

明 細 書

発明の名称：

ウェブサイトの自動巡回を行うためのシステム、方法、及びプログラム
技術分野

[0001] 本発明は、ウェブサイトの自動巡回を行うためのシステム、方法、及びプログラムに関するものである。

背景技術

[0002] 従来、様々な目的でウェブサイトの自動巡回（「クローリング」と呼ばれることもある。）が行われている（例えば、下記特許文献1を参照）。こうした自動巡回では、例えば、ウェブページに含まれる、リンク及びスクリプト等が設定されたボタン等の要素が選択されることによって、次のウェブページへの遷移が行われる。また、ウェブページに入力項目（例えば、入力可能なフォーム要素等）が存在する場合には、適切な値を入力して送信する必要がある、こうした入力値には、例えば、予め準備された値が設定される。

先行技術文献

特許文献

[0003] 特許文献1：特開2022-134094号公報

発明の概要

発明が解決しようとする課題

[0004] しかしながら、入力項目に対して設定される入力値のチェック方法は、ウェブサイトによって様々であり、全ての入力項目にとって適切な入力値を予め準備することは難しい。そして、不適切な入力値の設定によって予期しないエラーが発生してしまうと、自動巡回が中断してしまい、手動でのフォローが必要となってしまう。

[0005] 本発明の実施形態は、ウェブサイトの自動巡回における入力項目に対する適切な入力値の設定を支援することを目的の一つとする。本発明の実施形態の他の目的は、本明細書全体を参照することにより明らかとなる。

課題を解決するための手段

- [0006] 本発明の一実施形態に係るシステムは、1又は複数のコンピュータプロセッサを備え、ウェブサイトの自動巡回を行うためのシステムであって、前記1又は複数のコンピュータプロセッサは、前記ウェブサイトに含まれる第1ウェブページを取得するステップと、前記第1ウェブページにおける入力項目を特定し、前記入力項目に対して入力値を設定するステップと、設定した前記入力値の送信を伴って、前記第1ウェブページから遷移した第2ウェブページを取得するステップと、を実行し、前記入力値を設定するステップは、前記第1ウェブページに関する第1ページ情報の少なくとも一部を含む第1入力情報を第1学習済みモデルに対して入力し、前記第1学習済みモデルから出力される第1出力情報に基づいて、前記入力値を設定することを含む。
- [0007] 本発明の一実施形態に係る方法は、1又は複数のコンピュータによって実行され、ウェブサイトの自動巡回を行うための方法であって、前記ウェブサイトに含まれる第1ウェブページを取得するステップと、前記第1ウェブページにおける入力項目を特定し、前記入力項目に対して入力値を設定するステップと、設定した前記入力値の送信を伴って、前記第1ウェブページから遷移した第2ウェブページを取得するステップと、を備え、前記入力値を設定するステップは、前記第1ウェブページに関する第1ページ情報の少なくとも一部を含む第1入力情報を第1学習済みモデルに対して入力し、前記第1学習済みモデルから出力される第1出力情報に基づいて、前記入力値を設定することを含む。
- [0008] 本発明の一実施形態に係るプログラムは、ウェブサイトの自動巡回を行うためのプログラムであって、1又は複数のコンピュータに、前記ウェブサイトに含まれる第1ウェブページを取得するステップと、前記第1ウェブページにおける入力項目を特定し、前記入力項目に対して入力値を設定するステップと、設定した前記入力値の送信を伴って、前記第1ウェブページから遷移した第2ウェブページを取得するステップと、を実行させ、前記入力値を

設定するステップは、前記第1ウェブページに関する第1ページ情報の少なくとも一部を含む第1入力情報を第1学習済みモデルに対して入力し、前記第1学習済みモデルから出力される第1出力情報に基づいて、前記入力値を設定することを含む。

