明　細　書

発明の名称 ：

情報処理装置、検索サーバ、ウェブサーバ及びコンピュータプログラム

技術分野

[0001] 本発明は、情報処理装置、検索サーバ、ウェブサーバ及びコンピュータプログラムに関する。

背景技術

[0002] 近年はインターネットが急速に普及し、膨大な情報がインターネットを介して入手可能となっている。そのような膨大な情報量を効率的に扱うための手段として最も有効なのが、検索エンジンである。

[0003] 検索エンジンには、不特定多数のインターネット上の情報を検索語を頼りに探索するものや、特定のウェブサイト内の情報を同様に検索語に従って検索するものがある。

[0004] いずれにおいても検索結果はリスト表示され、リストに含まれるリンク先をユーザがクリックすると、該当ページが表示されたり（特許文献１を参照）、該当ページの画像が表示されたりする（特許文献２を参照）。

先行技術文献

特許文献

[0005] 特許文献1：特開２０１１－２３８２７３号公報

特許文献2：特開２０１０－１１３７３０号公報

発明の概要

発明が解決しようとする課題

[0006] しかしながら、上記文献において開示されている結果表示の形態では、検索語が検索結果のリンク先の文書等のどこに存在するかはユーザが自ら改めて探さざるを得なかった。また、特許文献１の図８に示されるような表示形態では、「Ｃａｃｈｅｄ」で示されたキャッシュ先をクリックするとサーバにキャッシュされたページの情報が表示され、当該ページ上では検索語がハ

イライトされることもあり得るが、当該ページはサーバに格納された時点の内容を示すものであって、現在のページの内容を示しているわけではなく、また、当該キャッシュページは検索結果に付随する参考情報であって、検索結果の文書そのものではない。

[0007] 本発明は上記問題に鑑みなされたものであり、検索結果の文書における検索語の位置をユーザが容易に特定可能として、ユーザが効率的かつ迅速に目的の文書に到達できるようにすることを目的とする。

課題を解決するための手段

[0008] 上記課題を解決するための情報処理装置の発明は、

検索語の入力を受け付ける受付手段と、

検索語に基づく検索指示を外部の検索サーバに対して送信する送信手段と、

前記検索指示に基づく検索結果を受信する受信手段と、

検索結果を表示する表示手段とを備え、

前記検索結果には複数のリンク情報が含まれ、いずれかのリンク情報が選択された場合に前記送信手段が該リンク情報に従って宛先サーバにウェブページの要求を送信し、

前記情報処理手段は、前記要求に応じて前記受信手段が受信した前記ウェブページにつき、該ウェブページ内で前記検索語を検索し、かつ、該検索語を強調表示するためのタグを該ウェブページ内に挿入して加工する処理手段を更に備え、

前記表示手段は前記強調表示のためのタグが挿入されたウェブページを、該ウェブページ内の前記検索語を強調して表示することを特徴とする。

[0009] 上記課題を解決する上記情報処理装置から検索語に基づく検索指示を受信し、検索処理を実行する検索サーバの発明は、

前記情報処理装置から検索語を受信する受信手段と、

前記検索語を用いて全文検索データベースを検索して検索結果を生成する検索手段と、

前記検索結果を前記情報処理装置へ送信する送信手段と

を備え、

前記検索手段は、前記情報処理装置において前記検索結果におけるいずれかのリンク情報が選択された場合に、該情報処理装置に該リンク情報に対応するウェブページであって検索語の強調表示のためのタグが挿入されたウェブページを、検索結果を表示するウィンドウ内の領域で表示させるためのスクリプトを前記検索結果に含めることを特徴とする。

[0010] 上記課題を解決するウェブサーバの発明は、

上記情報処理装置から、該情報処理装置が取得した所定の検索語に基づく検索結果に含まれるウェブページの要求を受信する受信手段と、

該要求されたウェブページを該情報処理装置に送信する送信手段とを備え、

前記送信手段が送信する前記ウェブページ内に、前記情報処理装置により実行されるスクリプトであって、該情報処理装置に前記ウェブページ内で前記検索語を検索させ、かつ、該検索語を強調表示するためのタグを該ウェブページ内に挿入させて加工を行わせるためのスクリプトを記述することを特徴とする。

発明の効果

[0011] 本発明によれば、検索結果の文書における検索語の位置をユーザが容易に特定可能として、ユーザが効率的かつ迅速に目的の文書に到達できるようになる。

図面の簡単な説明

[0012] [図1]発明の実施形態に対応する、インターネット上のウェブサイトの検索システムの全体構成を示すブロック図である。

[図2]発明の実施形態に対応する、クライアント１０１のハードウェア構成の一例を示すブロック図である。

[図3]発明の実施形態に対応する、検索サーバ１０３、ウェブサーバ１０４の機器構成を示すブロック図である。

[図4]発明の実施形態に対応する、検索サーバ１０３における検索用のテキストデータと画像データを生成する処理の一例を示すフローチャートである。

[図5]発明の実施形態に対応する、図４のフローチャートに対応する処理により生成された、全文検索データベース１０５におけるテキスト検索用テーブルのデータ構成の一例を示す図である。

[図6]発明の実施形態に対応する、図４のフローチャートに従った処理により生成される画像データベース１０６の画像登録テーブルのデータ構成の一例を示す図である。

[図7a]発明の実施形態に対応する検索処理の一例を示すフローチャートである。

[図7b]発明の実施形態に対応する検索処理におけるクライアント１０１の初期状態での処理の例を示すフローチャートである。

[図7c]発明の実施形態に対応する検索処理におけるクライアント１０１の検索画面表示待ち状態での処理の例を示すフローチャートである。

[図7d]発明の実施形態に対応する検索処理におけるクライアント１０１の検索画面表状態での処理の例を示すフローチャートである。

[図7e]発明の実施形態に対応する検索処理におけるクライアント１０１の検索結果受信待ち状態での処理の例を示すフローチャートである。

[図7f]発明の実施形態に対応する検索処理におけるクライアント１０１の検索結果表示状態での処理の例を示すフローチャートである。

[図7g]発明の実施形態に対応する検索処理におけるクライアント１０１のプレビュー情報受信待ち状態での処理の例を示すフローチャートである。

[図7h]発明の実施形態に対応する検索処理におけるクライアント１０１のプレビュー情報表示状態での処理の例を示すフローチャートである。

[図7i]発明の実施形態に対応する検索処理における検索サーバ１０３の要求待ち状態での処理の例を示すフローチャートである。

[図7j]発明の実施形態に対応する検索処理におけるウェブサーバ１０４の要求待ち状態での処理の例を示すフローチャートである。

[図8]発明の実施形態に対応する、図７ｉのステップＳ７９４における検索処理の詳細を示すフローチャートである。

[図9]発明の実施形態に対応する、検索画面の一例を示す図である。

[図10A]発明の実施形態に対応する、検索結果のウィンドウ表示の一例を示す図である。

[図10B]発明の実施形態に対応する、検索結果のプレビュー表示の一例を示す図である。

[図10C]発明の実施形態に対応する、検索語を強調表示した検索結果のプレビュー表示の一例を示す図である。

[図10D]発明の実施形態に対応する、検索語を強調表示した検索結果のプレビュー表示の他の一例を示す図である。

[図11]発明の実施形態に対応する、検索語を強調表示するための加工を行ったウェブページのソースファイルの一例を示す図である。

発明を実施するための形態

[0013] ［実施形態１］

以下に、添付する図面を参照して本発明の１つの実施形態を説明する。ただし、この実施の形態に記載されている構成要素はあくまでも例示であり、この発明の範囲をそれらのみに限定する趣旨のものではない。

[0014] ＜システム構成＞

図１は、本実施形態に対応する検索システムの全体構成を示すブロック図である。ユーザが操作してインターネット上で取得可能な文書ファイルを対象とした検索を検索サーバ１０３に行わせ、その検索結果を取得するクライアント１０１ａ、１０１ｂ、１０１ｃ（以下、これらを総称して「クライアント１０１」という。）、インターネット検索を行う検索サーバ１０３と、ウェブサイトの情報を提供するウェブサーバ１０４が、それぞれインターネット１０２に接続されている。検索サーバ１０３には、全文検索データベース１０５、画像データベース１０６が接続されている。ウェブサーバ１０４にはサイト情報データベース１０７が接続されている。

