(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2006-23936
(P2006-23936A)
(43)公開日　平成18年1月26日(2006.1.26)

(51)Int.Cl.		F I			テーマコード（参考）
G06F 17/30	(2006.01)	G06F	17/30	340A	5B075
		G06F	17/30	180Z	

審査請求 未請求　請求項の数 6　OL　（全 10 頁）

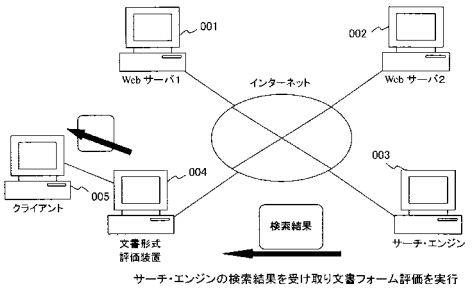
(21)出願番号　特願2004-200717 (P2004-200717)
(22)出願日　平成16年7月7日 (2004.7.7)

(71)出願人　000233055
日立ソフトウエアエンジニアリング株式会社
神奈川県横浜市鶴見区末広町一丁目１番４３
(74)代理人　100091096
弁理士　平木　祐輔
(74)代理人　100105463
弁理士　関谷　三男
(74)代理人　100102576
弁理士　渡辺　敏章
(74)代理人　100108394
弁理士　今村　健一

最終頁に続く

(54)【発明の名称】コンテンツ検索システム

(57)【要約】
【課題】　ユーザーが指定した文書形式条件を満たすWebコンテンツを提供する。
【解決手段】　インターネットによって接続される第１Webサーバ００１、第２Webサーバ００２と、サーチ・エンジン００３と、文書形式評価装置００４と、クライアント端末００５と、を有している。クライアント端末００５は、例えばパーソナルコンピュータであり、ネットワークを介して第１又は第２Webサーバ００１、００２上のWebコンテンツを閲覧する。クライアント端末００５は、検索する語句及び文書形式に関する検索条件を、本実施の形態による文書形式評価装置００４に送信する。文書形式評価装置００４は、検索対象語句をサーチ・エンジン００３に送信する。サーチ・エンジンは検索対象語句に基づいてコンテンツの検索を実行し、検索対象語句を含むWebコンテンツの一覧を文書形式評価装置００４に返す。この一覧に含まれるWebコンテンツに対して文書形式を評価し、ユーザーが指定した条件に適合するコンテンツのみをユーザーへ返す。
【選択図】　図１



【特許請求の範囲】
【請求項１】
Webコンテンツを検索するコンテンツ検索システムであって、
コンテンツ検索の際にコンテンツの文書形式に基づいてフィルタリングを行う文書形式評価装置を有することを特徴とするコンテンツ検索システム。
【請求項２】
クライアント端末及びコンテンツを提供するWebサーバとネットワークを介して接続され、Webコンテンツを検索するコンテンツ検索システムであって、
コンテンツの検索エンジンを備える検索装置と、
該コンテンツの検索結果として得られたコンテンツに関して文書形式条件に基づいて文書形式の評価を行う文書形式評価装置であって前記文書形式条件に合致したコンテンツを前記クライアント端末に提供する文書形式評価装置と
を有するコンテンツ検索システム。
【請求項３】
クライアント端末及びコンテンツを提供するWebサーバとネットワークを介して接続され、Webコンテンツを検索するコンテンツ検索システムであって、
コンテンツの検索エンジンと該検索エンジンによるコンテンツの検索結果として得られたコンテンツに関して文書形式条件に基づいて文書形式の評価を行う文書形式評価部とを備える検索装置であって前記文書形式条件に合致したコンテンツを前記クライアント端末に提供する検索装置を、有するコンテンツ検索システム。
【請求項４】
クライアント端末及びコンテンツを提供するWebサーバとネットワークを介して接続され、
文書形式条件に基づいて文書形式の評価を行う文書形式評価プログラムを保持する文書形式評価装置と、該文書形式評価装置における評価結果を蓄積するデータベースと、を備え、
前記文書形式評価プログラムが、前記ネットワークを介して複数の前記Webサーバ間を巡回しWebコンテンツの文書形式について評価を実行することを特徴とするコンテンツ検索システム。
【請求項５】
前記評価結果を、前記検索を実行する前に前記評価を実行して前記データベースに保存しておくことを特徴とする請求項４に記載のコンテンツ検索システム。
【請求項６】
請求項１から５までのいずれか１項に記載のコンテンツ検索システムであって、前記文書検索条件をコンテンツの検索キーに変換する変換部を備えるコンテンツ検索システム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、コンテンツ検索システムに関し、特に、ユーザーが指定した文書形式条件を満たすWebコンテンツを検索するシステムに関する。
【背景技術】
【０００２】
インターネットによって接続されたＷｅｂサーバにおいては、非常に多くの情報を検索可能である反面、多くの情報の中から本当に閲覧したい情報を選択するのが難しいという問題がある。そこで、各種のフィルタ技術を用いて、不要な情報を検索対象から除外することがなされている。
【０００３】
従来から、ユーザーがWebコンテンツを検索する場合に、使用するフィルタ項目としては、以下の項目が用いられていた。１）興味対象となる語句、２）Webコンテンツのドメインで認識される組織、３）ファイルタイプ、４）ページ全体又はタイトルやリンク等の

