

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4868245号
(P4868245)

(45) 発行日 平成24年2月1日(2012.2.1)

(24) 登録日 平成23年11月25日(2011.11.25)

(51) Int.Cl.

F 1

G 06 F 13/00 (2006.01)
G 06 F 17/30 (2006.01)G 06 F 13/00 540 R
G 06 F 17/30 180 Z
G 06 F 17/30 340 B

請求項の数 10 (全 33 頁)

(21) 出願番号	特願2007-213169 (P2007-213169)
(22) 出願日	平成19年8月17日 (2007.8.17)
(65) 公開番号	特開2009-48380 (P2009-48380A)
(43) 公開日	平成21年3月5日 (2009.3.5)
審査請求日	平成21年3月27日 (2009.3.27)

前置審査

(73) 特許権者	500257300 ヤフー株式会社 東京都港区赤坂9丁目7番1号
(74) 代理人	100109553 弁理士 工藤 一郎
(72) 発明者	前川 英之 東京都港区六本木六丁目10番1号 ヤフー株式会社内

審査官 千本 潤介

最終頁に続く

(54) 【発明の名称】検索システム、検索装置、および検索方法

(57) 【特許請求の範囲】

【請求項 1】

クライアント端末に設けられ、ブラウザとともに動作するよう構成されたツールバー装置と、検索装置と、からなる検索システムであって、
前記ツールバー装置は、

ブラウザから少なくとも利用者が閲覧したリソースのURLである閲覧URLとリソースごとの閲覧時間とを含む閲覧情報を取得する閲覧情報取得部と、

取得した閲覧情報を所定の閲覧管理サーバ装置に電気通信回線を介して送信する閲覧情報送信部と、を有し、

前記検索装置は、

前記閲覧情報送信部から送信される閲覧情報を収集し、収集した閲覧情報に基づいて、
クローラの収集により生成したインデックスファイルに含まれていないURLであって、
検索エンジンにて新規に検索対象として利用するURLである新規検索対象URLを抽出するとともに、抽出したURLのコンテンツをインデックスファイルにインデクシングする第一インデクサ部を有する第一閲覧管理サーバ装置と、

収集した閲覧情報を蓄積する閲覧情報蓄積部と、

閲覧情報蓄積部に蓄積されている閲覧情報に含まれるリソースごとの閲覧時間に基づいてURL毎にその閲覧時間の長短を示す視聴度指数であるトラフィックランクを算出するトラフィックランクスコアリング部を有する第二閲覧管理サーバ装置と、を含むとともに

10

—

20

トラフィックランクスコアリング部で算出されたトラフィックランクが所定値以上のＵＲＬを、検索エンジンにて新規に検索対象として利用するＵＲＬである新規検索対象ＵＲＬとして抽出するとともに、抽出したＵＲＬのコンテンツをインデックスファイルにインデクシングする第二インデクサ部を有する検索システム。

【請求項 2】

前記検索装置は、

クローラ部と、

トラフィックランクスコアリング部で算出されたトラフィックランクに基づいてクローラ部のクローリングスケジュールを決定するスケジュール決定部と、
を有する請求項 1 に記載の検索システム。 10

【請求項 3】

前記検索装置は、トラフィックランクスコアリング部で算出されたトラフィックランクに基づいて検索結果をソートしてクライアントに対して出力するランクソート出力部を有する請求項 1 または 2 に記載の検索システム。

【請求項 4】

複数の請求項 1 に記載のツールバー装置から閲覧情報を収集する検索装置であって、

収集した閲覧情報に基づいて検索エンジンにて新規に検索対象として利用するＵＲＬである新規検索対象ＵＲＬを抽出するとともに、抽出したＵＲＬのコンテンツをインデクシングする第一インデクサ部を有する第一閲覧管理サーバ装置と、

収集した閲覧情報を蓄積する閲覧情報蓄積部と、 20

閲覧情報蓄積部に蓄積されている閲覧情報に含まれるリソースごとの閲覧時間に基づいてＵＲＬ毎にその閲覧時間の長短を示す視聴度指数であるトラフィックランクを算出するトラフィックランクスコアリング部を有する第二閲覧管理サーバ装置と、含む検索装置。

【請求項 5】

トラフィックランクスコアリング部で算出されたトラフィックランクに基づいて検索結果をソートしてクライアントに対して出力するランクソート出力部を有する請求項 4 に記載の検索装置。

【請求項 6】

クローラ部と、

トラフィックランクスコアリング部で算出されたトラフィックランクに基づいてクローラ部のクローリングスケジュールを決定するスケジュール決定部と、
を有する請求項 4 または 5 に記載の検索装置。 30

【請求項 7】

トラフィックランクスコアリング部で算出されたトラフィックランクに基づいて検索エンジンにて新規に検索対象として利用するＵＲＬである新規検索対象ＵＲＬを抽出するとともに、抽出したＵＲＬのコンテンツをインデクシングする第二インデクサ部を有する請求項 4 から 6 のいずれか一に記載の検索装置。

【請求項 8】

検索装置において、

閲覧ＵＲＬとリソースごとの閲覧時間とを含む閲覧情報を収集する閲覧情報収集ステップと、 40

収集した閲覧情報に基づいて、クローラの収集により生成したインデックスファイルに含まれていないＵＲＬであって、検索エンジンにて新規に検索対象として利用するＵＲＬである新規検索対象ＵＲＬを、ＣＰＵの演算処理によって抽出するステップと、

抽出したＵＲＬのコンテンツを、ＣＰＵの演算処理によってインデクシングする第一インデクシングステップと、

収集した閲覧情報を保持するため閲覧情報保持部に格納する閲覧情報格納ステップと、

閲覧情報に含まれるリソースごとの閲覧時間に基づいて、ＣＰＵの演算処理によってＵＲＬ毎にその閲覧時間の長短を示す視聴度指数であるトラフィックランクを算出するトラフィックランクスコアリングステップと、 50

トラフィックランクスコアリングステップで算出されたトラフィックランクが所定値以上のURLを、検索エンジンにて新規に検索対象として利用するURLである新規検索対象URLとして、CPUの演算処理によって抽出するステップと、

抽出したURLのコンテンツを、CPUの演算処理によってインデクシングする第二インデクシングステップと、

を計算機に実行させる検索方法。

【請求項9】

トラフィックランクスコアリングステップで算出されたトラフィックランクに基づいて検索結果をCPUの演算処理によってソートしてクライアントに対して出力するランクソート出力ステップをさらに計算機に実行させる請求項8に記載の検索方法。 10

【請求項10】

クローラステップと、

トラフィックランクスコアリングステップで算出されたトラフィックランクに基づいて、CPUの演算処理によってクローラステップのクローリングスケジュールを決定するスケジュール決定ステップと、

を計算機に実行させる請求項8または9に記載の検索方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、WWW(World Wide Web)の検索システムにおいて、Webページなどのリソースの検索を効率的に行うための技術に関する。 20

【背景技術】

【0002】

インターネットユーザーがWWW上のWebページやドキュメント、画像データなどのリソースを検索しようとする場合、通常Web検索システムを利用する。そして当該検索システムは、そのユーザーからの検索キーワードなどに応じた検索処理を実行するために、以下のようにしてWWW上のリソースに関するインデックスを生成する。すなわち検索システムは、リソース間のハイパーリンクを辿りWWW上を巡回する「クローラ」と呼ばれるプログラムを実行し、そのクローラがWWW上を巡回して収集したリソースのURL(Uniform Resource Locator)やコンテンツ内容に関する情報を取得する。そしてそれら収集情報から、WWW上のリソースに関するインデックス情報を構築、蓄積する、という具合である。そしてユーザーからの検索キーワードなどを取得した場合には、検索システムは検索対象としてそのインデックス情報を参照し、検索結果をユーザー端末に返信する。 30

【0003】

ところで、これらWWW上のリソースは更新などされることが常である。また、インターネットは情報(リソース)の即時提供性が特徴の一つになっている。そのためユーザーは更新された最新の提供情報を入手するために、時事関連ニュースページなど更新頻度が高いWebページなどを特に中心的に検索する傾向がある。

【0004】

したがってユーザーのこのようなニーズに応えるため、検索システムでは「クローラ」によるリソース情報の収集処理ならびにインデックス情報の生成蓄積処理を、リソースの更新頻度などに応じて定期的に行い、その情報鮮度を保つよう構成されていることが望ましい。 40

【0005】

しかしインデックス情報の鮮度を保つために、例えば単にクローラの巡回頻度を上げてインデックス情報の鮮度を保つような構成とすると、WWW上のトラフィックを圧迫することになってしまう。そこで頻繁にインデックス情報を更新すべきリソースを特定し当該リソースに対する巡回優先度を上げることで、WWW上のトラフィックをあまり圧迫することなくインデックス情報の情報鮮度を保つよう構成する技術も提供されている。

【0006】

50

具体的には、前述のようにインターネットユーザーは更新頻度が高いWebページを中心的に検索しアクセスする傾向がある。つまり逆に言えば、アクセス頻度の高いリソースは頻繁に更新などされている可能性が高い。したがって、例えばユーザーのアクセス数を集計することで「クローラ」のリソースに対する巡回優先度を決定し、優先度の高いリソースの巡回回数を上げることで、WWW上のトラフィック負荷をあまり増大させることなくインデックス情報の情報鮮度を保つことができる、といった技術である。

【特許文献1】特開2006-277732号公報

【発明の開示】

【発明が解決しようとする課題】

【0007】

10

しかし従来の検索システムには、次のような課題が普遍的に存在する。先ず、インデックス情報の鮮度を保つためには、前述のように「クローラ」はリソース間のハイパーリンクを辿りWWW上を巡回するため、インデックス化の対象となるリソースは別のリソースにハイパーリンクされている必要がある。ところが、新規に開設されたばかりのWebページなどはいずれのリソースにもハイパーリンクされていないことがある。すると、そのようなリソースにはクローラが辿り着けないためインデックス化の対象から外れてしまう。つまり極めて高い情報鮮度を有するはずの新規開設Webページなどが検索の際に参照されるインデックスファイルに含まれていない可能性があり、そのために検索結果に反映することが出来ないことがあった。

【0008】

20

次に、従来の検索システムでは、上記ユーザーのアクセス数を集計して「クローラ」の巡回優先度を決定し、ネットワーク負荷をあまりかけずに情報鮮度を保つが、その集計されたアクセスは、ユーザーの意思による積極的なアクセスには限られない。例えば、操作ミスなどによってリンクがクリックされたことによるアクセスや、リンクを辿って目的のウェップページに辿り着くための経由点としてなされたアクセスなどが含まれている場合がある。そのため集計されたリソースアクセス数は、実効的なアクセス実体（ユーザーの閲覧意思等による積極的なアクセスの数）を示していない可能性がある。そのため、その誤った集計結果から立てられたクローラの巡回スケジュールでは、正確な巡回優先順位を算出できていない可能性があった。

【0009】

30

更に、検索結果の一覧を表示させる際の各検索結果の表示ランキング（順序）は、ページ間のリンク数に応じて行うページランクによるものその他、ユーザーによるクリックされたページを上位にランキングする方法も知られている。しかし、この場合、上述と同様に、ユーザーのクリックに実効的なアクセス実体を示していない可能性があるため、その誤った集計結果から立てられた検索結果の表示の方法では、実用に沿った正確なランキングを算出できていない可能性があった。

【課題を解決するための手段】

【0010】

以上の課題を主に解決するために、本発明は以下のような検索システムを提供する。すなわちユーザーの端末などにおいて、ブラウザからツールバー装置を介して、利用者が閲覧したリソースのURLをはじめとする閲覧情報を取得し、閲覧管理サーバ装置に送信する。そして閲覧管理サーバ装置にて管理されている各端末の閲覧情報を、WWW上のリソースを検索するためのインデクシング（インデックス化）機能を備える検索装置等にて収集し、検索装置にて利用することができるよう構成された検索システムを提供する。特に上記構成においては、ツールバー装置において利用者が閲覧したリソースのURLをはじめとする閲覧情報を取得することを特徴とする。

【0011】

そしてこの検索システムではツールバー装置を介して取得した閲覧情報をを利用して、例えば第一に、他のページにハイパーリンクされていないリソースのコンテンツをインデックス化のために取得したり、さらに「クローラ」の巡回先として指定したりすることができます。

40

50

きるよう構成されていることをさらなる特徴とする。また第二に、前記のような誤アクセスなどを排除して集計したリソースの実効的なアクセス指標（視聴度指数）であるトラフィックランクを算出し、「クローラ」の巡回スケジュールに利用することができるようにも構成されていることをさらなる特徴とする。

【0012】

具体的には、ブラウザから少なくとも閲覧URLを含む閲覧情報を取得する閲覧情報取得部と、取得した閲覧情報を所定の閲覧管理サーバ装置に送信する閲覧情報送信部と、を有するツールバー装置と、前記閲覧情報送信部から送信される閲覧情報を収集する検索装置と、からなる検索システムである。