発明の効果

[0009] 本発明の様々な実施形態は、ウェブサイトの自動巡回における入力項目に対する適切な入力値の設定を支援する。

図面の簡単な説明

[0010] [図1]本発明の一実施形態に係るウェブサイト検査サーバ10を含むネットワークの構成を概略的に示す構成図。

[図2]検査管理テーブル152が管理する情報を例示する図。

[図3]ページ管理テーブル154が管理する情報を例示する図。

[図4]エラー管理テーブル156が管理する情報を例示する図。

[図5]ウェブサイトの脆弱性を検査する際にウェブサイト検査サーバ10が実行する処理を例示するフローチャート。

[図6]対象サイト設定画面50を例示する図。

[図7]ウェブサイトの自動巡回の際にウェブサイト検査サーバ10が実行する処理を例示するフローチャート。

[図8]入力項目に対する入力値を生成可能な大規模言語モデルにおける入出力情報の一例を示す図。

[図9]大規模言語モデルにおける入出力情報の他の例を示す図。

[図10]大規模言語モデルにおける入出力情報のさらに他の例を示す図。

[図11]検査結果確認画面60を例示する図。

[図12]入力値のみが出力される場合の大規模言語モデルにおける入出力情報の一例を示す図。

[図13]エラーの拡張情報を出力する大規模言語モデルにおける入出力情報の一例を示す図。

発明を実施するための形態

- [0011] 以下、図面を参照しながら、本発明の実施形態について説明する。各図面において、同一の又は類似する構成要素に対しては同一の参照符号が付され得る。
- [0012] 図1は、本発明の一実施形態に係るウェブサイト検査サーバ10を含むネットワークの構成を概略的に示す構成図である。サーバ10は、図示するように、インターネット等の通信ネットワーク20を介してユーザ端末30と通信可能に接続されている。図1においては、1つのユーザ端末30のみが図示されているが、サーバ10は、複数のユーザ端末30と通信可能に接続されている。ウェブサイト検査サーバ10は、ユーザ端末30を操作するユーザに対して、当該ユーザによって指定されるウェブサイト22を検査するウェブサイト検査サービスを提供する。ウェブサイト検査サーバ10は、本発明のシステムの全部又は一部を実装する装置の一例である。
- [0013] まず、ウェブサイト検査サーバ10のハードウェア構成について説明する。ウェブサイト検査サーバ10は、一般的なコンピュータとして構成されており、図1に示すように、コンピュータプロセッサ11と、メインメモリ12と、入出力I/F13と、通信I/F14と、ストレージ（記憶装置）15とを備え、これらの各構成要素が図示しないバス等を介して電氣的に接続されている。
- [0014] コンピュータプロセッサ11は、CPU又はGPU等として構成され、ストレージ15等に記憶されている様々なプログラムをメインメモリ12に読み込んで、当該プログラムに含まれる各種の命令を実行する。メインメモリ12は、例えば、DRAM等によって構成される。
- [0015] 入出力I/F13は、操作者等との間で情報をやり取りするための各種の入出力装置を含む。入出力I/F13は、例えば、キーボード、ポインティングデバイス（例えば、マウス、タッチパネル等）等の情報入力装置、マイクロフォン等の音声入力装置、カメラ等の画像入力装置を含む。また、入出力I/F13は、ディスプレイ等の画像出力装置、スピーカー等の音声出力装置を含む。

- [0016] 通信 I / F 1 4 は、ネットワークアダプタ等のハードウェア、各種の通信ソフトウェア、又はこれらの組み合わせとして実装され、通信ネットワーク 2 0 等を介した有線又は無線の通信を実現できるように構成されている。
- [0017] ストレージ 1 5 は、例えば、磁気ディスク又はフラッシュメモリ等によって構成される。ストレージ 1 5 は、オペレーティングシステムを含む様々なプログラム及び各種データ等を記憶する。例えば、ストレージ 1 5 は、図 1 に示すように、検査に関する情報を管理する検査管理テーブル 1 5 2 と、検査対象のウェブサイト 2 2 に含まれるウェブページに関する情報を管理するページ管理テーブル 1 5 4 と、ウェブサイト 2 2 の自動巡回の際に発生したエラーに関する情報を管理するエラー管理テーブル 1 5 6 とを有する。これらのテーブルは、その一部が 1 つのテーブルに統合されてもよいし、複数のテーブルに分割されてもよい。
- [0018] また、例えば、ストレージ 1 5 は、本発明の一実施形態に係るサーバ側プログラム 4 0 を記憶する。当該プログラム 4 0 は、サーバ 1 0 を、ウェブサイト検査サービスを提供するためのシステムの全部又は一部として機能させるためのプログラムである。サーバ側プログラム 4 0 の少なくとも一部は、後述する端末側プログラム 4 2 を介して、ユーザ端末 3 0 側において実行されるように構成され得る。
- [0019] 本実施形態において、ウェブサイト検査サーバ 1 0 は、それぞれが上述したハードウェア構成を有する複数のコンピュータを用いて構成され得る。例えば、サーバ 1 0 は、複数のサーバ装置によって構成される。
- [0020] このように構成されたウェブサイト検査サーバ 1 0 は、ウェブサーバ及びアプリケーションサーバとしての機能を有するように構成することができ、ユーザ端末 3 0 からの要求に応答して各種の処理を実行し、当該処理の結果に応じた画面データ（例えば、HTML データ）及び制御データ等をユーザ端末 3 0 に対して送信する。ユーザ端末 3 0 では、受信したデータに基づくウェブページ又はその他の画面が出力される。
- [0021] 次に、ユーザ端末 3 0 のハードウェア構成について説明する。ユーザ端末

30は、一般的なコンピュータとして構成されており、図1に示すように、コンピュータプロセッサ31と、メインメモリ32と、入出力I/F33と、通信I/F34と、ストレージ（記憶装置）35とを備え、これらの各構成要素が図示しないバス等を介して電氣的に接続されている。

[0022] コンピュータプロセッサ31は、CPU又はGPU等として構成され、ストレージ35等に記憶されている様々なプログラムをメインメモリ32に読み込んで、当該プログラムに含まれる各種の命令を実行する。メインメモリ32は、例えば、DRAM等によって構成される。

[0023] 入出力I/F33は、操作者等との間で情報をやり取りするための各種の入出力装置を含む。入出力I/F33は、例えば、キーボード、ポインティングデバイス（例えば、マウス、タッチパネル等）等の情報入力装置、マイクロフォン等の音声入力装置、カメラ等の画像入力装置を含む。また、入出力I/F33は、ディスプレイ等の画像出力装置、スピーカー等の音声出力装置を含む。

[0024] 通信I/F34は、ネットワークアダプタ等のハードウェア、各種の通信用ソフトウェア、及びこれらの組み合わせとして実装され、通信ネットワーク20等を介した有線又は無線の通信を実現できるように構成されている。