ページ内容の一部が最終更新された時期等である。
【０００４】
また、弱視や色盲等の視覚障害を持つユーザーや、キーボード及びマウス操作が困難なユーザーが、各々にとってのアクセシビリティ（利用容易性）を判断する基準項目をフィルタとして使用し、基準を満たすWebコンテンツを検索する方法が提案されている（特許文献１参照）。この特許文献１に記載の「利用容易性」とは、「障害や特別の必要性を持つユーザーが個別に定義するもの又はＷＷＷコンソーシアムが公表するアクセシビリティ（利用容易性）・ガイドラインにおいて体系化されている定義」を指す。例えば、１）全ての画像について代替のテキストを提供する。２）前景色及び背景色は相互に十分なコントラストを有する。３）もし画像が、その代替テキストの情報以外に重要な情報を伝えるならば、拡張された記述を提供する。４）イベント・ハンドラは、マウスの使用を必要としないことを確認する等のフィルタ項目がある。本技術においては、視覚情報に依存するコンテンツや、マウス操作が不可能である場合に、ユーザーにとって利用し難いコンテンツを、ユーザーに提示する検索対象から除外するか、或いは、利用容易ではないことを示すインジケーターとともに検索結果に表示することが可能である。
【０００５】
【特許文献１】特開２００２－３３４０３４号公報
【発明の開示】
【発明が解決しようとする課題】
【０００６】
実際にＷｅｂ上で閲覧可能なコンテンツには画像などを主体としたコンテンツも多くなってきているが、実際上、文書形式で記載されているコンテンツも非常に多いことは周知の通りである。さらに、一般のユーザーが閲覧する際に、画像データなどに関しては、サイズと容量とのみが変数となるため特に問題とならないが、文書データに関しては、種々の表示形式が存在し、ユーザーにとって見やすい文書形式ではない場合が多いのも事実である。
【０００７】
上記特許文献１に記載の技術によってコンテンツが利用容易になる場合であっても、ユーザーが好む文書形式で提供されない場合には、閲覧を止めてしまうこともある。
【０００８】
本発明は、コンテンツをユーザーの希望する文書形式で提供する技術に関する。
【課題を解決するための手段】
【０００９】
本発明は、ユーザーが指定した文書形式条件を満たすコンテンツをユーザーへ提供する方法、又はプログラムによって構成されるWeb検索システムである。クライアント端末及びコンテンツを提供するWebサーバとはネットワークを介して接続可能な状態にあることが前提となる。本Web文書形式検索システムは、サーチ・エンジンが実行して得た検索結果を受け取り、検索結果に含まれるコンテンツが、ユーザーの指定した文書形式条件を満たすかどうかを評価する。条件を満たしていないと判断したコンテンツを、サーチ・エンジン検索結果から除外し、残りのコンテンツを結果セットとしてユーザーに返す。
【発明の効果】
【００１０】
本発明により、ユーザーは、検索システムを使用して得られる莫大な量の検索結果から、好まない文書形式を持つコンテンツを除外した結果セットを得ることが可能となる。つまり、本発明によれば、ユーザーが検索結果として受け取るコンテンツの量を少なくすることで、結果としてネットワーク転送量を削減し、検索効率を向上させることが可能である。
【発明を実施するための最良の形態】
【００１１】
本発明の実施の形態について説明する前に、発明者の行った考察について説明する。ユ

