待ちテキスト情報と前記システムプレリサーチモデル情報とに対して特徴情報マッチングし、第一マッチング結果を取得する第一マッチングモジュールと、
前記前処理された遮断待ちテキスト情報と前記ユーザーフィードバックモデル情報とに対して特徴情報マッチングし、前記第一マッチング結果から独立した第二マッチング結果を取得する第二マッチングモジュールと、
第一マッチング結果と第二マッチング結果とが一致しているか否かに基いて、前記遮断待ちテキスト情報を遮断する遮断モジュールと、
を備えることを特徴とする装置。
【請求項１１】
前記情報取得モジュールは、更に、前記ユーザーフィードバックモデル情報のコーパスを取得することを特徴とする請求項１０に記載の装置。
【請求項１２】
前記ユーザーフィードバックモデル情報のコーパスには、ユーザーフィードバックコーパス及び/または被遮断コーパスが含まれることを特徴とする請求項１１に記載の装置。
【請求項１３】
更に、
前記ユーザーフィードバックモデル情報のコーパス量及びそれに対応する閾値を取得する閾値取得モジュールと、
前記ユーザーフィードバックモデル情報のコーパス量及びそれに対応する閾値に基づいて、前記ユーザーフィードバックモデル情報を更新する更新モジュールと、
を備えることを特徴とする請求項１２に記載の装置。
【請求項１４】
前記前処理モジュールは、
前記遮断待ちテキスト情報に対してセグメント処理をするセグメントサブモジュールと、
前記セグメント処理された特徴項候補量を統計する統計サブモジュールと、
を備えることを特徴とする請求項１１、１２または１３に記載の装置。
【請求項１５】
前記第一マッチングモジュールは、
前記前処理された遮断待ちテキスト情報及び前記システムプレリサーチモデル情報を取得する情報取得サブモジュールと、
前記前処理された遮断待ちテキスト情報と前記システムプレリサーチモデル情報とをマッチングし、特徴項を取得するマッチングサブモジュールと、
前記特徴項のコーパス情報の得点を計算する統計サブモジュールと、
コーパス情報の得点に基づいて、前記特徴項に対応する遮断待ちテキスト情報が有害情報であるかどうかを判断する判断サブモジュールと、
判断結果に基いて、第一マッチング結果を取得する結果出力サブモジュールと、
を備えることを特徴とする請求項１４に記載の装置。
【請求項１６】
前記第二マッチングモジュールは、
前記前処理された遮断待ちテキスト情報及び前記システムプレリサーチモデル情報を取得する情報取得サブモジュールと、
前記前処理された遮断待ちテキスト情報と前記ユーザーフィードバックモデル情報とをマッチングし、特徴項を取得するマッチングサブモジュールと、
前記特徴項のコーパス情報の得点を統計する統計サブモジュールと、
コーパス情報の得点に基づいて、前記特徴項に対応する遮断待ちテキスト情報が有害情報であるかどうかを判断する判断サブモジュールと、

判断結果に基づいて、第二マッチング結果を取得する結果出力サブモジュールと、
を備えることを特徴とする請求項１５に記載の装置。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、コンピューター情報処理及び情報遮断の技術に関し、特に、統計と規則に基づくインターネットにおける有害情報の遮断方法と装置に関する。
【背景技術】
【０００２】
インターネットが迅速に発展するにつれて、情報を伝播するスピードも速くなる。インターネットにおいて、様々なコンテンツが混在しているため、広告、色情、暴力などの有害情報を禁止することが難しい。そして、このような有害情報はますますもっと隠蔽の形で拡散されているため、有害情報の拡散を抑制し、インターネット空間を浄化することが非常に重要である。インターネットにおける膨大なデータ情報は、人工的にはインターネットにおける有害情報を遮断する場合、極めて大量の労力と財力を必要とする。そのため、近年、インターネットにおける有害情報を自動的に遮断する技術の研究が注目されている。
【０００３】
現在、一般的には、インターネットにおける有害情報を自動的に遮断する技術としては、下記の二つの方法が取り上げられる。
（１）キーワードマッチングに基づく遮断方法
判定プロセスにおいて、この方法は精確なマッチング法でキーワードがあるテキストを遮断する。当該方法が採用される場合に、インターネットにおける有害情報を速く遮断でき、簡単で使いやすい。
（２）統計のテキスト分類モデルに基づく遮断方法
この方法において、本質的には、統計に基づく有害テキストの遮断モデルはテキストを二種類に分類する。テキスト分類は自然言語の処理領域における重要な研究方向であり、大量のモデルが参考にできる。理論上、統計のテキスト分類モデルは、効果的であるはずであるが、実際の適用時には性能が望ましくない。誤判断の場合がよくあり、主な原因が下記で示され、
（１）順方向コーパス（corpus）と逆方向コーパスはバランスが取れていない。その中、順方向コーパスに少量の種類しか含まれていなく、例えば、広告、色情、暴力など、ユーザーが関心を持っている有害情報はメインである。一方、逆方向コーパスには、大量の種類が含まれており、テキスト内容によって分類すると、経済、体育、政治、医薬、アート、歴史、政治、文化、環境、交通、コンピューター、教育、軍事などが分けられている。
（２）有害情報の内容の表現は非常に隠蔽で変わりやすい。伝播者は常に通常の言語をわざと避け、代わりに、同音字、漢字分解、略字、造語などが使用されている。
（３）ユーザー辞書にキーワードを精確にマッチングする方法しか提供されないため、判定方法は機械的で融通性がなくなる。しかも、単一のキーワードの単語感情極性は代表的なものではなく、誤判断率が高い。例えば、「免費(無料)」と「発票(インボイス)」が同時にコンテキストに現れる場合は、単一の「発票(インボイス)」より説得力がある。
（４）従来の中国語情報処理方法はテキスト分類に基づく有害情報の遮断には適用できない。例えば、一定規模の禁止用語の使用や、特徴項に二文字以上の語彙しか含まれないなど。
（５）広告、色情、暴力などの有害情報を総合的に遮断するための統一的モデルがない。
【０００４】
上記したインターネットにおける有害情報の自動的遮断を実現するプロセスにおいて、従来の技術では、現在のインターネットからの要請を満足できなく、そして、自動的な更