【0013】

そして上記第一又は第二のさらなる特徴点を実現するため、本発明の検索システムは上記構成に加えて、第一には、前記検索装置において、収集した閲覧情報に基づいて検索エンジンにて新規に検索対象として利用するURLである新規検索対象URLを抽出するとともに、抽出したURLのコンテンツをインデクシングする第一インデクサ部を有する第一閲覧管理サーバ装置を含むよう構成しても良い。

10

【0014】

また、第二には、前記検索装置において、収集した閲覧情報を蓄積する閲覧情報蓄積部と、閲覧情報蓄積部に蓄積されている閲覧情報に基づいてURL毎に視聴度指数であるトラフィックランクを算出するトラフィックランクスコアリング部を有する第二閲覧管理サーバ装置を含むよう構成し、さらに、クローラ部と、トラフィックランクスコアリング部で算出されたトラフィックランクに基づいてクローラ部のクローリングスケジュールを決定するスケジュール決定部と、を有するよう構成しても良い。

20

【発明の効果】

【0015】

以上のような構成をとる第一の本発明によって、ブラウザからツールバー装置を介して送信された、利用者が閲覧したリソースのURLをはじめとする閲覧情報を、閲覧管理サーバ装置を介して検索装置にて収集し、様々に利用することができる。具体的には、第一に他のページにハイパーアリンクされていないリソースのコンテンツを取得し検索用のインデックスファイル化処理を行うことができる。また「クローラ」の巡回先として指定することができる。そしてそれによって、極めて高い情報鮮度を有すると思われる新規開設Webページ等のリソースを「クローラ」にて巡回取得して検索対象としてインデックス化することができる。

30

【0016】

また、第二に誤アクセスなどを排除して集計したリソースのアクセス指標（視聴度指数）であるトラフィックランクを算出し、「クローラ」の巡回スケジュールに利用することができる。したがって、ユーザーのリソースの実効的なアクセス実体に応じて、クローラの巡回スケジューリングを立てることができる。

【発明を実施するための最良の形態】

【0017】

以下に、図を用いて本発明の実施の形態を説明する。なお、本発明はこれら実施の形態に何ら限定されるものではなく、その要旨を逸脱しない範囲において、種々なる態様で実施しうる。

40

【0018】

なお実施例1および実施例2は、主に請求項1、6、11などについて説明する。また実施例3は、主に請求項2、7、12などについて説明する。また実施例4は、主に請求項3、8、13などについて説明する。また実施例5は、主に請求項4、9、14などについて説明する。また実施例6は、主に請求項5、10、15などについて説明する。

【0019】

実施例1

50

【 0 0 2 0 】

<概要>

【 0 0 2 1 】

図1は、本実施例の検索システムにおける閲覧情報収集の一例を説明するための概念図である。この図にあるように、クライアント端末1にてブラウザが起動されている。また、このブラウザにはツールバーが組み込まれており、ブラウザの起動に合わせてツールバーもブラウザ上の所定領域(0101)に表示されている。

【 0 0 2 2 】

ここで、本実施例では、ユーザーがブラウザやツールバーに入力した検索キーやURLを、ツールバーのプログラムによって収集する。またブラウザでのWebページの表示時間や、移動先や移動元などWebページの遷移経路を示す情報も、ツールバーによって取得される。そして、このようにしてネットワーク上の各クライアント端末1, 2, 3, ···にて取得された情報が、閲覧管理サーバ装置を含む検索装置(0102)に送信される。

10

【 0 0 2 3 】

図2は、このようにクライアント端末のツールバーにて取得され、閲覧管理サーバ装置を介して検索装置に収集される閲覧情報の一例を表す図である。この図にあるように、ツールバーは閲覧情報として、ユーザーIDやブラウザにてアクセスしたWebページのURL、そのWebページの閲覧開始時刻(アクセス時刻)、閲覧時間(Webページなどの表示時間)、あるいは図示しない移動先や移動元で示されるWebページの遷移経路を示す情報などを取得する、という具合である。

20

【 0 0 2 4 】

そして検索装置では、各クライアント端末から閲覧管理サーバ装置を介して収集した上記のような閲覧情報をを利用して、例えばユーザーのWebブラウジング行動を解析したり、その解析結果からユーザーにとって人気の高いサイトなどを迅速に把握したりすることができる。そして、例えばいわゆる「クローラ」によって検索用インデックス情報を収集する際には、その人気サイトをクロールする頻度を上げて当該人気サイトの最新の情報が検索によって提供されるよう構成する、など様々に活用することができる、という具合である。

【 0 0 2 5 】

30

<機能的構成>

【 0 0 2 6 】

図3は、本実施例の検索システムにおける機能ブロックの一例を表す図である。この図にあるように、本実施例の「検索システム」(0300)は、「ツールバー装置」(0310)と、「検索装置」(0320)と、からなる。

【 0 0 2 7 】

なお、「ツールバー装置」(0310)とは、クライアント端末上のブラウザの機能を補助、拡張、代替などするため、ブラウザとともに動作するよう構成されたアプリケーションが組み込まれた装置をいう。またツールバー装置の一般的な機能としては、例えば、検索窓に入力されたキーワードによるWeb上のリソースの検索機能、RSS文書データの収集や表示機能、ショートカットアイコンによるランチャー機能などが挙げられる。そして、本実施例の検索システムにおいては、後述する構成によってブラウザにて閲覧されたURLを含む閲覧情報を取得し、閲覧管理サーバ装置に送信するというさらなる機能を有していることを特徴とする。

40

【 0 0 2 8 】

また、「検索装置」(0320)とは、前述のように「クローラ」が収集したWeb上のリソースの情報をを利用して所定の検索エンジンの処理によってインデックス情報を生成し、ユーザーからの検索キーワードなどに応じてインデックス情報の検索処理を実行するサーバ装置をいう。そして本実施例の検索システムにおいては、クライアント端末のツールバー装置にて取得され閲覧管理サーバ装置に送信された閲覧情報を収集し、その検索処理を実行する。

50

理やインデックス情報の生成処理などに活用することを特徴とする。

【0029】

なお本実施例では、図3(a)に示すように閲覧管理サーバ装置がこの検索装置に含まれる構成を例に挙げ説明するが、もちろん本実施例の検索システムの構成はそれに限定されるものではない。例えば図3(b)に示すように閲覧管理サーバ装置が検索装置とは別個のサーバ装置としてネットワーク上に設けられ、検索装置は当該閲覧管理サーバ装置にて管理されている閲覧情報を取得する、といった構成であっても良い。

【0030】

ここで、まず検索装置とともに本実施例の検索システムを構成する「ツールバー装置」(0301)の機能的構成について以下に説明する。図3にあるように「ツールバー装置」は、「閲覧情報送信部」(0303)と、「閲覧情報取得部」(0304)とを有する。

10

【0031】

なお、以下に記載する本システムのツールバー装置や検索装置の各機能ブロックは、ハードウェア、ソフトウェア、又はハードウェア及びソフトウェアの両方として実現され得る。具体的には、コンピュータを利用するものであれば、CPUや主メモリ、バス、あるいは二次記憶装置(ハードディスクや不揮発性メモリ、CDやDVDなどの記憶メディアとそれらメディアの読み取りドライブなど)、印刷機器や表示装置、その他の外部周辺装置などのハードウェア構成部、またその外部周辺装置用のI/Oポート、それらハードウェアを制御するためのドライバプログラムやその他アプリケーションプログラム、情報入力に利用されるユーザーインターフェースなどが挙げられる。

20

【0032】

そして主メモリ上に展開したプログラムに従ったCPUの演算処理によって、インターフェースを介して入力されメモリやハードディスク上に保持されているデータなどが加工、蓄積されたり、上記各ハードウェアやソフトウェアを制御するための命令が生成されたりする。また、この発明はシステムとして実現できるのみでなく、方法としても実現可能である。また、このような発明の一部をソフトウェアとして構成することができる。さらに、このようなソフトウェアをコンピュータに実行させるために用いるソフトウェア製品、及び同製品を記録媒体に固定した記録媒体も、当然にこの発明の技術的な範囲に含まれる(本明細書の全体を通じて同様である)。

30

【0033】

「閲覧情報取得部」(0311)は、ブラウザから少なくとも閲覧URL含む閲覧情報を取得する機能を有する。「閲覧URL」とは、このツールバー装置が組み込まれたクライアント端末のブラウザにて表示されているリソースのWeb上の所在を示す情報をいい、例えば「http://.../001.bmp」等の情報を挙げられる。

【0034】

そして本実施例の検索システムでは、ネットワーク上の各クライアント端末のツールバーにて取得され、閲覧管理サーバ装置を介して検索装置にて収集されるこの閲覧URLを利用した各種処理、例えば「クローラ」の巡回対象を決定するなどの処理を行うことを特徴とする。

40

【0035】

また閲覧情報には、上記閲覧URL以外にもブラウザでの閲覧に係る各種情報が含まれていて良い。例えば、ユーザーID、ブラウザIDやツールバーID、閲覧開始時刻情報、閲覧時間情報、あるいはWeb上の2以上のリソースを連続で閲覧した際の「移動元リソース」や「移動先リソース」などの遷移情報などが挙げられる。

【0036】

ここでユーザーIDとはユーザーを識別するための情報をいい、例えばクライアント端末へのログイン時のIDやサービス単位で割当てられるIDなどが挙げられる。また、ブラウザIDやツールバーIDも同様に当該ブラウザやツールバーを識別するための情報をいい、例えば予めブラウザプログラムやツールバープログラムに記録されている製品ID

50

などを利用すると良い。そして、これら識別情報を利用することでクライアント端末を識別することができるので、クライアント端末（ユーザー）単位での閲覧情報の管理、分析を行うことができる。なお、これらID情報の保持は、例えばツールバー装置自身が保持する方法や、ブラウザ装置が有するcookie機能を利用する方法などが挙げられる。

【0037】

また、閲覧開始時刻情報とは、ブラウザにてWeb上のリソースにアクセスされた時刻、あるいはそのアクセスによってブラウザにリソースが取得された時刻や表示された時刻などをいう。具体的に当該時刻は、例えばブラウザにてアクセスリクエストしたリソースのロード完了を検出した時刻や、Webサーバが「200：OK」などのHTTPステータスコードを応答した時刻等が挙げられる。

10

【0038】

また閲覧時間情報とは、ブラウザにてWeb上のリソースが表示されている時間をいい、例えば、ブラウザのUIウィンドウが、オープンしていた時間や、当該ウィンドウがアクティブとなっている時間、またウィンドウ領域にマウスポインタが存在する時間などが挙げられる。またさらに、一定時間の操作がないためクライアント端末のOSが自動的にログオフ処理を行ったり、スクリーンセーバ等が起動したりした場合には、ログオフされている時間やスクリーンセーバが起動している時間を除いた後の時間としてもよい。また閲覧時間情報は、前記閲覧開始時刻と閲覧終了時刻との差分を算出することで取得されても良い。

【0039】

20

また遷移情報とは、Web上の2以上のリソースを連続で閲覧した際の「移動元リソース」と「移動先リソース」とを示す情報をいい、例えばハイパーアリンクのクリックによる遷移のほか、ブラウザやツールバーのURL入力欄に連続で入力された2以上のURLを「移動先」「移動元」とする遷移情報などが挙げられる。

【0040】

そしてツールバー装置は上記閲覧情報を、例えばAPI（Application Program Interface）やDDE（Dynamic Data Exchange）やファイルシステム等を用いてクライアント端末のブラウザから取得する、という具合である。

【0041】

30

「閲覧情報送信部」（0312）は、取得した閲覧情報を所定の閲覧管理サーバ装置に送信する機能を有する。具体的には、上記のように取得した閲覧情報を、クライアント端末の通信IFを使用し通信回線等を介して、所定の閲覧管理サーバ装置に送信する機能を有する。なお、閲覧情報の送信先となる所定の閲覧管理サーバ装置の送信先アドレスは、例えばツールバー装置のプログラム内に予め記述されることで特定する方法などが挙げられる。また、この閲覧情報送信部から閲覧情報が送信されるタイミングはブラウザにて表示されているリソースが切換わるごとに送信するといったタイミングや、一定の時間、又は時刻が到来した際にバッチ処理で送信するタイミング、あるいはツールバー起動時または終了時に送信するタイミングなどが挙げられる。

【0042】

40

そして、このように各クライアント端末上ブラウザでの閲覧に応じてツールバー装置にて取得された閲覧情報が閲覧管理サーバ装置に送信される。そして後述するよう閲覧管理サーバ装置を介して検索装置にて当該閲覧情報が収集されることで、検索装置ではクライアント端末の閲覧リソースの特定などブラウジング（閲覧）履歴を容易に把握することができる、という具合である。

【0043】

続いて、上記ツールバー装置とともに本実施例の検索システムを構成する「検索装置」（0320）の機能的構成について、同様に図3を用いて以下に説明する。図3にあるように「検索装置」は、「閲覧管理サーバ装置」（0330）を備える。