[0025] ストレージ35は、例えば、磁気ディスク又はフラッシュメモリ等によって構成される。ストレージ35は、オペレーティングシステムを含む様々なプログラム及び各種データ等を記憶する。ストレージ35が記憶するプログラムは、アプリケーションマーケット等からダウンロードされてインストールされ得る。また、ストレージ35は、上述した端末側プログラム42を記憶する。当該プログラム42は、ウェブブラウザ、又は、その他のアプリケーション（例えば、本実施形態のウェブサイト検査サービス用の端末側アプリケーション等）として構成され、上述したように、サーバ側プログラム40の少なくとも一部を実行するように構成され得る。

[0026] 本実施形態において、ユーザ端末30は、スマートフォン、タブレット端末、又はパーソナルコンピュータ等として構成され得る。

- [0027] このように構成されたユーザ端末30を操作するユーザ（例えば、ウェブサイト22の管理者）は、ストレージ35等にインストールされている端末側プログラム42を介したサーバ10との通信を実行することによって、当該サーバ10が提供するウェブサイト検査サービスを利用することができる。
- [0028] 次に、このように構成されたウェブサイト検査サーバ10が有する機能について説明する。サーバ10のコンピュータプロセッサ11は、図1に示すように、メインメモリ12に読み込まれたプログラム（例えば、サーバ側プログラム40の少なくとも一部）に含まれる命令を実行することによって、管理機能制御部112、自動巡回制御部114、及び、検査制御部116として機能するように構成されている。
- [0029] 管理機能制御部112は、ウェブサイト検査サービスの管理機能の制御に関する様々な処理を実行するように構成されている。例えば、管理機能制御部112は、当該管理機能に関する様々な画面の画面データ及び制御データ等をユーザ端末30に対して送信し、ユーザ端末30において出力される当該画面を介したユーザによる操作入力に応答して様々な処理を実行し、当該処理の結果に応じた画面データ及び制御データ等をユーザ端末30に対して送信する。管理機能制御部112によって制御される管理機能は、例えば、ログイン処理（ユーザ認証）、課金制御、及び、ユーザアカウントの管理等を含む。
- [0030] 自動巡回制御部114は、ウェブサイト22の自動巡回の制御に関する様々な処理を実行するように構成されている。例えば、自動巡回制御部114は、ウェブサイト22に含まれるウェブページを取得し、当該ページに含まれる遷移対象要素（例えば、リンク又はスクリプト等が設定された様々な種類のオブジェクト（ボタン、画像、及び、テキスト等）を含む。）を特定し、当該遷移対象要素を選択（クリック）して遷移先のウェブページを取得し、これらを繰り返すことによって、ウェブサイト22を巡回する。
- [0031] 本実施形態において、自動巡回制御部114は、ウェブページに入力項目

（例えば、入力可能なフォーム要素）が存在する場合に、当該入力項目に対して入力値を設定し、当該入力値の送信を伴って、遷移先のウェブページに遷移するように構成されている。つまり、自動巡回制御部 114 は、ウェブページにおける入力項目を特定し、当該入力項目に対して入力値を設定し、設定した入力値の送信を伴って、当該ウェブページから遷移した遷移先のウェブページを取得する。

[0032] 本実施形態において、入力項目に対する入力値の設定は、こうした入力値を設定するための第 1 学習済みモデルを用いて行われる。例えば、自動巡回制御部 114 は、ウェブページに関するページ情報の少なくとも一部を含む入力情報を第 1 学習済みモデルに対して入力し、当該第 1 学習済みモデルから出力される出力情報に基づいて、当該ウェブページにおける入力項目に対する入力値を設定するように構成されている。こうした学習済みモデルは、多量の訓練データを用いた機械学習を介して生成される。

[0033] 検査制御部 116 は、ウェブサイト 22 の検査の制御に関する様々な処理を実行するように構成されている。例えば、検査制御部 116 は、ウェブサイト 22 の自動巡回によって取得された各ウェブページにおいて、予め定められた検査項目を検査するための様々なリクエストを送信し、当該ウェブサイト 22 から受信したレスポンスの内容に基づいて、対応する検査項目について判断する。

[0034] このように、本実施形態におけるウェブサイト検査サーバ 10 は、ウェブサイト 22 を自動巡回する際に、ウェブページにおける入力項目を特定し、当該入力項目に対して設定する入力値が、学習済みモデルから出力される情報に基づいて設定されるから、予め準備された入力値を利用する場合と比較して、様々なウェブサイトにとって適切な入力値の利用が促進される。このように、ウェブサイト検査サーバ 10 は、ウェブサイト 22 の自動巡回における適切な入力値の設定を支援する。

[0035] 本実施形態において、ウェブサイト 22 の検査は、様々な種類の検査を含む。例えば、当該検査は、ウェブサイト 22 の脆弱性の検査（診断）を含む。

。また、例えば、当該検査は、ウェブサイト 22 の性能、機能、その他の検査を含む。

[0036] 本実施形態において、入力項目に対する入力値を設定するための第 1 学習済みモデルは、様々な情報を出力するように構成され得る。例えば、第 1 学習済みモデルは、当該入力値自体を出力するように構成される。こうした学習済みモデルは、例えば、ウェブページの HTML データと共に、当該ウェブページにおける入力項目に対する入力値の提示を指示する情報が入力されると、当該入力値自体を提示する情報を出力する大規模言語モデル（LLM）として構成される。こうした構成は、学習済みモデルから出力される入力値自体の利用を可能とする。

