

【公報種別】特許法第 17 条の 2 の規定による補正の掲載

【部門区分】第 7 部門第 3 区分

【発行日】平成24年11月8日 (2012.11.8)

【公表番号】特表2012-504920(P2012-504920A)

【公表日】平成24年2月23日 (2012.2.23)

【年通号数】公開・登録公報2012-008

【出願番号】特願2011-530236(P2011-530236)

【国際特許分類】

H 0 4 L 9/36 (2006.01)

G 0 6 F 17/30 (2006.01)

【F I】

H 0 4 L 9/00 6 8 5

G 0 6 F 17/30 1 2 0 A

【手続補正書】

【提出日】平成24年9月19日 (2012.9.19)

【手続補正 1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項 1】

秘密情報を検出する方法であって、  
格納データを読み取るステップと、

上記格納データ内の文字列を識別するステップであって、各文字列は一連の連続バイトを含み、当該バイトの全てが、存在し得る値の予め定められたサブセット中にある値を有する、ステップと、

上記文字列の少なくとも幾つかの各々について、上記文字列が、1つ以上のフォーマット整合を表すバイトを含むかどうかを特定するステップであって、フォーマット整合が、秘密情報に関連する予め定められたフォーマットに整合する値のセットを含む、ステップと、

各フォーマット整合について、上記秘密情報に関連するルールをセットを用いて上記予め定められたフォーマットに整合する値をテストして、上記フォーマット整合が、1つ以上の無効な値を含む無効なフォーマット整合であるか否かを特定するステップと、

無効なフォーマット整合のカウントの他のフォーマット整合のカウントに対する比に少なくとも部分的には基づいて、上記格納データについての第 1 のスコアを計算するステップと、

を含む方法。

【請求項 2】

請求項 1 に記載の方法であって、

上記秘密情報がクレジット・カード番号である、方法。

【請求項 3】

請求項 2 に記載の方法であって、

上記文字列において検出された数字を表す値を有するバイトの数がクレジット・カード番号についての標準的なフォーマットにおける数字の数に等しい場合にフォーマット整合が起こるように決められている、方法。

【請求項 4】

請求項 3 に記載の方法であって、

クレジット・カード番号に関連するルールが、有効な発行者識別番号のリストの仕様を含む、方法。

【請求項 5】

請求項 3 に記載の方法であって、

クレジット・カード番号に関連するルールが、チェック・サム・アルゴリズムの仕様を含む、方法。

【請求項 6】

請求項 1 に記載の方法であって、

上記秘密情報が社会保障番号である、方法。

【請求項 7】

請求項 6 に記載の方法であって、

上記文字列において検出された数字を表す値を有するバイトの数が 9 に等しい場合にフォーマット整合が起こるように決められている、方法。

【請求項 8】

請求項 7 に記載の方法であって、

社会保障番号に関連するルールが、社会保障番号の最初の 5 桁によって表される数についての値の有効なサブセットの仕様を含む、方法。

【請求項 9】

請求項 1 に記載の方法であって、

上記秘密情報が電話番号である、方法。

【請求項 10】

請求項 9 に記載の方法であって、

上記文字列において検出された数字を表す値を有するバイトの数が 10 に等しいか、又は上記文字列において検出された数字の数が最初の数字が「1」である 11 個の数字に等しい場合にフォーマット整合が起こるように決められている、方法。

【請求項 11】

請求項 10 に記載の方法であって、

電話番号に関連するルールが、有効な市外局番のリストの仕様を含む、方法。

【請求項 12】

請求項 10 に記載の方法であって、

電話番号に関連するルールが、市外局番の後の最初の数字が 1 又は 0（ゼロ）であってはないという仕様を含む、方法。

【請求項 13】

請求項 1 に記載の方法であって、

上記秘密情報が郵便番号である、方法。

【請求項 14】

請求項 13 に記載の方法であって、

一連のバイトが、数字を表す値を有する 5 個のバイト又は 5 番目の数字と 6 番目の数字との間にハイフンを有する 9 個の数字を表す値を有する 10 個のバイトの何れかからなると検出された場合にフォーマット整合が起こるように決められている、方法。

【請求項 15】

請求項 14 に記載の方法であって、

郵便番号に関連するルールが、有効な 5 桁の郵便番号のリストの仕様を含む、方法。

【請求項 16】

請求項 1 に記載の方法であって、

各文字列について、当該文字列が名前と整合する 1 つ以上の単語を含むかどうかを特定するステップであって、単語とは、文字列内の一連の連続バイトであって、当該バイトの全てが、文字 - 数字のキャラクタを表す値を有し、そして名前とは、個々の人々を呼ぶのに一般に使用されるキャラクタの連りのリストからのキャラクタの連りである、ステップと、

