

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2018-41442

(P2018-41442A)

(43) 公開日 平成30年3月15日(2018.3.15)

(51) Int.Cl.	F I	テーマコード (参考)
G06F 21/55 (2013.01)	G06F 21/55	5B084
G06F 13/00 (2006.01)	G06F 13/00 530A	5B089
G06F 17/30 (2006.01)	G06F 13/00 351Z	
G06F 21/12 (2013.01)	G06F 17/30 210A	
G06F 21/44 (2013.01)	G06F 17/30 340A	

審査請求 有 請求項の数 20 O L 外国語出願 (全 22 頁) 最終頁に続く

(21) 出願番号 特願2017-126050 (P2017-126050)
 (22) 出願日 平成29年6月28日 (2017.6.28)
 (31) 優先権主張番号 2016136226
 (32) 優先日 平成28年9月8日 (2016.9.8)
 (33) 優先権主張国 ロシア (RU)
 (31) 優先権主張番号 15/437,828
 (32) 優先日 平成29年2月21日 (2017.2.21)
 (33) 優先権主張国 米国 (US)

(特許庁注：以下のものは登録商標)

1. J A V A S C R I P T

(71) 出願人 515348585
 エーオー カスペルスキー ラボ
 AO Kaspersky Lab
 ロシア、125212、モスクワ、レニン
 グラドスコ ショス
 Leningradskoe shoss
 e, Moscow, Russian
 Federation 125212

(74) 代理人 110001139

S K 特許業務法人

(74) 代理人 100130328

弁理士 奥野 彰彦

(74) 代理人 100130672

弁理士 伊藤 寛之

最終頁に続く

(54) 【発明の名称】 Webページの異常要素を検出するためのシステム及び方法

(57) 【要約】 (修正有)

【課題】 Webページの異常要素を検出するためのシステム及び方法を提供する。

【解決手段】 Webサーバを介してWebサイトに紐づいた前記Webページを要求し、クライアントコンピューティングデバイスによって前記Webサイトへのアクセスを取得するステップ300と、前記クライアントコンピューティングデバイスによって、前記Webページに関するデータを収集するために前記Webページを実行するステップ310と、少なくとも前記収集したデータに基づいて、少なくとも1つのN次元ベクトルを決定するステップ320と、前記少なくとも1つのN次元ベクトルに基づいて、N次元空間内の前記Webページの少なくとも1つの要素についてのベクトルの座標の値の組を含む少なくとも1つのクラスタを作成するステップ350と、前記少なくとも1つのクラスタに基づいて、前記Webページの統計モデルを作成するステップ360と、を備える。

【選択図】 図3

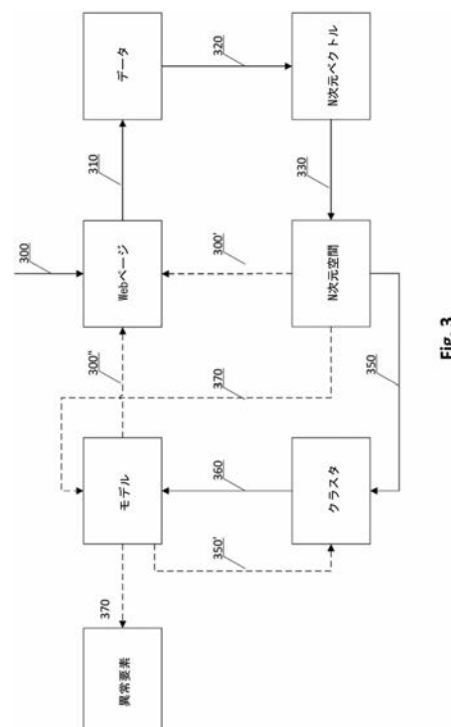


Fig. 3

【特許請求の範囲】

【請求項 1】

Webページの異常要素を検出するためのコンピュータ実装方法であって、

Webサーバを介してWebサイトに紐づいた前記Webページを要求し、クライアントコンピューティングデバイスによって前記Webサイトへのアクセスを取得するステップと、

前記クライアントコンピューティングデバイスによって、前記Webページに関するデータを収集するために前記Webページを実行するステップと、

少なくとも前記収集したデータに基づいて、少なくとも1つのN次元ベクトルを決定するステップと、

前記少なくとも1つのN次元ベクトルに基づいて、N次元空間内の前記Webページの少なくとも1つの要素についてのベクトルの座標の値の組を含む少なくとも1つのクラスタを作成するステップと、

前記少なくとも1つのクラスタに基づいて、前記Webページの統計モデルを作成するステップと、

を備える、コンピュータ実装方法。

【請求項 2】

前記少なくとも1つのN次元ベクトルを格納するステップと、

前記少なくとも1つのクラスタを作成する前に、少なくとも1つのN次元ベクトルに関する追加のデータを収集するため他のクライアントコンピューティングによりWebページを取得して実行するステップと、を備える、請求項 1 に記載のコンピュータ実装方法。

【請求項 3】

前記Webサーバは、前記Webページを要求するクライアントコンピューティングデバイスからの要求を受信したことに応じて、前記Webページに少なくとも1つのスクリプトを追加するよう構成される、請求項1に記載のコンピュータ実装方法。

【請求項 4】

前記クライアントコンピューティングデバイスによって前記Webページを実行するステップは、前記少なくとも1つのスクリプトを実行して前記Webページの前記少なくとも1つの要素に関するデータを収集するステップを有する、請求項 3 に記載のコンピュータ実装方法。

【請求項 5】

前記Webページのクラスタの前記少なくとも1つのN次元ベクトルと、先に構築された少なくとも1つの前記Webページの統計モデルのN次元ベクトルとを比較するステップと、

前記Webページの前記少なくとも1つの要素を、下記(1)～(3)のうち少なくとも1つを検出することにより異常であると判定し識別するステップと、

を備える、請求項 1 に記載のコンピュータ実装方法。

(1) N次元空間における、前記Webページの前記少なくとも1つの要素の前記少なくとも1つのN次元ベクトルと、前記Webページの統計モデルクラスタの中心との距離が、前記クラスタの半径よりも大きい。

(2) N次元空間における、前記Webページの前記少なくとも1つの要素の前記少なくとも1つのN次元ベクトルと、前記Webページの統計モデルのクラスタの中心との間の近接度が、第1の選択された閾値よりも大きい。

(3) N次元空間における、前記少なくとも1つの要素の前記少なくとも1つのN次元ベクトルと、前記Webページの統計モデルのクラスタのN次元ベクトルのうちクラスタ中心から最も離れたベクトルとの間の近接度が、第2の選択された閾値よりも大きい。

【請求項 6】

前記少なくとも1つの要素が異常でないことを検出すると、前記Webページの前記少なくとも1つの要素の前記少なくとも1つのN次元ベクトルを前記Webページの統計モデルに追加するステップをさらに備える、請求項 5 に記載のコンピュータ実装方法。

【請求項 7】

前記Webサーバは、

Webページの少なくとも1つの要素の異常の検出に応答して前記クライアントコンピューティングデバイスとの接続を無効にし、

前記少なくとも1つの要素の統計的有意性を閾値に関して少なくとも決定するために前記Webページの前記少なくとも1つの要素に対してアンチウィルススキャンを実行し、

前記少なくとも1つの要素の統計的有意性が閾値よりも大きいことの検出に応答して、前記少なくとも1つの要素が安全であると識別し前記クライアントコンピューティングデバイスとの接続を再確立するよう構成される、請求項5に記載のコンピュータ実装方法。

【請求項8】

前記Webページの前記少なくとも1つの要素の前記少なくとも1つのN次元ベクトルが前記Webページの統計モデルに対応していないことの検出に応答して、前記少なくとも1つの要素に関連するWebページの数と、前記Webページの統計モデルに用いられるWebページの総数との比率を決定するステップをさらに備える、請求項7に記載のコンピュータ実装方法。

【請求項9】

前記少なくとも1つの要素に関連するWebページの数には、前記少なくとも1つの要素又はそのN次元空間におけるN次元ベクトル間の距離が選択された閾値未満である近傍の要素を含むWebページの数が含まれる、請求項8に記載のコンピュータ実装方法。

