(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2007-58335
(P2007-58335A)
(43)公開日　平成19年3月8日(2007.3.8)

(51)Int.Cl.		F I			テーマコード（参考）
G06F 12/00	(2006.01)	G06F	12/00	546B	5B075
G06F 17/30	(2006.01)	G06F	17/30	419B	5B082

審査請求 未請求　請求項の数 7　ＯＬ　（全 17 頁）

(21)出願番号　特願2005-240254 (P2005-240254)
(22)出願日　平成17年8月22日 (2005.8.22)

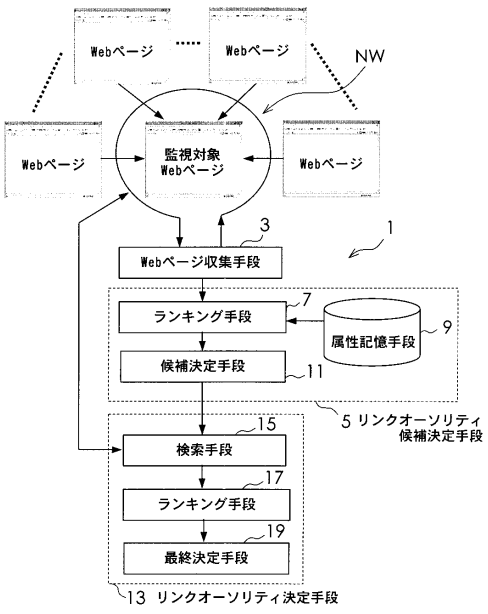
特許法第３０条第１項適用申請有り　平成１７年２月２２日　電子情報通信学会データ工学専門委員会主催の「電子情報通信学会　第１６回データ工学ワークショップ」のウエブサイト（ｈｔｔｐ：／／ｗｗｗ．ｄｉｇｉｔａｌｃｉｔｙ．ｇｒ．ｊｐ／～ｓａｔｏｈ／ＤＥＷＳ２００５／ｔｏｐ０４．ｈｔｍ）」にて発表

(71)出願人　504171134
国立大学法人　筑波大学
茨城県つくば市天王台一丁目１番１
(74)代理人　100091443
弁理士　西浦　▲嗣▼晴
(72)発明者　森嶋　厚行
茨城県つくば市天王台一丁目１番１　国立大学法人筑波大学内
(72)発明者　飯田　敏成
茨城県つくば市天王台一丁目１番１　国立大学法人筑波大学内
(72)発明者　杉本　重雄
茨城県つくば市天王台一丁目１番１　国立大学法人筑波大学内

最終頁に続く

(54)【発明の名称】リンクオーソリティ決定方法及び装置並びにプログラム

(57)【要約】
【課題】　できるだけ短い時間でしかも少ない労力で、適正なリンクオーソリティを決定できるリンクオーソリティの決定装置を提供する。
【解決手段】　リンクオーソリティ候補決定手段５は、Ｗｅｂページ収集手段３で収集した複数のＷｅｂページの中から予め定めた条件を満たす複数のＷｅｂページをリンクオーソリティ候補として定める。リンクオーソリティ決定手段１３のランキング手段１７は、候補となるＷｅｂページのリンクについてのリンク切れの割合が少ない順にランキングを行う。最終決定手段１９は、ランキング手段１７によるランキングの結果から上位のランクにある１以上のＷｅｂページをリンクオーソリティとして決定する。
【選択図】　図５