上記格納データにおいて検出された名前のカウントに少なくとも部分的には基づいて、上記格納データについての第2のスコアを計算するステップと、  
を更に含む、方法。

【請求項17】

請求項16に記載の方法であって、

名前の上記リストが、2個のサブセット（ファースト・ネーム及びラスト・ネーム）に分割されている、方法。

【請求項18】

請求項17に記載の方法であって、

各文字列について、当該文字列が1個以上のフル・ネームを含むかどうかを特定するステップであって、フル・ネームとは、ファースト・ネームのリストからの名前と、それに続くスペースと、それに続くラスト・ネームのリストからの名前と、からなるキャラクタの連なりである、ステップと、

検出されたフル・ネームのカウントに少なくとも部分的には基づいて、上記格納データについての第3のスコアを計算するステップと、  
を更に含む、方法。

【請求項19】

請求項16に記載の方法であって、

上記リスト中の名前が、それらに関連する頻度数を各々有し、上記格納データにおいて出現する名前についての平均頻度数が計算され、当該平均頻度数に少なくとも部分的には基づいて、上記格納データについての第2のスコアが計算される、方法。

【請求項20】

請求項19に記載の方法であって、

上記格納データにおいて検出される名前の数が閾値よりも少ない場合には上記平均頻度数が無視される、方法。

【請求項21】

請求項1に記載の方法であって、

各文字列について、2個の文字からなる単語の数をカウントするステップであって、単語とは、一連の連続バイトであって、当該バイトの全てが、文字・数字のキャラクタを表す値を有する、ステップ、

を更に含む、方法。

【請求項22】

請求項21に記載の方法であって、

各2文字単語について、当該2文字単語が有効な州の略称であるかどうかを特定するステップと、

有効な州の略称のカウント及び2文字単語のカウントに少なくとも部分的には基づいて、上記格納データについての第2のスコアを計算するステップと、  
を更に含む、方法。

【請求項23】

請求項1に記載の方法であって、

各文字列について、当該文字列が1個以上の州／郵便番号のペアを含むかどうかを特定するステップであって、州／郵便番号のペアとは、州の略称と、それに続くスペースと、次にそれに続く郵便番号と、からなるキャラクタの連なりである、ステップと、

検出された州／郵便番号のペアのカウントに少なくとも部分的には基づいて、上記格納データについての第2のスコアを計算するステップと、  
を更に含む、方法。

【請求項24】

請求項1に記載の方法であって、

電子的に格納されたファイル・システムにおける何れのファイルが最近更新されたのかを特定するステップと、

最近更新されたファイルの各々に、請求項 1 に記載の方法を適用するステップと、  
を更に含む、方法。

【請求項 25】

請求項 1 に記載の方法であって、

文字列を規定するバイト値のサブセットが、文字 - 数字のキャラクタ、括弧、ハイフン、及びスペースを表す、方法。

【請求項 26】

請求項 1 に記載の方法であって、

上記第 1 のスコアを閾値と比較するステップと、

上記第 1 のスコアが上記閾値を超える場合は、秘密情報を含む可能性有りとして上記格納データにフラグを立てるステップと、  
を更に含む、方法。

【請求項 27】

秘密情報を検出するシステムであって、

データ格納装置、及び

上記データ格納装置に接続されたランタイム環境を含み、

上記ランタイム環境が、

上記データ格納装置から格納データを読み取るステップと、

上記格納データ内の文字列を識別するステップであって、各文字列は一連の連続バイトを含み、当該バイトの全てが、存在し得る値の予め定められたサブセット中にある値を有する、ステップと、

上記文字列の少なくとも幾つかの各々について、上記文字列が、1つ以上のフォーマット整合を表すバイトを含むかどうかを特定するステップであって、フォーマット整合が、秘密情報に関連する予め定められたフォーマットに整合する値のセットを含む、ステップと、

各フォーマット整合について、上記秘密情報に関連するルールをセットを用いて上記予め定められたフォーマットに整合する値をテストして、上記フォーマット整合が、1つ以上の無効な値を含む無効なフォーマット整合であるか否かを特定するステップと、

無効なフォーマット整合のカウントの他のフォーマット整合のカウントに対する比に少なくとも部分的には基づいて、上記格納データについての第 1 のスコアを計算するステップと、

を実行するように構成されている、システム。

【請求項 28】

請求項 27 に記載のシステムであって、

上記秘密情報がクレジット・カード番号であり、

上記文字列において検出された数字を表す値を有するバイトの数がクレジット・カード番号についての標準的なフォーマットにおける数字の数に等しい場合にフォーマット整合が起こるように決められており、そして