ーザーが好まない文書形式を有するWebコンテンツを検索結果から除外しユーザーに表示する結果セットには含めないようにすることにより、より少数まで絞られた検索結果に基づいて、ユーザーが見やすいコンテンツに関する情報の提供を受けられることになる。このようなシステムを構築することにより、従来の方法による検索結果よりも、検索結果としてユーザーへ提供されるコンテンツ数及びコンテンツ量が削減される。従って、ネットワークの転送量が軽減されるという。また、ユーザーは「一見して閲覧を好まず閲覧を止める」ようなＷｅｂのページに、立ち入る機会自体が少なくなり検索効率を一層向上させることができる。

【００１２】

ユーザーが好まない文書形式であると判断する項目の例を以下に挙げる。１）文字のサイズが大き過ぎる、または、小さ過ぎる。２）文字の量に対して空白行の割合が少ない、または、行が一行又は数行おきに記述されていて空白が多すぎる。３）漢字の割合又はひらがなの割合が適正値から大きく外れている。或いは、カタカナが多過ぎたり、アルファベットが多過ぎたりする。４）１ページの内容量が多く縦方向のスクロールを何度もする必要がある。５）ファイルサイズが大きく、例えば５秒以内に内容が表示されない。６）Webコンテンツの閲覧に使用している表示画面のサイズが、Webコンテンツの分量に比べて小さすぎるため、横方向のスクロールが必要である。７）ページの左右におけるマージンが少ない。８）ポップアップが多い。９）背景に画像を使用している。１０）Webコンテンツの閲覧に使用しているブラウザーソフトウェアが、サポートしていない、または正常に表示することができない機能、ファイルのタイプを使用している。

【００１３】

以上の１）から１０）までの項目は、異なるユーザーの場合はむろんであるが、同じユーザーであっても、検索目的によって変化すると考える。例えば、語句の説明を得ることを検索目的とする場合は、簡潔な文章を検索結果とするために、フィルタ条件は多くなると考えられるが、語句に関する経験・感想等を趣味として閲覧する場合は、逆にフィルタ条件は少なくなることが好ましいと考えられる。また、検索に使用するパーソナルコンピュータのスペック（ネットワーク転送速度や画面サイズ、ブラウザーソフトウェアの性能）によっても、上記の項目は変化すると考えられる。従って、フィルタ条件として使用する文書形式条件は、上記に挙げたような項目毎に設定できるようにしておき、且つ、簡易に設定変更できるようにすることが望ましい。

【００１４】

以下、本発明の一実施の形態によるWeb文書形式検索システムについて、図面を参照しつつ説明を行う。

【００１５】

図１は、本発明の一実施の形態によるWeb文書形式検索システムの概略構成例を示す図である。図１に示すWeb文書形式検索システムは、文書形式評価装置００４がサーチ・エンジンとは独立に稼動する構成を示す図である。図１に示すように、本実施の形態によるWeb文書形式検索システムは、インターネットによって接続される第１Webサーバ００１、第２Webサーバ００２と、サーチエンジン００３と、文書形式評価装置００４と、クライアント端末００５と、を有している。クライアント端末００５は、例えばパーソナルコンピュータであり、ネットワークを介して第１又は第２Webサーバ００１、００２上のWebコンテンツを閲覧する。