新も実現できない。
【発明の概要】
【０００５】
本発明は、インターネットにおける有害情報の遮断方法と装置を提供することを目的とする。
本発明は、このような目的を達成するために、インターネットにおける有害情報の遮断方法であって、遮断待ちテキスト情報、システムプレリサーチ（pre-research）モデル情報及びユーザーフィードバックモデル情報を取得するステップと、　前記遮断待ちテキスト情報を前処理するステップと、　前記前処理された遮断待ちテキスト情報と前記システムプレリサーチモデル情報とに対して特徴情報マッチングし、第一マッチング結果を取得するステップと、前記前処理された遮断待ちテキスト情報と前記ユーザーフィードバックモデル情報とに対して特徴情報マッチングし、第二マッチング結果を取得するステップと、第一マッチング結果と第二マッチング結果に基いて、前記遮断待ちテキスト情報を遮断するステップと、を備えることを特徴とする。
【０００６】
また、本発明は、このような目的を達成するために、インターネットにおける有害情報の遮断装置であって、遮断待ちテキスト情報、システムプレリサーチモデル情報及びユーザーフィードバックモデル情報を取得する情報取得モジュールと、　前記遮断待ちテキスト情報を前処理する前処理モジュールと、　前記前処理された遮断待ちテキスト情報と前記システムプレリサーチモデル情報とに対して特徴情報マッチングし、第一マッチング結果を取得する第一マッチングモジュールと、前記前処理された遮断待ちテキスト情報と前記ユーザーフィードバックモデル情報とに対して特徴情報マッチングし、第二マッチング結果を取得する第二マッチングモジュールと、第一マッチング結果と第二マッチング結果に基いて、前記遮断待ちテキスト情報を遮断する遮断モジュールと、を備えることを特徴とする。
【０００７】
以上のように、本発明は、遮断待ちテキスト情報、システムプレリサーチモデル情報及びユーザーフィードバックモデル情報を取得し、前記遮断待ちテキスト情報を前処理し、前記前処理された遮断待ちテキスト情報と前記システムプレリサーチモデル情報とに対して特徴情報マッチングし、第一マッチング結果を取得し、前記前処理された遮断待ちテキスト情報と前記ユーザーフィードバックモデル情報とに対して特徴情報マッチングし、第二マッチング結果を取得し、第一マッチング結果と第二マッチング結果に基いて、前記遮断待ちテキスト情報を遮断する。二回のマッチングによって遮断処理が行われるため、システムにおいて有害情報を自動的に遮断する正確性が高く、システムの性能を向上することができる。また、本発明はユーザーフィードバックモデル情報を利用して有害情報を遮断するため、ユーザーフィードバック情報を適時に有害情報の自動的遮断プロセスに適用することが出来、システムモデル情報の自動的更新機能を実現できる。
【図面の簡単な説明】
【０００８】
【図１】本発明の実施例で、インターネットにおける有害情報の遮断方法を示すフローチャートである。
【図２】本発明の更なる実施例で、インターネットにおける有害情報の遮断方法を示すフローチャートである。
【図３】本発明の実施例で、インターネットにおける有害情報の遮断装置の構成を示す模式図である。
【図４】本発明の更なる実施例で、インターネットにおける有害情報の遮断装置の構成を示す模式図である。
【発明を実施するための形態】
【０００９】
以下、図面を参照しながら、実施例を使って本発明に係るインターネットにおける有害