[0015] クライアント１０１は、ユーザが操作して、ウェブサーバ１０４がインターネット上で公開するウェブサイト情報を閲覧可能な情報処理装置であり、ノートパソコン、デスクトップパソコン、携帯情報端末、携帯電話、スマートフォン、タブレット端末等を含む。クライアント１０１には、所謂インターネット・ブラウザのソフトウェアがインストールされているものとする。

[0016] クライアント１０１は、イーサネット（登録商標）等のネットワークケーブルを含むアクセス手段であるＬＡＮによってインターネット１０２に接続されている。尚、ネットワーク１０２へのアクセス手段はイーサネット（登録商標）ケーブルに限定されるものではなく、無線ＬＡＮ等の無線通信手段によって構成されていてもよい。後述する検索サーバ１０３も同様のアクセス手段を備えているものとする。

[0017] インターネット１０２は、世界中のネットワークが互いに接続されたネットワークであるが、例えば、イントラネットのように特定の組織内においてのみ接続可能なネットワークであってもよい。

[0018] 検索サーバ１０３は、クライアント１０１に対して、特定の企業、店舗或いは個人に関する情報を提供するウェブサイトに含まれるウェブページを含めたインターネット上で取得が可能な文書ファイルの検索のための検索サービスを提供する情報処理装置である。

[0019] 検索サーバ１０３は、全文検索データベース１０５、画像データベース１０６に格納されている情報を管理している。検索サーバ１０３は、例えば、ＬＡＮによって、全文検索データベース１０５、画像データベース１０６に接続される。ウェブサーバ１０４はサイト情報データベース１０７に格納されている情報を管理している。ウェブサーバ１０４は、例えば、ＬＡＮによって、サイト情報データベース１０７に接続される。

[0020] 全文検索データベース１０５、画像データベース１０６、サイト情報データベース１０７は、それぞれ所定のデータベース・ソフトウェアがインストールされた情報処理装置であり、各種データの管理を行う。

[0021] 全文検索データベース１０５は、後述のサイト情報データベース１０７に

格納されたウェブページの内容に基づき、単語とページのメタ情報(URL,タイトル,検索順位等)を紐付けたデータベースである。全文検索データベース１０５に登録される情報の一例については、図５を参照して後述する。検索サーバ１０３は、クライアント１０１から受信したユーザ指定の検索語を用いて、全文検索データベース１０５を利用して検索を行う。画像データベース１０６は、サイト情報データベース１０７に登録されている文書ファイルを変換して得られた画像データを格納するデータベースである。画像データベース１０６に登録される情報の一例については、図６を参照して後述する。

[0022] サイト情報データベース１０７は、ウェブサイトをインターネット１０２に公開するための文書ファイルの情報を管理する。文書ファイルには、ＨＴＭＬファイル、ＰＤＦファイルの他、特定アプリケーションで利用されるファイル（例えば、ＭＳワードなど）が含まれる。これらの文書ファイルには、ウェブサイトのオーナー及びオーナーが提供する商品などに関する情報を含むことができる。例えば、ウェブサイトが法人のホームページの場合、会社概要や投資家向け情報等の他に、法人が製造・販売している製品の情報や、その他法人が提供するあらゆるサービスに関する情報がサイト情報データベース１０７に格納される。例えば、法人が提供するサービスがブログ提供サービスの場合、当該サービスの提供を受けるユーザが作成したブログ情報等も含まれる。サイト情報データベース１０７は、随時更新され最新の情報が格納される。

[0023] 尚、本明細書では、便宜上、検索サーバ１０３、ウェブサーバ１０４、全文検索データベース１０５、画像データベース１０６及びサイト情報データベース１０７は、それぞれ物理的に独立した情報処理装置によって実現されるものとして説明するが、本発明の実施形態はこれに限られるものではない。例えば、これらが単一の情報処理装置によって実現されてもよい。その一方で、検索サーバ１０３等の各装置が複数台の情報処理装置により冗長構成、或いは分散構成されても良い。また、全文検索データベース１０５及び画像データベース１０６は、検索サーバ１０３とＬＡＮ等によって接続されて

いるものとして説明するが、例えば、インターネット１０２や非図示のイントラネット経由で検索サーバ１０３と通信可能な形態をとってもよい。ウェブサーバ１０４とサイト情報データベース１０７との関係についても同様である。

[0024] また本実施形態では、説明の簡単のために、検索サーバ１０３とウェブサーバ１０４とで検索とサイトの公開を個別に実行する場合を記載したが、システムの構成はこれに限定されるものではない。即ち、サイト公開のためのサーバと、検索のためのサーバを共通に設けることもできる。その際、検索の形態は、インターネット検索であってもよいし、自サーバで公開するサイトの情報を検索するサイト内検索であってもよい。なお、ウェブサイトとは、ひとまとまりに公開されているウェブページ群のことを言う。ウェブサイトには、例えば、法人や個人のホームページ等がある。

[0025] ＜情報処理装置の構成＞

次に、本実施形態に対応する検索システムを構成する情報処理装置の概略について説明する。図２は、クライアント１０１のハードウェア構成の一例を示すブロック図である。上述した情報処理装置としての全文検索データベース１０５、画像データベース１０６及びサイト情報データベース１０７も、同様或いは同等のハードウェア構成として構成しても良い。

[0026] 図２において、ＣＰＵ２００は、ハードディスク装置（以下、ＨＤと呼ぶ）２０５に格納されているアプリケーションプログラム、オペレーティングシステム（ＯＳ）や制御プログラム等を実行し、ＲＡＭ２０２にプログラムの実行に必要な情報、ファイル等を一時的に格納する制御を行う。なお、後述する図７ａから図７ｈにおける処理もＣＰＵ２００が対応する処理プログラムを実行することにより装置全体を制御して実現される。

[0027] ＲＯＭ２０１は、内部に基本Ｉ／Ｏプログラム等のプログラム、文書処理の際に使用するフォントデータ、テンプレート用データ等の各種データを記憶する。ＲＡＭ２０２は各種データを一時記憶し、ＣＰＵ２００の主メモリ、ワークエリア等として機能する。

[0028] 外部記憶ドライブ２０３は、記録媒体へのアクセスを実現するための外部記憶ドライブであり、メディア（記録媒体）２０４に記憶されたプログラム等を本コンピュータシステムにロードすることができる。尚、メディア２０４は、例えば、フロッピー（登録商標）ディスク（ＦＤ）、ＣＤ－ＲＯＭ、ＣＤ－Ｒ、ＣＤ－ＲＷ、ＰＣカード、ＤＶＤ、Ｂｌｕ－ｒａｙ、ＩＣメモリカード、ＭＯ、メモリスティック等を利用することができる。

[0029] 外部記憶装置２０５は、本実施形態では大容量メモリとして機能するＨＤ（ハードディスク）を用いている。ＨＤ２０５には、アプリケーションプログラム、ＯＳ、制御プログラム、関連プログラム等が格納される。なお、ハードディスクの代わりに、フラッシュ（登録商標）メモリ等の不揮発性記憶装置を用いても良い。

[0030] 指示入力装置２０６は、キーボードやポインティングデバイス（マウス等）、タッチパネル等がこれに相当する。指示入力装置２０６を用いて、ユーザは、クライアント１０１に対して、装置を制御するコマンド等を入力指示する。ディスプレイ２０７は、指示入力装置２０６から入力したコマンドや、それに対するクライアント１０１の応答出力等を表示したりするものである。システムバス２０９は、情報処理装置内のデータの流れを司る。インターフェイス（以下、Ｉ／Ｆという）２０８は、外部装置とのデータのやり取りを仲介する役割を果たす。

[0031] 尚、以上の各装置と同等の機能を実現するソフトウェアにより、ハードウェア装置の代替として構成することもできる。

[0032] 本実施形態に対応する処理を実行するために対応するプログラムを動作させる度に、既にプログラムがインストールされているＨＤ２０５からＲＡＭ２０２にロードするようにしてもよい。また、本実施形態に係るプログラムをＲＯＭ２０１に記録しておき、これをメモリマップの一部をなすように構成し、直接ＣＰＵ２００で実行することも可能である。さらに、メディア２０４から対応プログラム及び関連データを直接ＲＡＭ２０２にロードして実行させることもできる。