[0037] また、例えば、第 1 学習済みモデルは、入力項目、及び、入力値の組合せを出力するように構成される。こうした学習済みモデルは、例えば、ウェブページの HTML データと共に、当該ウェブページにおける入力項目及び入力値の組合せの提示を指示する情報が入力されると、当該組合せを提示する情報を出力する大規模言語モデル（LLM）として構成される。つまり、この場合、ウェブページにおける入力項目の特定は、第 1 学習済みモデルを用いて行われ、自動巡回制御部 114 は、第 1 学習済みモデルから出力される各入力項目に対して、同じく第 1 学習済みモデルから出力される、各入力項目に対応する入力値を設定するように構成される。こうした構成は、学習済みモデルから出力される入力項目及び入力値の組合せの利用を可能とする。

[0038] また、例えば、第 1 学習済みモデルは、入力項目、及び、入力値の組合せが入力されると、当該入力項目に対して当該入力値を設定した場合にエラーとなる可能性を示すエラー可能性情報を出力するように構成され得る。こうした学習済みモデルは、例えば、多量の訓練データ（例えば、ウェブページのページ情報（HTML データ等）、当該ウェブページにおける入力項目、当該入力項目に対して設定された入力値、及び、当該入力値の送信後のエラーの有無等）を用いた機械学習を介して生成され、エラー可能性情報として、エラーとなる可能性を示すスコア等を出力するように構成される。この場

合、自動巡回制御部 114 は、特定の入力項目に対して設定可能な、予め準備された入力値の複数の候補の各々について、第 1 学習済みモデルから出力されるエラー可能性情報に基づいて、エラーとなる可能性を判断し、エラーとなる可能性が低い候補が優先されるように、入力値を設定するように構成される。こうした構成は、学習済みモデルから出力されるエラー可能性情報に基づく入力値の設定を可能とする。

[0039] また、自動巡回制御部 114 は、入力値の送信を伴う遷移先のウェブページに関するページ情報に基づいて、エラーを判定し、エラーが有ると判定された場合に、遷移元のウェブページにおける入力項目に対して新たな入力値を設定するように構成され得る。こうした構成は、ページ情報に基づいてエラーが有ると判定された場合の入力値の再設定を可能とする。

[0040] 本実施形態において、遷移先のウェブページに関するページ情報に基づくエラーの判定もまた、機械学習を介して生成された学習済みモデルを用いて行われるようにしてもよい。例えば、自動巡回制御部 114 は、遷移先のウェブページに関するページ情報の少なくとも一部を含む入力情報を第 2 学習済みモデルに対して入力し、当該第 2 学習済みモデルから出力される出力情報に基づいて、エラーを判定するように構成され得る。例えば、第 2 学習済みモデルは、エラーの有無に関する情報（エラーである可能性を示すスコア等）を出力し、当該情報に基づいて、エラーの有無が判定される（例えば、エラーである可能性を示すスコアが閾値以上である場合にエラーであると判定される。）。こうした構成は、学習済みモデルを介したエラーの判定を可能とする。

[0041] 本実施形態において、ウェブページにおける複数の入力項目が特定され、これらの複数の入力項目の各々に対して対応する入力値が設定されるようにしてもよい。この場合、エラーを判定するための第 2 学習済みモデルは、これらの複数の入力項目のうち、エラーの原因である入力項目を特定するための情報（例えば、複数の入力項目の各々について、エラーの原因である可能性を示す情報、又は、当該入力項目自体を特定する情報等）を出力するよう

に構成され得る。こうした構成は、エラーの原因である入力項目を、学習済みモデルを介して特定することを可能とする。

[0042] また、自動巡回制御部 114 は、エラーが有ると判定された場合に、遷移元のウェブページにおける入力項目、及び、当該入力項目に対して設定されて送信された入力値の組合せを含むデータを、第2学習済みモデルの訓練データとして追加するように構成され得る。こうした構成は、エラーを判定するための学習済みモデルの適切な追加学習を支援する。

[0043] また、自動巡回制御部 114 は、エラーが有ると判定された場合の当該エラーに関する情報をユーザに提示するように構成され得る。例えば、自動巡回制御部 114 は、エラーに関する情報（例えば、エラー管理テーブル 156 によって管理される。）を含む画面の画面データ及び制御データ等をユーザ端末 30 に対して送信することによって、当該情報をユーザに対して提示する。こうした構成は、エラーに関する情報をユーザ（例えば、ウェブサイト 22 の管理者等）に知らせることを可能とする。

[0044] 次に、このような機能を有する本実施形態のウェブサイト検査サーバ 10 の一態様としての具体例について説明する。この具体例におけるサーバ 10 は、ウェブサイト検査サービスの一例として、ウェブサイトの脆弱性を検査する脆弱性検査サービスを提供する。

[0045] まず、この例において、各テーブルが管理する情報について説明する。図 2 は、この例において、検査管理テーブル 152 が管理する情報を例示する。この例における検査管理テーブル 152 は、個別の検査に関する情報を管理し、図示するように、個別の検査（案件）を識別する「検査ID」に対応付けて、検査対象のウェブサイトの検査の起点となるトップ画面（トップページ）のURLである「トップ画面URL」、画面遷移の設定日時である「画面遷移設定日時」、検査の実行日時である「検査実行日時」等の情報を管理する。