情報の遮断方法と装置を詳細に説明する。

【００１０】

図１で示されるように、本発明の一実施例はインターネットにおける有害情報の遮断方法であって、遮断待ちテキスト情報、システムプレリサーチモデル情報及びユーザーフィードバックモデル情報を取得するステップ１０１と、　前記遮断待ちテキスト情報を前処理するステップ１０２と、　前記前処理された遮断待ちテキスト情報と前記システムプレリサーチモデル情報とに対して特徴情報マッチングし、第一マッチング結果を取得するステップ１０３と、　前記前処理された遮断待ちテキスト情報と前記ユーザーフィードバックモデル情報とに対して特徴情報マッチングし、第二マッチング結果を取得するステップ１０４と、第一マッチング結果と第二マッチング結果に基いて、前記遮断待ちテキスト情報を遮断するステップ１０５と、を備える。

【００１１】

図２で示されるように、本発明の更なる実施例はインターネットにおける有害情報の遮断方法であって、下記の各ステップを備える。すなわち、

ステップ２０１：前記システムプレリサーチモデル情報のコーパス及びユーザーフィードバックモデル情報のコーパスを取得する。ここで、前記ユーザーフィードバックモデル情報のコーパスには、ユーザーフィードバックコーパス及び/または被遮断コーパスが含まれる。通常、前記システムプレリサーチモデル情報及びユーザーフィードバックモデル情報の学習コーパスには、順方向コーパスと逆方向コーパスとが含まれる。順方向コーパスとしては、例えば、広告、色情、暴力などの有害情報を含むテキストが１００００件用意される。一方、逆方向コーパスとしては、例えば、経済、政治、体育、文化、医薬、交通、環境、軍事、アート、歴史、コンピューター、教育、法律、不動産、科学技術、自動車、人材、娯楽などの非有害情報を含むテキストが３００００件用意される。

ここで、前記学習コーパスの収集において、順方向コーパスと逆方向コーパスはバランスが取れていない場合がよくあり、一方は範囲が広すぎるが、もう一方は範囲が狭すぎる。本発明において、このようなバランスが取れていないコーパスの分布が許容される。コーパス範囲が広い場合は、量ではなく、できるだけ多くの種類を確保しながら用意する。

【００１２】

ステップ２０２：遮断待ちテキスト情報、システムプレリサーチモデル情報及びユーザーフィードバックモデル情報を取得する。

【００１３】

ステップ２０３：前記遮断待ちテキスト情報を前処理する。

このステップにおいては、前記遮断待ちテキスト情報に対してセグメント処理をする。例えば、句読点と常用語に基づいて、コーパスを区切る。常用語とは、よく使用され判定には無意味な語彙であり、例えば「的」、「了」など。しかし、「

您

（貴方）」はよく順方向コーパスに、「我（私）」はよく逆方向コーパスに使用されるが、いずれも常用語に使用されない。

ここで、自然言語処理においては、よく用いられる禁止用語リストが常用語リストとして適用されない。通常、「方正智思分詞4.0（ペキンファンダー社が開発ソフトウェア）」によって、コーパスに対してセグメントや品詞分類をすることができる。前記セグメント処理されたセグメントユニットは後工程における最小の処理単位である。

前記セグメント処理された特徴項候補量を統計する。例えば、前記セグメント処理されたセグメントユニットにおける非漢字部分を統計し、前記セグメントユニットの合計をＮ１、非漢字部分の合計をＮ２とする場合、Ｎ２/Ｎ１が閾値より大きければ、当該特徴項候補に対応する遮断待ちテキスト情報は有害情報と判断される。判断の理由としては、大量のノイズ文字がこの情報に含まれ、広告などのスパムテキストであるかもしれない。もしくは、前記セグメントユニットにおける、広告によく用いられるＵＲＬ、電話番号、電子メールアドレス、ＱＱ等の連絡方法の数量num（ad）を統計し、デフォルトウェートsco

re_{ad}を与える。

【００１４】

ステップ２０４：前記前処理された遮断待ちテキスト情報と前記システムプレリサーチモデル情報とに対して特徴情報マッチングし、第一マッチング結果を取得する。このステップにおいては、

ステップ２０４１：前記前処理された遮断待ちテキスト情報及び前記システムプレリサーチモデル情報を取得する。前記システムプレリサーチモデル情報に規則索引データベースと前記システムリサーチモデルの特徴項情報とが含まれる。具体的には、前記規則索引データベースにおけるユーザー規則索引データベースとユーザーキーワード索引データベースが生成されるプロセスは以下のようである。すなわち、

ステップＳ１：キーワード解析。まず、常用漢字のピンインの索引を作成し、キーワードにおける各字のピンインの索引に基づいてキーワード全体の索引を生成する。それから、キーワードにおける各字に対して構造的に分解し、分解された結果に基いて、キーワードを再帰し再組合せする。最後、キーワードの索引と、分解の集合によってキーバリューペア（key value pair）を形成させ、全ての解析結果を保存し、ユーザーキーワードの索引データベースを生成する。例えば「法輪功」は、キーワード解析後に、一つの索引値が生成され、しかも幾つかの分解結果がある。具体的には、「三去車侖工力」、「法車侖功」などが含まれる。

