

(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2017-16446  
(P2017-16446A)

(43) 公開日 平成29年1月19日(2017.1.19)

(51) Int.Cl. F I テーマコード(参考)  
**G06F 17/30 (2006.01)** G06F 17/30 110F  
 G06F 17/30 414Z

審査請求 未請求 請求項の数 5 O L (全 16 頁)

(21) 出願番号	特願2015-133179 (P2015-133179)	(71) 出願人	000005496 富士ゼロックス株式会社 東京都港区赤坂九丁目7番3号
(22) 出願日	平成27年7月2日(2015.7.2)	(74) 代理人	100115129 弁理士 清水 昇
		(74) 代理人	100102716 弁理士 在原 元司
		(74) 代理人	100122275 弁理士 竹居 信利
		(72) 発明者	吉村 俊秀 神奈川県横浜市西区みなとみらい六丁目1番 富士ゼロックス株式会社内

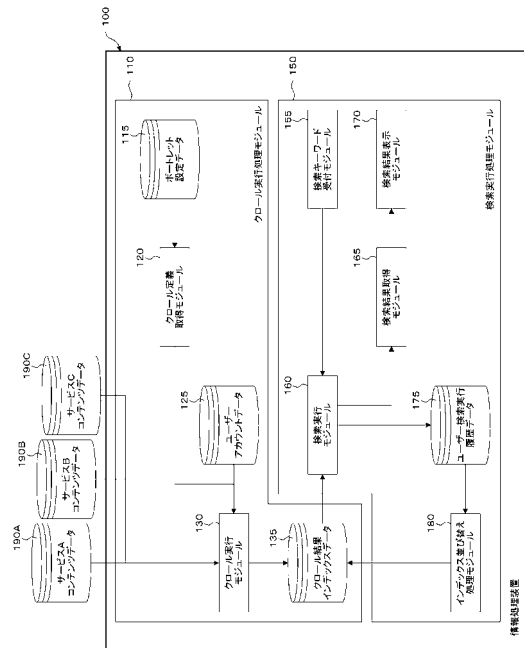
(54) 【発明の名称】 情報処理装置及び情報処理プログラム

(57) 【要約】

【課題】複数のWebサイトへの入り口を提供するポータルサイトにおいて、前記複数のWebサイトを対象とした検索を、ユーザーの前記複数のWebページそれぞれに対するアクセス権に応じて行う情報処理装置を提供する。

【解決手段】情報処理装置の記憶手段は、複数のWebサイトのそれぞれに対して、ユーザーのアクセス権に基づいてアクセスし、該複数のWebサイトに記載されている内容のインデックスを該ユーザーと対応づけて記憶し、制御手段は、前記複数のWebサイトへの入り口を表現したポータルサイトを表示するよう制御し、受付手段は、前記ポータルサイトにおいて、前記複数のWebサイトに対する検索条件の入力を前記ユーザーから受け付け、検索手段は、前記記憶手段に記憶された、前記ユーザーに対応づけられたインデックスに対して、前記入力された検索条件に基づいた検索を行う。

【選択図】 図1



**【特許請求の範囲】****【請求項 1】**

複数の Web サイトと、該複数の Web サイトに対するユーザーのアクセス権とを対応づけて記憶する第一の記憶手段と、

前記複数の Web サイトのそれぞれに対して、前記ユーザーのアクセス権に基づいてアクセスし、該複数の Web サイトに記載されている内容のインデックスを生成するインデキシング手段と、

前記インデキシング手段により生成されたインデックスを前記ユーザーと対応づけて記憶する第二の記憶手段と、

前記複数の Web サイトへの入り口を表現したポータルサイトを表示するよう制御する制御手段と、

前記ポータルサイトにおいて、前記複数の Web サイトに対する検索条件の入力を前記ユーザーから受け付ける受付手段と、

前記第二の記憶手段に記憶された、前記ユーザーに対応づけられたインデックスに対して、前記入力された検索条件に基づいた検索を行う検索手段と、

を有する情報処理装置。

10

**【請求項 2】**

ユーザーが入力した検索条件を履歴として記憶する第三の記憶手段と、

前記第二の記憶手段に記憶されている情報に対して、前記履歴に基づいて前記ユーザーのインデックスを並び替える並替手段

をさらに有する請求項 1 に記載の情報処理装置。

20

**【請求項 3】**

複数の Web サイトのそれぞれに対して、ユーザーのアクセス権に基づいてアクセスし、該複数の Web サイトに記載されている内容のインデックスを該ユーザーと対応づけて記憶する記憶手段と、

前記複数の Web サイトへの入り口を表現したポータルサイトを表示するよう制御する制御手段と、

前記ポータルサイトにおいて、前記複数の Web サイトに対する検索条件の入力を前記ユーザーから受け付ける受付手段と、

前記記憶手段に記憶された、前記ユーザーに対応づけられたインデックスに対して、前記入力された検索条件に基づいた検索を行う検索手段と、

を有する情報処理装置。

30

**【請求項 4】**

コンピュータを、

複数の Web サイトと、該複数の Web サイトに対するユーザーのアクセス権とを対応づけて記憶する第一の記憶手段と、

前記複数の Web サイトのそれぞれに対して、前記ユーザーのアクセス権に基づいてアクセスし、該複数の Web サイトに記載されている内容のインデックスを生成するインデキシング手段と、

前記インデキシング手段により生成されたインデックスを前記ユーザーと対応づけて記憶する第二の記憶手段と、

前記複数の Web サイトへの入り口を表現したポータルサイトを表示するよう制御する制御手段と、

前記ポータルサイトにおいて、前記複数の Web サイトに対する検索条件の入力を前記ユーザーから受け付ける受付手段と、

前記第二の記憶手段に記憶された、前記ユーザーに対応づけられたインデックスに対して、前記入力された検索条件に基づいた検索を行う検索手段

として機能させるための情報処理プログラム。

40

**【請求項 5】**

コンピュータを、

50

複数のWebサイトのそれぞれに対して、ユーザーのアクセス権に基づいてアクセスし、該複数のWebサイトに記載されている内容のインデックスを該ユーザーと対応づけて記憶する記憶手段と、

前記複数のWebサイトへの入り口を表現したポータルサイトを表示するよう制御する制御手段と、

前記ポータルサイトにおいて、前記複数のWebサイトに対する検索条件の入力を前記ユーザーから受け付ける受付手段と、

前記記憶手段に記憶された、前記ユーザーに対応づけられたインデックスに対して、前記入力された検索条件に基づいた検索を行う検索手段

として機能させるための情報処理プログラム。

10

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、情報処理装置及び情報処理プログラムに関する。

【背景技術】

【0002】

特許文献1には、ポータルサイトを起点とした情報検索において、利用者が目的を達成するために必要な情報を素早く簡単に入手できるように支援することを課題とし、ブラウザを実装するとともに、キーワード照会書式情報と、検索結果文書書式情報とを記憶したコンピュータにインストールされるプログラムであって、コンピュータに、検索キーワードを受け付けるステップと、検索キーワードをキーワード照会書式情報に基づく手順にしたがって所定のポータルサイトが提供する複数のサービスのそれぞれの検索エンジンに個別に与えるステップと、検索エンジンから返送されてくる検索結果文書を取得するステップと、検索結果文書に記載されているレコード件数を検索結果文書書式情報に基づいて取得するステップと、取得したレコード件数と検索結果文書を個別に表示させるための機能とを含んだ画面を表示するステップとを実行させることが開示されている。