【請求項10】

Webページの異常要素を検出するためのシステムであって、少なくとも1つのクライアントコンピューティングデバイスのプロセッサを備え、

当該プロセッサは、

Webサーバを介してWebサイトに紐づいた前記Webページを要求することで前記Webサイトへのアクセスを取得し、

前記クライアントコンピューティングデバイスによって、前記Webページに関するデータを収集するために前記Webページを実行し、

少なくとも前記収集したデータに基づいて、少なくとも1つのN次元ベクトルを決定し、

前記少なくとも1つのN次元ベクトルに基づいて、N次元空間内の前記Webページの少なくとも1つの要素についてのベクトルの座標の値の組を含む少なくとも1つのクラスタを作成し、

前記少なくとも1つのクラスタに基づいて、前記Webページの統計モデルを作成するよう構成される、システム。

【請求項11】

前記プロセッサはさらに、

前記少なくとも1つのN次元ベクトルを格納し、

前記少なくとも1つのクラスタを作成する前に、少なくとも1つのN次元ベクトルに関する追加のデータを収集するためWebページを取得して実行するよう構成される、請求項10に記載のシステム。

【請求項12】

前記Webサーバは、前記Webページを要求するクライアントコンピューティングデバイスからの要求を受信したことに応じて、前記Webページに少なくとも1つのスクリプトを追加するよう構成される、請求項10に記載のシステム。

【請求項13】

前記クライアントコンピューティングデバイスによって前記Webページを実行するため、前記少なくとも1つのプロセッサは、前記少なくとも1つのスクリプトを実行して前記Webページの前記少なくとも1つの要素に関するデータを収集するよう構成される、請求項12に記載のシステム。

【請求項14】

前記少なくとも1つのプロセッサはさらに、

前記Webページのクラスタの前記少なくとも1つのN次元ベクトルと、先に構築された少なくとも1つの前記Webページの統計モデルのN次元ベクトルとを比較し、

10

20

30

40

50

前記Webページの前記少なくとも1つの要素を、下記(1)～(3)のうち少なくとも1つを検出することにより異常であると判定し識別するよう構成される、請求項10に記載のシステム。

(1) N次元空間における、前記Webページの前記少なくとも1つの要素の前記少なくとも1つのN次元ベクトルと、前記Webページの統計モデルクラスタの中心との距離が、前記クラスタの半径よりも大きい。

(2) N次元空間における、前記Webページの前記少なくとも1つの要素の前記少なくとも1つのN次元ベクトルと、前記Webページの統計モデルのクラスタの中心との間の近接度が、第1の選択された閾値よりも大きい。

(3) N次元空間における、前記少なくとも1つの要素の前記少なくとも1つのN次元ベクトルと、前記Webページの統計モデルのクラスタのN次元ベクトルのうちクラスタ中心から最も離れたベクトルとの間の近接度が、第2の選択された閾値よりも大きい。

【請求項15】

前記少なくとも1つのプロセッサはさらに、前記少なくとも1つの要素が異常でないことを検出すると、前記Webページの前記少なくとも1つの要素の前記少なくとも1つのN次元ベクトルを前記Webページの統計モデルに追加するよう構成される、請求項14に記載のシステム。

【請求項16】

前記Webサーバは、

Webページの少なくとも1つの要素の異常の検出に応答して前記クライアントコンピューティングデバイスとの接続を無効にし、

前記少なくとも1つの要素の統計的有意性を閾値に関して決定するために前記Webページの前記少なくとも1つの要素に対してアンチウィルススキャンを実行し、

前記少なくとも1つの要素の統計的有意性が閾値よりも大きいことの検出に応答して、前記少なくとも1つの要素が安全であると識別し前記クライアントコンピューティングデバイスとの接続を再確立するよう構成される、請求項14に記載のシステム。

【請求項17】

前記少なくとも1つのプロセッサは、前記Webページの前記少なくとも1つの要素の前記少なくとも1つのN次元ベクトルが前記Webページの統計モデルに対応していないことの検出に応答して、前記少なくとも1つの要素に関連するWebページの数と、前記Webページの統計モデルに用いられるWebページの総数との比率を決定するよう構成され、

前記少なくとも1つの要素に関連するWebページの数には、前記少なくとも1つの要素又はそのN次元空間におけるN次元ベクトル間の距離が選択された閾値未満である近傍の要素を含むWebページの数が含まれる、請求項16に記載のシステム。

【請求項18】

Webページの異常要素を検出するためのコンピュータ実行可能命令が格納された非一時的コンピュータ可読媒体であって、

Webサーバを介してWebサイトに紐づいた前記Webページを要求し、クライアントコンピューティングデバイスによって前記Webサイトへのアクセスを取得させる命令と、

前記クライアントコンピューティングデバイスによって、前記Webページに関するデータを収集するために前記Webページを実行させる命令と、

少なくとも前記収集したデータに基づいて、少なくとも1つのN次元ベクトルを決定させる命令と、

前記少なくとも1つのN次元ベクトルに基づいて、N次元空間内の前記Webページの少なくとも1つの要素についてのベクトルの座標の値の組を含む少なくとも1つのクラスタを作成させる命令と、

前記少なくとも1つのクラスタに基づいて、前記Webページの統計モデルを作成させる命令と、

を備える、非一時的コンピュータ可読媒体。

【請求項19】

10

20

30

40

50

前記Webページのクラスタの前記少なくとも1つのN次元ベクトルと、先に構築された少なくとも1つの前記Webページの統計モデルのN次元ベクトルとを比較させる命令と、

前記Webページの前記少なくとも1つの要素を、下記(1)～(3)のうち少なくとも1つを検出することにより異常であると判定し識別させる命令と、

を備える、請求項18に記載の非一時的コンピュータ可読媒体。

(1) N次元空間における、前記Webページの前記少なくとも1つの要素の前記少なくとも1つのN次元ベクトルと、前記Webページの統計モデルクラスタの中心との距離が、前記クラスタの半径よりも大きい。

(2) N次元空間における、前記Webページの前記少なくとも1つの要素の前記少なくとも1つのN次元ベクトルと、前記Webページの統計モデルのクラスタの中心との間の近接度が、第1の選択された閾値よりも大きい。

(3) N次元空間における、前記少なくとも1つの要素の前記少なくとも1つのN次元ベクトルと、前記Webページの統計モデルのクラスタのN次元ベクトルのうちクラスタ中心から最も離れたベクトルとの間の近接度が、第2の選択された閾値よりも大きい。

【請求項20】

前記Webページの前記少なくとも1つの要素の前記少なくとも1つのN次元ベクトルが前記Webページの統計モデルに対応していないことの検出に応答して、前記少なくとも1つの要素に関連するWebページの数と、前記Webページの統計モデルで用いられるWebページの総数との比率を決定させる命令をさらに備える、請求項19に記載の非一時的コンピュータ可読媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、一般に、サイバーセキュリティの分野に関し、より具体的には、ウェブページの異常要素を検出するシステム及び方法に関する。

【背景技術】

【0002】

近年、銀行などの金融機関は、銀行業務の一環としてWebバンキングサービス（インターネットバンキング）を積極的に採用している。Webバンキングは、リモートバンキングサービス技術の一般的な用語であり、いつでも、インターネットにアクセス可能な任意のコンピュータから口座にアクセスし、（リモートバンキングサービス技術により）操作することができる。これらの操作を実行するには、典型的にはウェブクライアント（ブラウザ等）が使用される。