【特許請求の範囲】
【請求項１】
監視対象とするＵＲＬへのリンクを持つ複数のＷｅｂページを収集する収集ステップと、
前記複数のＷｅｂページの中から予め定めた条件を満たす複数のＷｅｂページをリンクオーソリティ候補として定めるリンクオーソリティ候補決定ステップと、
前記リンクオーソリティ候補に含まれる前記複数のＷｅｂページの中からリンク切れを修正するために利用可能なリンクオーソリティを決定するリンクオーソリティ決定ステップとからなるリンクオーソリティの決定方法であって、
前記リンクオーソリティ候補決定ステップでは、予めリンクオーソリティとなり得るＷｅｂページが有する複数の属性を定めて、前記複数の属性を基準にして前記複数のＷｅｂページをリンクオーソリティとして利用可能性が高いと推測される順にランキングを行い、前記ランキングの結果から上位のランクにある複数のＷｅｂページを前記リンクオーソリティ候補として定め、
前記リンクオーソリティ決定ステップでは、前記上位のランクにある複数のＷｅｂページのそれぞれについて、各Ｗｅｂページについてのリンク切れのリンクの数またはリンク切れではないリンクの数を求め、この数と前記各Ｗｅｂページにあるリンクの数とに基づいてリンク切れの割合が少ない順にランキングを行い、前記ランキングの結果から上位のランクにある１以上のＷｅｂページを前記リンクオーソリティとして決定することを特徴とするリンクオーソリティの決定方法。
【請求項２】
前記リンクオーソリティ決定ステップでは、前記リンクの数と前記リンク切れではないリンクの数の割合を反映した値をキーとして前記ランキングを行うことを特徴とする請求項１に記載のリンクオーソリティの決定方法。
【請求項３】
前記リンクの数と前記リンク切れではないリンクの数の割合の相乗平均を前記キーとすることを特徴とする請求項２に記載のリンクオーソリティの決定方法。
【請求項４】
監視対象とするＵＲＬへのリンクを持つ複数のＷｅｂページを収集する収集機能と、
前記複数のＷｅｂページの中から予め定めた条件を満たす複数のＷｅｂページをリンクオーソリティ候補として定めるリンクオーソリティ候補決定機能と、
前記リンクオーソリティ候補に含まれる前記複数のＷｅｂページの中からリンク切れを修正するために利用可能なリンクオーソリティを決定するリンクオーソリティ決定機能とをコンピュータに実現させるためのプログラムであって、
前記リンクオーソリティ候補決定機能は、予め定めたリンクオーソリティとなり得るＷｅｂページが有する複数の属性を基準にして、前記複数のＷｅｂページをリンクオーソリティとして利用可能性が高いと推測される順にランキングを行う機能と、前記ランキングの結果から上位のランクにある複数のＷｅｂページを前記リンクオーソリティ候補として定める機能を含み、
前記リンクオーソリティ決定機能は、前記上位のランクにある複数のＷｅｂページのそれぞれについて、各Ｗｅｂページについてのリンク切れのリンクの数またはリンク切れではないリンクの数を求める機能と、この数と前記各Ｗｅｂページにあるリンクの数とに基づいてリンク切れの割合が少ない順にランキングを行う機能と、前記ランキングの結果から上位のランクにある１以上のＷｅｂページを前記リンクオーソリティとして決定する機能とを含むことを特徴とするプログラム。
【請求項５】
前記リンクオーソリティ決定機能に含まれる前記ランキングを行う機能は、前記リンクの数と前記リンク切れではないリンクの数の割合を反映した値をキーとして前記ランキングを行うことを特徴とする請求項４に記載のプログラム。
【請求項６】