【００１６】

クライアント端末００５は、検索する語句及び文書形式に関する検索条件を、本実施の形態による文書形式評価装置００４に送信する。文書形式評価装置００４は、検索対象語句をサーチ・エンジン００３に送信する。サーチ・エンジンは検索対象語句に基づいてコンテンツの検索を実行し、検索対象語句を含むWebコンテンツの一覧を文書形式評価装置００４に返す。この一覧に含まれるWebコンテンツに対して文書形式を評価し、ユーザーが指定した条件に適合するコンテンツのみをユーザーへ返す。

【００１７】

尚、文書形式評価装置００４に対して１又はそれ以上の端末が接続されており、文書形式評価装置００４をプロキシ・サーバとして機能させる場合もある。

【００１８】

次に、本発明の第２の実施の形態によるWeb文書形式検索システムについて図面を参照しつつ説明を行う。図２は、本実施の形態によるシステムの構成例を示す図である。図２に示すように、本実施の形態によるシステムは、第１から第３までのＷｅｂサーバ００６～００８と、文書形式評価プログラムを搭載したサーチ・エンジン００９と、クライアント端末０１０と、を有している。図２が図１と異なる点は、文書形式評価機能がサーチ・エンジンの一機能として稼動する点である。

【００１９】

サーチ・エンジン００９に文書形式評価装置を搭載しているため、図１に示すシステムと比較して、サーチ・エンジンと文書形式評価装置間のネットワーク転送が不要となる。従って、図１に示すシステムに比べて、ネットワーク転送量が少なくなり、ユーザーの検索時間を短縮することが可能となる。

【００２０】

図３は、本発明の第１又は第２の実施の形態によるシステムにおいて、クライアント端末からの検索要求を受けて、クライアント端末に検索結果を返すまでの処理の流れを示すフローチャート図である。以下においては、図１の第１の実施の形態によるシステムを例に処理の流れを説明する。

【００２１】

図３に示すように、処理を開始すると、まずステップ１０１において、Webコンテンツを検索しようとするクライアントは、クライアント端末を用いてWeb文書形式検索システムへ、Ａ）検索対象となる語句、及び、Ｂ）検索条件とする文書形式（文書フォーム）に関する情報を送信する。ステップ１０２において、外部サーチ・エンジンが、Ａ）検索対象となる語句を含むWebコンテンツを検索し、ステップ１０３において、その検索結果をWeb文書形式検索システムの評価装置に渡す。

【００２２】

ステップ１０４において、Web文書形式検索システムは、検索結果であるコンテンツの文書形式を、検索条件として挙げられている項目について評価する。ステップ１０５において評価結果とユーザーが指定した条件とを比較し、Ｂ）検索条件とする文書形式を満たす場合にはステップ１０７に、満たさない場合にはステップ１０６に進む。ステップ１０６においてクライアント要求に該当しないコンテンツをサーチ・エンジン検索結果から除外し、ステップ１０７においてクライアント要求に該当するコンテンツのみをクライアントへ送信する。本実施の形態によるWeb文書形式検索システムは、サーチ・エンジンの一部であっても、独立したソフトウェアであっても良い。

【００２３】

以下に、検索対象外とする文書形式条件候補となる各項目の例を挙げる。以下に挙げる具体的な数値については、クライアントがカスタマイズできる。

１）文字のサイズが大き過ぎる（20px(ピクセル)以上、又はsize 5以上の文字が本文全体の20%以上。）。

２）文字のサイズが小さ過ぎる（10px以下、又はsize 2以下の文字が本文全体の40%以上。）。

３）文字が多過ぎる（段落内の文字数が300文字以上の段落に含まれる文字数が本文全体の50%以上。）。

４）行が1行間隔で空けられている（<p>タグ属性 line-height: 200% （行間がフォントの高さの２倍）以上の段落に含まれる文字数が本文全体の30%以上。）。