【００１５】

ステップＳ２：文法解析。コンピューターによって規則文法を、処理できる形に解析する。前記規則文法には、ＡＮＤと、ＯＲと、ＮＥＡＲと、ＮＯＴとが含まれる。例えば、「Ａ　ＡＮＤ　Ｂ」の場合、ＡとＢは解析待ちのキーワードであり、ＡＮＤ文法とはＡとＢが同時にコンテキストに現れる場合に、当該規則はマッチングに成功である。キーワードと規則文法に対してキーバリューペアを形成し、全ての解析結果を保存しユーザー規則索引データベースを生成する。

【００１６】

ここで、上記した規則索引データベースにおいては、ユーザーが設定した規則でも良いし、システムのプリセット規則でもよい。上記したステップはユーザー設定規則を解析し相応する索引データベースを生成するプロセスであり、当該索引データベースは以下のマッチングプロセスを最適化できる。

【００１７】

ステップ２０４２：前記前処理された遮断待ちテキスト情報と前記システムプレリサーチモデル情報とをマッチングし、特徴項を取得する。ここで、前記システムプレリサーチモデル情報には、規則索引データベースと前記システムプレリサーチモデル特徴項の情報とが含まれ、具体的には、システムプレリサーチモデル特徴項の情報を取得するプロセスは、以下のようである。すなわち、

ステップＳ１：前記セグメントユニットを文字列に組合せし、特徴項候補とする。

（例１）：連続的なセグメントユニットを文字列に組合せする場合。

各文のセグメントユニットに対して、一番目のセグメントユニットから、組合せウィンドウの最大値をＮとして組合せする。順序があるセグメントユニット「ＡＢＣＤ」を例として挙げると、組合せウィンドウの最大値が３である場合に、文字列の組合せはＡＢＣ、ＢＣＤ、ＡＢ、ＢＣ、ＣＤ、Ａ、Ｂ、Ｃ、Ｄとの九つがある。

（例２）：

非連続のセグメントユニットを文字列に組合せする場合。

例１で組合せの文字列に対してピンインの索引を計算し、前記ステップ２０４１におけるステップＳ１で生成されたユーザーキーワードの索引データベースに基づいてマッチングする。マッチング成功の集合があれば、マッチング成功の数量num（user）を統計する。それから、前記ステップ２０４１におけるステップＳ２で生成されたユーザー規則索引データベースに基づいてマッチングし、マッチング成功すれば、非連続のセグメントユニットに対して一つの文字列が生成される。例えば、例１における九つの文字列。ユーザー

キーワードの索引データベースにおいて、二つの文字列Ａ、Ｄがマッチング成功する。ユーザー規則索引データベースに規則「Ａ　ＮＥＡＲ２　Ｄ」がある場合に、特徴項ＡＤが新たに生成される。ここに、２はＡとＤの距離は２以下の意味とする。マッチング成功の数量num（user)を累計し、デフォルトウェートscore$_{user}$を与える。

【００１８】

ステップＳ２：前記特徴項候補を頻度によって遮断する。具体的には、学習コーパスに特徴項候補が現れる回数を統計し、頻度に従って遮断し、頻度が閾値以上の特徴項候補を残しておき、頻度が閾値未満の特徴項候補を削除し、閾値を調整することによって、残す範囲を制御する。

【００１９】

ステップＳ３：前記特徴項候補を頻度によって再遮断する。具体的な遮断プロセスは、

まず、改めて不適切の頻度を評価し、例えば、全てのＢが現れる時に、Ａも同時に現れ、ＡＢになる場合であれば、Ｂの頻度が０になる。頻度再評価式は：

$$\begin{cases} \log_2|a| * f(a) & a\text{が含まれていない場合} \\ \log_2|a| * (f(a) - \dfrac{\sum_{b \in T_a} f(b)}{P(T_a)}) & \text{その他} \end{cases}$$

ここで、aは特徴項であり、ｆ（ａ）はaのワード頻度であり、ｂはaが含まれる長い文字列の特徴項であり、Ｔ$_a$はｂの集合であり、Ｐ（Ｔ$_a$）は集合のサイズである。

【００２０】

それから、再評価された頻度に従って再遮断を行い、頻度が閾値未満の特徴項候補を削除し、閾値を調整することによって、残す範囲を制御する。

【００２１】

ステップＳ４：前記特徴項候補が自動的に選択されて、特徴項が抽出される。具体的には、当該ステップにおいて、前記ステップ３で順方向コーパスから取得される特徴項候補と前記ステップ３で逆方向コーパスから取得される特徴項候補とを組合せ、組合せによる特徴項候補は二つのワード頻度があり、それぞれ順方向頻度と逆方向頻度に対応する。統計学上のカイ2乗統計量によって特徴項を自動的に選択し、カイ2乗値が最大である前からのＮ個の特徴項候補を残して最終の特徴項情報として、Ｘ^{2}統計量の式は：

$$\chi^2(\omega_i, C_k) = \frac{N(AD-BC)^2}{(A+C)(A+B)(B+D)(C+D)}$$

【００２２】

その中、Ａ、Ｂ、Ｃ、Ｄの意味はそれぞれ下記で示され、

	Ｃ$_k$に属するテキスト	Ｃ$_k$に属しないテキスト	合計
ｗ特徴項が含まれる	Ａ	Ｂ	Ａ＋Ｂ
ｗ特徴項が含まれない	Ｃ	Ｄ	Ｃ＋Ｄ
合計	Ａ＋Ｃ	Ｂ＋Ｄ	Ｎ