【0044】

「閲覧管理サーバ装置」（0330）は、閲覧情報送信部（0312）から送信される

50

閲覧情報を受信、管理する機能を有する。そして、閲覧管理サーバ装置では閲覧情報を利用して、例えばユーザーのWebブラウジング行動を解析したり、その解析結果からユーザーにとって人気の高いサイトなどを迅速に把握したりすることができる。そして、検索装置での「クローラ」による検索用インデックス情報収集の際には、その人気サイトについてクロールする頻度を上げて当該人気サイトの最新の情報が検索によって提供されるようにする、など様々な処理を行うことができる。

【0045】

またその他にも、詳細は実施例2にて後述するように、例えば、閲覧情報に含まれる閲覧URLのうち自身が管理していないURLを他のページにハイパーリンクされていないリソースのURLとして「クローラ」の巡回先として新規に指定する処理を行っても良い。あるいは詳細は実施例3にて後述するように、閲覧情報に含まれる閲覧時間などの情報から、誤アクセスなどを排除して集計したリソースの実効的なアクセス指標（視聴度指数）であるトラフィックランクを算出し、「クローラ」の巡回スケジュールに利用する処理を行っても良い。10

【0046】

なお閲覧管理サーバ装置は前述のように、ネットワーク上にて検索装置とは別個のサーバ装置として存在しても良い。その場合、図3(b)に示すように閲覧管理サーバ装置にて管理されている閲覧情報を、「検索装置」が収集し、上記のような各種処理を行うよう構成されていても良い。

【0047】

<ハードウェア的構成>

【0048】

図4は、上記機能的な各構成要件をハードウェアとして実現した際の、検索システムにおける構成の一例を表す概略図である。この図を利用して本実施例の検索システムにおけるそれぞれのハードウェア構成部の働きについて説明する。この図にあるように、本実施例の検索システムのハードウェア構成は、クライアント端末に組み込まれた「ツールバー装置」(0410)と、サーバ装置としてネットワーク上に配置されている「検索装置」(0420)と、により構成される。なお本例では、閲覧管理サーバ装置を実現するための「閲覧管理サーバプログラム」が検索装置において実行される構成、すなわち検索装置が閲覧管理サーバ装置を含む構成を例に挙げて説明する。30

【0049】

そしてツールバー装置と検索装置は電気通信回線を介して相互に接続され、情報の送受信を行う。なお、電気通信回線はインターネットを含む。

【0050】

また「ツールバー装置」においては、閲覧情報取得部を実現し、またその他各種演算処理を行う「CPU(中央演算装置)」(0411)と、「RAM」(0412)や、閲覧情報送信部である「通信I/F」(0413)を備える。またキーボード、マウス等の入力装置である「UI」(ユーザーインターフェース)(0414)や、ブラウザプログラムにて処理されたリソースを表示するための「VRAM」(0415)や、ディスプレイなどの「表示装置」(0416)も備える。そしてそれらが「システムバス」などのデータ通信経路によって相互に接続され、情報の送受信や処理を行う。40

【0051】

またツールバー装置の「RAM」上には、ブラウザプログラムと、ツールバープログラムとが格納されており、これらプログラムに従い閲覧情報の取得や送信処理やその他処理に係る「CPU」の各種演算処理が実行される。また上記ツールバープログラムによって、RAM上の所定のアドレスには閲覧URLを格納する領域が確保されている。

【0052】

一方「検索装置」においては、各種演算処理を行う「CPU」(0421)と、「RAM」(0422)と、「通信I/F」(0423)と、大量の閲覧情報などを蓄積するためのハードディスクなどの「二次記憶装置」(0424)とを有している。そしてそれら

10

20

30

40

50

が「システムバス」などのデータ通信経路によって相互に接続され、情報の送受信や処理を行う。

【0053】

また検索装置の「RAM」上には、閲覧管理サーバプログラムが格納されており、当該プログラムに従い閲覧情報の管理処理やその他処理に係る「CPU」の各種演算処理が実行される。

【0054】

ここで、「ツールバー装置」においてユーザーの「UI」を介した操作入力を受け、「RAM」上のブラウザプログラムおよびツールバープログラムに従い以下のような処理が実行される。すなわち「UI」を介したブラウザ操作によってWeb上のリソースへのアクセス指示が入力されると、そのアクセス指示で指定されたリソースのURLが「RAM」の所定アドレスに格納される。するとブラウザプログラムに従い指定されたURLに対してHTTPリクエストが送信される。そしてそのHTTPリクエストに対するレスポンスをアクセスリソースのコンテンツとして「通信I/F」にて受信する。つづいて受信したコンテンツに含まれるHTMLファイル、イメージファイルなどを、ブラウザプログラムが有するHTMLレンダリングエンジンなどに従った「CPU」の演算処理によってレンダリング処理する。そしてその処理結果が「VRAM」に転送され「表示装置」上にWebコンテンツが表示される。そしてユーザーは表示された当該WWW上のWebページなどのリソースを閲覧する。

【0055】

また、それとともに前記ブラウザでのアクセス指示入力に応じてツールバープログラムは以下の処理を実行する。すなわち「RAM」の所定アドレスに格納されたURLを閲覧情報として、前述のような所定の各タイミングで「通信I/F」から「検索装置」に送信する、という具合である。

【0056】

また、ツールバープログラムは、その他に以下のようにしてユーザーIDや閲覧時間情報などを取得して、上記URLと同様に閲覧情報として送信しても良い。まず、ツールバープログラムが起動すると、インターネットキャッシュなどからユーザーIDを取得し、「RAM」上の所定のアドレスに格納する。また、ブラウザプログラムによりHTTPレスポンスの取得が完了するとOSのタイマ関数などより現在時刻を取得し「RAM」上の所定のアドレスに格納する。あるいは、ブラウザにおいて一のURLに関してウィンドウがアクティブ状態である時間を監視し、その時間を閲覧時間として「RAM」上の所定のアドレスに格納する。そして、上記「RAM」に格納した各情報を前述のような所定の各タイミングで「通信I/F」から「検索装置」に送信する、という具合である。

【0057】

そして、「検索装置」では、「通信I/F」にて各ツールバー装置から送信してきた閲覧情報を受信し「RAM」の所定アドレスに格納する。そして、閲覧管理サーバープログラムにしたがって、例えば閲覧情報に含まれるURLやユーザーIDなどパラメータとする閲覧履歴データベースなどを「CPU」の演算処理によって生成し、HDDなどの「二次記憶装置」に格納する、という具合である。

【0058】

そして、そのデータベースを利用して、例えば閲覧URLごとのアクセス数を「CPU」の演算処理によって集計する。そして「アクセス数 - 巡回頻度テーブル」などを参照し、集計結果の高い順にクローラの巡回頻度を決定する処理を行う、という具合である。

【0059】

<処理の流れ>

【0060】

図5は、本実施例の検索システムにおける処理の流れの一例を表すフローチャートである。なお、以下に示すステップは、媒体に記録され計算機を制御するためのプログラムを構成する処理ステップであっても構わない。

10

20

30

40

50

【0061】

この図にあるように、ツールバー装置にてツールバーが起動されると、まずURLを含む閲覧情報を取得する（ステップS0501）。次にユーザーのブラウザの操作により閲覧URLが更新（変更）されたか否かを判断する（ステップS0502）。更新されたと判断された場合には、閲覧URLを含む情報を閲覧情報として閲覧管理サーバ装置を含む検索装置に送信する（ステップS0503）。

【0062】

またツールバー装置では、閲覧情報の送信（ステップS0503）後、閲覧が終了したか否かの判断がなされる（ステップS0504）。ブラウザが終了されたなど、閲覧が終了したと判断されると処理は終了する。終了していないと判断された場合には、URLを含む閲覧情報を取得する（ステップS0501）処理に戻る。

10

【0063】

検索装置では、ステップS0503にてツールバー装置により送信されたURLを含む閲覧情報を受信し（ステップS0511）、記録媒体等に格納する（ステップS0512）。そして各ツールバー装置から取得した閲覧情報をを利用してデータベースを生成、蓄積し、各種処理に利用する。

【0064】

もちろん、前述のようにステップS0503では、検索装置とは別個の閲覧管理サーバ装置に閲覧情報を送信しても良い。そして、その場合には、検索措置にて閲覧管理サーバ装置で受信した閲覧情報を収集するステップが追加されると良い。

20

【0065】

<効果の簡単な説明>

【0066】

以上のように本実施例の検索システムによってクライアント端末のツールバー装置を利用してユーザーの閲覧情報を取得し、その閲覧情報を閲覧管理サーバ装置を介して検索装置に送信することが可能となる。そして検索装置では、ネットワーク上の複数のツールバー装置から送信される閲覧情報を幅広く収集しデータベースなどを構築することが容易に可能である。

【0067】

そしてその構築したデータベースを利用して、例えば閲覧URLごとのアクセス数を集計し、集計結果の高い順にクローラの巡回頻度を決定する処理を行ったりすることができる。

30

【0068】

実施例2

【0069】

<概要>

【0070】

図6は、本実施例の検索システムにおける処理の一例を説明するための概念図である。この図にあるように、あるユーザーAが、ツールバー装置を備えたクライアント端末（0601）を利用して、友人が新規に開設したWebサイトA（0602）にアクセスしている。ここで、新規WebサイトAは新規開設されたばかりであるため、他のWeb上のリソースからのハイパーリンクなどが無い状態である。そのため「クローラ」の巡回対象にはなっておらず、未だ検索装置（0603）の検索サービス用インデックスファイル（0604）には存在していない。

40

【0071】

しかし、本実施例の検索システムにおいては、上記クライアント端末によるWebサイトAへのアクセスによって、検索装置に対して当該URLを含む閲覧情報が送信される構成となっている。そこで、検索装置では送信されてきた当該URLと、検索サービス用インデックスファイルに含まれるURLとの比較処理を行う。そしてインデックスファイルに含まれていなければコンテンツを取得しインデックスファイルに追加格納する。またそ

50

れに加えて「クローラ」によるハイパーアクション以外の直接巡回対象URLとして設定する、という具合である。

【0072】

<機能的構成>

【0073】

図7は、本実施例の閲覧管理サーバ装置における機能ブロックの一例を表す図である。この図にあるように、本実施例の「検索システム」(0700)は、実施例1の構成を基本として、「ツールバー装置」(0710)と、「検索装置」(0720)とを有する。そして「ツールバー装置」は、実施例1同様に「閲覧情報取得部」(0711)と、「閲覧情報送信部」(0712)とを有する。なお、上記ツールバー装置と検索装置の機能的構成を含む詳細な説明は、実施例1にて記載済みであるので省略する。10

【0074】

そして本実施例の特徴点は、「検索装置」が「第一閲覧管理サーバ装置」(0730)を有し、その第一閲覧管理サーバ装置がさらに「第一インデクサ部」(0731)を有する点である。

【0075】

「第一インデクサ部」(0731)は、収集した閲覧情報に基づいて新規検索対象URLを抽出するとともに、抽出したURLのコンテンツをインデクシングする機能を有する。20

【0076】

「新規検索対象URL」とは、検索エンジンにて新規に検索対象として利用するURLをいう。具体的には、検索装置は「クローラ」が巡回して収集したWWW上のリソースに関する情報をまとめてインデックスファイルを生成、保持している。そして検索エンジンによる検索の際には、その保持しているインデックスファイルを参照し検索を行うよう構成されている。つまりインデックスファイルに含まれていないURLは検索対象とならない構成となっている。そこで、閲覧情報に含まれるURLとインデックスファイルに格納されているURLと差分情報を取得し、その差分であるURLを新規検索対象URLとして抽出する、という具合である。

【0077】

そして、このように抽出された新規検索対象のURL、および当該URLにあるリソースのコンテンツなどが検索装置にてインデクシング（インデックスファイルへの追加処理）されることで、本実施例の検索システムでは新規開設Webページなど本来検索対象となっていたいなかったようなWWW上のリソースについても検索対象とすることができます、という具合である。また、ここで「インデクシング」処理とは、詳細には検索エンジンがターゲットとなるキーワードを高速に検索しやすいようなデータ構造や、ファイル構造に再構成することである。例えば、抽出したキーワードをキーとし、これをハッシュテーブル化したインデックスを新たに付与するといった方法などが挙げられる。30

【0078】

また、本実施例の検索装置は上記インデクシング処理に加え、さらに抽出した新規検索対象URLを「クローラ」がハイパーアクションを辿らずに直接巡回する対象として追加する処理を行い、更新情報を取得するよう構成しても良い。40

【0079】

<ハードウェア的構成>

【0080】

図8は、上記機能的な各構成要素をハードウェアとして実現した際の、検索システムにおける構成の一例を表す概略図である。この図を利用して本実施例の検索システムにおけるそれぞれのハードウェア構成部の働きについて説明する。この図にあるように、本実施例の検索システムのハードウェア構成は、ツールバー装置(0810)と検索装置(0820)より構成される。そして両者はインターネットを含む電気通信回線を介して、相互に接続可能となっている。50