20

【0003】

特許文献2には、クライアントが、あるポータルサイトへの接続を試みた場合に、要求したポータルサイトにアクセス権があるデータのみをデータセンタから蓄積場所を取得することで、クライアントに対して表示させることが可能なWebコンテンツ配信方法を提供することを課題とし、クライアントが事業者ポータルサイトに接続すると、事業者毎に付与された事業者IDに基づいて、データセンタに対して、クライアントから要求のあったページ記述ファイルを構成する個々のデータについて、蓄積場所の問い合わせを行い、アクセス権がある場合に、事業者ポータルサイトを経由して、クライアントにデータの蓄積場所を通知し、事業者ポータルサイトにおいて、クライアントから要求のあったデータについて、コンピュータネットワークを経由して、データセンタからアクセス権のあるデータを取得して、クライアントに配信することが開示されている。

30

【先行技術文献】

【特許文献】

【0004】

40

【特許文献1】特開2008-015611号公報

【特許文献2】特開2002-007347号公報

【発明の概要】

【発明が解決しようとする課題】

【0005】

本発明は、複数のWebサイトへの入り口を提供するポータルサイトにおいて、前記複数のWebサイトを対象とした検索を、ユーザーの前記複数のWebページそれぞれに対するアクセス権に応じて行う情報処理装置及び情報処理プログラムを提供することを目的としている。

【課題を解決するための手段】

50

## 【 0 0 0 6 】

かかる目的を達成するための本発明の要旨とするところは、次の各項の発明に存する。

## 【 0 0 0 7 】

請求項 1 の発明は、複数の Web サイトと、該複数の Web サイトに対するユーザーのアクセス権とを対応づけて記憶する第一の記憶手段と、前記複数の Web サイトのそれぞれに対して、前記ユーザーのアクセス権に基づいてアクセスし、該複数の Web サイトに記載されている内容のインデックスを生成するインデキシング手段と、前記インデキシング手段により生成されたインデックスを前記ユーザーと対応づけて記憶する第二の記憶手段と、前記複数の Web サイトへの入り口を表現したポータルサイトを表示するよう制御する制御手段と、前記ポータルサイトにおいて、前記複数の Web サイトに対する検索条件の入力を前記ユーザーから受け付ける受付手段と、前記第二の記憶手段に記憶された、前記ユーザーに対応づけられたインデックスに対して、前記入力された検索条件に基づいた検索を行う検索手段と、を有する情報処理装置である。

10

## 【 0 0 0 8 】

請求項 2 の発明は、ユーザーが入力した検索条件を履歴として記憶する第三の記憶手段と、前記第二の記憶手段に記憶されている情報に対して、前記履歴に基づいて前記ユーザーのインデックスを並び替える並替手段をさらに有する請求項 1 に記載の情報処理装置である。

## 【 0 0 0 9 】

請求項 3 の発明は、複数の Web サイトのそれぞれに対して、ユーザーのアクセス権に基づいてアクセスし、該複数の Web サイトに記載されている内容のインデックスを該ユーザーと対応づけて記憶する記憶手段と、前記複数の Web サイトへの入り口を表現したポータルサイトを表示するよう制御する制御手段と、前記ポータルサイトにおいて、前記複数の Web サイトに対する検索条件の入力を前記ユーザーから受け付ける受付手段と、前記記憶手段に記憶された、前記ユーザーに対応づけられたインデックスに対して、前記入力された検索条件に基づいた検索を行う検索手段と、を有する情報処理装置である。

20

## 【 0 0 1 0 】

請求項 4 の発明は、コンピュータを、複数の Web サイトと、該複数の Web サイトに対するユーザーのアクセス権とを対応づけて記憶する第一の記憶手段と、前記複数の Web サイトのそれぞれに対して、前記ユーザーのアクセス権に基づいてアクセスし、該複数の Web サイトに記載されている内容のインデックスを生成するインデキシング手段と、前記インデキシング手段により生成されたインデックスを前記ユーザーと対応づけて記憶する第二の記憶手段と、前記複数の Web サイトへの入り口を表現したポータルサイトを表示するよう制御する制御手段と、前記ポータルサイトにおいて、前記複数の Web サイトに対する検索条件の入力を前記ユーザーから受け付ける受付手段と、前記第二の記憶手段に記憶された、前記ユーザーに対応づけられたインデックスに対して、前記入力された検索条件に基づいた検索を行う検索手段として機能させるための情報処理プログラムである。

30

## 【 0 0 1 1 】

請求項 5 の発明は、コンピュータを、複数の Web サイトのそれぞれに対して、ユーザーのアクセス権に基づいてアクセスし、該複数の Web サイトに記載されている内容のインデックスを該ユーザーと対応づけて記憶する記憶手段と、前記複数の Web サイトへの入り口を表現したポータルサイトを表示するよう制御する制御手段と、前記ポータルサイトにおいて、前記複数の Web サイトに対する検索条件の入力を前記ユーザーから受け付ける受付手段と、前記記憶手段に記憶された、前記ユーザーに対応づけられたインデックスに対して、前記入力された検索条件に基づいた検索を行う検索手段として機能させるための情報処理プログラムである。

40

## 【 発明の効果 】

## 【 0 0 1 2 】

請求項 1 の情報処理装置によれば、複数の Web サイトへの入り口を提供するポータル

50

サイトにおいて、前記複数のWebサイトを対象とした検索を、ユーザーの前記複数のWebページそれぞれに対するアクセス権に応じて行うことができる。

【0013】

請求項2の情報処理装置によれば、検索条件を履歴に基づいてユーザーのインデックスを並び替えることができる。

【0014】

請求項3の情報処理装置によれば、複数のWebサイトへの入り口を提供するポータルサイトにおいて、前記複数のWebサイトを対象とした検索を、ユーザーの前記複数のWebページそれぞれに対するアクセス権に応じて行うことができる。

【0015】

請求項4の情報処理プログラムによれば、複数のWebサイトへの入り口を提供するポータルサイトにおいて、前記複数のWebサイトを対象とした検索を、ユーザーの前記複数のWebページそれぞれに対するアクセス権に応じて行うことができる。

【0016】

請求項5の情報処理プログラムによれば、複数のWebサイトへの入り口を提供するポータルサイトにおいて、前記複数のWebサイトを対象とした検索を、ユーザーの前記複数のWebページそれぞれに対するアクセス権に応じて行うことができる。

