

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2014-502754  
(P2014-502754A)

(43) 公表日 平成26年2月3日(2014. 2. 3)

(51) Int.Cl.	F I	テーマコード (参考)
<b>G06F 13/00 (2006.01)</b>	G06F 13/00 540E	5B084
<b>G06F 17/27 (2006.01)</b>	G06F 17/27 Z	5B091

審査請求 未請求 予備審査請求 未請求 (全 20 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)  
 (85) 翻訳文提出日 平成25年8月20日 (2013. 8. 20)  
 (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)  
 (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)

(71) 出願人 507231932  
 北大方正集▲団▼有限公司  
 PEKING UNIVERSITY F  
 OUNDER GROUP CO., L  
 TD  
 中華人民共和国北京市▲海▼淀区成府路2  
 98号中▲関▼村方正大厦5▲層▼  
 5 Floor, Zhongguanc  
 un Founder Building  
 , No. 298, Chengfu R  
 oad, Haidian Distri  
 ct, 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) 【要約】

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開示し、コンピューター情報処理及び情報遮断技術に関する。本発明のインターネットにおける有害情報の遮断方法は、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を行い、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づいて、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。これによって、有害情報を自動的に遮断する性能を向上することができ、システム情報が自動的に更新されることも実現できる。

【選択図】 図1

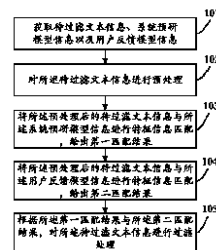


図1 / FIG. 1

101 ACQUIRING THE TEXT INFORMATION TO BE FILTERED, THE ADVANCED-RESEARCHED MODEL INFORMATION OF A SYSTEM AND THE USER FEEDBACK MODEL INFORMATION  
 102 PRE-PROCESSING THE TEXT INFORMATION TO BE FILTERED  
 103 IMPLEMENTING THE CHARACTERISTIC INFORMATION MATCHING BETWEEN THE PRE-PROCESSED TEXT INFORMATION TO BE FILTERED AND THE ADVANCED-RESEARCHED MODEL INFORMATION OF THE SYSTEM, AND GETTING A FIRST MATCHING RESULT  
 104 IMPLEMENTING THE CHARACTERISTIC INFORMATION MATCHING BETWEEN THE PRE-PROCESSED TEXT INFORMATION TO BE FILTERED AND THE USER FEEDBACK MODEL INFORMATION, AND GETTING A SECOND MATCHING RESULT  
 105 FILTERING THE TEXT INFORMATION TO BE FILTERED ACCORDING TO THE FIRST MATCHING RESULT AND THE SECOND MATCHING RESULT

**【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、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方法。

10

**【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。

20

**【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。

30

**【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。

40

**【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

50

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、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づいて、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】

10

20

30

40

50

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、

判断結果に基づいて、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。

10

#### 【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。

20

#### 【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。

30

#### 【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

40

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

50

、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正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「**發票(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「**發票(インボイス)**」より説得力がある。

(4) 従来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】

10

20

30

40

50

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、デフォルトウェイト $score_{ad}$ 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ものである。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づいて、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】

10

20

30

40

50

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、以下のものである。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「A B C D」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A B C、B C D、A B、B C、C D、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、デフォルトウェイトscore<sub>user</sub>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、A B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：

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

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<sub>a</sub>はbの集合であり、P(T<sub>a</sub>)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<sup>2</sup>統計量の式は：

10

20

30

40

50



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

【 0 0 2 2 】

その中、A、B、C、Dの意味はそれぞれ下記で示され、

	C <sub>k</sub> に属する テキスト	C <sub>k</sub> 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

10

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。

20

【 0 0 2 3 】

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。

【 0 0 2 4 】

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イトとして、各特徴項の逆方向頻度を特徴項の逆方向ウェイトとする。全ての特徴項の順方向/逆方向ウェイトに対して正規化を行い、これによってこそ、ウェイト値は比較する意味がある。正規化の式は：

30

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

【 0 0 2 5 】

生成された特徴項とそのウェイ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。

40

【 0 0 2 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：

50

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

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：

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

【 0 0 2 7 】

ステップ 2 0 4 4：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<sub>pos</sub>(doc) > score<sub>neg</sub>(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<sub>pos</sub>(doc) == score<sub>neg</sub>(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<sub>pos</sub>(doc) < score<sub>neg</sub>(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。

10

【 0 0 2 8 】

ステップ 2 0 4 5：判断結果に基づいて、第一マッチング結果を取得する。

【 0 0 2 9 】

ステップ 2 0 5：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 2 0 4におけるプロセスが大体同じである。

20

【 0 0 3 0 】

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 2 0 1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 2 0 6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。