クレジット・カード番号に関連するルールが、有効な発行者識別番号のリストの仕様、及びチェック・サム・アルゴリズムの仕様を含む、システム。

【請求項 29】

請求項 27 に記載のシステムであって、

上記秘密情報が社会保障番号であり、

上記文字列において検出された数字を表す値を有するバイトの数が 9 に等しい場合にフォーマット整合が起こるように決められており、そして

社会保障番号に関連するルールが、社会保障番号の最初の 5 桁によって表される数についての値の有効なサブセットの仕様を含む、システム。

【請求項 30】

請求項 27 に記載のシステムであって、

上記秘密情報が電話番号であり、

上記文字列において検出された数字を表す値を有するバイトの数が10に等しいか、又は上記文字列において検出された数字の数が最初の数字が「1」である11個の数字に等しい場合にフォーマット整合が起こるように決められており、そして

電話番号に関連するルールが、有効な市外局番のリストの仕様、及び市外局番の後の最初の数字が1又は0（ゼロ）であってはならないという仕様を含む、システム。

【請求項31】

請求項27に記載のシステムであって、

上記秘密情報が郵便番号であり、

一連のバイトが、数字を表す値を有する5個のバイト又は5番目の数字と6番目の数字との間にハイフンを有する9個の数字を表す値を有する10個のバイトの何れかからなると検出された場合にフォーマット整合が起こるように決められており、そして

郵便番号に関連するルールが、有効な5桁の郵便番号のリストの仕様を含む、システム

。

【請求項32】

請求項27に記載のシステムであって、

上記ランタイム環境が、

各文字列について、当該文字列が名前と整合する1つ以上の単語を含むかどうかを特定するステップであって、単語とは、文字列内の一連の連続バイトであって、当該バイトの全てが、文字・数字のキャラクタを表す値を有し、そして名前とは、個々の人々を呼ぶのに一般に使用されるキャラクタの連なりからのキャラクタの連なりであり、そして上記リストが、2個のサブセット（ファースト・ネーム及びラスト・ネーム）に分割されている、ステップと、

各文字列について、当該文字列が1個以上のフル・ネームを含むかどうかを特定するステップであって、フル・ネームとは、ファースト・ネームのリストからの名前と、それに続くスペースと、それに続くラスト・ネームのリストからの名前と、からなるキャラクタの連なりである、ステップと、

上記格納データにおいて検出された名前のカウントに少なくとも部分的には基づいて、且つ検出されたフル・ネームのカウントに少なくとも部分的には基づいて、上記格納データについての第2のスコアを計算するステップと、  
を実行するように更に構成されている、システム。

【請求項33】

請求項27に記載のシステムであって、

上記ランタイム環境が、

各文字列について、当該文字列が名前と整合する1つ以上の単語を含むかどうかを特定するステップであって、単語とは、文字列内の一連の連続バイトであって、当該バイトの全てが、文字・数字のキャラクタを表す値を有し、そして名前とは、個々の人々を呼ぶのに一般に使用されるキャラクタの連なりからのキャラクタの連なりであり、そして上記リスト中の名前が、それらに関連する頻度数を各々有する、ステップと、

上記格納データにおいて出現する名前についての平均頻度数を計算するステップと、

上記格納データにおいて検出された名前のカウントに少なくとも部分的には基づいて、且つ上記平均頻度数に少なくとも部分的には基づいて、上記格納データについての第2のスコアを計算するステップと、  
を実行するように更に構成されている、システム。

【請求項34】

請求項27に記載のシステムであって、

上記ランタイム環境が、

各文字列について、当該文字列が1個以上の州／郵便番号のペアを含むかどうかを特定するステップであって、州／郵便番号のペアとは、州の略称と、それに続くスペースと、次にそれに続く郵便番号と、からなるキャラクタの連なりである、ステップと、

検出された州／郵便番号のペアのカウントに少なくとも部分的には基づいて、上記格納

データについての第 2 のスコアを計算するステップと、  
を実行するように更に構成されている、システム。

【請求項 3 5】

請求項 2 7 に記載のシステムであって、

文字列を規定するバイト値のサブセットが、文字 - 数字のキャラクタ、括弧、ハイフン、及びスペースを表す、システム。

【請求項 3 6】

秘密情報を検出するためのコンピュータ・プログラムを格納するコンピュータ可読媒体であって、

上記コンピュータ・プログラムが、

格納データを読み取るステップと、

上記格納データ内の文字列を識別するステップであって、各文字列は一連の連続バイトを含み、当該バイトの全てが、存在し得る値の予め定められたサブセット中にある値を有する、ステップと、