５）漢字が多過ぎる（本文全体の60%以上。）。

６）カタカナが多過ぎる（本文全体の60%以上。）。

７）アルファベットが多過ぎる（本文全体の60%以上。）。

８）１ページの内容量が多く縦方向のスクロールが数十回必要になる（ファイルサイズが

100KBを超えている。）。
９）ファイルサイズが大きく、５秒以内に内容が表示されない（ファイルサイズが、Webコンテンツ閲覧に使用しているシステムのネットワーク転送速度において、5秒で伝送できる量を超えている。）。
１０） Webコンテンツの閲覧に使用しているパーソナルコンピュータの画面サイズが、Webコンテンツに比べて小さい為、横方向のスクロールが必要である（テーブル及び画像の横幅が600ドット以上。）。
１１）ページの左右にマージンが無い（bodyタグ属性margin）。
１２）ポップアップが多い（2個以上。）。
１３）背景に画像を使用している（bodyタグ属性backgroundが指定されている。）。
１４）背景に原色が使用されている（例：赤色　bodyタグ属性bgcolor=#ff0000）。
１５）背景が白色ではない（bodyタグ属性bgcolor=#ffffffではない。）。
１６）Webコンテンツの閲覧に使用しているブラウザーソフトウェアが、サポートしていない、又は、表示に不安がある機能やファイルのタイプを含む。
A.PNG（Portable Network Graphics）形式の画像を含む。
B.スタイルシートを使用している。
C.フレームを使用している。

【００２４】

クライアントは、検索を実行する前に、上記の項目のそれぞれについて、検索条件として使用するか否か、また、各項目の数値等を決定する。上記項目のうち、条件１）～７）は、コンテンツ内のテキスト本文の見易さ、読み易さに関する条件を示す項目である。以下、条件４）を設定した場合を例にし、図５を参照して具体的判定手順を説明する。まず、対象となるコンテンツのＨＴＭＬファイルを取得する（ステップ５０１）。全文字数、見難い文字数にゼロを初期設定した上で（ステップ５０２）、ＨＴＭＬファイル内のタグ（<>で囲まれた部分を、全部のタグが終了するまで順に処理する（Ｓ５０３、Ｓ５０４）。

【００２５】

次に現れたタグが タグで、かつline-height属性が２００％以上（行間がフォントの高さの２倍以上）の場合は、見難い文字であるとして、 タグが終了するまで（ が現れるまで）に出現する本文（<>で囲まれていない文字）の文字数をカウントし、全文字数と見難い文字数のそれぞれをカウントアップする（Ｓ５０６）。それ以外の場合（ タグ以外または タグであってもline-height属性が２００％未満と解釈される場合）は、タグが終了するまで全文字数のみをカウントアップする（Ｓ５０７）。

【００２６】

このようにして全部のタグの処理が終わった時点で、見難い文字数が全文字数に占める割合を計算し（Ｓ５０８）、３０％以上の場合は見難いコンテンツとして当該コンテンツを検索対象から除外し（Ｓ５１０）、３０％未満の場合は当該コンテンツを検索対象に含める（Ｓ５０９）。既に述べた通り、この２００％、３０％等の数値は、クライアント端末からクライアントが入力等できるカスタマイズ可能な値である。

【００２７】

条件４）以外の他の条件についても同様に、それぞれの条件に合致した「見難い文字数」をカウントし、本文の全文字数中に占める割合が一定以上の場合は見難いコンテンツであると判定して、当該コンテンツを検索対象から除外する。尚、条件３）の判定手順において、段落の開始、終了の判断は、<P>タグで判断する方法の他に、例えば
タグ（改行タグ）が出現したら、あるいは本文を挟まずに２つ以上続いたら段落が切り換わったというような判断をしてもよい。また、上記項目のうち、例えば条件８）～１６）は、クライアント端末、ネットワークの性能等に依存する見難さなど、文字自体の読み難さ以外の要素に依存する条件を示す項目である。