30

【 0 0 3 1 】

ステップ 2 0 6：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。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。

40

【 0 0 3 2 】

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

50

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づいて、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、判断結果に基づいて、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

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。

10

【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づいて、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。

20

【0043】

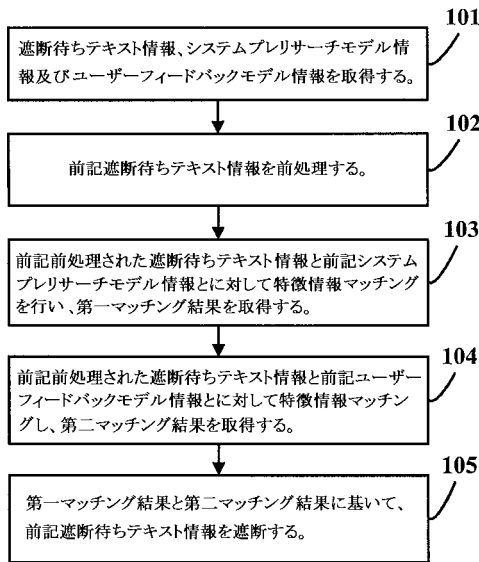
上記した説明から、当業者であれば分かるように、前記した実施例における全部または一部のステップは、プログラムによって関連するハードウェアで実行することができる。前記プログラムは例えば、ROM/RAM、磁気ディスク、光ディスクなどの記憶装置に記憶されてもよい。

30

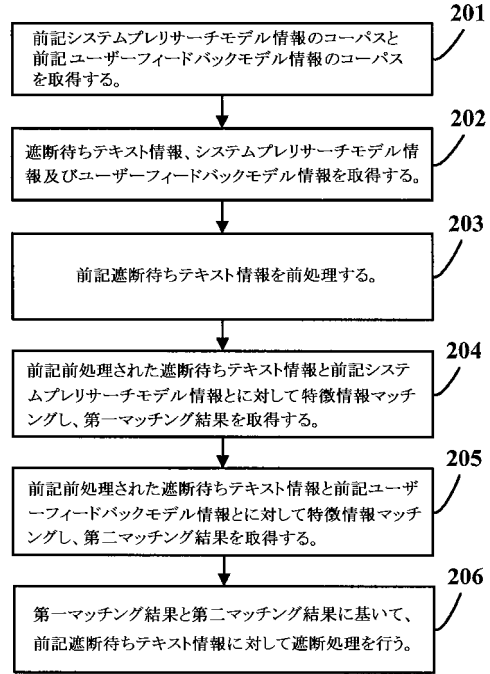
【0044】

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で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。