上記文字列の少なくとも幾つかの各々について、上記文字列が、1 つ以上のフォーマット整合を表すバイトを含むかどうかを特定するステップであって、フォーマット整合が、秘密情報に関連する予め定められたフォーマットに整合する値のセットを含む、ステップと、

各フォーマット整合について、上記秘密情報に関連するルールをセットを用いて上記予め定められたフォーマットに整合する値をテストして、上記フォーマット整合が、1 つ以上の無効な値を含む無効なフォーマット整合であるか否かを特定するステップと、

無効なフォーマット整合のカウントの他のフォーマット整合のカウントに対する比に少なくとも部分的には基づいて、上記格納データについての第 1 のスコアを計算するステップと、

をコンピュータに実行させるための命令を含む、コンピュータ可読媒体。

【請求項 3 7】

請求項 3 6 に記載のコンピュータ可読媒体であって、

上記秘密情報がクレジット・カード番号である、コンピュータ可読媒体。

【請求項 3 8】

請求項 3 7 に記載のコンピュータ可読媒体であって、

上記文字列において検出された数字を表す値を有するバイトの数がクレジット・カード番号についての標準的なフォーマットにおける数字の数に等しい場合にフォーマット整合が起こるように決められている、コンピュータ可読媒体。

【請求項 3 9】

請求項 3 8 に記載のコンピュータ可読媒体であって、

クレジット・カード番号に関連するルールが、有効な発行者識別番号のリストの仕様を含む、コンピュータ可読媒体。

【請求項 4 0】

請求項 3 8 に記載のコンピュータ可読媒体であって、

クレジット・カード番号に関連するルールが、チェック・サム・アルゴリズムの仕様を含む、コンピュータ可読媒体。

【請求項 4 1】

請求項 3 6 に記載のコンピュータ可読媒体であって、

上記秘密情報が社会保障番号である、コンピュータ可読媒体。

【請求項 4 2】

請求項 4 1 に記載のコンピュータ可読媒体であって、

上記文字列において検出された数字を表す値を有するバイトの数が 9 に等しい場合にフォーマット整合が起こるように決められている、コンピュータ可読媒体。

【請求項 4 3】

請求項 4 2 に記載のコンピュータ可読媒体であって、

社会保障番号に関連するルールが、社会保障番号の最初の 5 桁によって表される数についての値の有効なサブセットの仕様を含む、コンピュータ可読媒体。

【請求項 4 4】

請求項 3 6 に記載のコンピュータ可読媒体であって、  
上記秘密情報が電話番号である、コンピュータ可読媒体。

【請求項 4 5】

請求項 4 4 に記載のコンピュータ可読媒体であって、  
上記文字列において検出された数字を表す値を有するバイトの数が 1 0 に等しいか、又は上記文字列において検出された数字の数が最初の数字が「1」である 1 1 個の数字に等しい場合にフォーマット整合が起こるように決められている、コンピュータ可読媒体。

【請求項 4 6】

請求項 4 5 に記載のコンピュータ可読媒体であって、  
電話番号に関連するルールが、有効な市外局番のリストの仕様を含む、コンピュータ可読媒体。

【請求項 4 7】

請求項 4 5 に記載のコンピュータ可読媒体であって、  
電話番号に関連するルールが、市外局番の後の最初の数字が 1 又は 0（ゼロ）であってはないという仕様を含む、コンピュータ可読媒体。

【請求項 4 8】

請求項 3 6 に記載のコンピュータ可読媒体であって、  
上記秘密情報が郵便番号である、コンピュータ可読媒体。

【請求項 4 9】

請求項 4 8 に記載のコンピュータ可読媒体であって、  
一連のバイトが、数字を表す値を有する 5 個のバイト又は 5 番目の数字と 6 番目の数字との間にハイフンを有する 9 個の数字を表す値を有する 1 0 個のバイトの何れかからなると検出された場合にフォーマット整合が起こるように決められている、コンピュータ可読媒体。

【請求項 5 0】

請求項 4 9 に記載のコンピュータ可読媒体であって、  
郵便番号に関連するルールが、有効な 5 桁の郵便番号のリストの仕様を含む、コンピュータ可読媒体。

【請求項 5 1】

請求項 3 6 に記載のコンピュータ可読媒体であって、  
上記コンピュータ・プログラムが、  
各文字列について、当該文字列が名前と整合する 1 つ以上の単語を含むかどうかを特定するステップであって、単語とは、文字列内の一連の連続バイトであって、当該バイトの全てが、文字・数字のキャラクタを表す値を有し、そして名前とは、個々の人々を呼ぶのに一般に使用されるキャラクタの連りのリストからのキャラクタの連りである、ステップと、