[0033] 図３は、検索サーバ１０３及びウェブサーバ１０４の機器構成を示すブロック図である。ＣＰＵ２００、ＲＯＭ２０１、ＲＡＭ２０２、外部記憶ドライブ２０３、メディア２０４、ＨＤ２０５、指示入力装置２０６、ディスプレイ２０７、Ｉ／Ｆ２０８、システムバス２０９の機能及び用途、更には、これらの関係は、図２を用いて説明したものと同様或いは同等である。図３では、データベース３１０がシステムバス２０９に接続されていることに留意する。ここで、データベース３１０は、全文検索データベース１０５及び画像データベース１０６、又は、サイト情報データベース１０７を総称するものである。

[0034] ＜検索用データの生成＞

次に、発明の実施形態に対応する検索用データの生成処理について、図４のフローチャートと図５及び図６のテーブルとを参照して説明する。図４は、発明の実施形態に対応する、検索サーバ１０３における検索用のテキストデータと画像データを生成する処理の一例を示すフローチャートである。図４に対応する処理は、検索サーバ１０３がＨＤ３０５やデータベース３１０に保持する処理プログラムをＣＰＵ３００で実行することで実現される。図５は、図４のフローチャートに対応する処理により生成された、全文検索データベース１０５におけるテキスト検索用テーブルのデータ構成の一例を示す図である。図６は、図４のフローチャートに従った処理により生成される画像データベース１０６の画像登録テーブルのデータ構成の一例を示す図である。

[0035] 図４において、ステップＳ４０１では、ウェブサーバ１０４からインターネット１０２を介してサイト情報データベース１０７に格納されている処理対象のデータを取得する。ここで取得されるデータには、ＨＴＭＬ形式のページデータの他、ＰＤＦファイルや、ワードファイル、エクセルファイル、パワーポイントファイルといった文書ファイルも含まれる。

[0036] 続くステップＳ４０２では取得したデータから全文のテキスト情報を抽出する。ここで抽出されるデータは、文書ファイルの本文データだけでなく、

ＨＴＭＬ形式のページであればタグに含まれる情報も抽出対象となる。例えば、以下のようにページ内にタイトルやテーブル、画像のタグ、ページの情報を定義するメタタグが含まれる場合、引用符で囲まれた当該タグに付随するテキスト情報も抽出対象となる。

＜a title=”レース結果はこちら” href=”...”＞...＜／a＞

＜table summary=”2012年度レース結果”＞

＜img alt=”ABCロゴ” src=”...” ／＞

＜title＞Hondaホームページ：ABC工業株式会社＜／title＞

＜meta name=”keywords” content=”ABCホームページ,ABC,ABC工業株式会社,ABC工業”＞

＜meta name=”description” content=”ABCウェブサイトへようこそ。このサイトはABC工業株式会社のオフィシャルサイトです。”＞

[0037] また、ステップＳ４０２では、ＰＤＦファイルや、ワードファイル、エクセルファイル、パワーポイントファイルといったＨＴＭＬ以外のファイル形式のデータをＨＴＭＬ形式のファイル（ウェブページ）に変換する。

[0038] ステップＳ４０３では、ステップＳ４０２において抽出したテキスト情報を、ページ単位の情報に分割して、文書ファイルのＵＲＬと関連づけて全文検索データベース１０５内に登録する。その際、ＰＤＦファイル等を変換して得られたＨＴＭＬファイルも登録しておく。なお、本実施形態では、全文検索データベース１０５内に登録することとしたが、別途データベースを用意してもよい。

[0039] 全文検索データベース１０５におけるテキスト情報及びＨＴＭＬファイルの登録の様子は、図５のテキスト検索用テーブル５００に示す通りである。テキスト検索用テーブル５００において、取得元ＵＲＬ５０１には、文書ファイルを取得したＵＲＬが登録される。例えば「maindocument.html」や「letter.pdf」のような文書ファイルにつき、「http://www.abc.com/maindocument.html」や「http://www.abc.com/letter.pdf」 といったＵＲＬが取得元ＵＲＬとして登録される。

[0040] テキスト５０２には、取得元ＵＲＬの文書ファイルに対応するテキストデータが登録される。その際、文書ファイルに対しページ単位のアクセスが可能な場合にはファイル全体とページ単位のテキストデータとの両方を登録する。例えば、「letter.pdf」のファイルが５０ページで構成される場合、ファイル全体は「http://www.abc.com/letter.pdf」で指定されるが、当該ファイルうちの特定のページを表示させる場合には「http://www.abc.com/letter.pdf#page=N」で表示ページを指定することができる。ここでＮはページ番号を示し、「letter.pdf」の場合、N=1,2,3,...50となる。「letter.pdf」の場合にテキスト５０２には、文書ファイル全体のテキストファイルであるletter.txtと、各ページのテキストファイルであるletterN.txt (N=001,002,...050) が登録される。なお、「http://www.abc.com/maindocument.html」のような、元々がＨＴＭＬフォーマットのファイルは、ページ単位に分割せずに全体として一つのテキストファイル(document.txt)のみがテキスト５０２に登録される。

[0041] このテーブル５００を用いることで、ある文書ファイルと関連づけて全文のテキスト情報と、ページ単位のテキスト情報とを管理することが可能となる。

[0042] また、テーブル５００には、ＰＤＦファイル等の非ＨＴＭＬ形式の文書ファイルを変換して得られたＨＴＭＬファイルも登録される。当該ＨＴＭＬファイルは、後述するプレビュー情報としてクライアント１０１へ提供される。格納先ＵＲＬ５０３はソースとなったファイル単位に管理され、例えば「letter.pdf」を変換して得られたＨＴＭＬファイルについては、ＵＲＬとして「http://www.search.com/HTML/letter/letter.html」が与えられる。また本実施形態では、元のファイルから生成されたファイル全体のＨＴＭＬファイルと、各ページに対応するＨＴＭＬファイルとがそれぞれ登録される。例えば、「letter.pdf」のファイル全体については、「letter.html」が登録され、「letter.pdf」の途中のページについては、例えば２ページ目であれば「letter002.html」が登録される。この格納先ＵＲＬ５０３の情報は、クライア

ント１０１に対してプレビュー情報を提供する際のリンク情報として検索結果で使用することができる。なお、ＨＴＭＬファイルとして検索サーバ１０３側に登録されるのは、元々ＨＴＭＬ形式のファイルでなかったものなので、「maindocument.html」のようにＨＴＭＬ形式でウェブサーバ１０４から提供されるものについては、検索サーバ側でＨＴＭＬファイルとして管理はしなくてよい。なお、非ＨＴＭＬ形式の文書ファイルからＨＴＭＬファイルへの変換に際して、後述の図７ｈに示すプレビュー情報表示状態における処理をクライアント１０１に実行させるためのスクリプトをＨＴＭＬファイル内に記述しておく。

[0043] 図４のフローチャートの説明に戻り、続くステップＳ４０４では、ステップＳ４０１で取得した文書ファイルの各ページの画像データを生成する。具体的には、文書ファイルを表示するためのアプリケーションを用いて、ページ単位にディスプレイ表示可能な情報としてＲＡＭ３０２上に展開したのち、該ＲＡＭ３０２上の情報を画像データに変換する。画像データの形式は、例えばＪＰＥＧとすることができるが、これに限定されるものではない。また、文書ファイルの画像データへの変換方法は、上述の方法に限定されるものではなく、公知の他の技術を利用して実行することができる。また、ステップＳ４０４では、プレビュー用の画像データ（プレビュー画像）と検索結果の一覧表示用のデータ（サムネイル画像）との解像度の異なる複数の画像データを生成する。

[0044] 次に、ステップＳ４０５において、文書ファイルを構成する各ページのうち、代表的なページを決定する。代表ページは、例えば常に１ページ目を選択してもよいし、或いは、ページの中でもっても色数の多いページや、ページ毎に周波数解析を行い高周波数成分に高い係数を持つページを選択してもよい。

[0045] 続くステップＳ４０６では、ステップＳ４０４で生成した画像データを、文書ファイルのＵＲＬと関連づけて画像データベース１０６に登録する。このとき生成される画像データベース１０６内の画像登録テーブルの一例は、

図６に示す通りである。

[0046] 図６において、画像登録テーブル６００は、取得元ＵＲＬ６０１に、画像データに対応するＵＲＬが登録される。登録されるＵＲＬは、図５のテーブル５００と同様文書ファイルのＵＲＬが登録され、テーブル５００における取得元ＵＲＬ５０１の登録内容と対応している。