[0046] 図 3 は、この例において、ページ管理テーブル 154 が管理する情報を例示する。この例におけるページ管理テーブル 154 は、検査対象のウェブサ

イトに含まれる各ウェブページ（画面）に関する情報を管理し、図示するように、個別の検査を識別する「検査ＩＤ」、及び、当該検査において個別の画面を識別する「画面ＩＤ」の組合せに対応付けて、当該画面のＵＲＬである「画面ＵＲＬ」、当該画面の画面キャプチャに関する情報である「画面キャプチャ情報」、当該画面の「ＨＴＭＬデータ」、当該画面の遷移元の画面を識別する「遷移元画面ＩＤ」、当該画面の遷移元の要素（当該画面に遷移する際にクリックされた遷移対象要素）を識別する「遷移元要素ＩＤ」、遷移元の画面において、入力項目に対して入力された入力値に関する情報である「遷移元入力情報」、当該画面における検査結果に関する情報である「検査結果情報」等の情報を管理する。

[0047] 図４は、この例において、エラー管理テーブル１５６が管理する情報を例示する。この例におけるエラー管理テーブル１５６は、検査対象のウェブサイトを実行する際に発生したエラーに関する情報を管理し、図示するように、個別の検査を識別する「検査ＩＤ」、及び、当該検査において個別のエラーを識別する「エラーＩＤ」の組合せに対応付けて、エラーが発生した画面のＵＲＬである「画面ＵＲＬ」、当該画面の画面キャプチャに関する情報である「画面キャプチャ情報」、当該画面の「ＨＴＭＬデータ」、当該画面の遷移元の画面を識別する「遷移元画面ＩＤ」、当該画面の遷移元の要素（当該画面に遷移する際にクリックされた遷移対象要素）を識別する「遷移元要素ＩＤ」、遷移元の画面において、入力項目に対して入力された入力値に関する情報である「遷移元入力情報」、当該画面においてエラーが発生した可能性を示す「エラースコア」等の情報を管理する。

[0048] 以上、この例において、各テーブルが管理する情報について説明した。次に、この例において、ウェブサイト検査サーバ１０が実行する処理、及び、ユーザ端末３０において出力される画面等について説明する。

[0049] 図５は、この例において、ウェブサイトの脆弱性を検査する際にウェブサイト検査サーバ１０が実行する処理を例示するフローチャートである。サーバ１０は、まず、図示するように、脆弱性を検査する対象のウェブサイトの

設定を受け付ける（ステップS 1 0 0）。こうした設定の受付は、ユーザ端末3 0において出力される画面を介して行われる。

[0050] 図6は、ユーザ端末3 0において出力される、検査対象のウェブサイトの設定を受け付けるための対象サイト設定画面5 0を例示する。当該画面5 0は、図示するように、検査対象のウェブサイトのトップ画面（トップページ）のURLを入力するためのURL入力領域5 2と、当該ウェブサイトへのログインに用いられるログインIDを入力するためのログインID入力領域5 4と、当該ウェブサイトへのログインに用いられるパスワードを入力するためのパスワード入力領域5 6と、「巡回・検査開始」というテキストが付加された続行ボタン5 8とを有する。このように、この例では、検査対象のウェブサイトの設定情報として、トップ画面のURL、ログインID、及び、パスワードが受け付けられる。

[0051] 続行ボタン5 8は、検査対象のウェブサイトの自動巡回及び検査を指示するためのオブジェクトである。図5のフローチャートに戻り、当該ボタン5 8が選択されると、サーバ1 0は、次に、検査対象のウェブサイトを自動巡回し、検査の際の画面遷移を設定する（ステップS 1 1 0）。

[0052] 図7は、ウェブサイトの自動巡回の際にサーバ1 0が実行する処理を例示するフローチャートである。サーバ1 0は、まず、図示するように、検査対象のウェブサイトのトップ画面にアクセスし、当該トップ画面を画面遷移に追加する（ステップS 2 0 0）。具体的には、サーバ1 0において、ウェブブラウザが起動され、対象サイト設定画面5 0を介して入力されたURLによって特定されるトップ画面へのアクセスが行われる。また、ページ管理テーブル1 5 4において、当該トップ画面に関する情報が登録され、具体的には、検査ID、画面ID、画面URL、画面キャプチャ情報、及び、HTMLデータが設定される。このように、ページ管理テーブル1 5 4は、個別の検査において設定される画面遷移を管理するテーブルであるとも言える。

[0053] 続いて、サーバ1 0は、トップ画面に含まれる複数の遷移対象要素を特定する（ステップS 2 1 0）。具体的には、リンク又はスクリプト等が設定さ

れた画面構成要素（ボタン、画像、及び、テキスト等）が、遷移対象要素として特定される。こうした要素の特定は、対応するHTMLデータに基づいて行われ、例えば、`button`タグ、`img`タグ、及び、`input`タグ（`type = "button"`）等に対応する要素が、遷移対象要素の候補となる。

[0054] 次に、サーバ10は、入力項目（例えば、入力可能なフォーム要素）が存在する場合には（ステップS220においてYES）、入力項目に対する入力値を生成し、生成した入力値を設定する（ステップS230、S240）。入力項目の有無は、対応するHTMLデータに基づいて判定され、例えば、`input`タグ（`type = "text"`）、及び、`textarea`タグ等が、入力項目の候補となる。