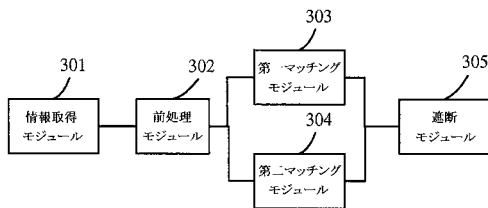
【 図 1 】



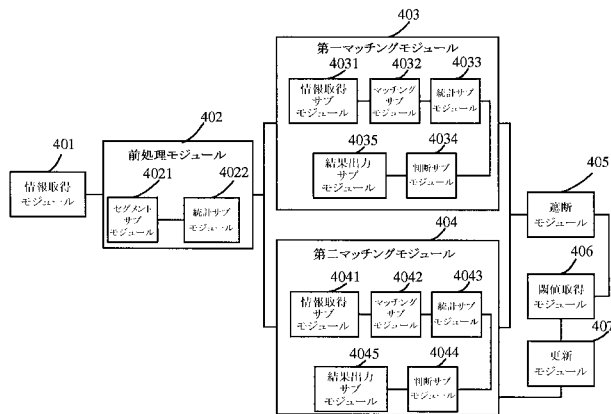
【 図 2 】



【 図 3 】



【 図 4 】



## 【 国际调查报告 】

<b>INTERNATIONAL SEARCH REPORT</b>		International application No. <b>PCT/CN2011/084699</b>
<b>A. CLASSIFICATION OF SUBJECT MATTER</b>		
See the extra sheet		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b>		
Minimum documentation searched (classification system followed by classification symbols)		
IPC: G06F 17/-, H04L 12/-, H04L 29/-, H04W 12/-, H04W 80/-		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
CNABS, VEN, CNTXT, CNKI: reactionary violent website shield stop firewall add up linguistic data positive negative recognition preset pre-research pre-store net+ advertisement bad erotic word match+ forbid+ filter+ rule delet+ predefin+ feedback		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5987457 A (ACCELERATION SOFTWAREINT CORP.), 16 November 1999 (16.11.1999), the whole document	1-16
A	CN 101477544 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.), 08 July 2009 (08.07.2009), the whole document	1-16
A	CN 101639824 A (BEIJING INSTITUTE OF TECHNOLOGY), 03 February 2010 (03.02.2010), the whole document	1-16
A	CN 101702167 A (SHANGHAI SECOND POLYTECHNIC UNIVERSITY), 05 May 2010 (05.05.2010), the whole document	1-16
A	CN 101877704 A (ZTE CORP.), 03 November 2010 (03.11.2010), the whole document	1-16
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family	
Date of the actual completion of the international search <b>05 March 2012 (05.03.2012)</b>	Date of mailing of the international search report <b>22 March 2012 (22.03.2012)</b>	
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer  <b>ZHAO, Xiaoning</b>  Telephone No.: (86-10) <b>62412071</b>	

**INTERNATIONAL SEARCH REPORT**  
Information on patent family members

International application No.

**PCT/CN2011/084699**

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
US 5987457 A	16.11.1999	None	
CN 101477544 A	08.07.2009	WO 2010078792 A1	15.07.2010
CN 101639824 A	03.02.2010	None	
CN 101702167 A	05.05.2010	None	
CN 101877704 A	03.11.2010	None	

**INTERNATIONAL SEARCH REPORT**

International application No.

**PCT/CN2011/084699**

**CLASSIFICATION OF SUBJECT MATTER**

G06F 17/27 (2006.01) i

G06F 17/30 (2006.01) i



国际检索报告		国际申请号 PCT/CN2011/084699
<b>A. 主题的分类</b>		
参见附加页		
按照国际专利分类(IPC)或者同时按照国家分类和 IPC 两种分类		
<b>B. 检索领域</b>		
检索的最低限度文献(标明分类系统和分类号)		
IPC:G06F17/-, H04L12/-, H04L29/-, H04W12/-, H04W80/-		
包含在检索领域中的除最低限度文献以外的检索文献		
在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) CNABS, VEN,CNTXT, CNKI 不良 反动 广告 色情 暴力 网络 网站 过滤 屏蔽 阻挡 防火墙 删除 统计 规则 语料 匹配 正向 负向 识别 预 设 预研 预存 反馈 词 net+ advertisement bad erotic word match+ forbid+ filter+ rule delet+ predefin+ feedback+		
<b>C. 相关文件</b>		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	US5987457A, (ACCELERATION SOFTWARE INT CORP), 16.11 月 1999(16.11.1999), 全文	1-16
A	CN101477544A, (腾讯科技(深圳)有限公司), 08.7 月 2009(08.07.2009), 全文	1-16
A	CN101639824A, (北京理工大学), 03.2 月 2010(03.02.2010), 全文	1-16
A	CN101702167A, (上海第二工业大学), 05.5 月 2010(05.05.2010), 全文	1-16
A	CN101877704A, (中兴通讯股份有限公司), 03.11 月 2010(03.11.2010), 全文	1-16
<input type="checkbox"/> 其余文件在 C 栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。		
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇 引用文件的公布日而引用的或者因其他特殊理由而引 用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件		“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了 理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的 发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件 结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件
国际检索实际完成的日期 5.3 月 2012(05.03.2012)		国际检索报告邮寄日期 22.3 月 2012 (22.03.2012)
ISA/CN 的名称和邮寄地址: 中华人民共和国国家知识产权局 中国北京市海淀区蓟门桥西土城路 6 号 100088 传真号: (86-10)62019451		受权官员  赵小宁  电话号码: (86-10) 62412071

**国际检索报告**  
关于同族专利的信息

国际申请号  
**PCT/CN2011/084699**

检索报告中引用的 专利文件	公布日期	同族专利	公布日期
US5987457A	16.11.1999	无	
CN101477544A	08.07.2009	WO2010078792A1	15.07.2010
CN101639824A	03.02.2010	无	
CN101702167A	05.05.2010	无	
CN101877704A	03.11.2010	无	

国际检索报告

国际申请号  
**PCT/CN2011/084699**

主题的分类

G06F17/27 (2006.01) i

G06F17/30 (2006.01) i

## フロントページの続き

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

(71) 出願人 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, Bei j i  
 ng 100871, China

(71) 出願人 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, Ha  
 idian District, Beijing 100085, China

(71) 出願人 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 10087  
 1, China

(74) 代理人 110000729  
 特許業務法人 ユニクス国際特許事務所

(72) 発明者 チェン、イェン  
 中華人民共和国 100085 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 100085 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 100085 ベイジン、北京市 海 淀区上地五街九号方正大厦

Fターム(参考) 5B084 AB04 BB17 BB19 CC05 CC15 DC25  
 5B091 AA15 CA12 EA01