【００２８】

例えば、ネットワーク転送速度３６ｋｂｐｓのモデムを介してネットワークに接続した

Ｂ５サイズのノート型パソコンを使用し、興味対象となる語句の意味を検索する目的で、Webコンテンツを閲覧しているユーザーがいるとする。ユーザーは、意味の説明として画像が検索結果に含まれることを期待しているが、使用しているブラウザーはPNG形式の画像表示をサポートしていないとする。このような場合には、まず、上記項目のうち以下の文書形式コンテンツを検索対象外に指定することで、ネットワーク転送量の負荷を抑えることが可能である。

９）ページの内容量が多く縦方向のスクロールが数十回必要になる（ファイルサイズが１００ＫＢを超えている。）。

１０）ポップアップが多い（例えば２個以上。）。

１４）背景に画像を使用している。

【００２９】

また、以下の項目を検索対象外に指定することで、表示できない画像を検索結果に含めないようにすることが可能である。

１７）A.PNG（Portable Network Graphics）形式の画像を含む。

【００３０】

以下の項目を検索対象外に指定することで、Ｂ５サイズのノートパソコンを使用しても横方向にスクロールする必要が無く、見易さを感じるコンテンツのみ、検索結果として受け取ることができる。

１１）テーブル及び画像の横幅が600ドット以上。

１６）背景が白色ではない（bodyタグ属性bgcolor=#ffffffではない。）。

Web文書形式検索システムは、プロキシ・サーバ上ではなくとも、クライアントパーソナルコンピューター上のソフトウェアとして実行されても良い。プロキシ・サーバ上のソフトウェアとしてWeb文書形式検索システムが稼動する場合、Webコンテンツの文書形式について評価を実行するタイミングとしては、前述のように「サーチ・エンジンからの検索結果を受け取った後」の他に、定期的に評価プログラムがWebサーバ上の全Webコンテンツを対象として評価を実行し、上記に挙げた全項目についての評価結果をデータベース化しておくという方法でも良い。この方法について第３の実施の形態として図４を参照しつつ説明する。

【００３１】

図４に示すように、本実施の形態によるWeb文書形式検索システムは、インターネットと、第１のWebサーバ０１１と、第２のWebサーバ０１２と、第３のWebサーバ０１３と、文書形式評価装置０１４と、データベース０１６と、を有する。

【００３２】

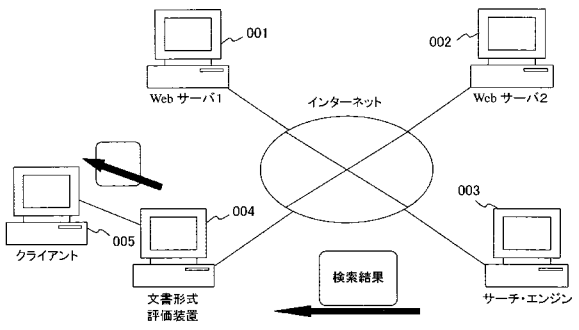
本実施の形態によるWeb文書形式検索システムにおいては、文書形式評価プログラム０１５は、定期的に第１のWebサーバ０１１、第２のWebサーバ０１２、第３のWebサーバ０１３の間を、これらを接続するネットワークを介して巡回し、Webコンテンツの文書形式について評価を実行する。評価結果を、文書形式評価サーバ０１４が所有するデータベース０１６に保存する。ユーザーが検索を実行した後ではなく、事前に評価を実行しておくことにおり、Web文書形式検索システムにおける検索速度が向上する。

【００３３】

以上、本発明について各実施の形態について説明したが、本発明はこれらに限定されるものではなく、種々の変形が可能である。例えば、ユーザーの要望に応じて文書形式を変換する変換部を有し、文書形式を変換した後にユーザー端末に提供するようにすることも可能である。或いは、文書形式に関してユーザーの要望（文書形式条件）と実際のコンテンツにおける文書形式との偏差を演算し、ある程度までの偏差を許容するようにすることも可能である。により、ユーザーは、従来の検索システムを使用して得られる莫大な量の検索結果から、好まない文書形式を持つコンテンツを除外した結果セットを得ることが可能となる。つまり、本発明は、ユーザーが検索結果として受け取るコンテンツの量を減少させ、結果としてネットワーク転送量を削減し、ユーザーの検索効率を向上させることが可能である。