【0081】

また、「ツールバー装置」(0810)は、実施例1と同様に「CPU」(0811)と「RAM」(0812)、「通信I/F」(0813)、「UI」(0814)、「VRAM」(0815)と「表示装置(ディスプレイなど)」(0816)と、を備える。

【0082】

そして「検索装置」(0820)も実施例1と同様に、「CPU」(0821)と「RAM」(0822)、「通信I/F」(0823)と、「二次記憶装置」(0824)と、を有する。

【0083】

ここで実施例1にて説明したように、ネットワーク上の複数のツールバー装置のツールバープログラムの処理によって、閲覧URLを含む閲覧情報が検索装置に送信され、「RAM」の所定アドレスに格納される。10

【0084】

検索装置では、第一閲覧管理サーバプログラムに含まれる第一インデクサプログラムに従い、以下のような処理を実行する。まず、予め「二次記憶装置」にて格納されている検索サービス用のインデックスファイルを参照し、「RAM」の所定アドレスに格納された閲覧URLをキーとした検索処理が「CPU」の演算によって実行される。そして、その検索処理の結果、インデックスファイルに当該URLが含まれていないとの判断結果が出力されれば、そのURLは新規検索対象のURLである判断する。すると第一インデクサプログラムに従い、「CPU」は「RAM」に格納されたそのURLに対して通信I/F(0913)を介してアクセスし、リソースコンテンツを取得する命令を出力する。そして取得したコンテンツに関し、「CPU」の処理によってインデクシング処理が実行され、新規のインデックス情報として「二次記憶装置」に保持されているインデックスファイルに記録される、という具合である。20

【0085】

また、第一インデクサプログラムに従って、「RAM」に格納された当該URLを、次回以降のクローラの巡回対象としてスケジュールに組み込む処理を実行しても良い。またさらに、当該URLに関しては情報鮮度の高いリソースであると仮定して、クローラの巡回頻度を上げたスケジュールを生成するよう処理しても良い。

【0086】

<処理の流れ>

【0087】

図9は、本実施例の検索システムにおける処理の流れの一例を表すフローチャートである。なお、以下に示すステップは、媒体に記録され計算機を制御するためのプログラムを構成する処理ステップであっても構わない。

【0088】

ここで、ツールバー装置にてツールバーが起動されると、まずURLを含む閲覧情報を取得する(ステップS0901)。次にユーザーのブラウザの操作により閲覧URLが更新(変更)されたか否かを判断する(ステップS0902)。そして更新されたと判断された場合には、閲覧URLを含む情報を閲覧情報として第一閲覧管理サーバ装置を含む検索装置に送信する(ステップS0903)。40

【0089】

つづいて検索装置において、ネットワーク上の各クライアント端末から送信されてきたn個のURLを含む閲覧情報を受信、収集する(ステップS0911)。次に、閲覧情報を記録媒体等に格納する(ステップS0912)。次にループ1に入り、格納されているn個のURLのそれぞれについて検索装置のインデックスファイル(データベース)を参照し、同一のURLが存在するか否かの判定を行う(ステップS0914)。

【0090】

そしてその判定処理の結果、インデックスファイルに存在しないと判定されたURLについては、新規検索対象URLとしてリソースコンテンツを取得しインデクシングする(50

ステップS0915)。一方、同一のURLが存在すると判定されたURLについては、インデクシングの処理はスキップする。そして、閲覧情報で示される全てのURLに対して上記判定処理が完了すると、処理を終了する。

【0091】

<効果の簡単な説明>

【0092】

以上のように本実施例の検索システムでは、新規開設Webページなどハイパリンクが張られていないため従来のクローラでは検索対象となっていなかつたようなWWW上のリソースについても検索用のインデックスファイルとしてインデックス化することができる。

10

【0093】

実施例3

【0094】

<概要>

【0095】

本実施例は、上記実施例1の検索システムを基本として、ネットワーク上の各クライアント端末のツールバー装置から閲覧情報を収集する。そしてWWW上のリソースの視聴度合いに関して、その閲覧情報をを利用して従来よりも精度の高い視聴度指数である「トラフィックランク」を算出することを特徴とする検索システムである。

【0096】

20

図10は、本実施例の検索装置における処理の一例を説明するための図である。この図にあるように、ツールバー装置を備えるクライアント端末にて、サイトBを経由したサイトAへのアクセスが実行された。ここで、サイトBへのアクセス時間は単なるリンクの経由であったため短い。一方、閲覧を目的としたサイトAへのアクセス時間は長いものとなっている。

【0097】

そして、そのようなアクセス(閲覧)時間を含む閲覧情報を検索装置に送信されると、検索装置ではそのアクセス時間や閲覧情報に含まれるその他情報を利用して「トラフィックランク」を算出する。そして、このトラフィックランクは、後述するように上記リンク経由によるアクセスや誤アクセスを排除した閲覧実体に即した指数となっているため、従来のサイト(WWW上のリソース)視聴度指数よりも精度の高い指数を算出することができる、という具合である。

30

【0098】

<機能的構成>

【0099】

図11は、本実施例の検索装置における機能ブロックの一例を表す図である。この図にあるように、本実施例の「検索システム」(1100)は、実施例1の構成を基本として、「ツールバー装置」(1110)と、「検索装置」(1120)とを有する。また「ツールバー装置」(1110)は、実施例1と同様に「閲覧情報送信部」(1111)と、「閲覧情報取得部」(1112)とを有する。なお、上記各構成は、実施例1にて既に記載済みであるのでその説明は省略する。

40

【0100】

そして本実施例の特徴は、「検索装置」(1120)が、「第二閲覧管理サーバ装置」(1130)を有し、その第二閲覧管理サーバ装置がさらに「閲覧情報蓄積部」(1131)と、「トラフィックランクスコアリング部」(1132)とを有する点である。

【0101】

「閲覧情報蓄積部」(1131)は、収集した閲覧情報を蓄積する機能を有する。具体的にはツールバー装置において取得され、閲覧管理サーバ装置を介するなどして検索装置にて収集された閲覧情報を入力として受け取り、HDDや不揮発性メモリなどの記憶媒体に格納する機能を有する。なお、ここで蓄積される閲覧情報は、後述するトラフィックラ

50

ンクスコアリング部におけるトラフィックランクの算出のため、例えばリソースごとの閲覧（アクセス）時間やユーザー ID（クライアント端末 ID、ツールバー装置 ID）、リソースの遷移情報などが含まれていることが望ましい。

【0102】

「トラフィックランクスコアリング部」(1132)は、閲覧情報蓄積部(1131)に蓄積されている閲覧情報に基づいて URL 毎にトラフィックランクを算出する機能を有する。「トラフィックランク」とは、WWW上のリソースの視聴度指数をいい、例えば以下のようにして算出する方法が挙げられる。

【0103】

すなわち、トラフィックランクの算出方法の一例としては、例えば1日における全ウェップページの合計表示時間に対するウェップページ A の合計表示時間のパーセンテージをウェップページのランクとして算出する方法などが挙げられる。このようにウェップページのランクを算出することで、クリックミスなどの表示時間が短いアクセスに関しては、そのランクを低いものとして算出することができる。したがって単純なアクセス数などでは推定することが困難な、クリックミスなどを排除して実体に即した広告効果などを示す指標としてウェップページのトラフィックランクを算出することができる。

10

【0104】

なお、上記ウェップページのランクの算出処理として、主に以下のような3つのパターンによる算出処理が考えられる。第一に、例えば一日を単位とし、集計された全端末あるいは一部端末からのブラウジング情報に含まれる表示時間を利用して、いわゆるウェップページごとの「全国ランク」を算出する処理が挙げられる。

20

【0105】

また第二に、閲覧情報に含まれるユーザー ID などから「ユーザー ID - 属性情報テーブル」などを参照して特定されるユーザーの属性情報をを利用して、年齢別、男女別、居住地別、職業別などの各種セグメントごとにウェップページの「属性別ランク」を算出するように構成しても良い。

【0106】

なお上記「ユーザー ID - 属性情報テーブル」は、一例として以下のようにして構築することができる。例えば、クライアント端末にツールバー装置をインストールする際には、上記ユーザーの属性情報の登録を必須とする。そして、登録された属性情報をCookieなどのユーザー ID と紐付けて検索装置にて管理することで、クライアント端末のツールバー装置から送信されたユーザー（属性情報）の特定を実行する、という具合である。

30

【0107】

また第三に、ユーザー ID などをを利用して閲覧情報をユーザー単位で分類し、「ユーザーにおけるサイト A のランク」、「ユーザーにおけるサイト B のランク」といった具合に、いわゆる「ユーザー別ランク」を算出するよう構成しても良い。

【0108】

具体的に、この第三の算出処理を行うための構成としては、本実施例の検索システムのツールバー装置と検索装置が以下の構成を備えていると良い。例えば、このツールバー装置は、その閲覧情報送信部にて送信する閲覧情報に、例えばCookieやその他ユーザーを識別するための情報であるユーザー識別情報を含み検索装置に送信するよう構成されている。

40

【0109】

そして、ここで送信されたユーザー識別情報をを利用して、検索装置の以下の構成によって「ユーザー別ランク」が算出されることになる。すなわち検索装置は、そのトラフィックランクスコアリング部がさらに「ユーザー別トラフィックランク算出手段」をさらに含むことを特徴とする。

【0110】

「ユーザー別トラフィックランク算出手段」は、ユーザー別にトラフィックランクを算

50

とする機能を有する。このユーザー別のトラフィックランクは、例えば前述の表示時間に応じたウェップページのトラフィックランク算出時に、ユーザー識別情報をを利用してユーザー単位で算出される、という具合である。

【0111】

また、上記ウェップページのランクの算出に際しては、同一ユーザーによる不正などを含めた同一リソースへの重複アクセスを排除してトラフィックランクを算出するよう構成しても良い。具体的には、閲覧情報に含まれるCookieなどをを利用して、同一ユーザーが同一ページへアクセスしたことを判断する。そして、そのように判断された表示時間に関しては、合計値ではなくその平均値などをを利用してトラフィックランクを算出する、という具合である。10

【0112】

また、上記のように集計されたトラフィックランクについて、前日のアクセス数からの増減率（バースト度）を算出し、その増減率をウェップページのランクに反映するような計算式を用いるよう構成しても良い。このような増減率を反映させることで、ウェップページの一一日ごとの盛り上がりを反映させることもできる。

【0113】

図12は、上記のようにして算出されたWebページ（リソース）ごとのトラフィックランクの一例を表す図である。この図にあるように、例えばその視聴回数やバースト度（前日のアクセス数からの増減率）を用いて、所定の関数によりリソースごとにトラフィックランクが「140」、「120」、「60」、「89」といった具合に算出される。そして、このようにして算出されたトラフィックランクに関して、本実施例の検索装置は当該リソースのURLと関連付けて図に示すようにテーブルデータなどとして保持する、という具合である。そしてそのトラフィックランクを利用して実施例4や5にて後述するよう「クローラ」の巡回頻度のスケジュール調整や検索結果の順位ソートなど様々な処理を実行することができる。20

【0114】

なお、例えばあるWebサイトの「トップページ」と「ページ1」「ページ2」、という具合に、通常は複数のリソースで一のWebコンテンツが構成されることが多い。したがってトラフィックランクスコアリング部で算出されるトラフィックランクは、URLごとでなくても構わない。例えば、同一Webサイトを構成するWebページであれば、IPアドレスやURLのドメインなどを比較することでそのことを判断し、URLは異なっていても一のトラフィックランクが算出されるよう構成しても良い。30

【0115】

<ハードウェア的構成>

【0116】

図13は、上記機能的な各構成要件をハードウェアとして実現した際の、検索システムにおける構成の一例を表す概略図である。この図を利用して本実施例の検索システムにおけるそれぞれのハードウェア構成部の働きについて説明する。

【0117】

この図にあるように、本実施例の検索システムのハードウェア構成は、ツールバー装置（1310）と検索装置（1320）とにより構成され、両者はインターネットを含む電気通信回線を介して、相互に接続可能となっている。40

【0118】

また、「ツールバー装置」（1310）は、実施例1と同様に「CPU」（1311）と「RAM」（1312）、「通信I/F」（1313）、「UI」（1314）、「VRAM」（1315）と「表示装置（ディスプレイなど）」（1316）と、を備える。

【0119】

そして「検索装置」（1320）も実施例1と同様に、「CPU」（1321）と「RAM」（1322）、「通信I/F」（1323）と、「二次記憶装置」（1324）と、を有する。50

【0120】

ここで実施例1にて説明したように、ツールバー装置のツールバープログラムの処理によって、閲覧URLがツールバー装置の「RAM」の所定アドレスに格納される。また、本実施例では、さらに以下のようにして取得された閲覧URLごとのアクセス時刻や閲覧(アクセス)時間などの情報が同様に「RAM」の所定アドレスに格納される。