【図面の簡単な説明】

【0017】

【図1】本実施の形態の構成例についての概念的なモジュール構成図である。

【図2】本実施の形態を利用したシステム構成例を示す説明図である。

【図3】本実施の形態による処理例を示すフローチャートである。

【図4】本実施の形態による処理例を示す説明図である。

【図5】クロール定義テーブルのデータ構造例を示す説明図である。

【図6】ユーザーアカウントテーブルのデータ構造例を示す説明図である。

【図7】インデックステーブルのデータ構造例を示す説明図である。

【図8】本実施の形態による処理例を示すフローチャートである。

【図9】インデックステーブルのデータ構造例を示す説明図である。

【図10】キーワード入力履歴テーブルのデータ構造例を示す説明図である。

【図11】本実施の形態を実現するコンピュータのハードウェア構成例を示すブロック図である。

【発明を実施するための形態】

【0018】

以下、図面に基づき本発明を実現するにあたっての好適な一実施の形態の例を説明する。

図1は、本実施の形態の構成例についての概念的なモジュール構成図を示している。

なお、モジュールとは、一般的に論理的に分離可能なソフトウェア（コンピュータ・プログラム）、ハードウェア等の部品を指す。したがって、本実施の形態におけるモジュールはコンピュータ・プログラムにおけるモジュールのことだけでなく、ハードウェア構成におけるモジュールも指す。それゆえ、本実施の形態は、それらのモジュールとして機能させるためのコンピュータ・プログラム（コンピュータにそれぞれの手順を実行させるためのプログラム、コンピュータをそれぞれの手段として機能させるためのプログラム、コンピュータにそれぞれの機能を実現させるためのプログラム）、システム及び方法の説明をも兼ねている。ただし、説明の都合上、「記憶する」、「記憶させる」、これらと同等の文言を用いるが、これらの文言は、実施の形態がコンピュータ・プログラムの場合は、記憶装置に記憶させる、又は記憶装置に記憶させるように制御するという意味である。また、モジュールは機能に一対一に対応していてもよいが、実装においては、1モジュールを1プログラムで構成してもよいし、複数モジュールを1プログラムで構成してもよく、逆に1モジュールを複数プログラムで構成してもよい。また、複数モジュールは1コンピュータによって実行されてもよいし、分散又は並列環境におけるコンピュータによって1

10

20

30

40

50

モジュールが複数コンピュータで実行されてもよい。なお、1つのモジュールに他のモジュールが含まれていてもよい。また、以下、「接続」とは物理的な接続の他、論理的な接続（データの授受、指示、データ間の参照関係等）の場合にも用いる。「予め定められた」とは、対象としている処理の前に定まっていることをいい、本実施の形態による処理が始まる前はもちろんのこと、本実施の形態による処理が始まった後であっても、対象としている処理の前であれば、そのときの状況・状態に応じて、又はそれまでの状況・状態に応じて定まることの意を含めて用いる。「予め定められた値」が複数ある場合は、それぞれ異なった値であってもよいし、2以上の値（もちろんのことながら、全ての値も含む）が同じであってもよい。また、「Aである場合、Bをする」という意味を有する記載は、「Aであるか否かを判断し、Aであると判断した場合はBをする」の意味で用いる。ただし、Aであるか否かの判断が不要である場合を除く。

10

また、システム又は装置とは、複数のコンピュータ、ハードウェア、装置等がネットワーク（一対一対応の通信接続を含む）等の通信手段で接続されて構成されるほか、1つのコンピュータ、ハードウェア、装置等によって実現される場合も含まれる。「装置」と「システム」とは、互いに同義の用語として用いる。もちろんのことながら、「システム」には、人為的な取り決めである社会的な「仕組み」（社会システム）にすぎないものは含まない。

また、各モジュールによる処理毎に又はモジュール内で複数の処理を行う場合はその処理毎に、対象となる情報を記憶装置から読み込み、その処理を行った後に、処理結果を記憶装置に書き出すものである。したがって、処理前の記憶装置からの読み込み、処理後の記憶装置への書き出しについては、説明を省略する場合がある。なお、ここでの記憶装置としては、ハードディスク、RAM（Random Access Memory）、外部記憶媒体、通信回線を介した記憶装置、CPU（Central Processing Unit）内のレジスタ等を含んでいてもよい。

20

#### 【0019】

本実施の形態である情報処理装置100は、複数のWebサイトへの入り口を提供するポータルサイトを構築するものであって、図1の例に示すように、クロール実行処理モジュール110、検索実行処理モジュール150を有している。

なお、ポータル（又は、ポータルサイト）とは、さまざまなWebサイトにおける入り口的な扱いのサイトのことである。本実施の形態の説明では、複数のサービスサイトへのアクセスを集約した入り口となるサイトを指す。特に、本実施の形態では、会員制ポータルサイトを主に対象とする。つまり、ポータルサイトを利用するのに、ユーザーはログインすることが必要であり、ポータルサイトを利用するユーザー（具体的には、ユーザーID）を、本実施の形態は、利用することができる。

30

ポートレットは、ポータルサイト（のトップページ）に機能や情報を追加することができるプログラムである。一般的には、Java（登録商標）等の小さなプログラムで生成される。本実施の形態の説明では、各サービスが提供するページをポータルサイトの一部として表示する機能を実現するものを指す。

クローラーは、インターネット上のあらゆるWebサイトの情報を取得して検索用データベース・インデックスを作成する自動巡回プログラムであり、ロボットともいわれる。本実施の形態の説明では、ポートレット経由で提供される各サービスのコンテンツを巡回するプログラムを指す。

40

#### 【0020】

クロール実行処理モジュール110は、ポートレット設定データ115、クロール定義取得モジュール120、ユーザーアカウントデータ125、クロール実行モジュール130、クロール結果インデックスデータ135を有している。クロール実行処理モジュール110は、ログインユーザー（以下、単にユーザーともいう）がアクセス可能なWebサイト（サービスAコンテンツデータ処理装置190A、サービスBコンテンツデータ処理装置190B、サービスCコンテンツデータ処理装置190C）から検索に用いるインデックスを生成する。

50

検索実行処理モジュール150は、検索キーワード受付モジュール155、検索実行モジュール160、検索結果取得モジュール165、検索結果表示モジュール170、ユーザー検索実行履歴データ175、インデックス並び替え処理モジュール180を有している。検索実行処理モジュール150は、複数のWebサイトへの入り口を表現したポータルサイトを表示するよう制御する。そして、ユーザーの検索指示に基づいて、クロール実行処理モジュール110が作成したインデックスを用いて検索を行う。

#### 【0021】

ポートレット設定データ115は、クロール定義取得モジュール120と接続されている。ポートレット設定データ115は、ポートレット毎に保持されているクロール関連情報等のデータを記憶している。クロール定義取得モジュール120がクロールするたびに参照される。例えば、クロール定義テーブル500を記憶している。図5は、クロール定義テーブル500のデータ構造例を示す説明図である。クロール定義テーブル500は、属性として、例えば、URL欄512、クロール対象URL欄514、クロール除外URL欄516、検索対象URL欄518、検索対象除外URL欄520、巡回深度欄522、巡回間隔欄524、巡回スレッド数欄526等を有している。URL欄512は、クロール対象のWebサイトのURLを記憶している。クロール対象URL欄514は、クロール対象のWebサイトのURLを記憶している。URL欄512によるURLをさらに限定する場合に用いられる。クロール除外URL欄516は、クロール対象から除外するWebサイトのURLを記憶している。検索対象URL欄518は、検索対象のWebサイトのURLを記憶している。検索対象除外URL欄520は、検索対象から除外するWebサイトのURLを記憶している。巡回深度欄522は、巡回深度(クロール対象のWebサイトのリンクを辿る回数)を記憶している。巡回間隔欄524は、巡回間隔(クロールを行う間隔)を記憶している。巡回スレッド数欄526は、巡回スレッド数(同時に(又は並列的に)クロール処理を行う数)を記憶している。