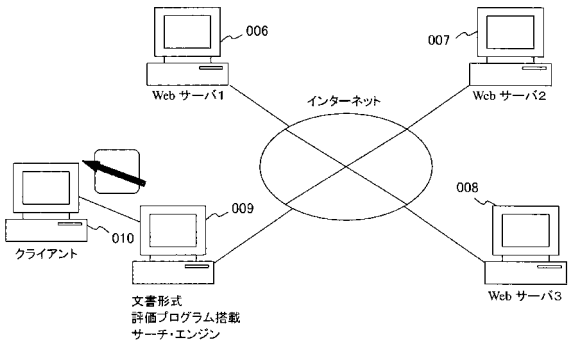
【図面の簡単な説明】
【００３４】
【図１】本発明の第１の実施の形態によるWebコンテンツ検索システムWebコンテンツ検索システムがサーチ・エンジンと独立して稼動する場合の構成例を示す図である。
【図２】本発明の第２の実施の形態によるWebコンテンツ検索システムの構成例を示す図であり、Webコンテンツ検索システムがサーチ・エンジンの一機能として稼動する場合の例を示す図である。
【図３】本発明の実施の形態によるWebコンテンツ検索システムにおける処理の流れを示すフローチャート図である。
【図４】本発明の第３の実施の形態によるWebコンテンツ検索システムであって、評価プログラムがWebサーバ上の全Webコンテンツを対象として、定期的に文書形式についての評価を実行し、その評価結果を記録しておくシステムの概念である。
【図５】行が１行間隔で空けられている（<p>タグ属性 line-height: ２００％ （行間がフォントの高さの２倍以上）の段落に含まれる文字数が本文全体の３０％以上。）という条件に基づくコンテンツ検索処理の流れを示すフローチャート図である。
【符号の説明】
【００３５】
００１/００２…Webサーバ、００３…サーチ・エンジン、００４…文書フォーム評価装置、００５…クライアントパーソナルコンピューター、００６/００７/００８…Webサーバ、００９…文書形式評価プログラム搭載サーチ・エンジン、０１０…クライアントパーソナルコンピューター、０１１/０１２/０１３…Webサーバ、０１４…文書形式評価装置、０１５…Webコンテンツの文書形式に関する評価実行プログラム、０１６…０１５による評価結果を格納するデータ-ベース。