[0047] 次にページ画像６０２には、文書ファイルを変換して得られたページ毎の画像データが登録される。当該画像データは、検索結果の一覧において選択された場合に、文書ファイルの代わりにクライアント１０１のディスプレイ２０７に表示される画像である。画像データのファイル名は、例えば、図５のページテキスト５０４に対応したものとすることができる。ページ画像６０２は取得元ＵＲＬ６０１と関連づけられているので、テキスト検索用テーブル５００における取得元ＵＲＬ５０１に基づき、テキストと画像とを一対一で対応づけることができる。

[0048] また、画像登録テーブル６００において、サムネイル画像６０３には、ページ画像６０２に登録されたページ毎の画像データの縮小画像が、対応するＵＲＬと関連づけられて登録される。該サムネイル画像は検索結果の一覧表示の際に検索結果のイメージを提供するために利用される。代表フラグ６０４は、文書ファイル全体を代表するページ画像とサムネイル画像を指定するためのフラグであって、１つの文書ファイルについて１対のページ画像とサムネイル画像について「１」の値が与えられ、他の画像は「０」となる。例えば、「http://www.abc.com/letter.pdf」の取得元ＵＲＬの文書ファイルについては、「letter001.jpg」と「letterS001.jpg」がそれぞれ代表画像となる。なお、「maindocument.html」のようにページ単位に画像が生成されないものについても生成された画像にフラグ値を設定しておいてよい。また、代表ページとして、例えば常に１ページ目を選択する場合などは代表フラグの項目そのものを設けなくてもよい。

[0049] 次に格納先ＵＲＬ６０５は、ページ画像及びサムネイル画像の格納先のＵＲＬを示す。各ページ画像及びサムネイル画像はソースとなったファイル単

位に管理され、例えば「letter.pdf」の画像については、ＵＲＬとして「http://www.search.com/JPGL/letter/」が与えられる。ページ画像６０２又はサムネイル画像６０３のファイル名と格納先ＵＲＬ６０５のＵＲＬとを組み合わせることにより、各画像を一意に特定するリンク情報として検索結果で使用することができる。以上により、所望のＵＲＬに格納されている文書ファイルについて、検索用のテキスト情報と画像データとを生成することができる。

[0050] 次に、図１乃至図６に示した構成に基づいて、本発明の検索システムで実行される検索処理の詳細を説明する。

[0051] 図７ａから図７ｊは、発明の実施形態に対応する検索処理の一例をSDL(Specification and Description Language）の記述法に従って示す状態遷移図である。図７ａから図７ｊに対応する処理は、クライアント１０１、検索サーバ１０３及びウェブサーバ１０４がそれぞれ保持する処理プログラムを各々が実行することにより実現される。また、クライアント１０１、検索サーバ１０３、ウェブサーバ１０４間の通信は、クライアント１０１において実行されるウェブブラウザの通信機能、或いは、ウェブブラウザのプラグイン(拡張プログラム)の通信機能を利用して実現することができる。例えばＪａｖａ（登録商標）Scriptを用いたHTTPプロトコルに従って実現することができる。その他に、Flash等を使ってもよいし、HTTPではない別のプロトコルに従った通信でもよい。

[0052] まず、図７ａは、クライアント１０１、検索サーバ１０３及びウェブサーバ１０４の状態遷移を示す図である。図７ａにおいて、クライアント１０１は、ステップＳ７１０からステップＳ７８０までの状態を取り、それぞれの状態において図７ｂ乃至図７ｈに示す処理が行われる。また、検索サーバ１０３はクライアント１０１からの要求待ち状態にあり、クライアント１０１から受信した要求に応じて、後述する図７ｉに示すような処理を実行する。ウェブサーバ１０４はクライアント１０１からの要求待ち状態にあり、クライアント１０１から受信した要求に応じて、後述する図７ｊに示すような処

理を実行する。

[0053] まず、クライアント１０１は、ステップＳ７１０において初期状態にある。この初期状態では、クライアント１０１のディスプレイ２０７には検索画面が表示されておらず、図７ｂに示すように、ステップＳ７１１において検索サーバ１０３に対して検索画面表示要求を送信する。その後、ステップＳ７２０の検索画面表示待ち状態に移行する。

[0054] ステップＳ７２０の検索画面表示待ち状態では、図７ｃに示すように、クライアント１０１は、ステップＳ７２１において検索サーバ１０３から検索画面情報を受信し、ステップＳ７２２においてディスプレイ２０７に検索画面を表示し、ステップＳ７３０における検索画面表示状態に移行する。

[0055] 検索画面の一例は図９に示す通りである。図９において検索画面９００には、検索語入力領域９０１と、検索実行ボタン９０２が含まれる。検索語入力領域９０１に検索語を構成する文字が漢字、平仮名、片仮名、アルファベット、記号、数字、或いは、合成語により入力することができる。図９では検索語として「ＡＢＣ」が入力されている。

[0056] ステップＳ７３０の検索画面表示状態では、図７ｄに示すように、クライアント１０１はステップＳ７３１においてユーザから検索語の入力を受け付ける。続くステップＳ７３２では図９の検索実行ボタン９０２の操作に基づくユーザからの検索実行指示を受け付け、検索語を指定した検索実行指示を検索サーバ１０３に送信する。その後、ステップＳ７４０の検索結果受信待ち状態に移行する。

[0057] ステップＳ７４０の検索語結果受信待ち状態では、図７ｅに示すように、ステップＳ７４１において検索結果を検索サーバ１０３から受信する。続いて、ステップＳ７４２において受信した検索結果をディスプレイ２０７に表示して、ステップＳ７５０の検索結果表示状態に移行する。検索結果の表示状態の一例は図１０Ａに示すとおりである。図１０Ａにおいて、ウィンドウ１０００は、ウェブブラウザの表示画面全体を示している。ウェブブラウザの構成は、一般的であるので特に説明しないが、例えば、インターネット・

エクスプローラ、クローム、ファイヤーフォックス、サファリなどの各種ウェブブラウザを利用することができる。

[0058] 検索結果はリンク毎に区別して表示され、与えられた領域毎に検索結果の本体のＵＲＬに接続するためのリンク１００１、リンク１００１の接続先のＵＲＬを示すＵＲＬ表示１００２、ページ画像にリンクしたサムネイル画像１００３を含む。なお、ＵＲＬ表示１００２は、ページ単位にアクセスが可能なファイル、例えばＰＤＦファイルについては該当ページを示す内容であってもよい。たとえば、５０ページからなる「letter.pdf」の４ページ目が検索結果に含まれる場合、４ページ目を指定するＵＲＬ「http://www.abc.com/letter.pdf#page=4」がＵＲＬ表示１００２に表示される。この場合、サムネイル画像１００３には当該ページに対応する画像（letterS004.jpg）が表示される。また、サムネイル画像がページ単位に用意できたものであっても、文書ファイルを扱うアプリケーションに基づく制約によりＵＲＬでページを指定できない文書ファイル（例えば、ＭＳワードなど）については、ＵＲＬ表示１００２はファイル全体に対するリンクとし、サムネイル画像１００３は該当ページのサムネイル画像を表示する。なお、この場合でもプレビュー画像やプレビュー情報もページ単位で提供が可能である。その他、関連箇所の要約テキストを表示することもできるが、図１０Ａでは図面の簡単のために省略している。なお、リンク先の文書の画像が画像登録テーブル６００に登録されていない場合には、サムネイル画像１００３は表示されないか、代替となる画像が表示される。

[0059] ステップＳ７５０の検索結果表示状態では、図７ｆに示すように、ステップＳ７５１において検索結果でリスト表示された項目のうち、プレビュー表示を行う項目の選択をユーザから受け付ける。当該項目の選択は例えば図１０Ａの結果表示の例において、サムネイル画像１００３上にカーソルを移動させるか、或いは、クリックすること等により行うことができる。なお、サムネイル画像はクリックを受け付けて選択項目の文書ファイル（ウェブページ）のＵＲＬへ接続したり、カーソルが上に位置した場合にプレビュー画像

やプレビュー情報を取得するためのリンク画像として利用されるので、これらに対するリンク情報が埋め込まれている。続いてステップＳ７５２において、選択された項目を指定してプレビュー画像の要求を検索サーバ１０３に送信する。さらにステップＳ７５３において、埋め込まれたリンクの接続先のＵＲＬに基づき選択された文書ファイル（ウェブページ）の要求を検索サーバ１０３又はウェブサーバ１０４に対して送信する。検索サーバ１０３に対してウェブページの要求が送信される場合は、図５で格納先ＵＲＬ５０３に登録されるような、ＰＤＦファイル等の非ＨＴＭＬファイルをＨＴＭＬファイルに変換して検索サーバ１０３で保存している場合である。一方、元々ＨＴＭＬファイルとしてウェブサーバ１０４が保有しているデータ（例えば、maindocument.html等）についてはウェブサーバ１０４に対してウェブページ要求が送信される。その後、ステップＳ７６０のプレビュー情報受信待ち状態に移行する。