上記格納データにおいて検出された名前のカウントに少なくとも部分的には基づいて、  
上記格納データについての第 2 のスコアを計算するステップと、  
をコンピュータに実行させるための命令を更に含む、コンピュータ可読媒体。

【請求項 5 2】

請求項 5 1 に記載のコンピュータ可読媒体であって、  
名前の上記リストが、2 個のサブセット（ファースト・ネーム及びラスト・ネーム）に分割されている、コンピュータ可読媒体。

【請求項 5 3】

請求項 5 2 に記載のコンピュータ可読媒体であって、  
上記コンピュータ・プログラムが、  
各文字列について、当該文字列が 1 個以上のフル・ネームを含むかどうかを特定するス

テップであって、フル・ネームとは、ファースト・ネームのリストからの名前と、それに続くスペースと、それに続くラスト・ネームのリストからの名前と、からなるキャラクタの連なりである、ステップと、

検出されたフル・ネームのカウントに少なくとも部分的には基づいて、上記格納データについての第3のスコアを計算するステップと、

をコンピュータに実行させるための命令を更に含む、コンピュータ可読媒体。

【請求項54】

請求項51に記載のコンピュータ可読媒体であって、

上記リスト中の名前が、それらに関連する頻度数を各々有し、

上記コンピュータ・プログラムが、

上記格納データにおいて出現する名前についての平均頻度数を計算するステップと、

当該平均頻度数に少なくとも部分的には基づいて、上記格納データについての第2のスコアを計算するステップと、

をコンピュータに実行させるための命令を更に含む、コンピュータ可読媒体。

【請求項55】

請求項54に記載のコンピュータ可読媒体であって、

上記コンピュータ・プログラムが、コンピュータに

上記格納データにおいて検出される名前の数が閾値よりも少ない場合には上記平均頻度数が無視させるための命令を更に含む、コンピュータ可読媒体。

【請求項56】

請求項36に記載のコンピュータ可読媒体であって、

上記コンピュータ・プログラムが、

各文字列について、2個の文字からなる単語の数をカウントするステップであって、単語とは、一連の連続バイトであって、当該バイトの全てが、文字・数字のキャラクタを表す値を有する、ステップ、

をコンピュータに実行させるための命令を更に含む、コンピュータ可読媒体。

【請求項57】

請求項56に記載のコンピュータ可読媒体であって、

上記コンピュータ・プログラムが、

各2文字単語について、当該2文字単語が有効な州の略称であるかどうかを特定するステップと、

有効な州の略称のカウント及び2文字単語のカウントに少なくとも部分的には基づいて、上記格納データについての第2のスコアを計算するステップと、

をコンピュータに実行させるための命令を更に含む、コンピュータ可読媒体。

【請求項58】

請求項36に記載のコンピュータ可読媒体であって、

上記コンピュータ・プログラムが、

各文字列について、当該文字列が1個以上の州/郵便番号のペアを含むかどうかを特定するステップであって、州/郵便番号のペアとは、州の略称と、それに続くスペースと、次にそれに続く郵便番号と、からなるキャラクタの連なりである、ステップと、

検出された州/郵便番号のペアのカウントに少なくとも部分的には基づいて、上記格納データについての第2のスコアを計算するステップと、

をコンピュータに実行させるための命令を更に含む、コンピュータ可読媒体。

【請求項59】

請求項36に記載のコンピュータ可読媒体であって、

上記コンピュータ・プログラムが、

電子的に格納されたファイル・システムにおける何れのファイルが最近更新されたのかを特定するステップと、

最近更新されたファイルの各々に、請求項1に記載の方法を適用するステップと、

をコンピュータに実行させるための命令を更に含む、コンピュータ可読媒体。



**【請求項 60】**

請求項 36 に記載のコンピュータ可読媒体であって、  
文字列を規定するバイト値のサブセットが、文字 - 数字のキャラクタ、括弧、ハイフン、及びスペースを表す、コンピュータ可読媒体。

**【請求項 61】**

請求項 36 に記載のコンピュータ可読媒体であって、  
上記コンピュータ・プログラムが、  
上記第 1 のスコアを閾値と比較するステップと、  
上記第 1 のスコアが上記閾値を超える場合は、秘密情報を含む可能性有りとして上記格納データにフラグを立てるステップと、  
をコンピュータに実行させるための命令を更に含む、コンピュータ可読媒体。