【0121】

具体的に、例えば「UI」の操作入力に応じてブラウザプログラムが当該URLのリソース(ここではWebページ)のコンテンツを「通信I/F」にて受信した時刻を内蔵時計などで取得し、アクセス時刻として「RAM」に格納する。またブラウザプログラムによる処理を監視することで、当該URLで示されるウェップページが例えばディスプレイ上の最前面に表示されるよう制御されている時間や、ポインティングデバイスが当該ウェップページのウィンドウ上にあるよう制御されている時間を図示しないカウンタや内蔵時計などで計測し、閲覧時間として「RAM」に格納する、という具合である。10

【0122】

そしてツールバー装置ではツールバープログラムに従って、このように取得された閲覧URLや、当該URLで示されるリソースへのアクセス時刻や閲覧時間に、例えばCookieなどで示されるユーザーIDを加えて図2に示すようにテーブル化する。そしてそれら情報を閲覧情報として「通信I/F」から検索装置に対して送信する。

【0123】

検索装置では、ツールバー装置のツールバープログラムにしたがった処理により送信されてきた閲覧情報を「通信I/F」にて受信し「RAM」の所定アドレスに格納する。また閲覧情報は必要に応じて「二次記憶装置」に記録、蓄積されてもよい。20

【0124】

そして検索装置にて閲覧情報が格納されると、トラフィックランク算出プログラムに従い、以下のような処理が実行される。すなわち「RAM」上などに格納されている上記閲覧情報を参照し、例えば以下のようないくつかの関数を利用した演算処理を「CPU」において実行しトラフィックランクを算出する。

【数1】

$$R_a = \frac{t_a}{T} \times B_a \times N_I \times \frac{n_a}{N} \times r_x$$
30

R_a : ウエップページa のトラフィックランク

t_a : 全ユーザーによるページa の視聴時間合算値

T : 全ユーザーによる全ウエップページの視聴時間合算値

B_a : バースト値

N_I : 総閲覧ID数

n_a : 全ユーザーによるページa の閲覧回数合算値

N : 全ユーザーによる全ウエップページの閲覧回数合算値

r_x : ページa の関連(例えばリンク先など)ページx のトラフィックランク

【0125】

ここで、例えばウェップページaのトラフィックランク R_a は数1のような関数を用いて算出される。 t_a は全ユーザーIDによるページaの視聴時間の合算値である。 T は、全ユーザ50

- IDによる全URLの視聴時間合算値である。 B_a は次の数2で示された、バースト度である。なおバースト度については後述する。 N_1 は、総閲覧ID数である。また、 n_a は全ユーザーIDによるページaの閲覧回数の合算値である。 N は全ユーザーによる全ウェップページの閲覧回数合算値である。そして、 r_x はページaの関連ページxのトラフィックランクである。

【0126】

このように数1では、全ページの視聴時間(T)のうちのページaの視聴時間(t_a)の割合(t_a / T)を算出する。この数値によってクリックミスなどの表示時間が短いアクセスに関しては、そのトラフィックランクを低いものとして算出することができる。したがって前述のように、単純なアクセス数などでは推定することが困難なクリックミスなどを排除して、実体に即した指標としてリソースの視聴度指数を算出することができる。

【数2】

$$B_a = \frac{r_{at}}{r_{ay}}$$

B_a : ウェップページa のバースト度

20

r_{at} : 本日のページa の閲覧数

r_{ay} : 昨日のページa の閲覧数

【0127】

また、WebページaのURLのバースト度 B_a は、例えば数2で示す関数式で求めると良い。 r_{at} は、Webページaの本日の閲覧数である。また r_{ay} は、Webページaの昨日の閲覧数である。つまり数1で算出されるバースト度は、昨日の閲覧数を基準とした本日の閲覧数の増減率となる。したがって、通常時のアクセス数に比べて急激な伸びが認められる場合にはこの値が高くなる。つまり、そのページの注目度の高まりをバースト度によって表し、当該Webページ(リソース)のトラフィックランクに反映させることができる。

30

【0128】

また、その他にも総閲覧ID数(N_1)を関数の変数として利用することで以下のような効果が期待できる。すなわち、例えばあるサイトのトラフィックランクを不正に上げようとして、一ユーザーがあるサイトへ複数アクセスを実行した。しかし、そのような場合でも変数 N_1 を利用することで、一ユーザーによる複数アクセスよりも複数ユーザーによる複数アクセスの方がトラフィックランクを高く算出することができる、という具合である。

40

【0129】

また、トラフィックランクの算出対象であるWebページaのリンク先などの関連ページ x_1, x_2, \dots のトラフィックランク(r_x)を、上記関数の変数として利用しても良い。

【0130】

また、ここで利用される関連ページのトラフィックランク r_x は、例えば算出の対象となっているURLからハイパーリンクが設けられているページについてのみ、加重平均等でトラフィックランクを算出するものでもよい。また、あるURLから遷移する確率が一定値以上であるURLを取り出し、各URLが有するトラフィックランクを加重平均により算出したものを関連URLのトラフィックランクとしてもよい。

50

【0131】

また上記算出されるランクは、いわゆる「全国ランク」以外にも、前述のように「属性別ランク」や「個人ランク」であっても良い。具体的には、受信した閲覧情報に含まれるCookieなどを「RAM」に格納する。そしてそのCookieをキーとして、予め保持しているユーザー登録情報などからその閲覧情報で示される表示時間に関してセグメントや個人を特定する処理を「CPU」の演算処理によって実行する。そして、上記例えばURLおよび全URLの表示時間の合算処理において、特定された個人やセグメント別にその合算値を算出し、その表示時間の割合を「CPU」の演算処理によって算出する、という具合である。

【0132】

そして、このようにして閲覧URLごとに算出したトラフィックランクを、「二次記憶装置」にトラフィックランクデータベースとして記録、保持する。そして実施例4や5で後述するように、「クローラ」の巡回頻度のスケジュール調整や検索結果の順位ソートなど際に当該トラフィックランクデータベースを参照する、という具合である。

【0133】

<処理の流れ>

【0134】

図14は、本実施例の検索システムにおける処理の流れの一例を表すフローチャートである。なお、以下に示すステップは、媒体に記録され計算機を制御するためのプログラムを構成する処理ステップであっても構わない。

20

【0135】

この図にあるように、ツールバー装置にてツールバーが起動されると、まず閲覧URLなどを取得する(ステップS1401)。次にユーザーのブラウザの操作により閲覧URLが更新(変更)されたか否かを判断する(ステップS1402)。更新されたと判断された場合には当該URLのアクセス時間を取得する(ステップS1403)。そして取得した閲覧URLやアクセス時間を含む情報を閲覧情報として閲覧管理サーバ装置を含む検索装置に送信する(ステップS1404)。

【0136】

次に検索装置において、この図にあるように、n個のURLと、当該URLに関連付けたユーザーIDや閲覧時間などの情報を含む閲覧情報をツールバー装置より受信する(ステップS1411)。次に受信した閲覧情報を主メモリなどに格納する(ステップS1412)。その後、ループ1(ステップS1413)を開始とする以下の処理を実行する。まず蓄積されたn個の閲覧情報を関数の変数として利用し、トラフィックランクを算出する(ステップS1414)。そして、n個のURLごとに算出したトラフィックランクを、例えばHDDなどにデータベースとして記録する。そして例えば当該データベースを参照し「クローラ」の巡回頻度のスケジュール調整を実行したり、検索結果の順位ソートを実行したりする。

30

【0137】

<効果の簡単な説明>

【0138】

このように本実施例の検索システムによって、誤アクセスなどを排除して集計した実効的なリソースのアクセス指標(視聴度指数)であるトラフィックランクを算出することができる。また、上記のような関数を利用すれば、さらに単にWebページのURL毎のアクセス数や、ユニークユーザー数などだけではなく、注目度の高まりなどを含めた総合的な評価指標によりWebページを評価することができるトラフィックランクを算出することができる。

40

【0139】

実施例4

【0140】

<概要>

50

【0141】

本実施例は、上記実施例3を基本として、算出したトラフィックランクをリソースの人気度の指標として利用し、クローラの巡回優先度の決定に利用することを特徴とする検索システムである。

【0142】

図15は、本実施例の検索システムにおける処理の一例を説明するための概念図である。この図にあるように、検索装置は上記実施例3で記載した構成、処理によって、トラフィックランクを算出し、データベースとして保持している。そしてさらに、本実施例の検索装置では、例えば「TR(トラフィックランク)：60」のサイトCに対しては「巡回優先度：12」、一方「TR：20」のサイトDであれば「巡回優先度：4」という具合に、トラフィックランクに応じて巡回優先度を算出していることを特徴とする。10

【0143】

そして、このトラフィックランクに応じた優先度を利用して、本実施例の検索装置では、クローラの巡回スケジュールを、例えばサイトCであれば「3回／週」としてスケジューリングし、一方優先度(トラフィックランク)の低いサイトDは「1回／週」としてスケジューリングする、という具合である。

【0144】

つまり、前述のようにインターネットユーザーは更新頻度が高いWebページを中心的に検索しアクセスする傾向がある。これは逆に言えば、アクセス頻度の高いリソースは頻繁に更新などされている可能性が高い、ということである。そして、上記実施例3で説明したように「トラフィックランク」は、ユーザーの実効的なアクセス数を示す指標である。したがって、このトラフィックランクによってユーザーの実効的なアクセス数に応じてクローラの巡回優先度の決定ができる、ということである。20

【0145】

<機能的構成>

【0146】

図16は、本実施例の検索システムにおける機能ブロックの一例を表す図である。この図にあるように、本実施例の「検索システム」(1600)は、実施例3を基本として「ツールバー装置」(1610)と、「検索装置」(1620)とを有する。また「ツールバー装置」(1610)は、実施例3と同様に「閲覧情報送信部」(1611)と、「閲覧情報取得部」(1612)とを有する。30

【0147】

また「検索装置」(1620)も、実施例3を基本として「第二閲覧管理サーバ装置」(1630)を有し、その第二閲覧管理サーバ装置が「閲覧情報蓄積部」(1631)と、「トラフィックランクスコアリング部」(1632)とを有する。なお、上記ツールバー装置及び検索装置の各構成は、実施例1や3にて既に記載済みであるのでその説明は省略する。

【0148】

そして、本実施例の特徴点は、検索装置が、さらに「クローラ部」(1621)と、「スケジュール決定部」(1622)と、を有する点である。40

【0149】

「クローラ部」(1621)は、インターネット上のリソース間に張られたハイパーインクをたどりWWW上を巡回することで各リソースにアクセスしコンテンツを収集するいわゆる「クローラ」プログラムによって実現することができる。

【0150】

また、このクローラプログラムは、通常その巡回対象となるリソースのURLやその巡回開始時刻などを予め定めたスケジュールに従ってコンテンツを収集する機能を有する。そして、本実施例では、下記の構成によってそのスケジューリングにおける巡回頻度を、トラフィックランクに応じて立案することを特徴とする。

【0151】

「スケジュール決定部」(1622)は、トラフィックランクスコアリング部(1632)で算出されたトラフィックランクに基づいてクローラ部(1621)のクローリングスケジュールを決定する機能を有する。

【0152】

具体的には、例えばトラフィックランクを変数とする比例関数 $y = f(TR)$ を利用したCPUの演算処理によって当該URLに対するクローラプログラム実行の巡回優先度yを算出する。そして算出した巡回優先度に基づいて、例えば「3回/週(巡回優先度12)」、「1回/週(巡回優先度4)」といった具合にクローラプログラムの巡回頻度を含むスケジュールを決定する、という具合である。

【0153】

このようにして、ユーザーの実効的なアクセス数を示すトラフィックランクによってクローラの巡回優先度の決定し、更新頻度が高く情報鮮度が高いと思われるWWW上のリソースの情報を検索装置にて好適なタイミングで更新取得することができる。

【0154】

<ハードウェア的構成>

【0155】

図17は、上記機能的な各構成要件をハードウェアとして実現した際の、検索システムにおける構成の一例を表す概略図である。この図を利用して本実施例の検索システムにおけるそれぞれのハードウェア構成部の働きについて説明する。

【0156】

この図にあるように、本実施例の検索システムのハードウェア構成は、ツールバー装置(1710)と検索装置(1720)より構成され、両者はインターネットを含む電気通信回線を介して、相互に接続可能となっている。

【0157】

なお「ツールバー装置」(1710)は、実施例3のツールバー装置と同様の構成、及び処理を行うため、ここでの説明は省略する。そして「検索装置」(1720)の構成は、上記実施例と同様に、「CPU」(1721)と「RAM」(1722)、「通信I/F」(1723)と、「二次記憶装置」(1724)と、を有し、以下のようないくつかの処理を実行する。

【0158】

すなわちツールバー装置から上記実施例3にて説明したような処理によって送信されてきた、閲覧URL、及びその閲覧URLごとのアクセス時刻や閲覧(アクセス)時間などの情報を含む閲覧情報を、検索装置は「通信I/F」にて受信し、「RAM」の所定アドレスに格納する。