【0003】

このような技術の普及は、自然とハッカー、とりわけリモートサービスシステムのユーザの口座から資金を盗むことに関心を持つ者の興味をかきたてる。Webバンキングのユーザに対する一般的な攻撃の一つは、ユーザに対して表示されるWebページのコンテンツの代わりに悪意のあるソフトウェアを配置することである。悪意のあるソフトウェアは、Webページ内にHTMLコードを埋め込む。この攻撃は、しばしば「マン・イン・ザ・ブラウザ」又は「Webコードインジェクション」攻撃と呼ばれる。この攻撃は、例えばトロイの木馬アプリケーションにより開始される。トロイの木馬は、被害者のブラウザに悪質な拡張子のファイルをインストールし、再起動したときに当該ファイルをブラウザで実行する。その後、特定のWebサイト（大抵は銀行サイト）にルーティングされるユーザのトラフィックを傍受する。そして、ユーザに表示されるWebページを（ロード中又はオープン中に）改ざんすることで、Webページの特定の要素の外観を変更したり、被害者の認証データを盗み出したり、ユーザが送金した資金を第三者の口座に転送したりするのである。

【0004】

現存する解決策は、外部コードをWebページに導入する攻撃に対し、ネット上においてユーザの安全な作業を増やすことを目指すものである。しかしながら、これらの解決策は、悪意のあるソフトウェアによってWebページが改ざんされたかどうかを効果的に判断は

10

20

30

40

50

せず、追加のソフトウェアをインストールしない限り、ユーザ側のWebページのバージョンの異常な要素を特定することもない。同時に、各種のセキュリティクライアント、シンクライアント（ライトエージェント）、その他のアンチウイルス手段等の追加のソフトウェアは、必ずしもユーザ側でインストール可能なわけではなく、最終的には、アンチウイルスアプリケーションの動作に過誤を発生させる。例えば、第1の種類の過誤は、データ転送チャンネルを捕捉して転送される全ての情報にアクセスするため、コンピューティングシステム（コンピュータ）に「マン・イン・ザ・ブラウザ」タイプの攻撃を許してしまうものであり、第2の種類の過誤は、ユーザ側での正当なWebページの変更が異常であると誤って判断されることである。

【発明の概要】

10

【0005】

Webページの異常要素を検出するためのシステム及び方法が開示される。

【0006】

1つの例示的な態様において、Webページの異常要素を検出するためのコンピュータ実装方法は、Webサーバを介してWebサイトに紐づいた前記Webページを要求し、クライアントコンピューティングデバイスによって前記Webサイトへのアクセスを取得するステップと、前記クライアントコンピューティングデバイスによって、前記Webページに関するデータを収集するために前記Webページを実行するステップと、少なくとも前記収集したデータに基づいて、少なくとも1つのN次元ベクトルを決定するステップと、前記少なくとも1つのN次元ベクトルに基づいて、N次元空間内の前記Webページの少なくとも1つの要素についてのベクトルの座標の値の組を含む少なくとも1つのクラスタを作成するステップと、前記少なくとも1つのクラスタに基づいて、前記Webページの統計モデルを作成するステップと、を備える。

20

【0007】

1つの例示的な態様において、方法は、前記少なくとも1つのN次元ベクトルを格納するステップと、前記少なくとも1つのクラスタを作成する前に、少なくとも1つのN次元ベクトルに関する追加のデータを収集するため他のクライアントコンピューティングによりWebページを取得して実行するステップと、を備える。

【0008】

1つの例示的な態様において、前記Webサーバは、前記Webページを要求するクライアントコンピューティングデバイスからの要求を受信したことに応じて、前記Webページに少なくとも1つのスクリプトを追加するよう構成される。

30

【0009】

1つの例示的な態様において、前記クライアントコンピューティングデバイスによって前記Webページを実行するステップは、前記少なくとも1つのスクリプトを実行して前記Webページの少なくとも1つの要素に関するデータを収集するステップを有する。

【0010】

1つの例示的な態様において、方法は、前記Webページのクラスタの前記少なくとも1つのN次元ベクトルと、先に構築された少なくとも1つの前記Webページの統計モデルのN次元ベクトルとを比較するステップと、前記Webページの前記少なくとも1つの要素を、下記（1）～（3）のうち少なくとも1つを検出することにより異常であると判定し識別するステップと、を備える。

40

（1）N次元空間における、前記Webページの前記少なくとも1つの要素の前記少なくとも1つのN次元ベクトルと、前記Webページの統計モデルクラスタの中心との距離が、前記クラスタの半径よりも大きい。

（2）N次元空間における、前記Webページの前記少なくとも1つの要素の前記少なくとも1つのN次元ベクトルと、前記Webページの統計モデルのクラスタの中心との間の近接度が、第1の選択された閾値よりも大きい。

（3）N次元空間における、前記少なくとも1つの要素の前記少なくとも1つのN次元ベクトルと、前記Webページの統計モデルのクラスタのN次元ベクトルのうちクラスタ中心から

50

最も離れたベクトルとの間の近接度が、第2の選択された閾値よりも大きい。

【0011】

1つの例示的な態様において、方法は、前記少なくとも1つの要素が異常でないことを検出すると、前記Webページの前記少なくとも1つの要素の前記少なくとも1つのN次元ベクトルを前記Webページの統計モデルに追加するステップをさらに備える。

【0012】

1つの例示的な態様において、前記Webサーバは、Webページの少なくとも1つの要素の異常の検出に応答して前記クライアントコンピューティングデバイスとの接続を無効にし、前記少なくとも1つの要素の統計的有意性を閾値に関して少なくとも決定するために前記Webページの前記少なくとも1つの要素に対してアンチウィルススキャンを実行し、前記少なくとも1つの要素の統計的有意性が閾値よりも大きいことの検出に応答して、前記少なくとも1つの要素が安全であると識別し前記クライアントコンピューティングデバイスとの接続を再確立するよう構成される。

10

【0013】

1つの例示的な態様において、方法は、前記Webページの前記少なくとも1つの要素の前記少なくとも1つのN次元ベクトルが前記Webページの統計モデルに対応していないことの検出に応答して、前記少なくとも1つの要素に関連するWebページの数と、前記Webページの統計モデルに用いられるWebページの総数との比率を決定するステップをさらに備える。

【0014】

20

1つの例示的な態様において、前記少なくとも1つの要素に関連するWebページの数には、前記少なくとも1つの要素又はそのN次元空間におけるN次元ベクトル間の距離が選択された閾値未満である近傍の要素を含むWebページの数が含まれる。

【0015】

1つの例示的な態様において、Webページの異常要素を検出するためのシステムは、少なくとも1つのクライアントコンピューティングデバイスのプロセッサを備え、当該プロセッサは、Webサーバを介してWebサイトに紐づいた前記Webページを要求することで前記Webサイトへのアクセスを取得し、前記クライアントコンピューティングデバイスによって、前記Webページに関するデータを収集するために前記Webページを実行し、少なくとも前記収集したデータに基づいて、少なくとも1つのN次元ベクトルを決定し、前記少なくとも1つのN次元ベクトルに基づいて、N次元空間内の前記Webページの少なくとも1つの要素についてのベクトルの座標の値の組を含む少なくとも1つのクラスタを作成し、前記少なくとも1つのクラスタに基づいて、前記Webページの統計モデルを作成するよう構成される。

30

【0016】

1つの例示的な態様において、Webページの異常要素を検出するためのコンピュータ実行可能命令が格納された非一時的コンピュータ可読媒体は、Webサーバを介してWebサイトに紐づいた前記Webページを要求し、クライアントコンピューティングデバイスによって前記Webサイトへのアクセスを取得させる命令と、前記クライアントコンピューティングデバイスによって、前記Webページに関するデータを収集するために前記Webページを実行させる命令と、少なくとも前記収集したデータに基づいて、少なくとも1つのN次元ベクトルを決定させる命令と、前記少なくとも1つのN次元ベクトルに基づいて、N次元空間内の前記Webページの少なくとも1つの要素についてのベクトルの座標の値の組を含む少なくとも1つのクラスタを作成させる命令と、前記少なくとも1つのクラスタに基づいて、前記Webページの統計モデルを作成させる命令と、を備える。

40

【0017】

1つの例示的な態様において、本発明の例示的な態様の上記簡略化した概要は、本発明の基本的な理解を提供する役割を果たす。この概要は、全ての熟考された形態を含んだ広範な概要ではなく、全ての態様の重要な要素を特定することも、本発明の任意の又は全ての態様の範囲を定めることも意図していない。本発明の1つ以上の態様は、前述の目的を達成するため、特に請求の範囲において挙げられている特徴を含んでいる。