監視対象とするＵＲＬへのリンクを持つ複数のＷｅｂページを収集する収集手段と、
前記複数のＷｅｂページの中から予め定めた条件を満たす複数のＷｅｂページをリンクオーソリティ候補として定めるリンクオーソリティ候補決定手段と、
前記リンクオーソリティ候補に含まれる前記複数のＷｅｂページの中からリンク切れを修正するために利用可能なリンクオーソリティを決定するリンクオーソリティ決定手段とからなるリンクオーソリティ決定装置であって、
前記リンクオーソリティ候補決定手段は、予め定めたリンクオーソリティとなり得るＷｅｂページが有する複数の属性を記憶する属性記憶手段と、前記属性記憶手段に記憶された前記複数の属性を基準にして、前記複数のＷｅｂページをリンクオーソリティとして利用可能性が高いと推測される順にランキングを行うランキング手段と、前記ランキング手段によるランキングの結果から上位のランクにある複数のＷｅｂページを前記リンクオーソリティ候補として定める候補決定手段を含み、
前記リンクオーソリティ決定手段は、前記上位のランクにある複数のＷｅｂページのそれぞれについて、リンク切れのリンクの数またはリンク切れではないリンクの数を求める検索手段と、前記検索手段により求めた数と前記各Ｗｅｂページにあるリンクの数とに基づいてリンク切れの割合が少ない順にランキングを行うランキング手段と、前記ランキング手段によるランキングの結果から上位のランクにある１以上のＷｅｂページを前記リンクオーソリティとして決定する最終決定手段とを含むことを特徴とするリンクオーソリティ決定装置。
【請求項７】
前記ランキング手段は、前記リンクの数と前記リンク切れではないリンクの数の割合を反映する値をキーとして前記ランキングを行うことを特徴とする請求項６に記載のリンクオーソリティ決定装置。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、Ｗｅｂページのリンク切れを自動的に修正するために用いられるリンクオーソリティを決定するために用いられるリンクオーソリティ決定方法及び装置並びにプログラムに関するものである。
【背景技術】
【０００２】
近年、Ｗｏｒｌｄ　Ｗｉｄｅ　Ｗｅｂ（以下Ｗｅｂ）は、社会における重要なメディアの一つである。そしてＷｅｂの特徴の一つに、分散管理が挙げられる。即ち、Ｗｅｂコンテンツは多くの組織・個人により独立して追加・削除・更新が行われている。この特徴はＷｅｂを便利なツールとする一方で、Ｗｅｂコンテンツの一貫性の維持を困難としている要因でもある。コンテンツの一貫性が損なわれる一例として、Ｗｅｂページのリンク切れがある。そこで従来から、リンク切れが発生したときに代わりとなるリンク候補を探す技術が種々提案されている。例えば、特開平09-081446号公報「ハイパーテキストシステム」（特許文献１）には、リンク切れが起こった場合に、代わりとなるリンク先ページ候補を探すことが記載されている。この公報に記載の技術では、代わりとなるリンク先ページの発見に、アドレス(URL)の情報とＷｅｂページの内容のみを利用している。また特開平11-039327号公報「リンク情報自動修復方法および装置」（特許文献２）には、リンク切れが起こった場合に，代わりとなるリンク先ページ候補を探すために、「同じノード(ページのこと)」を探すことが記載されている。さらに、特開2001-273185号公報「ホームページアドレス登録装置及びホームページアドレス登録処理プログラムを記憶した記憶媒体」（特許文献３）には、リンク切れが起こった場合に、代わりとなるリンク先Ｗｅｂページ候補を探すために、Ｗｅｂページの内容のみを利用する技術が開示されている。
【０００３】
これらの技術では、リンク先ページの探索精度が必ずしも高くない。そこで本出願の発明者等は、Ｗｅｂのリンク切れを発見すると、変更先と考えられるリンクの候補を自動的

に発見しリンクの訂正を試みるシステムの開発を行い、実験を行ってきた（非特許文献１：中溝昌佳、森嶋厚行、杉本重雄及び北川博之著「ＷＷＷリンク一貫性維持支援システムにおけるリンク切れ自動修復」日本データベース学会Ｌｅｔｔｅｒｓ、Ｖｏｌ．３、Ｎｏ．２、２００４年１２月。非特許文献２：中溝昌佳、森嶋厚行、有山智洋、杉本重雄及び北川博之著「ＷＷＷコンテンツ一貫性維持のためのリンク更新機構の提案」日本データベース学会Ｌｅｔｔｅｒｓ、Ｖｏｌ．２、Ｎｏ．２、６５頁－６８頁、２００３年１０月）。このシステムでは、Ｗｅｂのリンク切れはページの移動に伴って生じたものであると仮定し移動先の探索を行う。