10

20

#### 【0022】

クロール定義取得モジュール120は、ポートレット設定データ115、クロール実行モジュール130と接続されている。クロール定義取得モジュール120は、ポートレット設定データ115内からクロールの定義を取得する。具体的には、クロールに必要なクロール定義テーブル500を取得して、クロール実行モジュール130に渡す。

30

#### 【0023】

ユーザーアカウントデータ125は、クロール実行モジュール130と接続されている。ユーザーアカウントデータ125は、複数のWebサイトと、その複数のWebサイトに対するユーザーのアクセス権とを対応づけて記憶する。つまり、サービスコンテンツデータ処理装置190に対するユーザーのアカウントのデータを記憶している。クロールする際に、偽装ログインするために、クロール実行モジュール130によって参照される。例えば、ユーザーアカウントテーブル600を記憶している。クロール実行モジュール130は、ユーザーアカウントテーブル600内のデータを用いて、各サービスコンテンツデータ処理装置190にログインして、そのサービスコンテンツデータ処理装置190内のデータにアクセスする。図6は、ユーザーアカウントテーブル600のデータ構造例を示す説明図である。ユーザーアカウントテーブル600は、ユーザーID欄610、サービスA URL欄620、ログインID欄622、パスワード欄624、サービスB URL欄630、ログインID欄632、パスワード欄634等を有している。ユーザーID欄610は、本実施の形態において、ユーザー(情報処理装置100におけるログインユーザー)を一意に識別するための情報(ユーザーID: Identification)を記憶している。サービスA URL欄620は、サービスA(サービスAコンテンツデータ処理装置190A)のURLを記憶している。ログインID欄622は、サービスAのログインID(サービスAコンテンツデータ処理装置190AにおけるログインユーザーのID)を記憶している。パスワード欄624は、サービスAのパスワードを記憶している。サービスB URL欄630は、サービスB(サービスBコンテンツデータ処理装置190B)のURLを記憶している。ログインID欄632は、サービスBのログイン

40

50

ンID (サービスBコンテンツデータ処理装置190BにおけるログインユーザーのID)を記憶している。パスワード欄634は、サービスBのパスワードを記憶している。もちろんのことながら、サービスコンテンツデータ処理装置190にアクセスするための情報であるならば、ログインID、パスワード以外の情報であってもよいし、暗号化しておいてもよいし、例えば、複数のパスワードを必要とする場合は、その複数のパスワードを記憶している。

#### 【0024】

クロール実行モジュール130は、クロール定義取得モジュール120、ユーザーアカウントデータ125、クロール結果インデックスデータ135、サービスAコンテンツデータ処理装置190A、サービスBコンテンツデータ処理装置190B、サービスCコンテンツデータ処理装置190Cと接続されている。クロール実行モジュール130は、インデクサとしての機能を有しており、複数のWebサイトのそれぞれに対して、ユーザーアカウントデータ125によって記憶されているユーザーのアクセス権に基づいてアクセスし、複数のWebサイトに記載されている内容のインデックスを生成する。一般的に、クロールするページ毎にURL (Uniform Resource Locator) が異なっており、そのURL毎にインデクサが検索結果をデータベース化する。つまり、クロール定義取得モジュール120が取得したクロールの定義にしたがって、サービスコンテンツデータ処理装置190にクロールを実行する。その際に、サービスコンテンツデータ処理装置190にアクセスするために、ユーザーアカウントテーブル600を用いて偽装ログインする。また、インデックスの生成は、既存の技術を用いればよい。例えば、形態素解析、tf-idf等の技術を用いて、キーワードを抽出すればよい。

なお、検索実行処理モジュール150による検索が実行される前に、クロールを行って、インデックスを生成し、クロール結果インデックスデータ135にそのインデックスを記憶させていけばよい。また、ポータルサイトにログインしたことを契機として、クロール実行モジュール130がクロールを開始するようにしてもよい。

また、クロール実行モジュール130は、アクセス権についてクロール時にアクセス不可のURLをマークしキャッシュしておくことで、都度確認は行わないようにし、キャッシュはURL先のページ更新時に更新するようにしてもよい。

#### 【0025】

クロール結果インデックスデータ135は、クロール実行モジュール130、検索実行処理モジュール150の検索実行モジュール160、インデックス並び替え処理モジュール180と接続されている。クロール結果インデックスデータ135は、クロール実行モジュール130により生成されたインデックスをユーザーと対応づけて記憶する。つまり、複数のWebサイトのそれぞれに対して、ユーザーのアクセス権に基づいてアクセスし、その複数のWebサイトに記載されている内容のインデックスをそのユーザーと対応づけて記憶する。つまり、クロールした結果のデータを記憶しており、検索実行処理モジュール150による検索実行時に参照される。例えば、インデックステーブル700を記憶している。図7は、インデックステーブル700のデータ構造例を示す説明図である。インデックステーブル700は、UID欄710、URL欄720、キーワード欄730、ステータス欄740、更新日時欄750、優先ポイント欄760を有している。UID欄710は、UID (ユーザーID)を記憶している。URL欄720は、そのユーザーIDでアクセス (クロール) したURLを記憶している。キーワード欄730は、そのURLのWebサイト内のキーワード (インデックス) を記憶している。ステータス欄740は、そのURLにアクセスできたか否かを示すステータスを記憶している。例えば、アクセス可能を示す情報として「200」を用い、アクセスできなかったことを示す情報として「403」を用いる。更新日時欄750は、そのURLの更新日時 (年、月、日、時、分、秒、秒以下、又はこれらの組み合わせであってもよい) を記憶している。優先ポイント欄760は、そのURLの優先して表示すべき度合いを示す情報である優先ポイントを記憶している。

#### 【0026】

10

20

30

40

50

検索キーワード受付モジュール155は、検索実行モジュール160と接続されている。検索キーワード受付モジュール155は、ポータルサイトにおいて、複数のWebサイトに対する検索条件の入力をユーザーから受け付ける。例えば、ユーザーの操作に基づいて、検索のために入力されたキーワードを受け付ける。

検索実行モジュール160は、検索キーワード受付モジュール155、検索結果取得モジュール165、ユーザー検索実行履歴データ175、クロール実行処理モジュール110のクロール結果インデックスデータ135と接続されている。検索実行モジュール160は、クロール結果インデックスデータ135に記憶された、ユーザーに対応づけられたインデックスに対して、検索キーワード受付モジュール155によって入力された検索条件に基づいた検索を行う。例えば、クロール結果インデックスデータ135内のインデックス内を検索実行する。実行後は検索履歴としてユーザー検索実行履歴データ175に記録する。

10

検索結果取得モジュール165は、検索実行モジュール160、検索結果表示モジュール170と接続されている。検索結果取得モジュール165は、検索実行モジュール160による検索した結果を取得する。

検索結果表示モジュール170は、検索結果取得モジュール165と接続されている。検索結果表示モジュール170は、検索した結果を、検索指示を行ったユーザーが確認することができる液晶ディスプレイ等の表示装置に表示する。一般的には、検索指示を行った画面に表示する。

20

#### 【0027】

ユーザー検索実行履歴データ175は、検索実行モジュール160、インデックス並び替え処理モジュール180と接続されている。ユーザー検索実行履歴データ175は、ユーザーが入力した検索条件を履歴として記憶する。例えば、ユーザーによる検索条件、そして、その検索条件にしたがって検索した結果であるデータを記憶している。また、クロールインデックスの並び替えによる最適化処理を行うために、インデックス並び替え処理モジュール180から参照される。