【図１】



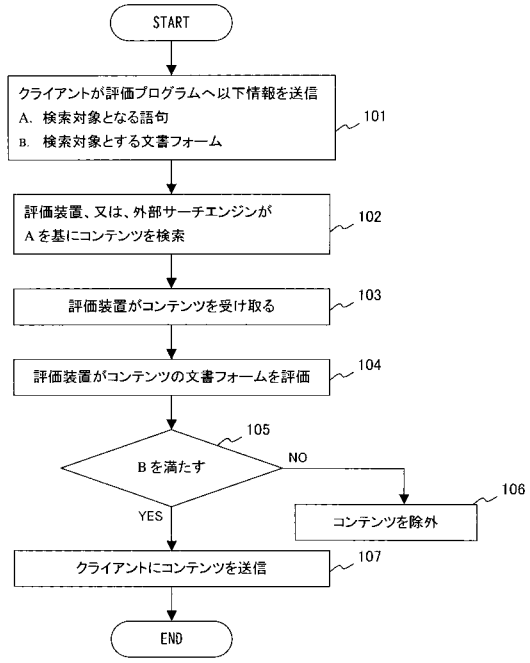
サーチ・エンジンの検索結果を受け取り文書フォーム評価を実行

【図２】

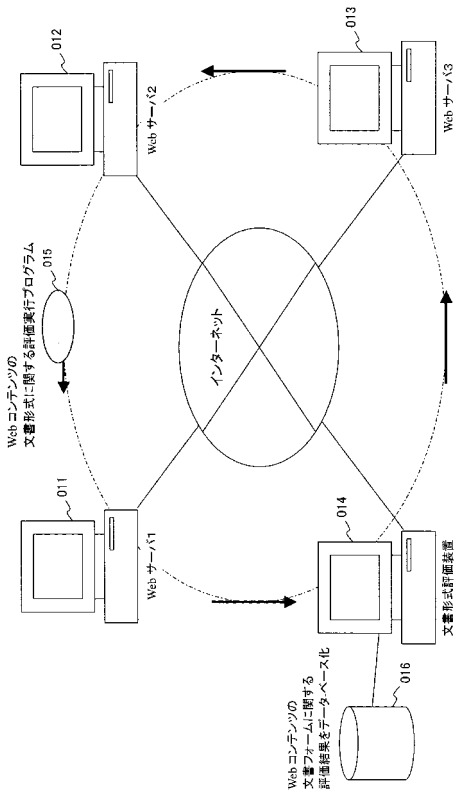


サーチ・エンジンの一機能として文書フォーム評価を実行

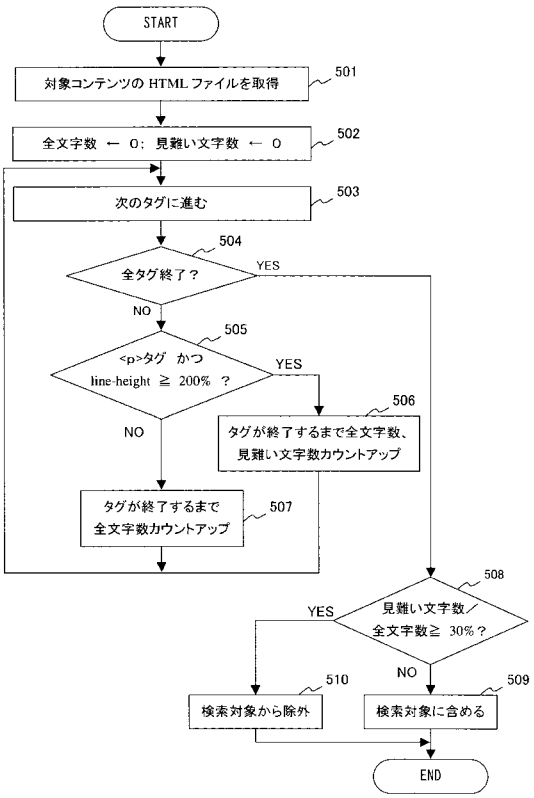
【図３】


START
クライアントが評価プログラムへ以下情報を送信
A. 検索対象となる語句
B. 検索対象とする文書フォーム
101
評価装置、又は、外部サーチエンジンが
A を基にコンテンツを検索
102
評価装置がコンテンツを受け取る
103
評価装置がコンテンツの文書フォームを評価
104
105
B を満たす
NO
106
コンテンツを除外
YES
クライアントにコンテンツを送信
107
END

【図４】


012
Web サーバ2
013
Web サーバ3
Web コンテンツの
文書形式に関する評価実行プログラム
015
インターネット
011
Web サーバ1
014
文書形式評価装置
016
Web コンテンツの
文書フォームに関する
評価結果をデータベース化

【図５】


START
対象コンテンツの HTML ファイルを取得
501
全文字数 ← 0; 見難い文字数 ← 0
502
次のタグに進む
503
504
全タグ終了？
YES
NO
505
<p>タグ かつ
line-height ≧ 200% ？
YES
NO
506
タグが終了するまで全文字数、
見難い文字数カウントアップ
タグが終了するまで
全文字数カウントアップ
507
508
見難い文字数／
全文字数≧ 30%？
YES
NO
510
検索対象から除外
509
検索対象に含める
END

フロントページの続き

(72)発明者　松下　智子
東京都品川区東品川４丁目１２番７号　日立ソフトウエアエンジニアリング株式会社内

Ｆターム(参考) 5B075 KK02　KK07　ND03　ND16　NR02　PQ02　PQ42　PR08