50

【図面の簡単な説明】

【0018】

添付図面は、本明細書に組み込まれ、本明細書の一部を構成する。添付図面は、本願の1つ以上の例示的な態様を示しており、発明の詳細な説明とともに、それらの態様の原理及び実装を説明する役割を果たす。

【0019】

図1は、本発明の態様に係る、Webページの異常要素を検出するための例示的なシステムを示す図である。

【0020】

図2は、本発明の態様に係るN次元空間の例を、統計モデル及びクラスタのメトリックとともに示す図である。

10

【0021】

図3は、本発明の態様に係る、Webページの異常要素を検出するための例示的な方法を示す図である。

【0022】

図4a、図4b、図4cは、本発明の態様に係る、統計モデルの視覚化の例を示す図である。

【0023】

図5は、Webページの異常要素を検出するシステム及び方法の態様を実施可能な汎用コンピュータシステムの一例を示す図である。

20

【発明を実施するための形態】

【0024】

以下では、Webページの異常要素を検出するためのシステム、方法及びコンピュータプログラム製品についての、発明の例示的な態様を説明する。当業者であれば、以下の記載はあくまでも例示であり、あらゆる制限をも意図するものではないことが理解されよう。本発明の利益を享受する当業者であれば、他の態様も容易に示唆されることであろう。添付の図面では、その詳細を示すための参照符号が付されている。図面と以下の記載において、同一の又はそれに相当する項目については、同一の参照符号を用いるものとする。

【0025】

以下の用語が、本発明の例示的な態様を説明する際に使用される。

30

【0026】

Webページは、Webクライアント（ブラウザ）によって処理されるようWebサーバによって生成されたデータ含み、ハイパーテキストマークアップ言語（HTML、XHTML、XML、WML、VML、PGML、SVG、XBRL等）やスクリプト言語（JScript、JavaScript、ActionScript、Tcl、Lua、Perl、PHP、Python、REBOL、Ruby等）を用いて構成される。

【0027】

本明細書において、コンテンツはWebページのコンテンツのことを示すこととする。

【0028】

スクリプトは、スクリプト言語で書かれた実行中のプロシージャを含み、厳密に決定されたWebページを表示する際に到着する要求によって、サーバ側又はクライアント側で実行されてもよい。

40

【0029】

インラインスクリプトは、その実行形式のコード（body）がWebページのコンテンツの一部となったスクリプトを含む。例えば、インラインスクリプトはタグ<script> </script>の間に配置される。

【0030】

タグ（マーカ）は、ハイパーテキストマークアップ言語（HTML）の特別な構成を含む。タグは、山カッコ<name_tag>で囲まれたテキストを構成する。各タグは、タグ及びそれに続くコンテンツをどのように表示するか特定の命令をブラウザにもたらし。特定の場合において、タグは、タグを明確にし、タグの可能性を拡張してより柔軟な制御を可能にす

50

る属性を有する（例えば、コンテナタグの中身）。例えば、`<script src = "URL"> ... </script>`などである。src属性は、スクリプトのbodyの場所を示す。

【0031】

コンテナタグは、開始タグおよび終了タグを有するペアタグを含む。コンテナタグは、テキストとハイパーテキスト言語の他の要素の両方を含む。

【0032】

Webページの要素（マークアップ言語の要素）は、開始タグ及び終了タグ（ある場合、例えば`<br>`タグの場合には、開始タグと終了タグは一致する。）の組み合わせと、タグ間のコンテンツを含むことができる。Webページの要素の全体は、Webページのコンテンツを形成する。要素には、対応するタグの名前によって区別される少なくとも次の種類のものが存在する：ハイパーリンク、テキストブロック、テキストフォーマット、リスト、オブジェクト（例えば、メディアファイル、アプレット、スクリプト、ネイティブコード等）、イメージ、イメージマップ、表、フォーム、文字等。

【0033】

要素のN次元ベクトルには、順序付けられたn個の実数の組が含まれ、その数にはベクトルの座標が含まれる。ベクトルの座標の個数は、ベクトルの次元数として知られている。座標は、N次元空間における、Webページの対応する要素（スクリプトなど）又はWebページの同じ種類の要素グループ（フォームの要素など）の位置を定める（図2は2次元空間の例を示す）。ベクトルは、要素又は要素グループのコンテンツに関する情報を変換することによって得られる。ベクトルは、要素又は要素グループのコンテンツに関する特定の情報を表す。一例では、各座標が要素のコンテンツの特性の1つを表す。例えば、1つの座標がスクリプト内の演算子の数を示し、別の座標は評価演算子の数を示す。数値に、要素のコンテンツの文字列パラメータの辞書式順序、又は異なる要素の文字列パラメータ間のレーベンシュタイン距離を反映させることもできる。例えば図2は、ベクトル、特に、座標（1666,1889）及び（1686,1789）を有する2次元ベクトルの例を示す。

【0034】

クラスタは、N次元空間において、厳密に定義された要素又は要素グループについてのベクトルの座標の許容値の集合を含有する。1つの例示的な態様によれば、ある選択された要素又は要素グループは、その要素のN次元ベクトルからあるクラスタの中心までの距離がそのクラスタの半径よりも小さい場合に、当該クラスタに割り当てられる。図2は、クラスタ210'の一例を示す。一例において、ある要素のN次元ベクトルとあるクラスタの要素のうち最も近いN次元ベクトルとの間の距離（図2の「d'」）の値が許容できる最大距離（距離[d']の閾値）よりも小さいか、又は、ある要素のN次元ベクトルからあるクラスタの中心までの距離（図2の「d」）の値がこのクラスタの半径より小さい場合、その要素は当該クラスタに割り当てられる。例えば、ベクトル（1666,1889）からクラスタの中心までの距離はクラスタの半径よりも小さいので、そのコンテンツがベクトルに表された要素又は要素グループは、当該クラスタに属することになる。一方、ベクトル（1686,1789）からクラスタの中心までの距離はクラスタの半径よりも大きく、最も近いN次元ベクトルまでの距離は閾値よりも大きいので、そのコンテンツがベクトルに表された要素又は要素グループは、当該クラスタに属さない。近接度を評価するための距離の種類には、直線距離、ユークリッド距離、ユークリッド距離の2乗、一般化されたミンコフスキー指数距離、チェビシェフ距離、マンハッタン距離等があるが、これらに限定されるわけではない。

【0035】

近接度（類似度、類似度係数）は、Webページの要素の類似性を判定するための無次元パラメータを含む。近接度を定めるために用いる方法（基準）には、落合係数、Jaccard係数、Sokal-Sneath係数、Kulczynski係数、対称なDice係数などが含まれるが、これらに限定されない。

【0036】

クラスタの中心（重心）は、N次元空間における複数のN次元ベクトルの幾何平均位置で

ある。1つのベクトルからなるクラスタの場合は、そのベクトルをクラスタの中心とすることができる。

【0037】

クラスタの半径（図2の「R」）は、クラスタ内に存在するN次元ベクトルのクラスタ中心からの最大距離である。

【0038】

クラスタリングには、階層型（凝集型及び分割型）と非階層型を含む様々な既知のアルゴリズム及び手法を使用することができる。

【0039】

Webページの要素の統計モデル（Webページの要素のモデル）は、1種類の要素又は1種類の要素グループについてのクラスタ210の集合を含む。例えば、Webページのスク립トの統計モデルや、Webページのフォームの統計モデルである。図2に示すように、Webページの要素の統計モデルは220と表される。1つのクラスタで構成されるモデルの場合、そのクラスタが要素のモデルとなる。

【0040】

Webページの統計モデル（Webページのモデル）は、Webページの全ての種類の要素及び／又は要素グループ（異なる種類の要素を備えた要素グループを含む）のクラスタの集合を含む。例えば、認証ページの統計モデルである。言い換えると、Webページの統計モデル230は、Webページの要素のモデル220の集合を含む。同様に、Webサイトの統計モデルは、Webサイトの全種類の要素及び／又は要素グループのクラスタの集合を含む。つまり、Webサイトの統計モデル（図示せず）は、Webページのモデル230の集合を含むことになる。