インデックス並び替え処理モジュール180は、ユーザー検索実行履歴データ175、クロール実行処理モジュール110のクロール結果インデックスデータ135と接続されている。インデックス並び替え処理モジュール180は、クロール結果インデックスデータ135に記憶されている情報に対して、ユーザー検索実行履歴データ175に記憶されている履歴に基づいてユーザーのインデックスを並び替える。

30

#### 【0028】

サービスコンテンツデータ処理装置190（サービスAコンテンツデータ処理装置190A、サービスBコンテンツデータ処理装置190B、サービスCコンテンツデータ処理装置190C）は、クロール実行処理モジュール110のクロール実行モジュール130と接続されている。サービスコンテンツデータ処理装置190は、サービスサイトのコンテンツデータを記憶している。つまり、ポートレット内に表示されるデータを記憶している。ユーザーがアクセス可能であるサービスコンテンツデータ処理装置190だけでなく、アクセスできないサービスコンテンツデータ処理装置190が含まれていてもよい。

図1の例では、3つのサービスコンテンツデータ処理装置190を示しているが、もちろんのことながら、2つ以上のサービスコンテンツデータ処理装置190があればよい。また、1つのポータルサイト内には、全てのサービスコンテンツデータ処理装置190に対応したポートレットがある必要はなく、ユーザーによる設定（又はデフォルトの設定）で選択される。また、1つのサービスコンテンツデータ処理装置190に対して複数のポートレットが割り当てられるようにしてもよい。この場合、一般的には、各ポートレットに割り当てられているURLは、サービスコンテンツデータ処理装置190内の異なるURLである場合が多い。例えば、サービスコンテンツデータ処理装置190として、そのユーザーが管理している複写機、ファックス、スキャナ、プリンター、複合機（スキャナ、プリンター、複写機、ファックス等のいずれか2つ以上の機能を有している画像処理装置）に関する情報（例えば、プリント枚数、故障情報等）を表示するものであってもよい

40

50

## 【 0 0 2 9 】

情報処理装置 1 0 0 による処理の概要を説明する。

ポータルサイトの各ポートレット毎にクローラーでインデックスが生成される。ポータルサイトからポートレット経由で提供されたサービス情報を検索する場合、ログインユーザーがアクセス可能なサービス情報に対して検索を行う。つまり、ログインユーザーから見ると、ポータルサイト内をあたかも検索しているかの如くに、各ポートレットに対応づけられている Web サイトを検索することとなる。また、ログインユーザー毎に検索実行履歴を元に検索実行順序の決定を行う。

## 【 0 0 3 0 】

図 2 は、本実施の形態を利用したシステム構成例を示す説明図である。

サービス A コンテンツデータ処理装置 1 9 0 A、サービス B コンテンツデータ処理装置 1 9 0 B、サービス C コンテンツデータ処理装置 1 9 0 C、情報処理装置 1 0 0、クロール実行処理装置 2 1 0、検索実行処理装置 2 5 0、ユーザー端末 2 8 0 A、ユーザー端末 2 8 0 B、ユーザー端末 2 8 0 C は、通信回線 2 9 0 を介してそれぞれ接続されている。通信回線 2 9 0 は、無線、有線、これらの組み合わせであってもよく、例えば、通信インフラとしてのインターネット、イントラネット等であってもよい。

予め情報処理装置 1 0 0 は、クロール処理によって各ユーザーがアクセス可能なサービスコンテンツデータ処理装置 1 9 0 から検索用のインデックスを生成しておく。

ユーザーの操作に応じて、ユーザー端末 2 8 0 からポータルサイトにログインされ、検索指示が行われる。情報処理装置 1 0 0 は、その検索指示を行ったユーザーに対応するインデックスに対して、検索処理を行って、結果を表示する。したがって、検索結果には、そのユーザーがアクセス可能な情報が表示されることになる。

また、情報処理装置 1 0 0 内のクロール実行処理モジュール 1 1 0 と検索実行処理モジュール 1 5 0 は、別々の情報処理装置として構築されていてもよい。例えば、クロール実行処理装置 2 1 0 は、クロール実行処理モジュール 1 1 0 を有しており、検索実行処理装置 2 5 0 は、検索実行処理モジュール 1 5 0 を有している形態であってもよい。

情報処理装置 1 0 0、クロール実行処理装置 2 1 0、検索実行処理装置 2 5 0 による機能は、クラウドサービスとして実現してもよい。

## 【 0 0 3 1 】

図 3 は、本実施の形態による処理例（クロール実行処理モジュール 1 1 0 によるクロール実行処理例）を示すフローチャートである。

サービスコンテンツデータ処理装置 1 9 0 に偽装ログインして、ポートレット毎に定義されたクロールを実行する。巡回時には HTTP のリターンコード（具体的には、「 2 0 0 」、「 4 0 3 」等）をインデックスに記録しておき、そのユーザーがアクセス可能な URL が否かを後の検索実行時に利用できるようにしておく。

図 3 の例に示す処理をユーザー毎にクロール定義で定められた間隔で定期的に行うことで、クロールのインデックスデータをポートレット毎に作成する。

## 【 0 0 3 2 】

ステップ S 3 0 2 では、クロール定義取得モジュール 1 2 0 は、ポートレット設定からクロール定義を取得する。

ステップ S 3 0 4 では、クロール実行モジュール 1 3 0 は、ユーザーに偽装ログイン（*impersonate*）する。

ステップ S 3 0 6 では、クロール実行モジュール 1 3 0 は、サービスの URL を巡回し、キーワードを抽出する。

ステップ S 3 0 8 では、クロール実行モジュール 1 3 0 は、巡回すべきサービスの URL が残っているか否かを判断し、残っている場合はステップ S 3 1 0 へ進み、それ以外の場合は処理を終了する（ステップ S 3 9 9 ）。

ステップ S 3 1 0 では、クロール実行モジュール 1 3 0 は、当該 URL は前回（前回）巡回時から更新されているか否かを判断し、更新されている場合はステップ S 3 1 4 へ進

10

20

30

40

50

み、それ以外の場合はステップ S 3 1 2 へ進む。

ステップ S 3 1 2 では、クロール実行モジュール 1 3 0 は、次の URL に処理を進め、ステップ S 3 0 6 以降の処理を繰り返す。

ステップ S 3 1 4 では、クロール実行モジュール 1 3 0 は、URL へアクセスを試みた（偽装ログインした）結果、アクセス可能であるか否かを判断し、アクセス可能である場合はステップ S 3 1 6 へ進み、それ以外の場合はステップ S 3 1 8 へ進む。

ステップ S 3 1 6 では、クロール実行モジュール 1 3 0 は、アクセス可能な URL としてインデックスに記録する。具体的には、インデックステーブル 7 0 0 のステータス欄 7 4 0 に「2 0 0」を設定する。

ステップ S 3 1 8 では、クロール実行モジュール 1 3 0 は、アクセス不可な URL としてインデックスに記録する。具体的には、インデックステーブル 7 0 0 のステータス欄 7 4 0 に「4 0 3」を設定する。