【０００４】

そして発明者等は、信頼できるリンクを含むページである「リンクオーソリティ（Ｌｉｎｋ　Ａｕｔｈｏｒｉｔｙ）」を求めるための仕組みと、それを実装したＬＡサーバ（Ｌｉｎｋ　Ａｕｔｈｏｒｉｔｙ　Ｓｅｒｖｅｒ）を提案し、ＬＡサーバを利用することによりＷｅｂページの移動先を効率よく発見できる可能性があることを指摘した（非特許文献３：中溝昌佳、森嶋厚行、杉本重雄及び北川博之著「ＷＷＷにおける信頼度の高いリンクの発見」情報処理学会研究報告、Ｖｏｌ．２００４、Ｎｏ．７２（２００４－ＤＢＳ－１３４（II）、３９７頁－４０２頁。非特許文献４：中溝昌佳、森嶋厚行、杉本重雄及び北川博之著「ＷＷＷにおける信頼度の高いリンクの発見」電子情報通信学会技術研究報告、Ｖｏｌ．１０４、Ｎｏ．１７７（ＤＥ２００４－６３）、８７頁－９２頁、２００４年７月）。

【０００５】

ここで発明者等が定義したＷｅｂページのリンクオーソリティとは、Ｗｅｂページが移動したときに、全Ｗｅｂページ中で十分に大きな確率でリンクが更新されるページを意味する。直観的には、リンクオーソリティとは、「リンク先の内容が変化したときに、全Ｗｅｂページ中で十分大きな確率でリンクが更新されるページ」のことである（ＧｏｏｇｌｅなどにおけるＡｕｔｈｏｒｉｔｙページとは異なる観念である）。なおここで全Ｗｅｂページとは、システムにおいて利用可能なＷｅｂページの全てという意味である。

【０００６】

例えば、図１に示す例で説明すると、ある大学Ａの研究室のＷｅｂページu_1（ｍ．ｌ．ｓ．ａｃ．ｊｐ）が存在し、このページは複数のページからリンクされているものとする。このうちＷｅｂページv_1（ｌ．ｓ．ａｃ．ｊｐ）はその研究室が所属する学科の研究室一覧ページである。このとき、一般には、Ｗｅｂページv_1はＷｅｂページu_1に関するリンクオーソリティである。したがって、Ｗｅｂページu_1がＷｅｂページu_2（ｍ．ｏｒｇ）に移動したとき、Ｗｅｂページu_1へリンクしているページはリンク切れを起こすが、Ｗｅｂページv_1はＷｅｂページu_1へのリンクをＷｅｂページu_2に貼り換えるはずである。そこで発明者等は、あるＷｅｂページｖは次の２つの条件を満たすとき、別のＷｅｂページｕのリンクオーソリティとなるものと定義した。（１）Ｗｅｂページｖがページｕへのリンクを持っており、且つ（２）Ｗｅｂページｕが別のＷｅｂページu_{new}に移動すると、Ｗｅｂページｖ中のＷｅｂページｕへのリンクがＷｅｂページu_{new}へのリンクに確実に変更されることが強く期待される。

【０００７】

あるＷｅｂページ中に含まれるＷｅｂページｕへのリンクがリンク切れになった場合に、もしＷｅｂページｕのリンクオーソリティとなるＷｅｂページｖを知っていれば、Ｗｅｂページｖを見ることにより新たなリンク先u_{new}を知ることができ、リンク先の修正が可能になる。したがってリンク先の修正をするためには、リンクオーソリティを高い精度で決定することが必要となる。

【特許文献１】特開平09-081446号公報
【特許文献２】特開平11-039327号公報
【特許文献３】特開2001-273185号公報
【非特許文献１】中溝昌佳、森嶋厚行、杉本重雄及び北川博之著「ＷＷＷリンク一貫性維持支援システムにおけるリンク切れ自動修復」日本データベース学会Ｌｅｔｔｅｒｓ、Ｖ