表におけるｋは「０」または「１」で、順方向タイプと逆方向タイプの二タイプを代表する。

【００２３】

ここで、前記特徴項は一文字単語（単一の文字からなる単語）と複数文字単語（複数の文字からなる単語）とを含む。一文字単語は逆方向テキストの判定に影響が大きい。特に、フォーラムテキスト情報の内容において、一文字単語に基づくセグメントユニットがよく用いられ、一文字単語を考えなければ、逆方向テキストが誤判断しやすくなる。

【００２４】

ステップ２０４３：前記特徴項のコーパス情報の得点を計算する。ステップＳ４で前記特徴項の頻度が既に保存され、特徴項ごとに順方向頻度と逆方向頻度をそれぞれ代表する二つの頻度を有する。例えば、「発票（インボイス）」の順方向頻度は逆方向頻度よりずっと大きく、「発票（インボイス）」は広告の有害情報によく用いられるからである。各特徴項の順方向頻度を特徴項の順方向ウェートとして、各特徴項の逆方向頻度を特徴項の逆方向ウェートとする。全ての特徴項の順方向/逆方向ウェートに対して正規化を行い、これによってこそ、ウェート値は比較する意味がある。正規化の式は：

$$score(\omega_i) = \frac{freq(\omega_i)}{\sum freq(\omega_i)}$$

【００２５】

生成された特徴項とそのウェートがシステムより準備しておく標準的二種類のコーパスに基づいて学習することによって取得されるため、生成された結果を保存しシステムプレリサーチモデル特徴項情報とする。

【００２６】

前記前処理された遮断待ちテキスト情報と前記システムプレリサーチモデル特徴項情報とに対して特徴情報マッチングし、遮断待ちテキスト特徴項情報を取得し、前記特徴項情報の順方向得点を計算し、その計算式は：

$$score_{pos}(doc) = \sum \log(score(\omega_i)_{pos})$$

前記特徴項情報の逆方向得点を計算し、その計算式は：

$$score_{neg}(doc) = \sum \log(score(\omega_i)_{neg})$$

なお、num（ad）とnum（user）も考慮すると、上記計算式の右側が下記のようになる：

$$\sum \log(score(\omega_i)_{neg})) + num(ad) * score_{ad} + num(user) * score_{user}$$

【００２７】

ステップ２０４４：コーパス情報の得点に基づいて、前記特徴項に対応する遮断待ちテキスト情報が有害情報であるかどうかを判断する。$score_{pos}(doc) > score_{neg}(doc)$の場合に、システムプレリサーチモデルはこの遮断待ちテキスト情報が有害テキストと判断する。また、$score_{pos}(doc) == score_{neg}(doc)$の場合に、このモデルが無効となり、判定が無効される。また、$score_{pos}(doc) < score_{neg}(doc)$の場合に、システムプレリサーチモデルはこの遮断待ちテキスト情報が通常テキストと判断する。