[0055] また、入力項目に対する入力値は、具体的には、こうした入力値を生成可能な学習済みモデルを用いて生成される。当該学習済みモデルは、この例では、大規模言語モデルとして構成され、対象画面のHTMLデータと共に、当該画面における入力項目及び入力値の組合せの提示を指示する情報が入力されると、当該画面に含まれる各入力項目が、対応する入力値を伴って出力する（つまり、入力項目及び入力値の組合せが出力される。）ように構成されている。こうした大規模言語モデルとして、例えば、米国OpenAI社が提供するGPT-4が適用され得る。

[0056] 図8は、大規模言語モデルにおける入出力情報の一例を示す。図示するように、入力情報において、JSON形式での出力を指示することによって、出力情報に基づく入力値の設定が可能となる。

[0057] 図9は、大規模言語モデルにおける入出力情報の他の例を示す。図示するように、HTMLデータにおけるパラメータ名（「`param1`」及び「`param2`」）がパラメータの用途を示していない場合であっても、大規模言語モデルは、ラベル等の周辺の要素に基づいて、パラメータの用途を判定することができる。また、入力規則が自然言語（「半角カナで指定し、苗字と氏名の間にスペースを空けないでください」との記載）で指定されている

場合であっても、大規模言語モデルは、こうした入力規則に従った入力値を提示することができる。

[0058] 図10は、大規模言語モデルにおける入出力情報のさらに他の例を示す。図示するように、HTMLデータにおけるパラメータ名に誤記がある（「`l s a t _ n a m e`」は「`l a s t _ n a m e`」の誤記である）場合であっても、大規模言語モデルは、こうした誤記を予測して、パラメータの用途を判定することができる。

[0059] 続いて、サーバ10は、対応する遷移対象要素を選択（クリック）する（ステップS250）。なお、入力項目が存在しない場合には（ステップS220においてNO）、直ちに、遷移対象要素の選択が行われる。

[0060] 次に、サーバ10は、遷移先の画面におけるエラーを判定する（ステップS260）。こうしたエラーは、具体的には、当該エラーを判定するための学習済みモデルを用いて判定される。当該学習済みモデルは、この例では、多量の訓練データを用いた機械学習を介して生成され、遷移元の画面のHTMLデータ、HTTPリクエスト（入力項目に対して設定された入力値を含む。）、及び、HTTPレスポンス（遷移先の画面のHTMLデータを含む。）が入力されると、エラーが発生している可能性を示すエラースコアを出力するように構成されている。そして、当該エラースコアの値が閾値以上である場合、エラーが発生していると判定され、当該エラースコアの値が閾値未満である場合、エラーが発生していないと判定される。

[0061] 例えば、エラーのメッセージが自然言語（予め列挙できない文言等）のみで表示される場合（例えば、「処理を完了することができませんでした」とのメッセージ等）には、ルールベースの手法でのエラーの判定は困難であるが、学習済みモデルを用いることによって、こうしたエラーの判定が可能となる。

[0062] そして、遷移先の画面においてエラーが発生していると判定されない場合には（ステップS265においてNO）、サーバ10は、遷移先画面への遷移を画面遷移に追加する（ステップS270）。具体的には、ページ管理テ

ーブル154において、遷移先の画面に関する情報が登録され、つまり、検査ID、画面ID、画面URL、画面キャプチャ情報、HTMLデータ、遷移元画面ID、遷移元要素ID、及び、遷移元入力情報が設定される。遷移元入力情報には、入力項目及び入力値の組合せが設定される。

[0063] ここで、エラーが発生していない場合であっても、遷移先の画面が、画面遷移への追加対象外の画面である場合には、画面遷移への追加が行われなようにしてもよい。例えば、検査対象のウェブサイトの外部のサイトの画面、及び、既に画面遷移に含まれる画面と重複する画面（こうした重複は、例えば、画面の類似度に基づいて判断される。）等については、画面遷移への追加対象外と判断される。

[0064] 一方、遷移先の画面においてエラーが発生していると判定される場合には（ステップS265においてYES）、サーバ10は、遷移先画面におけるエラーを登録する（ステップS280）。具体的には、エラー管理テーブル156において、検査ID、エラーID、画面URL、画面キャプチャ情報、HTMLデータ、遷移元画面ID、遷移元要素ID、遷移元入力情報、及び、エラースコアが設定される。なお、エラーが発生している場合、当該画面の画面遷移への追加は行われない。

[0065] そして、トップ画面に含まれる複数の遷移対象要素の全てについて、選択（クリック）に応じた遷移先画面の画面遷移への追加、又は、エラーの登録が行われると（ステップS290においてNO）、次に、サーバ10は、画面遷移に追加された各画面について、トップ画面と同様に、当該画面に含まれる複数の遷移対象要素の特定（ステップS310）、及び、当該要素の選択に応じた遷移先画面の画面遷移への追加、又は、エラーの登録を行う（ステップS220～S290）。こうした処理は、画面遷移への新たな画面の追加がなくなるまでの間（ステップS300においてNO）、繰り返し行われる。

[0066] 図5のフローチャートに戻り、こうして画面遷移が設定されると、続いて、サーバ10は、設定された画面遷移に従って、脆弱性の検査を実行する（

ステップS 1 2 0)。具体的には、ウェブブラウザを起動し、ページ管理テーブル1 5 4において管理されている情報に従って、画面遷移に含まれる各画面を順に訪れて、予め定められた脆弱性の検査項目の検査を行う。脆弱性の検査項目は、例えば、OWASP T o p 1 0において規定されている項目を含む。検査結果は、ページ管理テーブル1 5 4の検査結果情報に登録される。なお、ウェブサイトの自動巡回の際に（つまり、画面遷移の設定の際に）、各画面において、検査項目の検査を行うようにしてもよい。