[0060] ステップＳ７６０のプレビュー情報受信待ち状態では、図７ｇに示すように、ステップＳ７６１においてまずはプレビュー画像を検索サーバ１０３から受信する。続いて、ステップＳ７６２において受信したプレビュー画像をプレビュー領域に表示する。このときの表示状態は、例えば図１０Ｂのようになる。図１０Ｂにおいて、ページ画像１０１１を表示するプレビュー領域１０１０が検索結果を表示するウィンドウ１０００に対して重畳表示される。本実施形態では先にページ画像を送信して表示内容の確認を可能とする。

[0061] 次にステップＳ７６３においてプレビュー情報としてウェブページのページデータを検索サーバ１０３又はウェブサーバ１０４から受信する。続いて、ステップＳ７６４において受信したウェブページを表示済のページ画像１０１１と置き換えてプレビュー領域１０１０を更新し、ステップＳ７７０のプレビュー情報表示状態となる。このときの表示状態は、表示された情報が画像からウェブページに変わっただけで外観としては例えば図１０Ｂと同様になる。

[0062] 次にステップＳ７７０のプレビュー情報表示状態では図７ｈに示すように

検索語に基づきウェブページをプレビュー表示用の情報に加工してプレビュー情報の表示を行う。具体的に、ステップＳ７７１では、検索語を取得する。当該検索語は、ブラウザが内部変数として保持するリファラを参照して抽出することができる。当該リファラには、現在表示されているウェブページのリンク元のページの情報が含まれ、当該リンク元が検索結果の画面である場合には、検索に使用された検索語が含まれるからである。

[0063] 次にステップＳ７７２では、表示中のウェブページのうち、プレビュー領域に表示されるべき本文（テキスト）中の検索語を探索し、見つかった検索語にタグを埋め込み分離する。このとき埋め込むタグは例えば、＜ｓｐａｎ＞＜／ｓｐａｎ＞があり、これにより検索語の表示位置を特定することができる。また、例えば、本文が以下の文１で、検索語をＡＢＣとすると、埋め込みの例は文２に示すようになる。

「ＡＢＣのクルマを体感した、みなさまからご投稿いただいた感想の一部をご紹介。」・・・（文１）

「＜ｓｐａｎ＞ＡＢＣ＜／ｓｐａｎ＞のクルマを体感した、みなさまからご投稿いただいた感想の一部をご紹介。」・・・（文２）

[0064] このようにして分離した検索語に対して更に強調表示要素（例えば、吹き出し要素）を付加するためにタグ＜ｅｍ＞＜／ｅｍ＞を追加する。文２に対するタグの付加の例は文３に示すようになる。

「＜ｓｐａｎ＞ＡＢＣ＜／ｓｐａｎ＞＜ｅｍ＞ＡＢＣ＜／ｅｍ＞のクルマを体感した、みなさまからご投稿いただいた感想の一部をご紹介。・・・（文３）

上記の処理は本文中で見つかった検索語の全てについて行う。次にステップＳ７７３に移行して、表示要素の属性値について検索語の探索を行う。表示要素の属性値とは、例えばテーブル、画像などに付与された属性（メタデータ）をいう。例えば、画像であれば以下の文４のような例がある。

[0065] ＜img alt=”ABCロゴ” src=”...” ／＞ （文４）

これに対して、強調表示要素を付加すると文５に示すようになる。

＜img alt="ABCロゴ" src="..." ／＞＜ｅｍ＞ＡＢＣ＜／ｅｍ＞ ・・・（文５）

Ｓ７７３の上記の処理は表示要素の属性値について見つかった検索語の全てについて行う。当該表示要素の強調表示は、テキスト中の検索語に付与される吹き出し要素とは外観を異ならせてもよい。例えば、表示要素の強調表示要素は検索語を含めず所定のマーク（例えば、フラグ）を付与するだけでもよい。

[0066] 次に、ステップＳ７７４に移行して非表示要素の属性値について検索語の探索を行う。非表示要素の属性値とは、例えば以下に示すような、ページタイトルやページに付与されたメタタグの情報を言う。

＜title＞ＡＢＣホームページ：ＡＢＣ工業株式会社＜／title＞

＜meta name="keywords" content="ABCホームページ,ABC,ABC工業株式会社,ABC工業"＞

＜meta name="description" content="ABCウェブサイトへようこそ。このサイトはABC工業株式会社のオフィシャルサイトです。"＞

[0067] 上記のメタタグの情報は、クライアント１０１におけるページ表示においては視覚的に表示されることはないが検索語の探索範囲とはなっているので、本実施形態では、その存在をユーザに知らしめるために強調表示を行う。具体的には、例えば＜ｅｍ＞非表示エリアにマッチ＜／ｅｍ＞のようなタグを付与する。当該非表示要素の強調表示は、特定の要素との関連で表示されるわけではないので、非表示領域の検索語がヒットしていることをユーザに通知するための表示として、表示領域内の特定の位置に吹き出し等を表示させてもよい。

[0068] 次に、ステップＳ７７５に移行して、Ｓ７７２からＳ７７４までの処理結果に基づいて表示されているページの内容を更新し、その後Ｓ７８０の操作待ち状態に移行する。操作待ち状態においては、クライアント１０１が表示するプレビュー領域に対するユーザの操作を受け付け、当該操作に応じた表示を行う。

[0069] また、ステップＳ７７５におけるページ更新の際には、強調表示の表示態様を決定するためにスタイルシートを参照、或いは適用する。当該スタイルシートは、追加した強調表示の表示形態を定義するための情報であって、例えば、以下の情報を定義しておく。

・強調表示用に字体や色（テキスト、表示要素、非表示要素につき定義）

・強調表示の形状レイアウト（テキスト、表示要素、非表示要素につき定義）

・強調表示の表示／非表示（表示時間、表示消去の条件）

・オーバーレイの為の自由な座標位置の指定

・オリジナルの文書より手前に表示する指定

[0070] なお、スタイルシートは、当該処理を実行するためのスクリプトによりウェブページ内へ挿入されたものを参照するか、各タグにstyle属性の値としてスタイルを指定するか、或いはスクリプトにより直接適用されてもよい。ウェブページ内へ挿入される場合は、＜LINK＞タグで拡張子が 「．ｃｓｓ」の外部スタイルシートファイルを読み込むような記載がなされているか、或いは＜HEAD＞～＜／HEAD＞内等に＜STYLE＞タグを用いてスタイルシートを直接に記述されていてもよい。

[0071] なお、図７ｈに示すプレビュー情報表示状態における処理は、例えばウェブサーバ１０４から提供されるウェブページ内に当該処理を実行するためのスクリプトが記述されており、クライアント１０１においてブラウザが当該スクリプトを実行することにより実現することができる。また、ウェブページ内に当該スクリプトが埋め込まれていない場合であっても、ブラウザが図７ｈに対応する処理を実行する機能を有していれば、同様に実現することができる。

[0072] 以上により、ページ中に検索語の位置を示す強調表示が含まれたプレビュー情報が表示される。図１１は以上の処理に従い強調表示を行うためにクライアント１０１により加工されたウェブページのソースファイルの一例を示す。図１１において、点線で囲んだ領域１１０１は非表示要素のうちページ

タイトルを示す。同じく点線領域１１０２はページにスタイルシートを挿入した場合の一例を示す。点線領域１１０３は非表示領域のうちメタタグの情報を示し、非表示領域に対応する吹き出し要素が点線１１０４であり、ページ本体(＜ｂｏｄｙ＞タグ以降)に挿入されている。点線領域１１０５は表示要素である画像に対して加工を行った結果を示している。点線領域１１０６は本文中のテキストに対して加工を行った結果を示している。