【0159】

そして、同じく上記実施例3にて記載したような処理を実行し、「RAM」上に格納されている上記閲覧情報を用いてURLごとのトラフィックランクを算出し、トラフィックランクデータベースとして「二次記憶装置」に記録、保持する。

【0160】

つづいてスケジュール決定プログラムに従い以下のようないくつかの処理が実行される。すなわちクローラの巡回スケジュールを立てるべきURLをキーとして「二次記憶装置」に保持されているトラフィックランクデータベースを参照し、当該URLのトラフィックランクの値TRを取得する。そして例えば $y = f(TR)$ といった関数を利用した演算処理を「CPU」において実行し、当該トラフィックランクに係るURLを対象としたクローラプログラムの巡回優先度yを算出する。そして、「二次記憶装置」に保持されている「巡回優先度 - 巡回頻度テーブル」を参照し、算出された巡回優先度yと対応付けられた、例えば「3回/週」といった巡回頻度を決定する。

【0161】

そして決定された「3回/週」といった巡回頻度に応じてタスクスケジューラなどでクローラプログラムの実効日時を予約し、決定された頻度でのクローラプログラムによるリ

10

20

30

40

50

ソース巡回、コンテンツ取得処理が実行される、という具合である。

【0162】

<処理の流れ>

【0163】

図18は、本実施例の検索システムにおける処理の流れの一例を表すフローチャートである。なお、以下に示すステップは、媒体に記録され計算機を制御するためのプログラムを構成する処理ステップであっても構わない。

【0164】

まず、ツールバー装置における処理の流れについては、図14を用い実施例3で述べた処理(S1401～S1404)と同様であるためその説明は省略する。

10

【0165】

次に検索装置において、この図にあるように、n個のURLと、当該URLに関連付けたユーザーIDや閲覧時間などの情報を含む閲覧情報をツールバー装置より受信する(ステップS1811)。次に受信した閲覧情報を主メモリなどに格納する(ステップS1812)。その後、ループ1(ステップS1813)を開始とする以下の処理を実行する。まず蓄積されたn個の閲覧情報を、例えば前述の関数の変数として利用しトラフィックランクを算出する(ステップS1814)。そして、算出したトラフィックランクに基づいて当該URLに対するクローラ部のクローリングスケジュールを決定する(ステップS1815)。そして上記処理をn個のURLに関して繰り返して実行する(ステップS1816)。

20

【0166】

<効果の簡単な説明>

【0167】

このように本実施例の検索システムによって、ユーザーの実効的なアクセス数を示すトラフィックランクによってクローラの巡回優先度の決定することができる。したがって、更新頻度が高く情報鮮度が高いと思われるWWW上のリソースの情報を検索装置にて好適なタイミングで更新取得することができる。

【0168】

実施例5

【0169】

<概要>

【0170】

本実施例は、上記実施例3や4を基本として、算出したトラフィックランクを利用して、ツールバー装置からの検索リクエストに対する検索結果をソートすることを特徴とする検索システムである。

30

【0171】

図19は、本実施例の検索システムにおける検索処理の一例を説明するための概念図である。この図にあるように、クライアント端末が当該検索システムにおける検索用Webページにアクセスし、検索クエリの入力、送信を行う。すると検索装置は従来の検索システム同様に検索クエリを含むリソースを検索用のインデックスファイルから抽出する。そして、抽出されたリソースをリスト化などし、検索結果用画面を生成する。

40

【0172】

ここで、本実施例の検索装置は従来と異なる以下のようないくつかの処理をさらに実行する。すなわち、検索結果で示されるURL(のリスト)に係るトラフィックランク値の大小に応じて、例えばサイトBを検索結果(リスト)の1番上位にソートし、サイトAを2番目にソートする、といった検索結果の並び替えを実行する、という具合である。

【0173】

このようにして、本実施例の検索システムではユーザーの実効的なアクセス数を示すトラフィックランクに応じた検索結果をクライアント端末に返信することができる。

【0174】

50

<機能的構成>

【0175】

図20は、本実施例の検索システムにおける機能ブロックの一例を表す図である。この図にあるように、本実施例の「検索システム」(2000)は、実施例3を基本として「ツールバー装置」(2010)と、「検索装置」(2020)とを有する。また「ツールバー装置」(2010)は、実施例3と同様に「閲覧情報送信部」(2011)と、「閲覧情報取得部」(2012)とを有する。

【0176】

また「検索装置」(2020)も、実施例3を基本として「第二閲覧管理サーバ装置」(2030)を有し、その第二閲覧管理サーバ装置が「閲覧情報蓄積部」(2031)と、「トライフィックランクスコアリング部」(2032)とを有する。また、上記実施例4を基本として、検索装置が図示しない「クローラ部」や「スケジュール決定部」を有していても良い。なお、上記ツールバー装置及び検索装置の各構成は、上記各実施例にて既に記載済みであるのでその説明は省略する。10

【0177】

そして、本実施例の特徴点は、検索装置が、「ランクソート出力部」(2021)をさらに有する点である。

【0178】

「ランクソート出力部」(2021)は、トライフィックランクスコアリング部(2032)で算出されたトライフィックランクに基づいて検索結果をソートしてクライアントに対して出力する機能を有する。20

【0179】

具体的に「ランクソート出力部」による前記ソートの前に、本実施例の検索システムでは従来の検索システム同様の検索処理が実行される。すなわち、クライアント端末にて送信された検索クエリをキーとして、検索用インデックスファイルを検索する。そして検索クエリを含むリソースをインデックスファイルから抽出する。

【0180】

続いて、その抽出したリソースを並び替えて例えば箇条形式(リスト形式)などの検索結果としてクライアントに返信することになる。ここで通常の検索システムでは、検索結果の並び替えを、例えばデータ生成や取得の古い順/新しい順、アクセス数順、あいうえお順、あるいはリンクを利用して付されるリソースの再帰的な格付け、などに応じて実行する。しかし、本実施例の検索システムでは、実施例3で算出したユーザーの実体的なアクセス数を示すトライフィックランクを利用する事を特徴とする。すなわち、検索によって抽出されたリソースのトライフィックランクを、上記実施例3にて記載したトライフィックランクデータベースなどから取得する。そしてトライフィックランクの大小比較をCPUの演算処理によって実行し、例えばトライフィックランクの大きい順にリソースを並び替え(ソートして)検索結果とする、という具合である。30

【0181】

このようにして、ユーザーの実効的なアクセス数、すなわち実体的なアクセス人気度を示すトライフィックランクに応じた検索結果をクライアント端末に返信することができる。40

【0182】

<ハードウェア的構成>

【0183】

図21は、上記機能的な各構成要件をハードウェアとして実現した際の、検索システムにおける構成の一例を表す概略図である。この図を利用して本実施例の検索システムにおけるそれぞれのハードウェア構成部の働きについて説明する。

【0184】

この図にあるように、本実施例の検索システムのハードウェア構成は、ツールバー装置(2110)と検索装置(2120)より構成され、両者はインターネットを含む電気通信回線を介して、相互に接続可能となっている。50

【0185】

なお「ツールバー装置」(2110)は、実施例3や4のツールバー装置と同様の構成、及び処理を行うため、ここでの説明は省略する。そして「検索装置」(2120)の構成は、上記実施例と同様に、「CPU」(2121)と「RAM」(2122)、「通信I/F」(2123)と、「二次記憶装置」(2124)と、を有し、以下のような処理を実行する。

【0186】

すなわちツールバー装置から送信された閲覧URL、及びその閲覧URLごとのアクセス時刻や閲覧(アクセス)時間などの情報を含む閲覧情報をを利用して、検索装置にてトラフィックランクが算出され「二次記憶装置」に記録、保持される。

10

【0187】

その後、ネットワーク上のツールバー装置が検索用Webページにアクセスし、検索クエリの入力、送信を行うと、検索装置は「通信I/F」にてその検索クエリを含むHTTPリクエストを受信し、「RAM」の所定アドレスに格納する。すると、検索装置では検索サーバプログラムに従い以下の処理を実行する。すなわち「RAM」に格納されている検索クエリをキーとして、「二次記憶装置」に保持されている検索用のインデックスファイルを参照し、検索クエリを含むリソースを抽出する。

【0188】

つづいて、同じく「二次記憶装置」に保持されているトラフィックランクデータベースを参照し、抽出したリソースのトラフィックランクを取得する。そして、「CPU」の比較演算処理によってトラフィックランクの大小比較を実行し、例えばトラフィックランクの大きい(ランクが高い)順に検索結果の並び順を決定する。そして決定された並び順に従って抽出したリソースを並べて箇条形式とした検索結果を「CPU」の演算処理によって生成し、「通信I/F」より検索クエリの送信元のツールバー装置に対して返信する、という具合である。

20

【0189】

<処理の流れ>

【0190】

図22は、本実施例の検索システムにおける処理の流れの一例を表すフローチャートである。なお、以下に示すステップは、媒体に記録され計算機を制御するためのプログラムを構成する処理ステップであっても構わない。

30

【0191】

まず、ツールバー装置が組み込まれるなどしたクライアント端末における処理の流れについては、図14を用い実施例3で述べた処理(S1401～S1404)と同様であるためその説明は省略する。また検索装置におけるトラフィックランクの算出(ステップS2211)までの処理も実施例3で述べた処理(S1411～S1414)までと同様であるのでその説明は省略する。

【0192】

そして、この図にあるように、検索装置ではネットワーク上のクライアント端末(ツールバー装置が組み込まれていなくとも構わない)から送信された検索クエリを受信する(ステップS2212)と、検索クエリをキーとして検索用インデックスファイルの検索処理を実行する(ステップS2213)。そして検索クエリをコンテンツに含むリソースを抽出し検索結果として取得する(ステップS2214)。

40

【0193】

つづいて、抽出したリソースのトラフィックランクを取得し(ステップS2215)、そのトラフィックランクの例えれば大小順に応じて検索結果のリソースを並び替える(ステップS2216)。そして並び替えた検索結果を、検索クエリの送信元のクライアント端末に対して返信する(ステップS2217)。

【0194】

<効果の簡単な説明>

50

【0195】

このように本実施例の検索システムによって、ユーザーの実効的なアクセス数、すなわち実体的なアクセス人気度を示すトラフィックランクに応じた検索結果をクライアント端末に返信することができる。

【0196】

実施例 6

【0197】

<概要>

【0198】

本実施例は、実施例 2 と同様に新規の検索対象を抽出し、抽出した新規検索対象 URL を利用して検索用のインデックスファイルを追加更新する機能を備える。そして、実施例 2 との相違点は、その新規検索対象 URL の抽出において、実施例 3 などにて説明したトラフィックランクを利用する点である。10

【0199】

具体的には、検索装置において、クライアント端末のツールバー装置にて閲覧情報を取得し、上記実施例で記載したように URL ごとにトラフィックランクを算出する。ここで、実施例 2 と同じようにその URL が検索用のインデックスファイルに含まれるかを判断する。そして、インデックスファイルに含まれておらず、かつ例えばトラフィックランクが所定値以上であれば、新規にインデックスファイルに追加するに相応しいリソースであるとして、クローラの新規巡回検索対象として、実施例 2 同様にインデクシング処理を実行する、という具合である。20

【0200】

<機能的構成>

【0201】

図 23 は、本実施例の検索システムにおける機能ブロックの一例を表す図である。この図にあるように、本実施例の「検索システム」(2300)は、実施例 3 を基本として「ツールバー装置」(2310)と、「検索装置」(2320)とを有する。また「ツールバー装置」(2310)は、実施例 3 と同様に「閲覧情報送信部」(2311)と、「閲覧情報取得部」(2312)とを有する。

【0202】

また「検索装置」(2320)も、実施例 3 を基本として「第二閲覧管理サーバ装置」(2330)を有し、その第二閲覧管理サーバ装置が「閲覧情報蓄積部」(2331)と、「トラフィックランクスコアリング部」(2332)とを有する。また、上記実施例 4 や 5 を基本として、検索装置が図示しない「クローラ部」や「スケジュール決定部」、「ランクソート出力部」を有していても良い。なお、上記ツールバー装置及び検索装置の各構成は、上記各実施例にて既に記載済みであるのでその説明は省略する。30

【0203】

そして、本実施例の特徴点は、検索装置が、「第二インデクサ部」(2321)をさらに有する点である。

【0204】

「第二インデクサ部」(2321)は、トラフィックランクスコアリング部で算出されたトラフィックランクに基づいて検索エンジンにて新規に検索対象として利用する URL である新規検索対象 URL を抽出するとともに、抽出した URL のコンテンツをインデクシングする機能を有する。40