[0067] 図1 1は、ユーザ端末3 0において出力される、検査結果を確認するための検査結果確認画面6 0を例示する。当該画面6 0は、図示するように、脆弱性の検査結果を一覧表示する一覧表示領域6 2を有する。当該領域6 2は、各画面の脆弱性の検査結果（深刻度、脆弱性名、検出箇所、及び、脆弱箇所等）を一覧表示する。これらの情報は、ページ管理テーブル1 5 4の検査結果情報において管理されている。

[0068] 上述した例において、エラー管理テーブル1 5 6において管理されている情報は、遷移先の画面におけるエラーを判定するための上述した学習済みモデルの追加学習の際の訓練データに自動的に追加される。

[0069] 上述した例において、エラー管理テーブル1 5 6において管理されている情報を含む画面が、対応するユーザ（ウェブサイトの管理者等）のユーザ端末3 0において、出力されるようにしてもよい。

[0070] 上述した例では、大規模言語モデルとして構成される学習済みモデルから、入力項目及び入力値の組合せが出力されるようにしたが、入力値のみが出力されるようにしてもよい。この場合、対象画面のHTMLデータ等に基づいて入力項目が特定され、学習済みモデルに対して、当該HTMLデータと共に、特定された入力項目に対する入力値の提示を指示する情報が入力されると、当該学習済みモデルから、当該入力値自体を提示する情報が出力される。図1 2は、この場合の大規模言語モデルにおける入出力情報の一例を示す。この例において、「p a r a m 1」は、対象画面のHTMLデータに基づいて特定された入力項目（氏名）のパラメータであり、こうした入力情報

が、サーバ 10 によって自動的に生成される。

[0071] また、学習済みモデルが、HTML データ、並びに、入力項目及び入力値の組合せが入力されると、エラーとなる可能性を示す情報（エラースコア等）を出力するように構成されてもよい。この場合、対象画面の HTML データ等に基づいて入力項目が特定され、当該入力項目に対して設定可能な複数の入力値の候補（予め定められている。）の各々について、学習済みモデルを介してエラーとなる可能性を取得し、当該可能性が最も低い候補を入力値として適用するようにすればよい。

[0072] 上述した例では、遷移先の画面におけるエラーを判定するための学習済みモデルは、遷移元の画面の HTML データ、HTTP リクエスト、及び、HTTP レスポンスが入力されると、エラーが発生している可能性を示すエラースコアを出力するように構成されているが、複数の入力項目のうち、エラーの原因である入力項目を特定するための情報が出力されるようにしてもよい。例えば、当該学習済みモデルは、大規模言語モデルとして構成され、遷移元の画面の HTML データ、HTTP リクエスト、及び、HTTP レスポンスと共に、エラーの原因である入力項目の提示を指示する情報が入力されると、当該学習済みモデルから、当該入力項目を提示する情報が出力される。

[0073] さらに、エラーを判定するための学習済みモデルから、エラーの理由、及び、入力値の代替値等の拡張情報を取得するようにしてもよい。こうした学習済みモデルは、大規模言語モデルとして構成され、遷移元の画面の HTML データ、及び、入力項目に対する入力値と共に、こうした情報の提示を指示する情報が入力されると、当該学習済みモデルから、当該拡張情報を提示する情報が出力される。図 13 は、この場合の大規模言語モデルにおける入出力情報の一例を示す。こうして取得した拡張情報は、例えば、エラー管理テーブル 156 において管理される。

[0074] 上述した例では、ウェブサイトの脆弱性を検査する脆弱性検査サービスが提供されるようにしたが、本実施形態の他の例では、脆弱性以外の他の観点

での様々な検査が行われ、例えば、ウェブサイトのその他の性能、及び、機能等の検査が行われる。

[0075] 以上説明した本実施形態に係るウェブサイト検査サーバ10は、ウェブサイト22を自動巡回する際に、ウェブページにおける入力項目を特定し、当該入力項目に対して設定する入力値が、学習済みモデルから出力される情報（例えば、入力項目及び入力値の組合せ）に基づいて設定されるから、予め準備された入力値を利用する場合と比較して、様々なウェブサイトにとって適切な入力値の利用が促進される。このように、ウェブサイト検査サーバ10は、ウェブサイト22の自動巡回における適切な入力値の設定を支援する。

[0076] 本発明の他の実施形態において、上述した実施形態におけるウェブサイト検査サーバ10が有する機能の一部又は全部は、ウェブサイト検査サーバ10及びユーザ端末30が協働することによって実現され、又は、ユーザ端末30によって実現され得る。また、本発明の他の実施形態においては、ウェブサイト22の検査以外の様々な目的で、当該ウェブサイト22の自動巡回が行われる。

[0077] 本明細書で説明された処理及び手順は、明示的に説明されたもの以外にも、ソフトウェア、ハードウェアまたはこれらの任意の組み合わせによって実現される。例えば、本明細書で説明される処理及び手順は、集積回路、揮発性メモリ、不揮発性メモリ、磁気ディスク等の媒体に、当該処理及び手順に相当するロジックを実装することによって実現される。また、本明細書で説明された処理及び手順は、当該処理・手順に相当するコンピュータプログラムとして実装し、各種のコンピュータに実行させることが可能である。