[0073] 図１０Ｃは、ページ中に検索語の位置を示す強調表示が含まれたプレビュー情報の表示例を示す図である。ここで、検索語の強調表示を含むプレビュー領域１０２０が検索結果を表示するウィンドウ１０００に対して重畳表示される。なお、プレビュー情報は検索語及びプレビュー要求されたウェブページの内容に応じてクライアント１０１が生成し、表示されているオリジナルのウェブページの表示内容から更新していってもよい。なお、本実施形態ではウェブページのウェブサーバ１０４からの取得に時間がかかってプレビュー領域に空白が生じないように先にページ画像を表示しておき、ウェブページが取得できたらページ画像と置き換える。

[0074] これにより、ウェブページの取得やプレビュー情報生成を待機することによるユーザのフラストレーションを未然に防止できる。また、ユーザはプレビュー画像の内容により、当該ページが自分が探しているページかどうかを判断できるので、仮に不要なページであればプレビュー情報の受信を待たずに次の検索結果を選択することができる。なお、プレビュー画像の表示は、通信速度やクライアント１０１の処理速度に応じて省略されてもよい。これは、通信速度及び処理速度が十分に確保された環境では、プレビュー画像を表示せずともユーザにストレスを与えない状況でウェブページをプレビュー領域１０１０に表示させることが可能だからである。

[0075] 図１０Ｃでは、ウェブページ内の検索語に該当する部分に吹き出し１０２１による強調表示がなされる。図１０Ｃでは、吹き出し１０２１ａ及び１０２１ｂの２つの吹き出しが表示されている。いずれも検索語である「ＡＢＣ」に対応する文字（テキスト）に対して表示されたものである。次にウェブ

ページにおいて表示要素の属性値に検索語に該当する語が含まれる場合に、当該表示要素と関連づけてフラグ表示１０２２による強調表示がなされる。図１０Ｃではフラグ表示１０２２ａから１０２２ｃまでの３つが表示されている。これらのフラグ表示は、表示要素の１つである画像に対して付されたものであり、このような表示は画像のファイル名や、タグ内の引用符で囲まれた情報に該当する語が存在する場合に行われる。また「非表示エリアにマッチ」との表示１０２３は、ウェブページを表示した際に表示情報として表には現れないメタタグや、タイトルに検索語とマッチする語が含まれる場合に表示される。当該表示１０２３によれば、仮に本文中に検索語に該当する語が存在しない場合でも当該ページが検索結果に含まれることとなった理由を直ちに把握できるので、ユーザが当該ページがヒットした理由を探すために無駄な時間を費やすことが無くなる。

[0076] なお、上記の強調表示１０２１から１０２３はウェブページの表示が完了した後、所定時間経過の後に自動的に消えてもよいし、或いは、ユーザの手動による吹き出しの選択操作により消えてもよいし、さらには、吹き出しの表示／非表示ボタンを設けて当該ボタンにより表示の有無を操作可能としてもよい。

[0077] ＵＩ１０２４は、表示形態を変更するためのインタフェースである。図７ｈの操作待ち状態Ｓ７８０では、当該ＵＩ１０２４を介してユーザからの操作を受け付けることができる。ＵＩ１０２４内には検索語１０２４ｂ（図１０Ｃの場合は「ＡＢＣ」）が表示され、「＜」ボタン１０２４ａと「＞」ボタン１０２４ｃが含まれている。当該ボタンはプレビュー表示における強調表示を順に選択して行くための操作ボタンである。ボタン１０２４ａは逆方向、ボタン１０２４ｃは順方向にそれぞれ表示１０２１から１０２３を選択するためのボタンである。ボタン１０２４ｄはプレビュー領域１０２０の表示サイズを全画面表示に切り替えるためのボタンである。ボタン１０２４ｅはプレビュー領域１０２０の表示を消すためのボタンである。当該ボタン１０２４ｅが操作されるとプレビュー領域１０２０の表示が消えて、図１０Ａ

の元の検索結果のリストのみの表示状態に戻る。

[0078] 次に、検索サーバ１０３側の処理を説明する。図７ｉのステップＳ７９０において、検索サーバ１０３は当初要求待ち状態に置かれている。この要求待ち状態において検索サーバ１０３は、クライアント１０１からステップＳ７１１で送信された検索画面表示要求、ステップＳ７３２で送信された検索実行指示、及びステップＳ７５２で送信されたプレビュー画像要求のいずれかを受信する。

[0079] ステップＳ７９１において検索画面表示要求を受信すると、ステップＳ７９２において検索画面情報をクライアント１０１へ送信する。その後、ステップＳ７９０の要求待ち状態に戻る。さらに、ステップＳ７９３において検索実行指示を受信すると、ステップＳ７９４に移行して検索処理を実行する。この検索処理は、検索実行指示に含まれる検索語に基づき、全文検索データベース１０５を用いて行われる。続くステップＳ７９５では検索結果をクライアント１０１へ送信して、ステップＳ７９０の要求待ち状態に戻る。ステップＳ７９６においてプレビュー画像要求を受信すると、ステップＳ７９７に移行してプレビュー画像を送信する。当該プレビュー画像は、プレビュー画像要求に含まれるＵＲＬ等の文書を特定するための情報に基づき、画像データベース１０６の画像登録テーブル６００を参照し、取得したＵＲＬに対応するページ画像６０２を選択することで得られる。その後、ステップＳ７９０の要求待ち状態に戻る。ステップＳ７９８においてウェブページ要求を受信すると、ステップＳ７９９に移行してウェブページを送信する。当該ウェブページは、ウェブページ要求に含まれるＵＲＬ等のＨＴＭＬファイルを特定するための情報に基づき、全文検索データベース１０５のテーブル５００を参照し、取得したＵＲＬに対応するウェブページを選択することで得られる。その後、ステップＳ７９０の要求待ち状態に戻る。

[0080] 次に、図７ｉの検索サーバ１０３における処理につき、詳細な処理の内容を説明する。まず、図８は、図７ｉのステップＳ７９４における検索処理の詳細を示すフローチャートである。

[0081] ステップＳ８０１では、クライアント１０１から受信した検索実行指示から抽出した検索語に基づいて全文検索データベース１０５を検索する。このとき、ＰＤＦファイル等のページ分割可能な文書ファイルを検索対象とする場合には、文書ファイル全体のテキスト（例：letter.txt）とページ毎のテキスト(例：letter001.txt)との全てを検索する。この検索の結果、特定のページが文書ファイルの中で有意と判定され、検索結果に含めるとの判定が成される場合もあるし、特定のページに絞り込めないが文書ファイル全体としては検索語を含んでいるので検索結果に含めるとの判定が成される場合がある。

[0082] 例えば、「ＡＢＣ」と「工業株式会社」のどちらをも含む条件として検索語が設定された場合に、あるページに「ＡＢＣ」、「工業株式会社」の語がどちらも含まれている場合には、該当ページを検索結果に含めるとの判定が成される。一方、「ＡＢＣ」、「工業株式会社」それぞれが別々のページに含まれている場合は、検索結果として単一ページを特定しない場合もあるが、少なくとも文書ファイル全体としては検索語に関連性があると判断され、単一ページを検索結果に含めるとの判定が成される場合がある。

[0083] 次にステップ８０２において検索結果の表示候補を表示順に選択する。ステップＳ８０３では、表示候補のリンク情報を生成する。ここで生成されるリンク情報とは、検索結果の本体のＵＲＬに接続するための情報である。ここで本体のＵＲＬは、基本的にウェブサーバ１０４のサイト情報データベース１０７に登録されているデータを特定するための情報である。例えば、表示候補がウェブサーバ１０４においてＵＲＬ「http://www.abc.com/letter.pdf」に格納されている「letter.pdf」の４ページ目の場合には、検索結果がクライアント１０１のディスプレイ２０７に表示されユーザのクリック操作が行われた場合に、ＵＲＬ「http://www.abc.com/letter.pdf」の４ページ目へ接続するためのリンク情報「http://www.abc.com/letter.pdf#page=4」等が生成される。なお、ここでページ単位にリンクを指定可能な場合はページを指定したリンク情報として生成するが、文書ファイルの形式によってはペ

ージを指定できないものもあるので、そのような文書ファイルについてはファイル全体を指定する。

[0084] 次にステップＳ８０４では、ステップＳ８０２でリンク情報として記載したＵＲＬに対応する画像データが、画像登録テーブル６００に登録されているか否かを判定する。例えば、表示候補が「http://www.abc.com/letter.pdf」の４ページ目の場合、当該ＵＲＬを取得元ＵＲＬ６０１として持つページ画像６０２及びサムネイル画像６０３が登録されているか否かが判定される。もし画像が登録されていると判定された場合（ステップＳ８０４において「ＹＥＳ」）、ステップＳ８０５に移行する。一方、画像が登録されていないと判定された場合（ステップＳ８０４において「ＮＯ」）、ステップＳ８０６に移行する。