【0041】

Webページの異常要素には、Webページの要素であって、そのベクトルが特定のタイプの要素に対して構築されたいずれの統計モデルのクラスタにも割り当てられていないもの、又は、統計的有意性が閾値以下のものが含まれる。

【0042】

要素の統計的有意性は、モデル構築のために取得したWebページの総数又は所定のセクション（評価セクション）においてモデルの構築のために取得したWebページの数に対する、評価対象の要素がWebページのコンテンツ内で見つかった回数の比とする。ここで、セクションの範囲は、例えば要素の観測の開始時間のような特定の時間以降に、モデルの構築のため得られたWebページの数によって決定される。例えば、100ページを取得し、評価対象の要素が30回見つかった場合、統計的有意性は30%である。

【0043】

クラスタの統計的有意性は、モデル構築のために取得したWebページの総数又は所定のセクションにおいてモデルの構築のために取得したWebページの数に対する、Webページのコンテンツ内で評価され、そのベクトルがクラスタを形成する要素の数の比とする。ここで、セクションの範囲は、例えばクラスタの観測の開始時間のような特定の時間以降に、モデルの構築のために得られたWebページの数によって決定される。

【0044】

要素又はクラスタの統計的有意性の値を、統計的有意性の閾値とすることができる。この値を超えると、当該要素又はクラスタ（及びクラスタの要素）は統計的に有意であるとみなされる。そして、クラスタの要素の統計的有意性の値が設定された閾値よりも低い場合、要素又はクラスタは異常であるとみなされる。

【0045】

クラスタの形成には、階層的方法を用いることができ、例えば、凝集法により形成することができる。凝集法では、（距離の）最も近い要素のN次元ベクトルの組が取り出されてクラスタが生成され、又は、（距離の）最も近いクラスタの組が1つのクラスタに併合される。この方法を用いる場合、線形又はユークリッド距離や、一般化された指数関数的なMinkowski距離、Chebyshev距離又はManhattan距離を使用することができる。そして、

最小の相互距離を有するベクトルの組は最も近いものとされ、クラスタは、その半径ができるだけ半径の閾値に近づくよう分割される。ここで、最も近づいた半径は、クラスタを分離する次の回における半径の閾値を超えるものであっても良い。別の場合、許容可能な近接度が残っているクラスタ又はベクトルがなくなるまで、クラスタを分離することができる。この場合、許容可能な近接度は、設定された閾値を超えない値である。中心間の距離が最も短い2つのクラスタが最も近いクラスタである。

【0046】

あるいは、分割法によりクラスタを形成することもできる。分割法では、相互距離が最大許容距離よりも小さいベクトルの集合によってクラスタが形成される。距離の最大許容度は閾値によって決定され、クラスタは、例えば、その半径が当該半径の閾値と同じかそれ未満になるまで分割される。

【0047】

図1は、本発明の態様に係るWebページの異常要素を検出するための例示的なシステムを示す。このシステムは、限定はしないが、以下のデバイスを含むことができ、これらはネットワークを介して接続されている：Webクライアント110がインストールされたユーザデバイス120、Webサーバ130、コントロールサーバ150、データベース160。

【0048】

1つの例示的な態様では、WebブラウザなどのWebクライアント110をユーザのデバイス120にインストールすることができる。Webクライアント110は、Webサイトのコンテンツを要求、処理、操作及び表示するよう構成され、論理的に相互に関連するWebページ100のセットを含んでいる。Webクライアント110は、例えば、Webサーバ130に対して指定されたユニフォーム・リソース・ロケータ（URL）アドレスを用いてリソースを取得する要求を送信し、一般にはWebサーバ130のWebページ100又はWebページの要素とともに、応答を受信する。Webクライアント110からの要求に応じて、Webサーバ130は、準備されたWebページ100を送信したり、動的にページを生成したりすることができる。本発明のWebサーバ130によってクライアントに送られる各Webページには、通常のコンテンツに加えて、スクリプトが追加されている。スクリプト140の機能には、少なくともWebクライアント側110においてそのスクリプト140を含むWebページ100のデータ（Webページの要素又は要素グループに関する情報、1つの要素に関する情報、特定の場合には1つの要素のコンテンツの情報を含む）を収集することが含まれる。一例では、Webページ100の要素に関する情報には、その要素のコンテンツが含まれている。クライアント側100におけるWebページ100の要素及びWebページ100の要素のコンテンツは、サーバ側130におけるWebページの同じバージョンの要素及びWebページ100のこれらの要素のコンテンツとは異なる場合がある。この理由は、Webクライアント側110のWebページの動的リフレッシュによるものか、又は、「マン・イン・ザ・ブラウザ」攻撃によるものである。

【0049】

1つの例示的な態様では、コントロールサーバ150は、スクリプトによって収集されたWebページの要素又は要素グループに関する情報を受け取る。スクリプトは、収集されたデータを「生の」状態又は変換された状態のいずれかの状態で送信することができ、送信されたデータのフォーマットは、Webサーバ130がWebページ100に追加したスクリプト140の機能によって決定される。すなわち、実行中のスクリプトは、Webページ100の要素に関する厳密に定義された情報を厳密に定められた形式で送信し、スクリプトの機能により決定される。あるいは、スクリプトは、Webサーバ又はコントロールサーバに、クライアント側110においてスクリプトの開始が成功したことについてのデータを送信し、応答として、Webページ100のどの要素を収集する必要があるか、どのような形式によって受信者（Webサーバ130又はコントロールサーバ150に直接）に情報を送信するかについての命令を受信する。

【0050】

データ変換の主な方法には、量子化、ソート、マージ（貼り付け）、グループ化、データセットの設定、テーブルへの値の挿入、算出値、データコーディング、正規化（スケールリング）等がある。

【0051】

一例では、データは、データ変換の結果として情報の属性を取得する。

【0052】

スクリプトの変換方法の1つには、抽象構文木の構築、重要なオペレータのみの受信者（Webサーバ130又は直接コントロールサーバ150）への転送（送信）、スクリプト140の設定により又は受信者からの命令により予め決定された構成が含まれる。

【0053】

10

1つの例示的な態様では、スクリプト140によって収集された全てのデータをコントロールサーバ150に転送する。コントロールサーバ150は、Webクライアント110から直接データを取得するか、又はWebサーバ130を介してデータを取得するよう構成される。特定の場合、コントロールサーバ150は、Webサーバ130と同じネットワーク内に存在する。コントロールサーバ150によって収集されたデータは、Webページの統計モデル230を構築してWebページの異常要素を検出するために用いられる。多数のモジュール（図示せず）を、コントロールサーバ150上に実装することができる。コントロールサーバ150の処理モジュールは、スクリプト140によって収集されたデータをN次元ベクトルに変換し、データベース160に格納する。コントロールサーバ150の解析モジュールは、受信したベクトルからクラスタ210を形成し、異常な要素又は要素グループを検出する。なお、異常な要素又は要素グループのコンテンツは、受信したベクトルにより表されている。この機能は、N次元空間におけるN次元ベクトルと形成されたクラスタ210の相互比較によって実現される。

20

【0054】

データベース160は、構築されたモデル及びベクトルを記録するよう構成される。

【0055】

本願に開示されたシステムは、Webページの統計モデル230の構築方法及び、図3に示すWebページの構築されたモデル230を用いたWebページ100の異常要素の検出方法を含む、いくつかの方法を実行することができるが、これらの方法に限定されるわけではない。