ステップ S 3 2 0 では、クロール実行モジュール 1 3 0 は、次の URL に処理を進め、ステップ S 3 0 6 以降の処理を繰り返す。

#### 【0033】

図 4 は、本実施の形態による処理例を示す説明図である。

ポータル画面 4 0 0 は、ユーザーの操作に応じて、ポータルサイトにログインした場合の画面例である。

ポータル画面 4 0 0 には、ポートレット表示領域 4 1 0、ポートレット表示領域 4 2 0、ポートレット表示領域 4 3 0、検索領域 4 5 0 が表示されている。ポートレット表示領域 4 1 0、ポートレット表示領域 4 2 0、ポートレット表示領域 4 3 0 は、ユーザーによって設定されており、サービスコンテンツデータ処理装置 1 9 0 内の情報が表示されている。

検索領域 4 5 0 は、検索キーワード指定枠 4 5 2、検索ボタン 4 5 4 を有している。検索キーワード指定枠 4 5 2 に検索用のキーワードが入力され、検索ボタン 4 5 4 が選択されることによって、検索指示が行われる。

#### 【0034】

図 8 は、本実施の形態による処理例（検索実行処理モジュール 1 5 0 による検索処理例）を示すフローチャートである。

ユーザーが検索のために入力したキーワードを用いて、クロール結果のインデックスからそのキーワードにヒットする URL を抽出する。その中でアクセス可能なものを検索結果として採用する。

検索完了後は、検索結果からインデックスの並び替えを実施して、次回以降の検索効率を向上させる。

図 8 の例に示す処理をポートレット毎に実行し、全ての結果をマージしたものをポータルサイトの検索結果としてユーザーに表示する。

#### 【0035】

ステップ S 8 0 2 では、検索キーワード受付モジュール 1 5 5 は、ユーザーの操作によって検索ボタン 4 5 4 が選択された場合（検索指示が行われた場合）、ポータルサイトの検索領域 4 5 0 の検索キーワード指定枠 4 5 2 内の検索キーワードを受け付ける。

ステップ S 8 0 4 では、検索実行モジュール 1 6 0 は、クロール結果のインデックス（具体的には、インデックステーブル 9 0 0）からキーワードにヒットする URL を抽出する。図 9 は、インデックステーブル 9 0 0 のデータ構造例を示す説明図である。インデックステーブル 9 0 0 は、図 7 に例示したインデックステーブル 7 0 0 と同等のデータ構造を有しているが、そのユーザーにおけるインデックステーブル 9 0 0 としたものである。具体的には、インデックステーブル 9 0 0 の UID 欄 9 1 0 からログインユーザーにマッチするレコードを抽出する。つまり、UID 欄 9 1 0 がログインユーザー「taro」であるレコードだけから構成されているものとなる。そして、インデックステーブル 9 0 0 のキーワード欄 9 3 0 内で、キーワードにヒットしたものを検索結果として抽出する。

#### 【0036】

10

20

30

40

50

ステップ S 8 0 6 では、検索実行モジュール 1 6 0 は、URL が抽出されたか否かを判断し、抽出された場合はステップ S 8 0 8 へ進み、それ以外の場合はステップ S 8 1 4 へ進む。

ステップ S 8 0 8 では、検索実行モジュール 1 6 0 は、URL はアクセス可能であるかを判断し、アクセス可能である場合はステップ S 8 1 2 へ進み、それ以外の場合はステップ S 8 1 0 へ進む。具体的には、インデックステーブル 9 0 0 のステータス欄 9 4 0 を参照して、アクセス無効なステータス（「4 0 3」）は検索結果から除外する。

【0 0 3 7】

ステップ S 8 1 0 では、検索実行モジュール 1 6 0 は、次のインデックスの処理に進める。

ステップ S 8 1 2 では、検索実行モジュール 1 6 0 は、検索結果一覧の最後に追加する。例えば、キーワード入力履歴テーブル 1 0 0 0 に追加する。図 1 0 は、キーワード入力履歴テーブル 1 0 0 0 のデータ構造例を示す説明図である。

キーワード入力履歴テーブル 1 0 0 0 は、UID 欄 1 0 1 0、キーワード欄 1 0 2 0、入力日時欄 1 0 3 0 を有している。UID 欄 1 0 1 0 は、ユーザー ID を記憶している。キーワード欄 1 0 2 0 は、そのユーザーによる検索指示におけるキーワード（検索対象語）を記憶している。入力日時欄 1 0 3 0 は、その検索指示が行われた日時を記憶している。

【0 0 3 8】

ステップ S 8 1 4 では、インデックス並び替え処理モジュール 1 8 0 は、検索結果一覧の上位からインデックスの並び替えを実施する。具体的には、キーワード入力履歴の日時、頻度に応じてインデックスデータを並び替える。例えば、キーワード入力履歴テーブル 1 0 0 0 でのキーワード出現数で優先ポイントを計算して並び替える。具体例として、キーワード「プリンター」で 3 1 回、「F A X」で 1 1 回、「印刷」で 8 回の場合、以下の式で優先ポイントを計算する。

$$\text{プリンター} (31 \text{ 回}) + \text{F A X} (11 \text{ 回}) + \text{印刷} (8 \text{ 回}) = 50 \text{ pt}$$

つまり、過去に検索用のキーワードとして用いたものが多いほど、優先ポイントが高くなるようになり、他の URL よりも優先して表示する。具体的には、検索結果における表示順序を上位にするようにしてもよい。そして、表示形態（色、文字の太さ、点滅表示等）を他の検索結果と異ならせるようにしてもよい。

ステップ S 8 1 6 では、検索結果取得モジュール 1 6 5 は、検索結果を取得する。

ステップ S 8 1 8 では、検索結果表示モジュール 1 7 0 は、検索結果をポータルサイトに表示させる。

【0 0 3 9】

なお、本実施の形態としてのプログラムが実行されるコンピュータのハードウェア構成は、図 1 1 に例示するように、一般的なコンピュータであり、具体的にはパーソナルコンピュータ、サーバーとなり得るコンピュータ等である。つまり、具体例として、処理部（演算部）として CPU 1 1 0 1 を用い、記憶装置として RAM 1 1 0 2、ROM 1 1 0 3、HD 1 1 0 4 を用いている。HD 1 1 0 4 として、例えばハードディスク、SSD（Solid State Drive）を用いてもよい。クロール定義取得モジュール 1 2 0、検索キーワード受付モジュール 1 5 5、検索実行モジュール 1 6 0、検索結果取得モジュール 1 6 5、検索結果表示モジュール 1 7 0、インデックス並び替え処理モジュール 1 8 0 等のプログラムを実行する CPU 1 1 0 1 と、そのプログラムやデータを記憶する RAM 1 1 0 2 と、本コンピュータを起動するためのプログラム等が格納されている ROM 1 1 0 3 と、ポートレット設定データ 1 1 5、ユーザーアカウントデータ 1 2 5、クロール結果インデックスデータ 1 3 5、ユーザー検索実行履歴データ 1 7 5 等の機能を有する補助記憶装置（フラッシュメモリ等であってもよい）である HD 1 1 0 4 と、キーボード、マウス、タッチパネル、マイク等に対する利用者の操作に基づいてデータを受け付ける受付装置 1 1 0 6 と、CRT、液晶ディスプレイ、スピーカー等の出力装置 1 1 0 5 と、ネットワークインタフェースカード等の通信ネットワークと接続するための通信回線イ

10

20

30

40

50

ンタフェース 1107、そして、それらをつないでデータのやりとりをするためのバス 1108により構成されている。これらのコンピュータが複数台互いにネットワークによって接続されていてもよい。