【００２８】

ステップ２０４５：判断結果に基いて、第一マッチング結果を取得する。

【００２９】

ステップ２０５：前記前処理された遮断待ちテキスト情報と前記ユーザーフィードバックモデル情報とに対して特徴情報マッチングし、第二マッチング結果を取得する。当該ス

テップに含まれるプロセスとステップ２０４におけるプロセスが大体同じである。
【００３０】
ここで、前記ユーザーフィードバックモデル情報を取得するプロセスとシステムプレリサーチモデル情報を取得するプロセスについて、主な相違点はステップ２０１で学習コーパスの選択である。前記ユーザーフィードバックモデル情報の学習コーパスが下記の二つの方面から取得できる：
（１）ユーザーフィードバックメカニズム
実際の使用するプロセスにおいて、判定には問題があると発見され、主に有害情報が通常情報と誤判断される場合に、ユーザーはシステムにエラーを報告し、システムはユーザーからの標準回答を受けフィードバックコーパスとする。
（２）判断モデルメカニズム
処理待ちのテキストがステップ２０６での有害情報判定を受け、当該テキストの判定結果が出力される。結果は有害情報テキストか通常テキストである。信頼性を判定する状況に基づいて、処理待ちのテキストはフィードバック学習に用いられるかどうかを判断する。
【００３１】
ステップ２０６：前記第一マッチング結果と第二マッチング結果に基いて、前記遮断待ちテキスト情報に対して遮断処理を行う。具体的には、前記第一マッチング結果と第二マッチング結果、つまり、システムプレリサーチモデル情報とユーザーフィードバックモデル情報が一致するかどうかを判断する。一致と判定され、いずれも有害情報テキストまたは通常情報テキストであれば、判定結果の信頼性が高く、フィードバック学習に使用できる。一方、不一致と判定されれば、判定結果の信頼性がより低くなる。比較的厳しい遮断方針が採用される場合に、このテキストが遮断されるが、フィードバック学習に使用できない。その中には一つのモデルが無効である場合に、結果は残りのモデルの判定結果次第であり、ある程度の信頼性があり、フィードバック学習に使用できる。二つのモデルが両方とも無効である場合に、無効標識をリターンさせ、フィードバック学習に使用できない。
【００３２】
ここで、前記方法において、遮断待ちテキスト情報の判定プロセスの完成後に、
前記ユーザーフィードバックモデル情報のコーパス量及びそれに対応する閾値を取得する。具体的には、フィードバック学習に使用できるコーパス量を統計し、前記コーパス量はそれに対応する閾値を越えているかどうかを判断する。
【００３３】
前記ユーザーフィードバックモデル情報のコーパス量及びそれに対応する閾値に基づいて、前記ユーザーフィードバックモデル情報を更新する。コーパス量は閾値より大きい場合に、改めてフィードバックコーパスを学習し、ユーザーフィードバックモデル情報を更新する。閾値を調整することによって、更新の周期が調整される。
【００３４】
図３で示されるように、本発明の一実施例のインターネットにおける有害情報の遮断装置は、
遮断待ちテキスト情報、システムプレリサーチモデル情報及びユーザーフィードバックモデル情報を取得する情報取得モジュール３０１と、前記遮断待ちテキスト情報を前処理する前処理モジュール３０２と、前記前処理された遮断待ちテキスト情報と前記システムプレリサーチモデル情報とに対して特徴情報マッチングし、第一マッチング結果を取得する第一マッチングモジュール３０３と、前記前処理された遮断待ちテキスト情報と前記ユーザーフィードバックモデル情報とに対して特徴情報マッチングし、第二マッチング結果を取得する第二マッチングモジュール３０４と、第一マッチング結果と第二マッチング結果に基いて、前記遮断待ちテキスト情報を遮断する遮断モジュール３０５と、を備える。
【００３５】
図４で示されるように、本発明の一実施例のインターネットにおける有害情報の遮断装

置は、下記のモジュールを備える。すなわち、
情報取得モジュール４０１：遮断待ちテキスト情報、システムプレリサーチモデル情報及びユーザーフィードバックモデル情報を取得し、ユーザーフィードバックモデル情報の学習コーパスを取得する。ここで、前記ユーザーフィードバックモデル情報のコーパスには、ユーザーフィードバックコーパス及び/または被遮断コーパスが含まれる。
【００３６】
前処理モジュール４０２：前記遮断待ちテキスト情報を前処理する。このモジュールは、前記遮断待ちテキスト情報に対してセグメント化処理をするセグメントサブモジュール４０２１と、前記セグメント処理された特徴項候補量を統計する統計サブモジュール４０２２と、を備える。
【００３７】
第一マッチングモジュール４０３：前記前処理された遮断待ちテキスト情報と前記システムプレリサーチモデル情報とに対して特徴情報マッチングし、第一マッチング結果を取得する。このモジュールは、前記前処理された遮断待ちテキスト情報及び前記システムプレリサーチモデル情報を取得する情報取得サブモジュール４０３１と、前記前処理された遮断待ちテキスト情報と前記システムプレリサーチモデル情報とをマッチングし、特徴項を取得するマッチングサブモジュール４０３２と、前記特徴項のコーパス情報の得点を計算する統計サブモジュール４０３３と、コーパス情報の得点に基づいて、前記特徴項に対応する遮断待ちテキスト情報が有害情報であるかどうかを判断する判断サブモジュール４０３４と、判断結果に基いて、第一マッチング結果を取得する結果出力サブモジュール４０３５と、を備える。
【００３８】
第二マッチングモジュール４０４：前記前処理された遮断待ちテキスト情報と前記ユーザーフィードバックモデル情報とに対して特徴情報マッチングし、第二マッチング結果を取得する。このモジュールは、前記前処理された遮断待ちテキスト情報及び前記システムプレリサーチモデル情報を取得する情報取得サブモジュール４０４１と、前記前処理された遮断待ちテキスト情報と前記ユーザーフィードバックモデル情報とをマッチングし、特徴項を取得するマッチングサブモジュール４０４２と、前記特徴項のコーパス情報の得点を統計する統計サブモジュール４０４３と、コーパス情報の得点に基づいて、前記特徴項に対応する遮断待ちテキスト情報が有害情報であるかどうかを判断する判断サブモジュール４０４４と、判断結果に基づいて、第二マッチング結果を取得する結果出力サブモジュール４０４５と、を備える。
【００３９】
遮断モジュール４０５：前記第一マッチング結果と第二マッチング結果に基づいて、前記遮断待ちテキスト情報に対して遮断処理を行う。
【００４０】
閾値取得モジュール４０６：前記ユーザーフィードバックモデル情報のコーパス量及びそれに対応する閾値を取得する。
【００４１】
更新モジュール４０７：前記ユーザーフィードバックモデル情報のコーパス量及びそれに対応する閾値に基づいて、前記ユーザーフィードバックモデル情報を更新する。前記閾値取得モジュールが取得したユーザーフィードバックモデル情報のコーパス量はそれに対応する閾値に達する場合に、前記更新モジュールが前記ユーザーフィードバックモデル情報のコーパス量及びそれに対応する閾値に基づいて、前記ユーザーフィードバックモデル情報を更新する。
【００４２】
以上のように、本発明の実施例に提供されるインターネットにおける有害情報の遮断方法と装置は、遮断待ちテキスト情報、システムプレリサーチモデル情報及びユーザーフィードバックモデル情報を取得し、前記遮断待ちテキスト情報を前処理し、前記前処理された遮断待ちテキスト情報と前記システムプレリサーチモデル情報とに対して特徴情報マッ