30

【0056】

1つの例示的な態様では、Webページの統計モデル230の構築方法は、以下のように実施することができる。ステップ300において、ユーザは自身のデバイスからWebサイトへアクセスする。Webクライアント110は、Webサーバ130へ送信された要求により、Webサーバ130からサイトのWebページを取得するよう構成される。そのプロセスの間に、Webサーバ130は、Webページにスクリプト140を追加する。ステップ310において、スクリプトをWebクライアント側110で実行し、Webページ100に含まれるデータを収集する。スクリプト140によって収集されたデータには様々な情報が含まれる。特に、スクリプト140は、Webページの少なくとも1つの要素（スクリプト、フォームなど）のコンテンツを収集することができる。スクリプト140によって収集されたデータは、必要に応じて変換される。データは、スクリプト140自体によって、又はコントロールサーバ150上の処理手段によって変換される。ステップ320において、収集されたデータを、少なくとも1つのN次元ベクトルに変更し、ステップ330において保存する。ステップ350において、少なくとも1つのベクトルから、少なくとも1つのクラスタ210を形成する。ステップ360において、形成された少なくとも1つのクラスタ210に基づいて、Webページの統計モデル230を生成する。

40

【0057】

一例では、ステップ300'において、取得したN次元ベクトルを保存したあと、別のWebクライアント110にWebページ100を取得させ、ステップ320において、このWebページから収集したデータに基づいてN次元ベクトルを追加的に取得させ、この後にクラ

50

スタが形成されてもよい。

【0058】

別の例では、クラスタ210を作成しモデル230を構築した後、ステップ300'において、別のWebクライアント110によりWebページ100を取得する。このWebページからスクリプト140により収集されたデータに基づいて、N次元ベクトルを取得し、取得したN次元ベクトルに基づいて、先に作成されたクラスタ210を修正（リフレッシュ）する（クラスタの半径、中心／重心を変更する）か、新たなクラスタ210を作成する。その結果、Webページの統計モデル230は、（クラスタ210の修正により）精緻化され、（新たなクラスタ210の作成により）補完される。スクリプト140が収集するデータは、それ以前のバージョン（イテレーション）において収集したデータと異なってもよく、例えば、Webページ100の他の要素の情報を収集しても良い。

10

【0059】

1つの例示的な態様では、ウェブページの統計モデル230に基づいて異常要素を検出する方法は、以下のように実施することができる。ステップ300において、ユーザは自身のデバイスからWebサイトへアクセスする。Webクライアント110は、Webサーバ130へ送信された要求により、Webサーバ130からサイトのWebページを取得するよう構成される。そのプロセスの間に、Webサーバ130はWebページにスクリプト140を追加する。ステップ310において、スクリプトをWebクライアント側110で実行し、Webページ100に含まれるデータを収集する。スクリプト140によって収集されたデータには様々な情報が含まれる。特に、スクリプト140は、Webページの少なくとも1つの要素（スクリプト、フォームなど）のコンテンツを収集することができる。スクリプト140によって収集されたデータは、必要に応じて変換される。データは、スクリプト140自体によって、又はコントロールサーバ150上の処理手段によって変換される。ステップ320において、収集されたデータを、少なくとも1つのN次元ベクトルに変更し、ステップ330において保存する。ステップ370において、得られたベクトルを、（例えば取得したベクトルとクラスタの中心との相互距離を測定することによって、）Webページの構築された統計モデル230のクラスタ及び／又は特定の統計モデルのN次元ベクトルと比較する。ステップ370において、比較の結果として、分析した要素を、例えば以下の場合に異常であると識別する。

20

（i）N次元空間において、その要素のN次元ベクトルとモデルの各クラスタの中心との距離が、これらのクラスタの半径よりも大きい。

30

（i i）N次元空間において、その要素のN次元ベクトルとモデルの全てのクラスタの中心との間の近接度が、閾値よりも大きい。

（i i i）N次元空間において、その要素のN次元ベクトルとモデルのクラスタのN次元ベクトルのうちクラスタ中心から最も離れたベクトルとの間の近接度が、閾値よりも大きい。

【0060】

一例では、要素が異常であると認識されない場合、ステップ350'において、その要素のN次元ベクトルをWebページの統計モデル230に追加する。

【0061】

別の例では、Webサーバ130は、Webページ100の異常要素を検出すると、Webクライアント110及びユーザのデバイス120との接続を無効にするか、又は、その接続は維持するもののWebサーバ130がクライアント110の要求に対する応答を停止する（接続によるデータ送信を停止する）よう構成される。データ送信が停止している間、検出されたWebページの異常要素は、悪意のある機能（危険性）の存在を理由に、コントロールサーバ150のアンチウィルス手段（図示せず）によりスキャンされるか、又は監視される。閾値を超える統計的有意性を有するクラスタが検出した要素の周囲に形成される場合には、検出された異常要素は安全であると識別され、接続が再確立され、セッションが継続される。

40

【0062】

50

特定の場合、異常要素が含まれているかどうか事前に分かっていないWebページに基づいてモデルを構築すると、競合が発生する可能性がある。すなわち、要素のN次元ベクトルがモデルのいずれのクラスタにも現れず、ジレンマ—当該ベクトルに基づいて新たなクラスタを作成するのか、又は当該ベクトルによってそのコンテンツが表されている要素を異常なものともみなすのか—が発生する。競合は、評価セクションで評価されている要素と似た（近い）要素に基づいて作成された、要素又はクラスタの統計的有意性の評価に基づいて解決される。すなわち、評価対象の要素（又は、近傍の要素、つまりN次元空間内のN次元ベクトル間の距離がある閾値よりも小さい要素）を含むWebページの数と、セクションで評価されるモデルの構築に使用されているWebページの総数との比に基づいて、セクションの範囲がページ数又はバージョン（イテレーション）数として測定される。

評価セクションにおける評価されたWebページの要素の統計的有意性の値が、同一セクションにおける他の要素の統計的有意性の値（又は統計的有意性の平均値）に近い（近接性は閾値によって決定される）か、又はある閾値を超える（例えば、20%）場合、その要素は統計的に有意であると認識され、そうでなければ（それが閾値を超えなければ）異常であると認識される。例えば、モデル構築ステップにおいて、そのベクトルが以前に形成されたクラスタ210のいずれにも現れないWebページ100の要素が出現する可能性がある。したがって、例えばそのWebページ100が含まれるセクションにおける統計的有意性を定めることによって、その要素が異常であるかどうかを決めることが必要である。一例では、所定のタイプの要素の統計的有意性の閾値は20%である。この検証は、要素の特定の範囲において、評価された要素に近い要素が4回出現したことを示し、統計的有意性の値が閾値より小さい2%であることに対応する。したがって、評価された要素とそれに近い要素（評価された要素の周囲に形成されたクラスタ）は異常である可能性がある。特定の場合、統計的有意性の閾値を、同じ種類の要素についてのクラスタの統計的有意性の最小値と定めることもできる。例えば、モデルが25%、32%、47%及び95%の統計的有意性を有するスクリプトのクラスタを含む場合、同じ種類の要素の閾値を25%に設定することができる。

【0063】

1つの例示的な態様では、統計的有意性はWebページの異常要素の検出にも用いられる。例えば、これは、統計モデルが構築されていない場合、又は上述のジレンマの解決中に使用される。第1のステップにおいて、Webクライアント110により、Webサーバ130からWebページ100を取得し、ユーザのデバイス120上で実行する。スクリプト140を含むWebページは、スクリプト140が実行されると、Webページ100の少なくとも1つの要素のコンテンツに関する情報をWebクライアント側110で収集し、収集された情報をユーザのデバイスから送信する。前述のスクリプトは、Webクライアント110の助けを借りて実行され、Webページ100の要素の少なくとも1つのコンテンツの情報をWebクライアント側110で収集し、収集した情報を、Webクライアント110がWebページ100を受信したユーザのデバイス120から送信する。コントロールサーバ150において、デバイス120から受信したコンテンツに関する情報は、要素のN次元ベクトルに変換されるよう構成されており、得られたN次元ベクトルは任意の適切な方法によってクラスタ化される。

N次元ベクトルは、Webページの各要素について、要素のグループについて、あるいは同じ種類の要素のグループについて形成される。また、異なる種類の要素がグループを形成していてもよい。クラスタ210が少なくとも1つのベクトルを含む場合、クラスタ210が形成されてから、得られたクラスタ210の統計的有意性を定めても良い。ここで、統計的有意性は、クラスタ210内のN次元ベクトルの数とWebページ100の数との比と定めることができる。なお、ここでのWebページ100の数は、その要素のコンテンツの情報が収集され、コントロールサーバ150又はWebサーバ130に送信されたWebページの数のことである。閾値の有意性は、上述した方法によって定められ、要素の種類、クラスタリングの方法、評価セクションの範囲等に依存させることができる。