【0205】

なお、算出されたトラフィックランクに係る URL を、クローラの新規巡回検索の対象とするための処理については、実施例 2 にて記載したものと同様であるのでその説明は省略する。また、本実施例においては、インデックスファイル中の URL とトラフィックランクに係る URL の差分のみならず、その URL で示されるリソースの視聴度指数であるトラフィックランクを利用するため、実施例 2 で記載した処理に加えさらに以下のような50

処理を実行しても良い。

【0206】

すなわち、トラフィックランクが所定値以上であるか否かの比較判断処理を実行し、所定値以上であればクローラの新規巡回検索対象として相応しいリソースである、と判断する。一方、トラフィックランクが所定値以下であると判断された場合には、クローラの新規巡回検索対象としてインデックスファイルには追加しない、という具合である。

【0207】

<ハードウェア的構成>

【0208】

図24は、上記機能的な各構成要件をハードウェアとして実現した際の、検索システムにおける構成の一例を表す概略図である。この図を利用して本実施例の検索システムにおけるそれぞれのハードウェア構成部の働きについて説明する。10

【0209】

この図にあるように、本実施例の検索システムのハードウェア構成は、ツールバー装置(2410)と検索装置(2420)より構成され、両者はインターネットを含む電気通信回線を介して、相互に接続可能となっている。

【0210】

なお「ツールバー装置」(2410)は、実施例3や4、5のツールバー装置と同様の構成、及び処理を行うため、ここでの説明は省略する。そして「検索装置」(2420)の構成は、上記実施例と同様に、「CPU」(2421)と「RAM」(2422)、「通信I/F」(2423)と、「二次記憶装置」(2424)と、を有し、以下のようないくつかの処理を実行する。20

【0211】

まず、上記実施例と同様の処理により、ツールバー装置から送信された閲覧URL、及びその閲覧URLごとのアクセス時刻や閲覧(アクセス)時間などの情報を含む閲覧情報を利用して、検索装置にてトラフィックランクが算出され「二次記憶装置」に記録、保持される。

【0212】

つづいて、予め「二次記憶装置」にて格納されている検索サービス用のインデックスファイルを参照し、「二次記憶装置」の保持されたトラフィックランクに係るURLをキーとした検索処理を実行する。そして、その検索処理の結果、インデックスファイルに当該URLが含まれていないとの判断結果が出力されれば、そのURLは新規検索対象のURL候補であると判断する。30

【0213】

つづいて、第二インデクサプログラムに従い、「CPU」は前記候補URLに係るトラフィックランクを「RAM」に格納し、予め「二次記憶装置」に保持されている所定値との大小比較処理を実行する。そしてその大小比較処理の結果トラフィックランクが所定値以上であると判断されれば、当該URLは新規検索対象のURLとしてインデクシング処理(二次記憶装置に保持されているインデックスファイルへの追加更新処理)が実行される、という具合である。40

【0214】

<処理の流れ>

【0215】

図25は、本実施例の検索システムにおける処理の流れの一例を表すフローチャートである。なお、以下に示すステップは、媒体に記録され計算機を制御するためのプログラムを構成する処理ステップであっても構わない。

【0216】

まず、ツールバー装置における処理の流れについては、図14を用い実施例3で述べた処理(S1401～S1404)と同様であるためその説明は省略する。

【0217】

10

20

30

40

50

次に検索装置において、この図にあるように、n個のURLと、当該URLに関連付けたユーザーIDや閲覧時間などの情報を含む閲覧情報をツールバー装置より受信する(ステップS2511)。次に受信した閲覧情報を主メモリなどに格納する(ステップS2512)。その後、ループ1(ステップS2513)を開始とする以下の処理を実行する。まず蓄積されたn個の閲覧情報を関数の変数として利用し、トラフィックランクを算出する(ステップS2514)。

【0218】

つづいて検索装置のインデックスファイルに当該URLが存在するか否かの第一の判断処理を行う(ステップS2515)。そして第一の判断処理の結果、インデックスファイルに当該URLが存在しないとの判断結果が出力された場合、当該URLに係るトラフィックランクが所定値以上であるか否かの判断処理を実行する(ステップS2516)。10

【0219】

そして第二の判断処理の結果、トラフィックランクが所定値以上であるとの判断結果が出力されれば、クローラの新規巡回検索対象として相応しいリソースである、として抽出したURLを、クローラの新規検索対象URLとしてインデクシング処理を実行する(ステップS2517)。

【0220】

<効果の簡単な説明>

【0221】

このように本実施例の検索システムによって、インデックスファイルへの追加(インデクサ)を行うに相応しいリソースであるかを判断した上で、新規開設Webページなどハイパーアリンクが張られていないため従来のクローラでは検索対象となっていたようなWWW上のリソースについてもインデクサを実行することができる。20

【図面の簡単な説明】

【0222】

【図1】実施例1の検索システムにおける閲覧情報収集の一例を説明するための概念図

【図2】実施例1の検索システムの検索装置に収集される閲覧情報の一例を表す図

【図3】実施例1の検索システムにおける機能ブロックの一例を表す図

【図4】実施例1の検索システムにおけるハードウェア構成の一例を表す図

【図5】実施例1の検索システムにおける処理の流れの一例を表すフローチャート

30

【図6】実施例2の検索システムにおける処理の一例を説明するための概念図

【図7】実施例2の検索システムにおける機能ブロックの一例を表す図

【図8】実施例2の検索システムにおけるハードウェア構成の一例を表す図

【図9】実施例2の検索システムにおける処理の流れの一例を表すフローチャート

【図10】実施例3の検索システムにおける処理の一例を説明するための概念図

【図11】実施例3の検索システムにおける機能ブロックの一例を表す図

【図12】実施例3の検索システムのトラフィックランクスコアリング部にて算出されたWebページごとのトラフィックランクの一例を表す図

【図13】実施例3の検索システムにおけるハードウェア構成の一例を表す図

【図14】実施例3の検索システムにおける処理の流れの一例を表すフローチャート

40

【図15】実施例4の検索システムにおける処理の一例を説明するための概念図

【図16】実施例4の検索システムにおける機能ブロックの一例を表す図

【図17】実施例4の検索システムにおけるハードウェア構成の一例を表す図

【図18】実施例4の検索システムにおける処理の流れの一例を表すフローチャート

【図19】実施例5の検索システムにおける処理の一例を説明するための概念図

【図20】実施例5の検索システムにおける機能ブロックの一例を表す図

【図21】実施例5の検索システムにおけるハードウェア構成の一例を表す図

【図22】実施例5の検索システムにおける処理の流れの一例を表すフローチャート

【図23】実施例6の検索システムにおける機能ブロックの一例を表す図

【図24】実施例6の検索システムにおけるハードウェア構成の一例を表す図

50

【図25】実施例6の検索システムにおける処理の流れの一例を表すフローチャート

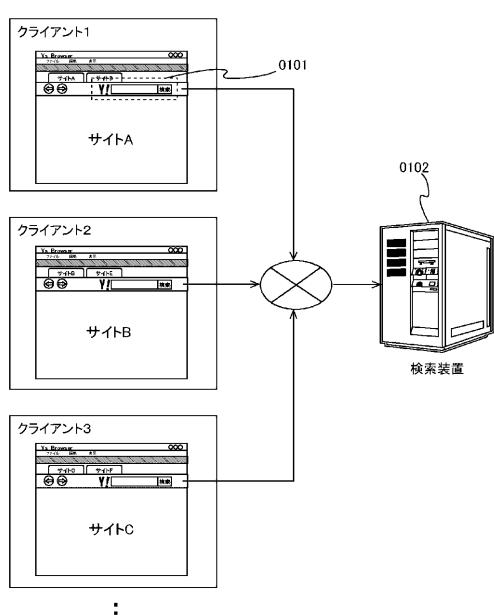
【符号の説明】

【0223】

- 0300 検索システム
- 0310 ツールバー装置
- 0311 閲覧情報取得部
- 0312 閲覧情報送信部
- 0320 検索装置
- 0730 第一閲覧管理サーバ装置
- 0731 第一インデクサ部
- 1130 第二閲覧管理サーバ装置
- 1131 閲覧情報蓄積部
- 1132 トラフィックランクスコアリング部

10

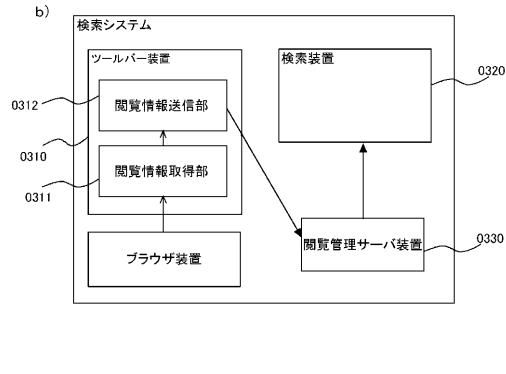
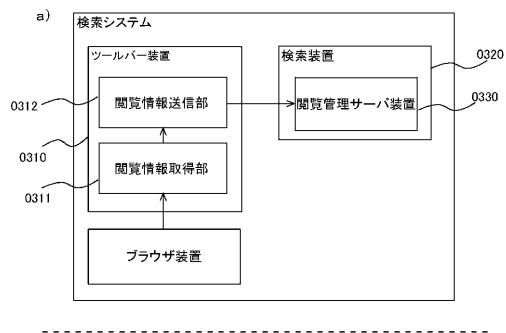
【図1】



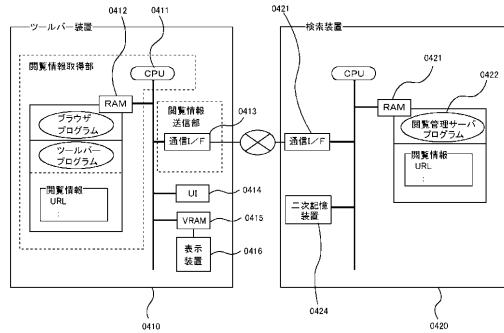
【図2】

ユーザID	URL	閲覧開始時刻	閲覧時間
0X02A	http://yahoo.co.jp/	16:47:04:05	00:02:20:03
0X02A	http://list3.auctions.yahoo.co.jp/jp/2336-category.html	16:49:24:08	00:03:48:05
0X02A	http://aaaxx.com/blog/070425/	16:53:12:13	00:02:13:07
0X02A	http://bbb.yyy.net/blog/070426/	16:55:25:20	00:04:25:27
:	:	:	:

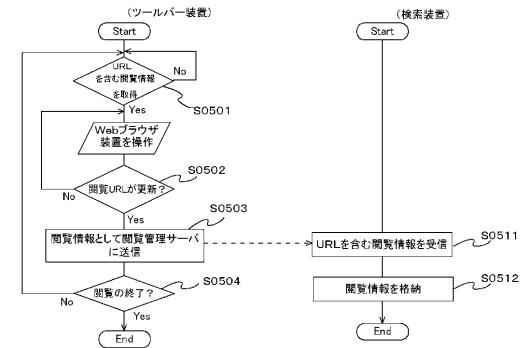
【図3】



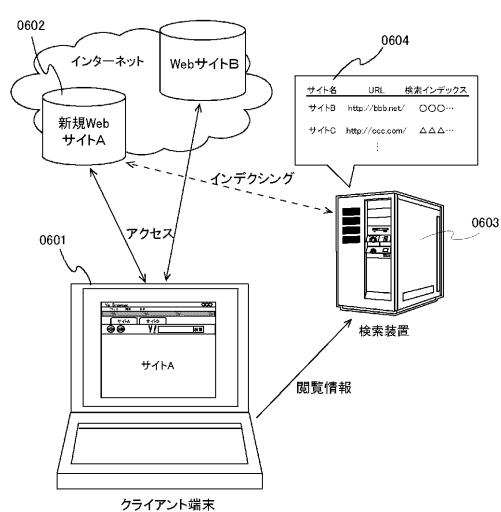
【図4】



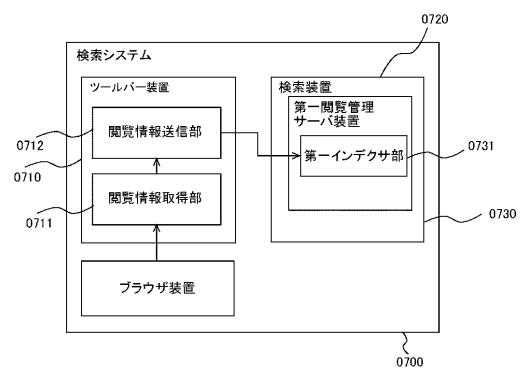
【図5】



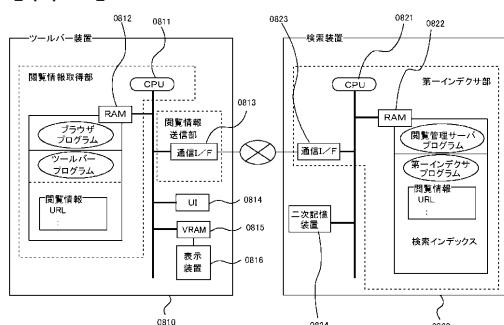
【図6】



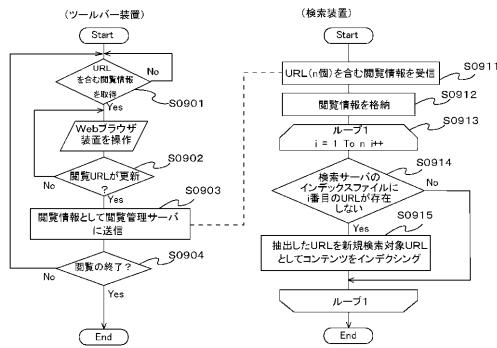
【図7】



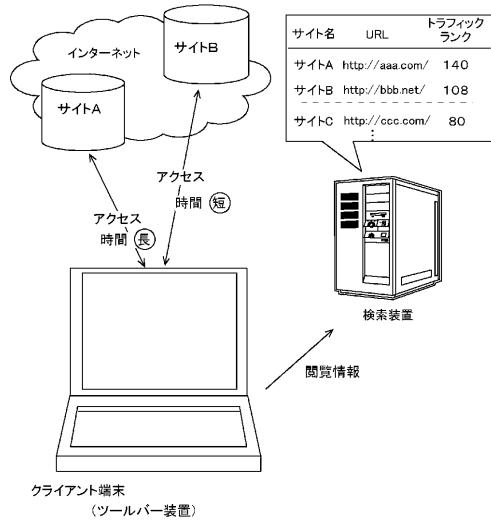
【図8】



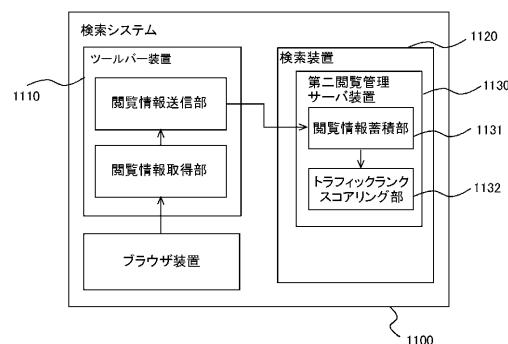
【図9】



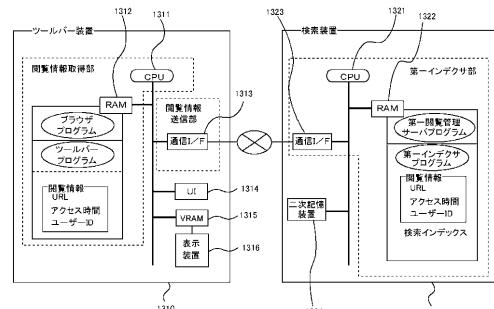
【図10】



【図11】



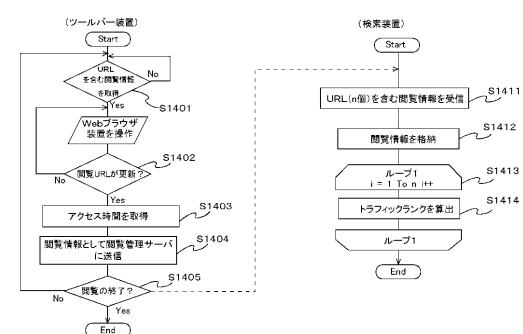
【図13】



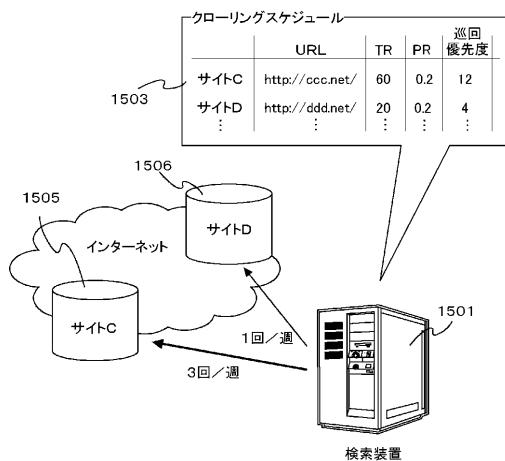
【図12】

URL	視聴回数				バースト度	TR
	日別	週別	月別	年別		
http://yahoo.co.jp/news/37487.html	1525	39874	655702	4952160	1. 63	140
http://list3.auctions.yahoo.co.jp/~2338-category.html	324	452	452	452	12. 4	120
http://aaaxx.com/blog/070425/	190	736	1925	1925	4. 7	60
http://bbb.yyyy.net/blog/070426/	467	480	480	480	35. 9	89
⋮	⋮	⋮	⋮	⋮	⋮	⋮

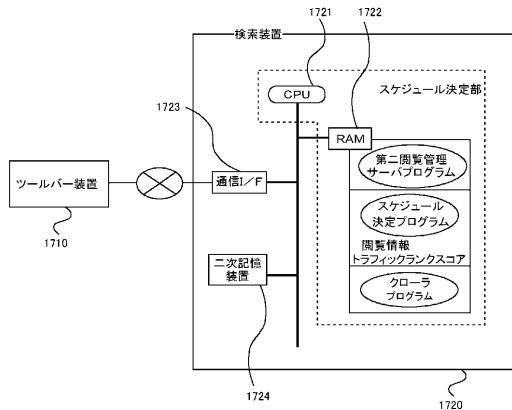
【図14】



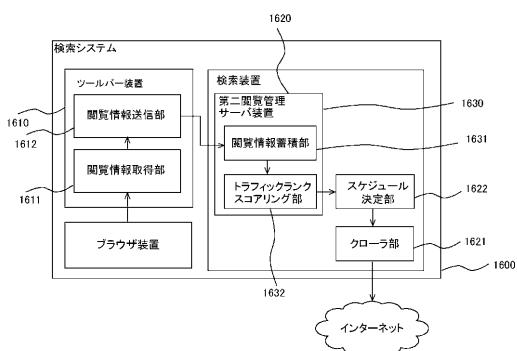
【図15】



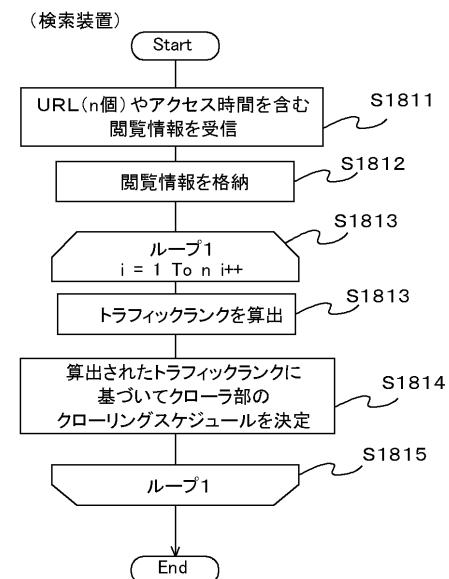
【図17】



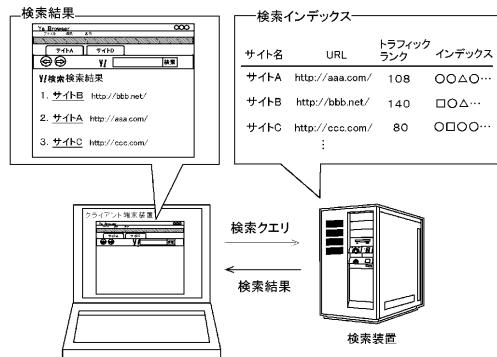
【図16】



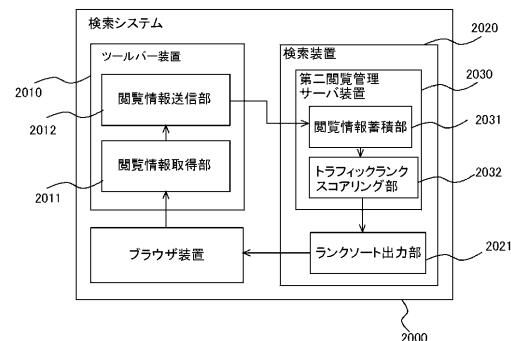
【図18】



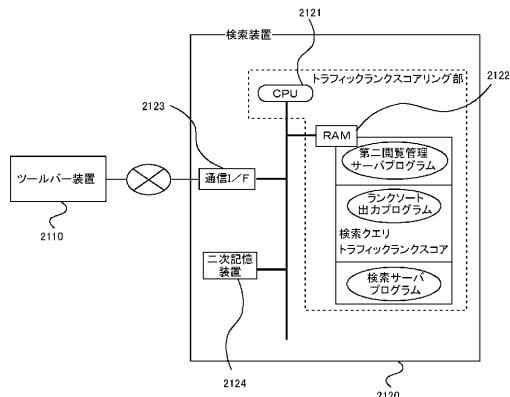
【図19】



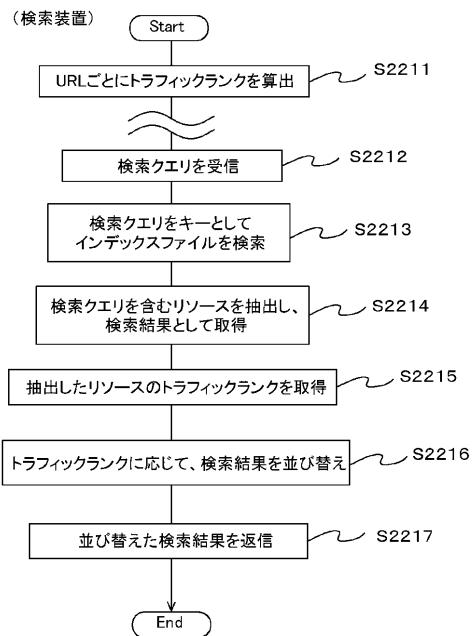
【図20】



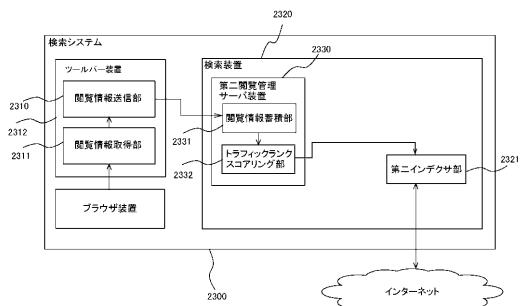
【図21】



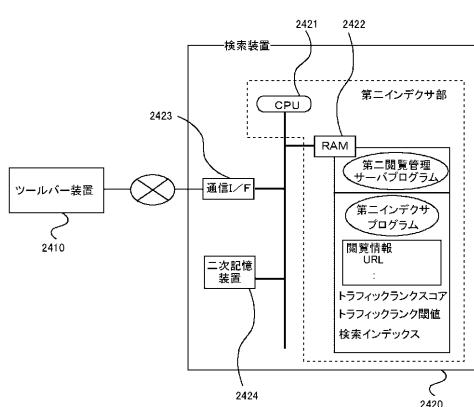
【図22】



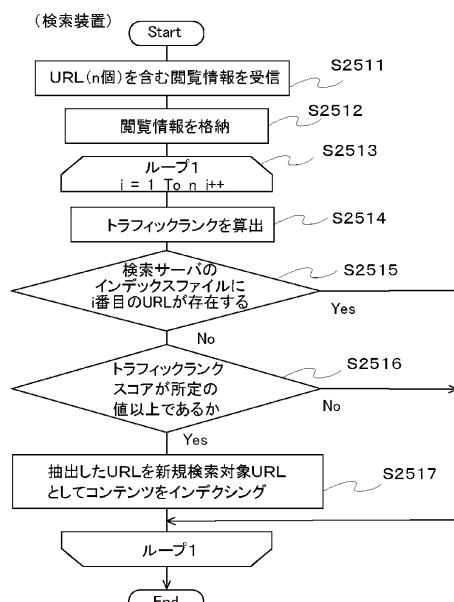
【図23】



【図24】



【図25】



フロントページの続き

(56)参考文献 特開2002-117206(JP,A)

特開2005-190065(JP,A)

特開2001-290843(JP,A)

特開2006-277732(JP,A)

木村 祥久 Yoshihisa Kimura, アクセス履歴を利用したWeb情報の共有手法, 第60回(

平成12年前期)全国大会講演論文集(3) データベースとメディア ネットワーク, 日本, 社団法人情報処理学会, 2000年 3月14日, 3-151~3-152ページ

森 隆宣 Takanobu MORI, Webキャッシュサーバーを用いたユーザー適応型情報収集エージェントWebProbe: User Adaptive Information Retrieval Agent with Web Cache Server, マルチメディア, 分散, 協調とモバイル(DICOMO)シンポジウム論文集

1997年~2006年版 Ver.1.1 [DVD-ROM] Multimedia,Distributed,Cooperative and Mobile Symposium, 日本, 社団法人情報処理学会, 2006年 8月31日, 第2000巻, 133~138ページ

(58)調査した分野(Int.Cl., DB名)

G 06 F 13 / 00

G 06 F 17 / 30