チングし、第一マッチング結果を取得し、前記前処理された遮断待ちテキスト情報と前記ユーザーフィードバックモデル情報とに対して特徴情報マッチングし、第二マッチング結果を取得し、第一マッチング結果と第二マッチング結果に基いて、前記遮断待ちテキスト情報を遮断する。二回のマッチングによって遮断処理が行われるため、システムにおいて有害情報を自動的に遮断する正確性が高く、システムの性能を向上することができる。また、本発明はユーザーフィードバックモデル情報を利用して有害情報を遮断するため、ユーザーフィードバック情報を適時に有害情報の自動的遮断プロセスに適用することが出来、システムモデル情報の自動的更新機能を実現できる。

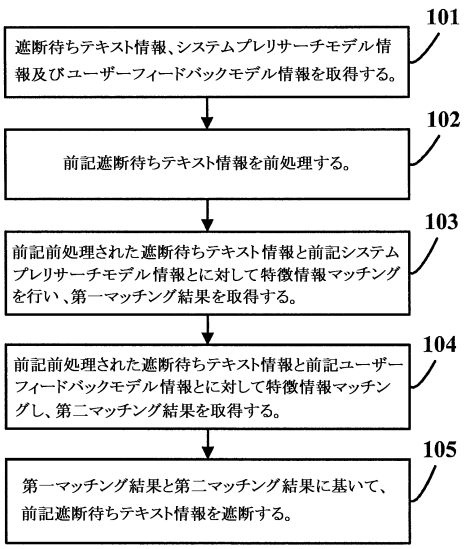
【００４３】

上記した説明から、当業者であれば分かるように、前記した実施例における全部または一部のステップは、プログラムによって関連するハードウェアで実行することができる。前記プログラムは例えば、ＲＯＭ/ＲＡＭ、磁気ディスク、光ディスクなどの記憶装置に記憶されてもよい。

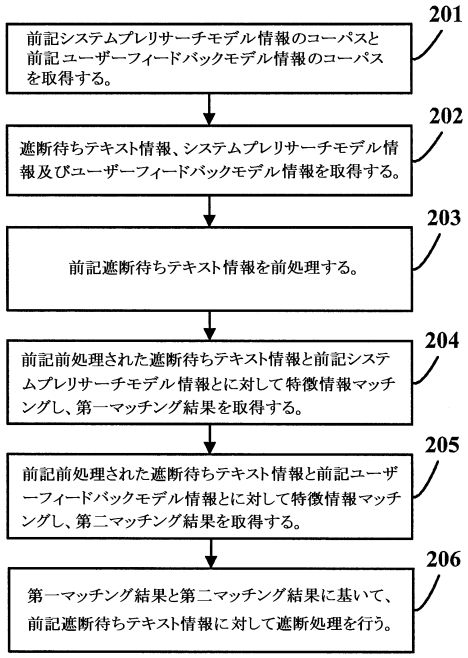
【００４４】

以上で説明した内容はただ本発明の各実施形態であり、本発明が保護しようとする範囲はここに限定されるものではなく、いかなる当業者は本発明より開示された技術範囲で容易に想到できる適宜組み合わせて得られる実施形態についても本発明の技術的範囲に含まれる。