【0064】

1つの例示的な態様では、ユーザがWebバンキングサイト、例えばhttps://my.KasperskyBank.ru/のWebページを要求すると、要求されたWebページはスクリプト140を追加し、ページ100がユーザのデバイス120上で実行されるWebクライアント110に送信される。ユーザ側のスクリプト140は、以下のようなWebページ上の<script>要素を収集する。

```
<script>document.documentElement.id="js";var .../Kasperskybank/;</script>
```

```
<script src="//static.kaspersky.ru/dist/kfs/kfs.js" crossorigin="anonymous"></script>
```

10

```
<script src="https://static.kaspersky.ru/ib/prod/2842c77095d860e412d7a8cf30231fd53c89fb4e/Kasperskybank/Kasperskybank.js" crossorigin="anonymous"></script>
```

```
<script async="" src="/kfs/kfs"></script>
```

```
<script>!function(){var e=document.getElementById("before-init__noscript");e&&(e.className="ui-browser__holder-block-hide");var o=function(){try{return"withCredentials"in new XMLHttpRequest}catch(e){return!1}}();if(o){var t=function(){if(navigator.cookieEnabled)return!0;document.cookie="cookietest=1";var e=-1!=document.cookie.indexOf("cookietest=");return document.cookie="cookietest=1; expires=Thu, 01-Jan-1970 00:00:01 GMT",e}();if(t)document.body.removeChild(document.getElementById("before-init"));else{var n=document.getElementById("before-init__nocookies");n&&(n.className="ui-browser__holder-block")}}else{var r=document.getElementById("before-init__old-browser");r&&(r.className="ui-browser__holder-block")}}();</script>
```

20

【0065】

src属性を持つ<script>要素の場合、スクリプト本体のロードと正規化が実行される。インラインスクリプトの場合は、正規化のみである。例えば、上述したインラインスクリプトの場合、正規化された形式は次のようになる（重要な言語構成と標準的なオブジェクト／メソッドのみを記述し、リテラルは「非人格化」している）。

30

```
document.documentElement.i0=v0;var i1=window.i1||{};i1.i2=v1,i1.i3=v2,i1.i4=v3,i1.i5=v4,i1.i6={i7:v5,i8:v6},i1.i9=v7;
```

```
!function(){var i0=document.getElementById(v0);i0&&(i0.i1=v1);var i2=function(){try{return v2 in new XMLHttpRequest}catch(i0){return!v3}}();if(i2){var i3=function(){if(navigator.i4)return!v4;document.cookie=v5;var i0=-v3!=document.cookie.indexOf(v6);return document.cookie=v7,i0}();if(i3)document.body.removeChild(document.getElementById(v8));else{var i5=document.getElementById(v9);i5&&(i5.i1=v10)}}else{var i6=document.getElementById(v11);i6&&(i6.i1=v10)}}();
```

40

【0066】

次に、スクリプト140は、ページ上に存在する<input>要素を収集する。

```
<input autocomplete="off" autocorrect="off" autocapitalize="off" class="m-login__form-field-input ng-pristine ng-invalid ng-invalid-required ng-touched" type="text" ... ng-blur="login.focus = false" placeholder="password">
```

【0067】

スクリプト140は、<input>要素の収集されたデータを、例えば以下のように正規化しつつ変換する（属性はアルファベット順にソートされ、タグ名は切り捨てられ、属性値の空白は切り取られ、属性は「;」によりリスト化される）。

```
<autocapitalize=off;autocomplete=off;autocorrect=off;class=m-login__form-field-inputng-pristineng-invalidng-invalid-requiredng-touched;name=lg;ng-blur=login.foc
```

50

```
us=false;warmUp();;ng-change=input(true);ng-disabled=false;ng-keydown=login.focus=true&&$event.keyCode===13&&authUser();ng-keyup=fix(login.form.lg,$event);ng-model=login.lg;placeholder=login;spellcheck=false;style=padding:0px;;type=text;ui-focus=login.setFocus;validator=validator.lg>
```

```
<autocapitalize=off;autocomplete=off;autocorrect=off;class=m-login__form-field-inputng-pristineng-untouchedng-invalidng-invalid-required;name=pw;ng-blur=login.focus=false;ng-change=input();ng-disabled=false;ng-keydown=login.focus=true&&$event.keyCode===13&&authUser();ng-keyup=fix(login.form.pw,$event);ng-model=login.pw;placeholder=password;spellcheck=false;type=password;validator=validator.pw>
```

10

【0068】

スクリプト140は、収集されたデータをコントロールサーバ150に送信する。コントロールサーバ150は、対応するモデル（全てのスクリプト要素の集合形—スクリプト型要素の統計モデル220）のコンテキスト内の<script>要素の収集されたデータを、以下のように処理することができる。

- ・各スクリプトについて、数値ベクトルを取得する（例えば、ベクトルは2次元とすることができる）。ここで、ベクトルは文字列のコード（ASCIIのような文字コードを得るため、任意の適切な符号化方法が用いられる）から計算され、収集されたデータを構成する（インラインスクリプトの場合、このデータには正規化されたスクリプトの内容が含まれ、その他の場合、データはsrc属性の内容である）。結果として得られるWebページ100

20

o 16314,10816

o 2254,2598

o 16084,15036

o 356,822

o 20010,51838

- ・各ベクトルは、モデル230の2次元空間に保存され、異常がない所定の場合、全てのベクトルは以前に形成されたクラスタ内に収まる（すなわち、全てのベクトルが、以前のWebページのそれらのバージョンのスクリプト140から来たデータと一致する）。図4aにおいて、ドットは分析された<script>要素を示し、着色された領域は、モデル230の一部として以前に作成されたモデル220のクラスタ210である。

30

【0069】

コントロールサーバ150は、<input>要素の収集されたデータを同様の方法で処理する。結果として、視覚化は図4bに示す形式となる。異常要素が検出されていないため、処理は終了する。

【0070】

1つの例では、同じページ<https://my.kasperskyBank.ru/>のユーザの1人に、追加の<script>要素の形で、悪意のあるインジェクションが現れる。

```
<script src="https://static.kasperskyBank.ru/ib/prod/bank/malware.js" crossorigin="anonymous"></script>
```

40

【0071】

上述した方法によって計算されたベクトルは(4560,3192)に等しく、モデルは図4C（インジェクションの内容を表した異常である現在のベクトルは、赤色にマークされる）に示される形となる。検出された異常要素は、コントロールサーバ150のアンチウィルス手段によって処理され、接続自体を無効にすることができる。また同時に、モデル空間内の要素をその統計的有意性を測定するために観察しても良い。

【0072】

本発明に開示されている分析モジュール及び処理モジュールが実装されたWebサーバ、Webクライアント、データベース、コントロールサーバは、実際のデバイス、システム、コンポーネント、統合されたマイクロサーキットなどのハードウェアを使用して実現される

50

コンポーネントのグループ（特定用途向け集積回路、ASIC）、又は、フィールドプログラマブルゲートアレイ（FPGA）を含み、例えば、マイクロプロセッサシステム及びプログラム命令のセットのようなソフトウェアとハードウェアの組み合わせの形で、又はニューロシナプティックチップ上で実行することができる。システムの示された要素の機能は、ハードウェアのみによって、また、組み合わせによって実現される。後者の場合、システムの要素の機能の一部はソフトウェアによって実現され、一部はハードウェアによって実現される。特定の変形例では、要素のいくつか、又は要素の全ては、汎用コンピュータ（図5に示すものなど）のプロセッサ上に実装することができる。