#### 【0040】

前述の実施の形態のうち、コンピュータ・プログラムによるものについては、本ハードウェア構成のシステムにソフトウェアであるコンピュータ・プログラムを読み込ませ、ソフトウェアとハードウェア資源とが協働して、前述の実施の形態が実現される。

なお、図11に示すハードウェア構成は、1つの構成例を示すものであり、本実施の形態は、図11に示す構成に限らず、本実施の形態において説明したモジュールを実行可能な構成であればよい。例えば、一部のモジュールを専用のハードウェア（例えば特定用途向け集積回路（Application Specific Integrated Circuit：ASIC）等）で構成してもよく、一部のモジュールは外部のシステム内にあり通信回線で接続しているような形態でもよく、さらに図11に示すシステムが複数互いに通信回線によって接続されていて互いに協調動作するようにしてもよい。また、特に、パーソナルコンピュータの他、携帯情報通信機器（携帯電話、スマートフォン、モバイル機器、ウェアラブルコンピュータ等を含む）、情報家電、ロボット、複写機、ファックス、スキャナ、プリンター、複合機などに組み込まれていてもよい。

#### 【0041】

前述の実施の形態においては、ポータルサイトとして、ポートレットを用いる例を示したが、HTMLにおけるiframe（インラインフレーム）等を用いて実現してもよい。

#### 【0042】

なお、説明したプログラムについては、記録媒体に格納して提供してもよく、また、そのプログラムを通信手段によって提供してもよい。その場合、例えば、前記説明したプログラムについて、「プログラムを記録したコンピュータ読み取り可能な記録媒体」の発明として捉えてもよい。

「プログラムを記録したコンピュータ読み取り可能な記録媒体」とは、プログラムのインストール、実行、プログラムの流通等のために用いられる、プログラムが記録されたコンピュータで読み取り可能な記録媒体をいう。

なお、記録媒体としては、例えば、デジタル・バーサタイル・ディスク（DVD）であって、DVDフォーラムで策定された規格である「DVD-R、DVD-RW、DVD-RAM等」、DVD+RWで策定された規格である「DVD+R、DVD+RW等」、コンパクトディスク（CD）であって、読出し専用メモリ（CD-ROM）、CDレコーダブル（CD-R）、CDリライタブル（CD-RW）等、ブルーレイ・ディスク（Blu-ray（登録商標）Disc）、光磁気ディスク（MO）、フレキシブルディスク（FD）、磁気テープ、ハードディスク、読出し専用メモリ（ROM）、電氣的消去及び書換可能な読出し専用メモリ（EEPROM（登録商標））、フラッシュ・メモリ、ランダム・アクセス・メモリ（RAM）、SD（Secure Digital）メモリーカード等が含まれる。

そして、前記のプログラム又はその一部は、前記記録媒体に記録して保存や流通等させてもよい。また、通信によって、例えば、ローカル・エリア・ネットワーク（LAN）、メトロポリタン・エリア・ネットワーク（MAN）、ワイド・エリア・ネットワーク（WAN）、インターネット、イントラネット、エクストラネット等に用いられる有線ネットワーク、又は無線通信ネットワーク、さらにこれらの組み合わせ等の伝送媒体を用いて伝送させてもよく、また、搬送波に乗せて搬送させてもよい。

さらに、前記のプログラムは、他のプログラムの一部分であってもよく、又は別個のプログラムと共に記録媒体に記録されていてもよい。また、複数の記録媒体に分割して記録されていてもよい。また、圧縮や暗号化等、復元可能であればどのような態様で記録されていてもよい。

#### 【符号の説明】

10

20

30

40

50

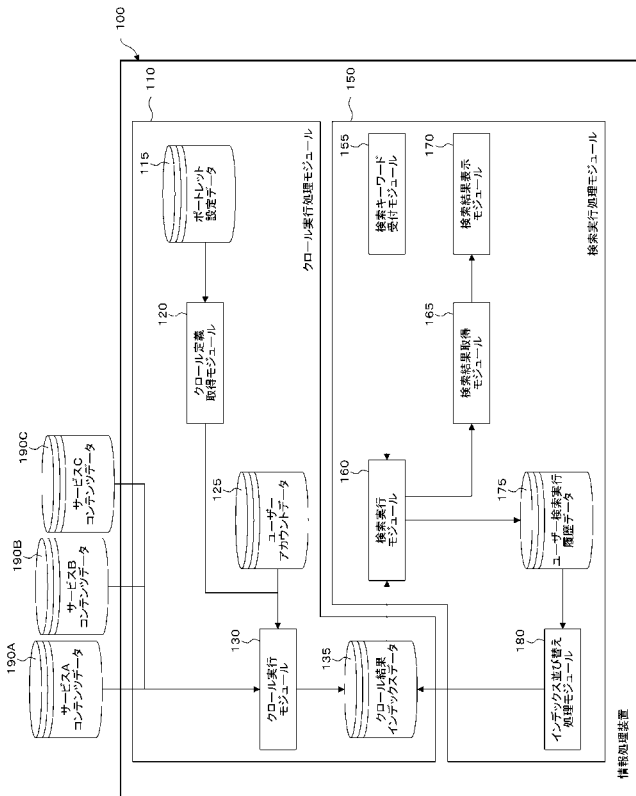
【 0 0 4 3 】

- 1 0 0 ... 情報処理装置
- 1 1 0 ... クロール実行処理モジュール
- 1 1 5 ... ポートレット設定データ
- 1 2 0 ... クロール定義取得モジュール
- 1 2 5 ... ユーザーアカウントデータ
- 1 3 0 ... クロール実行モジュール
- 1 3 5 ... クロール結果インデックスデータ
- 1 5 0 ... 検索実行処理モジュール
- 1 5 5 ... 検索キーワード受付モジュール
- 1 6 0 ... 検索実行モジュール
- 1 6 5 ... 検索結果取得モジュール
- 1 7 0 ... 検索結果表示モジュール
- 1 7 5 ... ユーザー検索実行履歴データ
- 1 8 0 ... インデックス並び替え処理モジュール
- 1 9 0 ... サービスコンテンツデータ処理装置
- 2 1 0 ... クロール実行処理装置
- 2 5 0 ... 検索実行処理装置
- 2 8 0 ... ユーザー端末
- 2 9 0 ... 通信回線