[0085] ステップＳ８０５では、表示候補のリンク情報に画像データ表示のためのタグを追加する。例えば、表示候補が「＜a href="http://abc.com/letter.pdf"＞letter＜／a＞」の４ページ目の場合、ＵＲＬ「http://abc.com/letter.pdf」と関連づけて登録されているページ画像６０２には、「letter004.jpg」があり、サムネイル画像６０３には「letterS004.jpg」がある。そして、これらが画像データベース１０６において格納されている位置を特定するためのＵＲＬは、格納先ＵＲＬに基づき、「http://search.com/JPG/letter/letter004.jpg」と、「http://search.com/JPG/letter/letterS004.jpg」と特定することができる。このうち、サムネイル画像「letterS004.jpg」はクリックを受け付けてウェブサーバ１０４に対してウェブページを要求するためのリンク画像として利用される他、ページ画像「letter004.jpg」は選択時のプレビュー表示に利用される。よって一例としては、

＜a href="http://www.abc.com/letter.pdf#page=4"
onMouseOver="preview('http://search.com/JPG/letter/letter004.jpg', 'http://search.com/HTML/letter/letter004.html')"＞
＜img src="http://search.com/JPG/letter/letterS004.jpg"＞＜／a＞ が記載される。

[0086] この場合、文書ファイルの本体の格納場所と、プレビュー情報の元となるＨＴＭＬファイルの格納場所とが異なっている。これは、文書ファイルがＰＤＦファイルであるため、本体のＰＤＦファイルはウェブサーバ１０４で管理されているが、プレビュー情報の元となるＨＴＭＬファイルは検索サーバ１０３で管理されているためである。よって、検索結果の文書ファイルが元々ＨＴＭＬ形式のウェブページであった場合は、文書ファイルの本体の格納場所とプレビュー情報の元となるＨＴＭＬファイルの格納場所が一致することとなる。例えば、図５の「maindocument.html」の場合、上記記載は以下のようになる。

＜a href="http://www.abc.com/maindocument.html"
onMouseOver="preview('http://search.com/JPG/document/document.jpg', '
http://www.abc.com/maindocument.html')"＞
＜img src="http://search.com/JPG/document/documentS.jpg"＞＜／a＞

[0087] なお、表示候補のリンクが「＜a href="http://abc.com/letter.pdf"＞letter＜／a＞」のように、文書ファイル全体を示す場合には、代表ページとして指定されたページのページ画像とサムネイル画像が使用される。

[0088] 以上により、１つの表示候補について、本来のＵＲＬにリンクするためのリンク情報と、選択時のプレビュー表示動作とを含めた検索結果表示情報を生成することができる。但し、選択時のプレビュー表示動作に関しては、以上のようにHTML内へ直接記述する以外にも、別途スクリプトによるプログラムから行ってもよい。以上の処理は、検索結果に含まれる表示候補の全てについて行われ、全てについてリンク情報を生成した場合には検索結果表示情報が完成したこととなる。

[0089] よって、ステップＳ８０６では、未処理の表示候補があるか否かを判定し、未処理の表示候補がある場合にはステップＳ８０２に戻って処理を継続する。一方、全ての表示候補についてリンク情報を生成した場合には、ステップＳ８０７に移行する。ステップＳ８０７では検索結果表示においてサムネイルが選択された場合に、図１０Ｂ及びＣに示すようなプレビュー領域を表

示するためのプレビュー表示用のスクリプト（JavaScript）を検索結果のデータに含める。その後、ステップＳ７９４における処理を終了して、ステップＳ７９５に移行する。

[0090] なお、検索結果表示情報には、各表示候補の要約テキストが添付されてもよいが、係る要約テキストの添付方法は公知の技術であるので本明細書での説明は省略している。また、上述の説明では、サムネイル画像が存在する場合を説明したが、ページ画像のみがテーブル６００に登録されている場合であっても、サムネイル画像の代わりにページ画像のリンクをテキスト表示することで、同様の処理によりリンク情報を生成することができる。

[0091] 次に、ウェブサーバ１０４側の処理を説明する。図７ｊのステップＳ７００において、ウェブサーバ１０４は当初要求待ち状態に置かれている。この要求待ち状態においてウェブサーバ１０４は、クライアント１０１からステップＳ７５３で送信されたウェブページ要求を受信する。

[0092] ステップＳ７０１において検索画面表示要求を受信すると、ステップＳ７０２において要求されたウェブページをサイト情報データベース１０７から取得し、ステップＳ７０３でクライアント１０１へ送信する。その後、ステップＳ７００の要求待ち状態に戻る。なお、ここでウェブサーバ１０４が送信するウェブページには、図７ｈに示すプレビュー情報表示状態における処理をクライアント１０１が実行するためのスクリプトが記述されていてもよい。

[0093] 以上の本実施形態によれば、クライアント１０１のウェブブラウザ検索結果を表示した際に、検索結果のリンク先の文書ファイルを利用して、プレビュー表示において検索語を強調表示し、検索結果における検索語の所在を視覚的に直ちに把握することが可能となる。当該プレビュー表示では、コンテンツ制作者の意図とするＨＴＭＬ等における固有のレイアウト情報・位置関係において、検索語のありかを示す情報を提供することができるので、ユーザが直ちに該当文書が自分が本当に探している文書であるかどうかを効率的に判断することが可能となる。

[0094] また、本実施形態で生成されるプレビュー情報は、サイト情報データベース１０７に格納されたオリジナルの文書ファイルに対して検索語に関連する強調表示を行うための加工を施したものであって、当該強調表示以外の情報は全てオリジナルのままであるから、プレビュー情報が表示された図１０Ｃの状態で、ページ内を自由に閲覧することが可能であり、またページ内に含まれるリンク情報についても当該リンクを操作することによりリンク先の情報を取得することもできる。また、ページに埋め込まれたJavaScriptを実行することも可能であるし、更に添付された動画を再生することも可能である。また、プレビュー表示において、先頭の吹き出し表示の位置に画面が自動的にスクロールするようにしてもよい。

[0095] さらに、プレビュー領域は検索結果を表示するウィンドウ内で表示されるため、検索結果の１つのリンクの内容を確認する度にページの表示を切り替えたり、元の画面に戻ったりという作業が不要となる。

[0096] ［実施形態２］

以下、本発明の他の実施形態を説明する。ただし、この実施の形態に記載されている構成要素はあくまでも例示であり、この発明の範囲をそれらのみに限定する趣旨のものではない。

[0097] 本実施形態におけるシステム構成及びハードウェア構成は図１から図３に示したものと同様であるが、本実施形態では、検索結果における画像提供機能及び検索結果から選択された項目についてプレビュー領域を表示させる機能が省略されているものとする。よって、検索サーバ１０３において、画像データベース１０６は不要である。また、検索用データの生成処理方法は基本的には図４に示すものと同様であるが、検索結果に添付する画像については生成しない。よってＳ４０４からＳ４０６は省略される。また、図６のテーブルも不要である。

[0098] 検索処理の詳細について、図７ａから７ｊのうち図７ａのＳ７１０からＳ７５０までの処理は基本的に同じであるが、図７ｆの検索結果表示状態の処理においてはＳ７５２を省略する。また、Ｓ７６０、Ｓ７７０のプレビュー

表示について図１０Ｂや図１０Ｃのようなプレビュー領域１０１０、１０２０をウィンドウ１０００とは別に表示することはせず、ウィンドウ１０００内に直接プレビュー領域を表示する。従って図７ｇではＳ７６１、Ｓ７６２は省略する。また、図７ｉの処理についてはＳ７９６からＳ７９９は省略する。図７ｊの処理は同じである。次に図８の検索処理においては、Ｓ８０４、Ｓ８０５、Ｓ８０７を省略する。

[0099] このようにして、例えば図１０Ａの検索結果の表示（但し、サムネイル画像１００３の表示はなし）においてリンク１００１をクリックすると、図１０Ｄのような表示に切り替えることが可能となる。図１０Ｄでは、図１０Ｃにおいて検索結果を表示するウィンドウ内のプレビュー領域で表示されていた表示内容が、ウィンドウ１０００の表示領域内に画面１０３０として直接表示されている。また、ページタイトル１０３１は、当該ウェブページのタイトルが示され（ＡＢＣホームページ：ＡＢＣ工業株式会社）、ＵＲＬも当該ウェブページのもの（http://abc.com/main.document.html）が表示されている。当該画面１０３０を生成する処理は図７ｈと同様である。