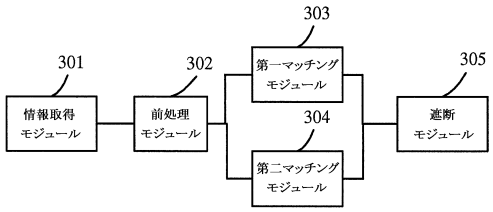
【図１】



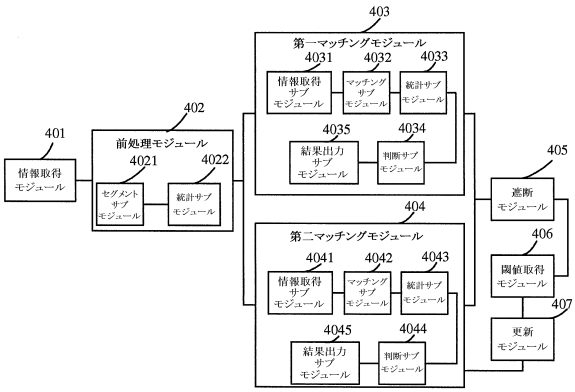
【図２】



【図３】


301
情報取得
モジュール
302
前処理
モジュール
303
第一マッチング
モジュール
304
第二マッチング
モジュール
305
遮断
モジュール

【図４】


401
情報取得
モジュール
402
前処理モジュール
4021
セグメント
サブ
モジュール
4022
統計サブ
モジュール
403
第一マッチングモジュール
4031
情報取得
サブ
モジュール
4032
マッチング
サブ
モジュール
4033
統計サブ
モジュール
4035
結果出力
サブ
モジュール
4034
判断サブ
モジュール
404
第二マッチングモジュール
4041
情報取得
サブ
モジュール
4042
マッチング
サブ
モジュール
4043
統計サブ
モジュール
4045
結果出力
サブ
モジュール
4044
判断サブ
モジュール
405
遮断
モジュール
406
閾値取得
モジュール
407
更新
モジュール

フロントページの続き

(73)特許権者 507232478
北京大学
ＰＥＫＩＮＧ ＵＮＩＶＥＲＳＩＴＹ
中華人民共和国北京市▲海▼淀区▲頤▼和▲園▼路５号
Ｎｏ．５， Ｙｉｈｅｙｕａｎ Ｒｏａｄ， Ｈａｉｄｉａｎ Ｄｉｓｔｒｉｃｔ， Ｂｅｉｊｉｎｇ １００８７１， Ｃｈｉｎａ
(73)特許権者 507232456
北京北大方正▲電▼子有限公司
ＢＥＩＪＩＮＧ ＦＯＵＮＤＥＲ ＥＬＥＣＴＲＯＮＩＣＳ ＣＯ．， ＬＴＤ．
中華人民共和国北京市▲海▼淀区上地五街９号方正大厦
Ｆｏｕｎｄｅｒ Ｂｕｉｌｄｉｎｇ， Ｎｏ．９， Ｓｈａｎｇｄｉｗｕ Ｓｔｒｅｅｔ， Ｈａｉｄｉａｎ Ｄｉｓｔｒｉｃｔ， Ｂｅｉｊｉｎｇ １０００８５， Ｃｈｉｎａ
(73)特許権者 513157039
北京北大方正技▲術▼研究院有限公司
ＰＥＫＩＮＧ ＵＮＩＶＥＲＳＩＴＹ ＦＯＵＮＤＥＲ Ｒ ＆ Ｄ ＣＥＮＴＥＲ
中華人民共和国北京市▲海▼淀区成府路２９８号中▲関▼村方正大厦４▲層▼
４ Ｆｌｏｏｒ， Ｚｈｏｎｇｇｕａｎｃｕｎ Ｆｏｕｎｄｅｒ Ｂｕｉｌｄｉｎｇ， Ｎｏ．２９８， Ｃｈｅｎｇｆｕ Ｒｏａｄ， Ｈａｉｄｉａｎ Ｄｉｓｔｒｉｃｔ， Ｂｅｉｊｉｎｇ １００８７１， Ｃｈｉｎａ
(74)代理人 110000729
特許業務法人 ユニアス国際特許事務所
(72)発明者 チェン、イェン
中華人民共和国 １００００８５ ベイジン、北京市▲海▼淀区上地五街九号方正大厦
(72)発明者 ユー、シャオミン
中華人民共和国 １００００８５ ベイジン、北京市▲海▼淀区上地五街九号方正大厦
(72)発明者 ヤン、チエンウー
中華人民共和国 １００００８５ ベイジン、北京市▲海▼淀区上地五街九号方正大厦

審査官 木村 雅也

(56)参考文献 特開２００５－２３５２０６（ＪＰ，Ａ）
特開２００７－１７９５４０（ＪＰ，Ａ）
特表２０１０－５０７１５３（ＪＰ，Ａ）
特開２００９－１４０４３７（ＪＰ，Ａ）
米国特許第０５９８７４５７（ＵＳ，Ａ）
中国特許出願公開第１０１４７７５４４（ＣＮ，Ａ）
中国特許出願公開第１０１６３９８２４（ＣＮ，Ａ）
中国特許出願公開第１０１７０２１６７（ＣＮ，Ａ）
中国特許出願公開第１０１８７７７０４（ＣＮ，Ａ）
米国特許第０５８６７７９９（ＵＳ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ １３／００
Ｇ０６Ｆ １７／２７