[0078] 本明細書中で説明された処理及び手順が単一の装置、ソフトウェア、コンポーネント、モジュールによって実行される旨が説明されたとしても、そのような処理または手順は複数の装置、複数のソフトウェア、複数のコンポーネント、及び／又は複数のモジュールによって実行され得る。また、本明細書において説明されたソフトウェアおよびハードウェアの要素は、それらを

より少ない構成要素に統合して、またはより多い構成要素に分解することによって実現することも可能である。

- [0079] 本明細書において、発明の構成要素が単数もしくは複数のいずれか一方として説明された場合、又は、単数もしくは複数のいずれとも限定せずに説明された場合であっても、文脈上別に解すべき場合を除き、当該構成要素は単数又は複数のいずれであってもよい。

請求の範囲

- [請求項1] 1又は複数のコンピュータプロセッサを備え、ウェブサイトの自動巡回を行うためのシステムであって、前記1又は複数のコンピュータプロセッサは、
- 前記ウェブサイトに含まれる第1ウェブページを取得するステップと、
- 前記第1ウェブページにおける入力項目を特定し、前記入力項目に対して入力値を設定するステップと、
- 設定した前記入力値の送信を伴って、前記第1ウェブページから遷移した第2ウェブページを取得するステップと、を実行し、
- 前記入力値を設定するステップは、前記第1ウェブページに関する第1ページ情報の少なくとも一部を含む第1入力情報を第1学習済みモデルに対して入力し、前記第1学習済みモデルから出力される第1出力情報に基づいて、前記入力値を設定することを含む、
- システム。
- [請求項2] 前記第1出力情報は、前記入力値を含む、
- 請求項1のシステム。
- [請求項3] 前記第1出力情報は、前記入力項目、及び、前記入力値の組合せを含み、
- 前記入力値を設定するステップは、前記第1出力情報に含まれる前記入力項目に対して、前記第1出力情報に含まれる前記入力値を設定することを含む、
- 請求項1のシステム。
- [請求項4] 前記第1入力情報は、前記入力項目、及び、前記入力値の組合せを含み、
- 前記第1出力情報は、前記入力項目に対して前記入力値を設定した場合にエラーとなる可能性を示すエラー可能性情報を含み、
- 前記入力値を設定するステップは、前記エラー可能性情報に基づい

て、前記入力値を設定するか否かを判断することを含む、
請求項 1 のシステム。

[請求項5] 前記 1 又は複数のコンピュータプロセッサは、さらに、
前記第 2 ウェブページに関する第 2 ページ情報に基づいて、エラー
を判定するステップと、
前記エラーが有ると判定された場合に、前記入力項目に対して新たな
入力値を設定するステップと、を実行する、
請求項 1 のシステム。

[請求項6] 前記エラーを判定するステップは、前記第 2 ページ情報の少なくとも
一部を含む第 2 入力情報を第 2 学習済みモデルに対して入力し、前
記第 2 学習済みモデルから出力される第 2 出力情報に基づいて前記エ
ラーを判定することを含む、
請求項 5 のシステム。

[請求項7] 前記第 2 出力情報は、前記エラーの有無に関する情報を含む、
請求項 6 のシステム。

[請求項8] 前記入力値を設定するステップは、前記第 1 ウェブページにおける
複数の入力項目を特定し、前記複数の入力項目の各々に対して対応す
る入力値を設定することを含み、
前記第 2 出力情報は、前記複数の入力項目のうち、前記エラーの原
因である入力項目に関する情報を含む、
請求項 6 のシステム。

[請求項9] 前記 1 又は複数のコンピュータプロセッサは、さらに、前記エラー
が有ると判定された場合に、前記入力項目、及び、前記入力項目に対
して設定されて送信された前記入力値の組合せを含むデータを、前記
第 2 学習済みモデルの訓練データとして追加するステップを実行する
、
請求項 6 のシステム。

[請求項10] 前記 1 又は複数のコンピュータプロセッサは、さらに、前記エラー

が有ると判定された場合の前記エラーに関する情報をユーザに提示するステップを実行する、

請求項5のシステム。

[請求項11] 1又は複数のコンピュータによって実行され、ウェブサイトの自動巡回を行うための方法であって、

前記ウェブサイトに含まれる第1ウェブページを取得するステップと、

前記第1ウェブページにおける入力項目を特定し、前記入力項目に対して入力値を設定するステップと、

設定した前記入力値の送信を伴って、前記第1ウェブページから遷移した第2ウェブページを取得するステップと、を備え、

前記入力値を設定するステップは、前記第1ウェブページに関する第1ページ情報の少なくとも一部を含む第1入力情報を第1学習済みモデルに対して入力し、前記第1学習済みモデルから出力される第1出力情報に基づいて、前記入力値を設定することを含む、

方法。

[請求項12] ウェブサイトの自動巡回を行うためのプログラムであって、1又は複数のコンピュータに、

前記ウェブサイトに含まれる第1ウェブページを取得するステップと、

前記第1ウェブページにおける入力項目を特定し、前記入力項目に対して入力値を設定するステップと、

設定した前記入力値の送信を伴って、前記第1ウェブページから遷移した第2ウェブページを取得するステップと、を実行させ、

前記入力値を設定するステップは、前記第1ウェブページに関する第1ページ情報の少なくとも一部を含む第1入力情報を第1学習済みモデルに対して入力し、前記第1学習済みモデルから出力される第1出力情報に基づいて、前記入力値を設定することを含む、

プログラム。