[0100] このように検索結果から直接にページを表示した場合であっても、本願発明に特有の該当する検索語の吹き出し表示がなされており、ユーザは直ちにページ中の検索語のありかを確認することができる。

[0101] なお、以上の実施形態においてプレビュー領域の生成は、検索結果からリンクしてウェブページがクライアント１０１へ提供される場合にのみ実行されるように制限してもよい。例えば、ブラウザが内部変数として保持する上記のリファラを参照して、取得したページの前のページがどこから提供されているかを判定することができる。そこで、前ページが検索結果を表示するページであるかどうかをＨＴＴＰリファラに所定の検索エンジンのＵＲＬが含まれるかどうかにより判定し、含まれる場合にのみ図７ｈの処理を行ない、当該ＵＲＬが含まれない場合には図７ｈの処理を実行しないようにしてもよい。

[0102] 以上の処理（例えば上記図４、図７ａ乃至図７ｊや図８に示したフローチ

ャートに従った処理等）をコンピュータプログラムを所定の記憶媒体に記憶させ、この記憶媒体に記憶されているプログラムをコンピュータに読み込ませる（インストール、もしくはコピーさせる）ことで、このコンピュータは以上の処理を行うことができる。よって、コンピュータプログラム及び記憶媒体も本発明の範疇にあることは明白である。

請求の範囲

[請求項1] 検索語の入力を受け付ける受付手段と、

検索語に基づく検索指示を外部の検索サーバに対して送信する送信手段と、

前記検索指示に基づく検索結果を受信する受信手段と、

検索結果を表示する表示手段と

を備える情報処理装置であって、

前記検索結果には複数のリンク情報が含まれ、いずれかのリンク情報が選択された場合に前記送信手段が該リンク情報に従って宛先サーバにウェブページの要求を送信し、

前記情報処理装置は、前記要求に応じて前記受信手段が受信した前記ウェブページにつき、該ウェブページ内で前記検索語を検索し、かつ、該検索語を強調表示するためのタグを該ウェブページ内に挿入して加工する処理手段を更に備え、

前記表示手段は前記強調表示のためのタグが挿入されたウェブページを、該ウェブページ内の前記検索語を強調して表示することを特徴とする情報処理装置。

[請求項2] 前記処理手段は、前記受信したウェブページが検索結果のリンク情報の選択に基づき取得されたページである場合にのみ、該ウェブページの前記加工を行うことを特徴とする請求項１に記載の情報処理装置。

[請求項3] 前記処理手段は、前記ウェブページに含まれる本文中のテキスト、及び、所定のタグと関連づけられたテキストにおいて前記検索語を検索し、

前記所定のタグには、少なくともウェブページのタイトルを定義するタグ、表示画像を指定するためのタグ、ページ情報を定義するためのタグが含まれることを特徴とする請求項１又は２に記載の情報処理装置。

[請求項4] 前記表示手段は、

前記表示手段に表示される前記ウェブページ内のテキストについて前記強調表示を行う場合に、第１の表示要素を用い、

前記表示手段に表示される前記ウェブページ内の画像について前記強調表示を行う場合に、前記第１の表示要素とは異なる第２の表示要素を用い、

前記表示手段に表示されない前記ウェブページ内のテキストについて前記強調表示を行う場合に、前記第１及び第２の表示要素とは異なる第３の表示要素を用いる

ことを特徴とする請求項１乃至３のいずれか１項に記載の情報処理装置。

[請求項5] 前記処理手段は、前記受信したウェブページ内に含まれるスクリプトを実行することにより該ウェブページを加工することを特徴とする請求項１乃至４のいずれか１項に記載の情報処理装置。

[請求項6] 前記表示手段は、前記強調表示のためのタグが挿入されたウェブページの表示において、表示形態を制御するためのインタフェースを併せて表示し、

前記インタフェースの操作に従って前記表示手段は、前記強調表示された検索語を登場順又は逆順に選択した表示、前記強調表示の表示及び非表示、前記強調表示のためのタグが挿入されたウェブページの全画面表示、及び、前記強調表示のためのタグが挿入されたウェブページの表示の消去のいずれかを実行することを特徴とする請求項１乃至５のいずれか１項に記載の情報処理装置。

[請求項7] 前記表示手段は、前記強調表示のためのタグが挿入されたウェブページにおける前記強調表示を、表示された後所定時間経過後に消滅させることを特徴とする請求項１乃至６のいずれか１項に記載の情報処理装置。

[請求項8] 前記表示手段は、前記強調表示のためのタグが挿入されたウェブペ

ージを表示している間にいずれかの前記強調表示が選択された場合、該選択された強調表示を消滅させることを特徴とする請求項１乃至７のいずれか１項に記載の情報処理装置。

[請求項9] 前記送信手段は、前記いずれかのリンク情報が選択された場合に前記検索サーバに該リンク情報に対応するウェブページのプレビュー画像の要求を更に送信し、

前記受信手段は、前記プレビュー画像を更に受信し、

前記表示手段は、前記ウェブページを前記受信手段が受信して表示可能となるまで前記プレビュー画像を表示し、前記ウェブページが表示可能となった場合に前記プレビュー画像と置き換えて前記ウェブページを表示する

ことを特徴とする請求項１乃至８のいずれか１項に記載の情報処理装置。

[請求項10] 前記表示手段は、前記強調表示のためのタグが挿入されたウェブページを前記検索結果を表示するウィンドウ内の領域で表示することを特徴とする請求項１乃至９のいずれか１項に記載の情報処理装置。

[請求項11] 前記ウィンドウ内の領域での表示は、前記検索サーバから受信した前記検索結果に含まれるスクリプトを前記処理手段が実行することにより行われることを特徴とする請求項１０に記載の情報処理装置。

[請求項12] 前記表示手段は、前記強調表示のためのタグが挿入されたウェブページを、前記検索結果を表示するウィンドウ内の領域に、該検索結果と置き換えて表示することを特徴とする請求項１乃至９のいずれか１項に記載の情報処理装置。

[請求項13] 請求項１乃至１１のいずれか１項に記載の情報処理装置から検索語に基づく検索指示を受信し、検索処理を実行する検索サーバであって、

前記情報処理装置から検索語を受信する受信手段と、

前記検索語を用いて全文検索データベースを検索して検索結果を生

成する検索手段と、

前記検索結果を前記情報処理装置へ送信する送信手段と

を備え、

前記検索手段は、前記情報処理装置において前記検索結果におけるいずれかのリンク情報が選択された場合に、該情報処理装置に該リンク情報に対応するウェブページであって検索語の強調表示のためのタグが挿入されたウェブページを、検索結果を表示するウィンドウ内の領域で表示させるためのスクリプトを前記検索結果に含めることを特徴とする検索サーバ。

[請求項14] 情報処理装置により実行されるスクリプトであって、該情報処理装置にウェブページ内で検索語を検索させ、かつ、該検索語を強調表示するためのタグを該ウェブページ内に挿入させて加工を行わせるためのスクリプトを、該ウェブページ内に記述する処理手段を更に備え、

前記情報処理装置から前記検索結果におけるいずれかのリンク情報に対応するウェブページの要求を前記受信手段が受信した場合に、前記送信手段は、要求されたウェブページに対応する前記スクリプトが記述されたウェブページを前記情報処理装置へ送信することを特徴とする請求項１３に記載の検索サーバ。

[請求項15] 請求項１乃至１２のいずれか１項に記載の情報処理装置から、該情報処理装置が取得した所定の検索語に基づく検索結果に含まれるウェブページの要求を受信する受信手段と、

該要求されたウェブページを該情報処理装置に送信する送信手段とを備えるウェブサーバであって、

前記送信手段が送信する前記ウェブページ内に、前記情報処理装置により実行されるスクリプトであって、該情報処理装置に前記ウェブページ内で前記検索語を検索させ、かつ、該検索語を強調表示するためのタグを該ウェブページ内に挿入させて加工を行わせるためのスクリプトを記述することを特徴とするウェブサーバ。

［請求項16］ コンピュータを、請求項１乃至１２のいずれか１項に記載の情報処理装置の前記処理手段として機能させるためのコンピュータプログラム。