【0073】

図5は、本発明の例示的な態様に係る、Webページの異常要素を検出するためのシステム及び方法の態様の一例を示す図である。図示のように、コンピュータシステム20（パーソナルコンピュータ又はサーバ）は、CPU21と、システムメモリ22と、CPU21に関連付けられたメモリを含む様々なシステムコンポーネントを接続するシステムバス23とを含む。当業者に理解されるように、システムバス23は、バスメモリ又はバスメモリコントローラ、周辺バス、及び他のバスアーキテクチャと相互作用するローカルバスを含むことができる。システムメモリは、永久メモリ（ROM）24及びランダムアクセスメモリ（RAM）を含む。基本入出力システム（BIOS）26は、ROM24を使用してオペレーションシステムをロードする際に、コンピュータシステム20の要素間で情報を転送するための基本的な手順を記憶する。

【0074】

ハードディスク27、磁気ディスクドライブ28、及び光学式のドライブ30は、ハードディスクインターフェース32、磁気ディスクインターフェース33、及び光学式のドライブインターフェース34それぞれを介してシステムバス23と接続される。ドライブ及び対応するコンピュータ情報メディアは、コンピュータ命令、データ構造体、プログラムモジュール、及びコンピュータシステム20の他のデータを蓄積するための電源依存のモジュールである。

【0075】

当業者であれば、コンピュータによって読み取り可能な形式でデータを記憶することのできる任意のタイプのメディア56（ソリッド・ステート・ドライブ、フラッシュメモリカード、デジタルディスク、ランダムアクセスメモリ（RAM）等）を使用できることが理解されよう。

【0076】

コンピュータシステム20は、オペレーティングシステム35が格納されるファイルシステム36と、追加のプログラムアプリケーション37と、他のプログラムモジュール38と、プログラムデータ39とを有する。コンピュータシステム20のユーザは、キーボード40、マウス42、又は当業者に知られた任意の他の入力デバイス（非限定的な例：マイクロフォン、ジョイスティック、ゲームコントローラ、スキャナ等）を用いて命令を入力する。このような入力デバイスは、典型的には、シリアルポート46を介してコンピュータシステム20に接続され、シリアルポートはシステムバスに接続される。ただし、当業者であれば、入力デバイスを、パラレルポート、ゲームポート、又はユニバーサル・シリアル・バス（USB）など（これらに限定されない）を介して接続することができることを理解するであろう。モニタ47又は他のタイプの表示装置もまた、ビデオアダプタ48などのインタフェースを介してシステムバス23に接続することができる。パーソナルコンピュータは、モニタ47に加えて、ラウドスピーカやプリンタなどの他の周辺出力装置（図示せず）を備えていても良い。

【0077】

コンピュータシステム20は、1つ又は複数のリモートコンピュータ49へのネットワーク接続により、ネットワーク環境で動作することができる。リモートコンピュータ（又は複数のコンピュータ）49は、コンピュータシステム20の性質を説明する上で前述した要素のほとんど又は全てを含むローカルコンピュータワークステーション又はサーバで

あっても良い。ルータ、ネットワークステーション、ピアデバイス、又は他のネットワークノードなどの他のデバイスも、コンピュータネットワークに存在することができるが、これに限定されるものではない。

【0078】

ネットワーク接続は、ローカルエリアコンピュータネットワーク（LAN）50及びワイドエリアコンピュータネットワーク（WAN）を形成することができる。このようなネットワークは、企業のコンピュータネットワーク及び社内ネットワークで使用され、一般にインターネットにアクセスする。LAN又はWANネットワーク内では、パーソナルコンピュータ20は、ネットワークアダプタ又はネットワークインタフェース51を介してローカルエリアネットワーク50に接続される。ネットワークが使用される場合、コンピュータ20は、モデム54又はインターネットなどのワイドエリアコンピュータネットワークとの通信を可能にする当業者に周知の他のモジュールを使用することができる。モデム54は、内部又は外部のデバイスとすることができ、シリアルポート46によってシステムバス23に接続される。上記のネットワークシステムが、通信モジュールを使用して1つのコンピュータが他のコンピュータへの接続を確立するための多くの周知の方法のうちの非限定的な例であることは、当業者には理解されるだろう。

【0079】

様々な実施形態において、ハードウェア、ソフトウェア、ファームウェア、又はこれらのあらゆる組み合わせにより、ここで説明されたシステム及び方法を実施し得る。ソフトウェアにおいて実装される場合は、方法は不揮発性コンピュータ可読メディアの1つ又は複数の命令又はコードとして保存され得る。コンピュータ可読メディアは、データストレージを含む。あくまでも例であり限定するものではないが、そのようなコンピュータ可読メディアは、RAM、ROM、EEPROM、CD-ROM、フラッシュメモリ、若しくは他のタイプの電気、磁気、光学式の記憶媒体、又はその他のメディアであってもよい。すなわち、これらによって命令又はデータ構造体という形で、要求されたプログラムコードを運ぶか又は保存することができ、汎用コンピュータのプロセッサによってアクセスすることができる。

【0080】

様々な実施形態で、本願のシステム及びメソッドはモジュールとして実装され得る。ここで用語「モジュール」は、実世界の機器、コンポーネント、又はハードウェアを用いて実装されたコンポーネント配置であり、例えばASIC（Application Specific Integrated Circuit）、FPGA（Field-Programmable Gate Array）等の、又は例えばモジュールの機能を実行するマイクロプロセッサシステムや命令セットによる等、ハードウェアとソフトウェアの組み合わせとして実装され得る。これらは、実行中にマイクロプロセッサシステムを特定の機器に変換する。モジュールは、ハードウェア単体により促進される一定の機能とハードウェア及びソフトウェアの組み合わせによって促進される他の機能という2つの組み合わせとして実施されてもよい。モジュールの少なくとも一部又は全部は、汎用コンピュータのプロセッサにおいて実行できる（図1～4を用いて上記において詳述したもの等）。したがって、各モジュールは様々な適当な構成で実現することができて、ここに例示した特定の実装に限られるものではない。

【0081】

なお、実施形態の通常機能のうちの全てをここで開示しているわけではない。本発明のいずれの実施形態を開発する場合においても、開発者の具体的な目標を達成するためには実装に係る多くの決定が必要であり、これらの具体的な目標は実施形態及び開発者ごとに異なることに留意されたい。このような開発努力は、複雑で時間を要するものであるが、本発明の利益を享受し得る当業者にとっては当然のエンジニアリングであると理解されたい。

【0082】

さらに、本明細書で使用される用語又は表現は、あくまでも説明のためであり、限定するものではない。つまり、本明細書の用語又は表現は、関連する技術分野の当業者の知識と組み合わせ、ここに示される教示及び指針に照らして当業者によって解釈されるべき

フロントページの続き

(51)Int.Cl.	F I	テーマコード (参考)
	G 0 6 F 21/12 3 8 0	
	G 0 6 F 21/44	
(72)発明者	オレグ ブイ. キュプリーヴ	
	ロシア 1 2 5 2 1 2 モスクワ, レニングラドスコ ショス 3 9 エー / 3, エーオー カ	
	スペルスキー ラボ	
(72)発明者	アントン ビー. ガルチェンコ	
	ロシア 1 2 5 2 1 2 モスクワ, レニングラドスコ ショス 3 9 エー / 3, エーオー カ	
	スペルスキー ラボ	
(72)発明者	ミハイル ブイ. ウスチノフ	
	ロシア 1 2 5 2 1 2 モスクワ, レニングラドスコ ショス 3 9 エー / 3, エーオー カ	
	スペルスキー ラボ	
(72)発明者	ヴィタリ ブイ. コンドラトフ	
	ロシア 1 2 5 2 1 2 モスクワ, レニングラドスコ ショス 3 9 エー / 3, エーオー カ	
	スペルスキー ラボ	
(72)発明者	ウラジミール エー. クスコフ	
	ロシア 1 2 5 2 1 2 モスクワ, レニングラドスコ ショス 3 9 エー / 2, カスペルスキ	
	ー ラボ エーオー	
F ターム (参考)	5B084 AA02 AA12 AB04 AB16 BB16 DB02 DC02 DC03	
	5B089 GA11 GA21 GB01 GB09 HA10 HB05 JA21 JB02 JB08 KA17	
	KB13	

【外国語明細書】
2018041442000001.pdf