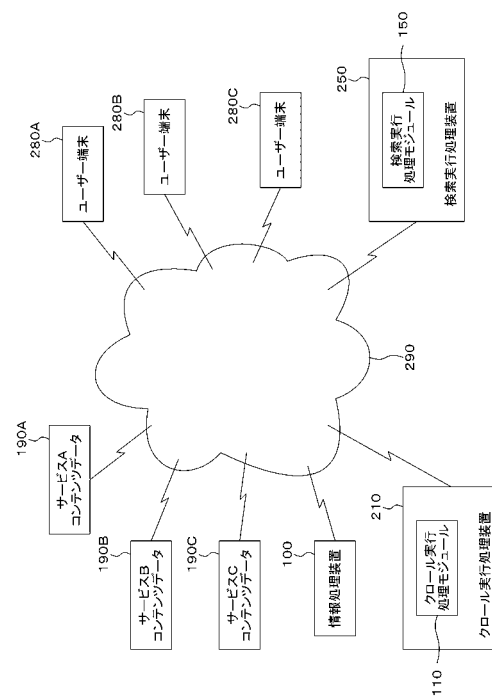
10

20

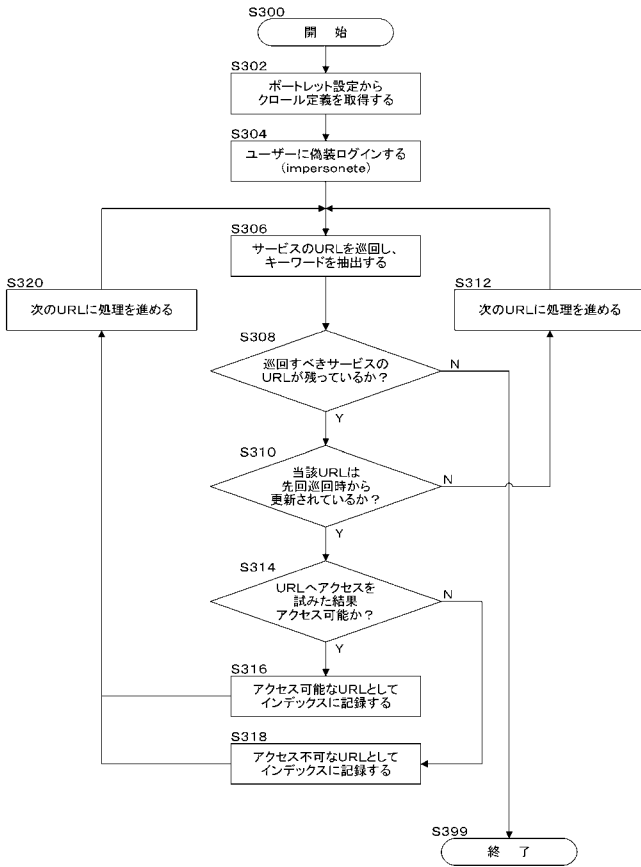
【 図 1 】



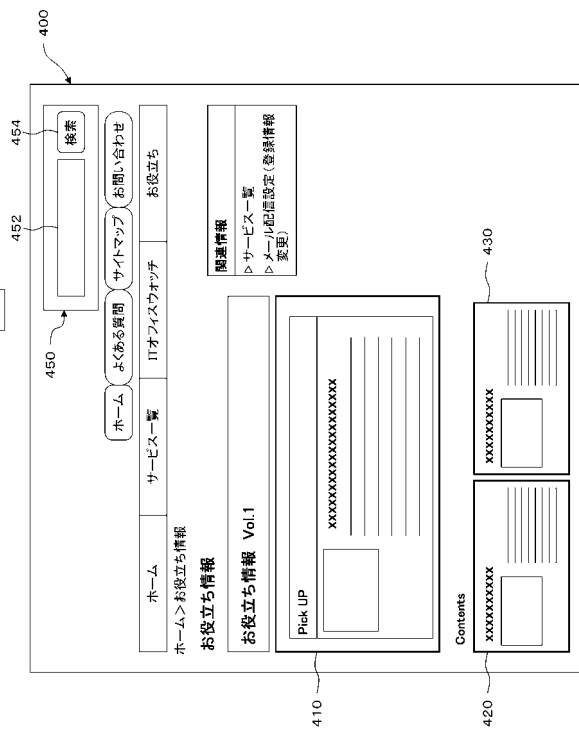
【 図 2 】



【 図 3 】



【 図 4 】



【 図 5 】

属性	値
512 URL	http://www.ffffxxxx.co.jp/xxx/yyy/zzz/
514 クローリング対象URL	http://www.ffffxxxx.co.jp/xxx/yyy/zzz/*
516 クローリング除外URL	
518 検索対象URL	http://www.ffffxxxx.co.jp/xxx/yyy/zzz/*
520 検索対象除外URL	http://www.ffffxxxx.co.jp/xxx/yyy/zzz/priv/*
522 巡回深度	10
524 巡回間隔(msec)	5000
526 巡回スレッド数	5

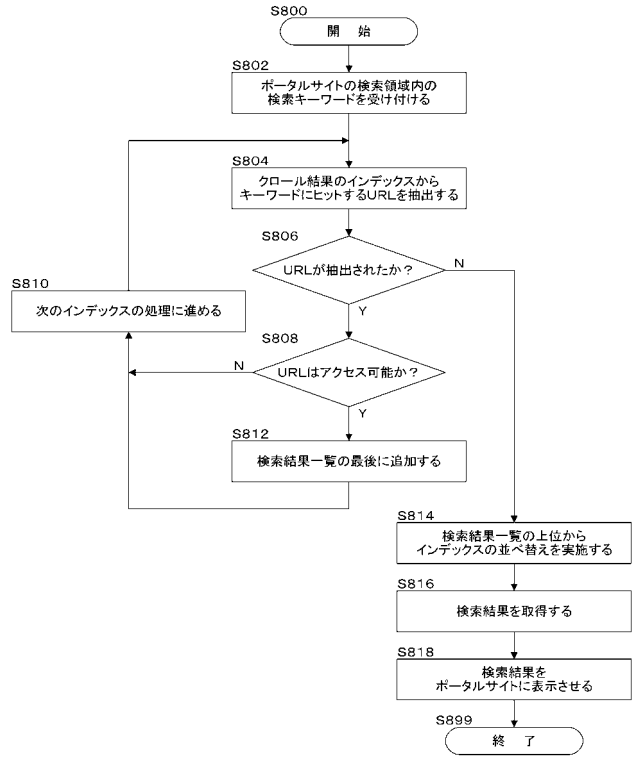
【 図 6 】

610	ユーザーID	50001					
620	サービスA URL	https://www.a.co.jp					
622	ログインID	a0001					
624	パスワード	xxxxx					
630	サービスB URL	https://www.b.co.jp					
632	ログインID	b0001					
634	パスワード	yyyyy					
600	...						

【図7】

710	UID							
720	URL	http://www.ffffxxxx.co.jp/xxx/yyy/zzz/aaa	プリンター-FAX印刷:	200	...	...	...	...
730	キーワード	http://www.ffffxxxx.co.jp/xxx/yyy/zzz/bbb	ソリューション課題: 解決方法:	403	...	...	...	...
740	ステータス	http://www.ffffxxxx.co.jp/xxx/yyy/zzz/ccc	...	200	...	...	...	...
750	更新日時	...	...	...	...	...	...	...
760	優先ポイント	...	...	...	...	...	...	...

【図8】



【図9】

910	UID							
920	URL	http://www.ffffxxxx.co.jp/xxx/yyy/zzz/aaa	プリンター-FAX印刷:	200	...	...	...	...
930	キーワード	http://www.ffffxxxx.co.jp/xxx/yyy/zzz/bbb	ソリューション課題: 解決方法:	403	...	...	...	...
940	ステータス	http://www.ffffxxxx.co.jp/xxx/yyy/zzz/ccc	...	200	...	...	...	...
950	更新日時	...	...	...	...	...	...	...
960	優先ポイント	...	...	50pt	...	47pt	...	30pt

【図10】

1010	UID	1020	キーワード	1030	入力日時
	taro		プリンター		2015/3/5 15:02
	taro		消耗品		2015/3/4 15:02
	taro		プリンター		2015/3/3 15:02
	...		...		...

【図11】