ｏｌ．３、Ｎｏ．２、２００４年１２月
【非特許文献２】中溝昌佳、森嶋厚行、有山智洋、杉本重雄及び北川博之著「ＷＷＷコンテンツ一貫性維持のためのリンク更新機構の提案」日本データベース学会Ｌｅｔｔｅｒｓ、Ｖｏｌ．２、Ｎｏ．２、６５頁－６８頁、２００３年１０月
【非特許文献３】中溝昌佳、森嶋厚行、杉本重雄及び北川博之著「ＷＷＷにおける信頼度の高いリンクの発見」情報処理学会研究報告、Ｖｏｌ．２００４、Ｎｏ．７２（２００４－ＤＢＳ－１３４（II）、３９７頁－４０２頁
【非特許文献４】中溝昌佳、森嶋厚行、杉本重雄及び北川博之著「ＷＷＷにおける信頼度の高いリンクの発見」電子情報通信学会技術研究報告、Ｖｏｌ．１０４、Ｎｏ．１７７（ＤＥ２００４－６３）、８７頁－９２頁、２００４年７月

【発明の開示】
【発明が解決しようとする課題】
【０００８】
リンクオーソリティを利用してリンク切れを修正する場合には、リンクオーソリティとしての機能ができるだけ発揮できるＷｅｂページをリンクオーソリティとして決定する必要がある。リンクオーソリティの決定方法は、種々考えられるものの、監視対象となるＷｅｂページが多くなればなるほど、決定に要するまでの時間をできるだけ短くすることができ、しかも少ない労力で実行できるものが望まれる。しかしながら従来は、この点に着目していなかったため、このような要望に答えることができる方法、装置及びプログラムはなかった。
【０００９】
本発明の目的は、できるだけ短い時間でしかも少ない労力で、適正なリンクオーソリティを決定できるリンクオーソリティの決定方法及び装置並びにプログラムを提供することにある。
【課題を解決するための手段】
【００１０】
本発明のリンクオーソリティ決定方法では、収集ステップと、リンクオーソリティ候補決定ステップと、リンクオーソリティ決定ステップとから構成される。収集ステップは、監視対象とするＵＲＬへのリンクを持つ複数のＷｅｂページを収集する。リンクオーソリティ候補決定ステップでは、収集した複数のＷｅｂページの中から予め定めた条件を満たす複数のＷｅｂページをリンクオーソリティ候補として定める。そしてリンクオーソリティ決定ステップでは、リンクオーソリティ候補に含まれる複数のＷｅｂページの中からリンク切れを修正するために利用可能なリンクオーソリティを決定する。
【００１１】
特に本発明では、リンクオーソリティ候補決定ステップで、まず予めリンクオーソリティとなり得るＷｅｂページが有する複数の属性を定める。ここで複数の属性の定め方は任意である。しかしながらできるだけリンクオーソリティになり得るＷｅｂページが集まるように属性を定める必要がある。例えば「監視対象のＷｅｂページと同一サイトの同一ディレクトリに存在する」、「監視対象となるＷｅｂページの同一サイトの上位ディレクトリに存在する」、「監視対象となるＷｅｂページの同一サイトの上位ディレクトリに存在する」、「監視対象のＷｅｂページとの間に直接的、もしくは間接的な相互リンクが存在する」「ファイル名にデフォルトファイル名が含まれている」等である。
【００１２】
このような複数の属性を基準にして、複数のＷｅｂページをリンクオーソリティとして利用可能性が高いと推測される順にランキングする。ランキングの手法（何をキーとしてランキング行うか）は任意である。そしてランキングの結果から上位のランクにある複数のＷｅｂページをリンクオーソリティ候補として定める。ここで上位のランクにある複数のＷｅｂページの定め方は任意である。例えば予め定めた順位（例えば１位及び２位）に属するＷｅｂページを候補と定める方法を採用してもよいし、また上位３０件以内に入るＷｅｂページを候補と定める方法を採用することができる。

【００１３】
そして本発明の方法では、リンクオーソリティ決定ステップで、各Ｗｅｂページについてのリンク切れのリンクの数またはリンク切れではないリンクの数を求める。そしてこの数と各Ｗｅｂページにあるリンクの数と基づいてリンク切れの割合が少ない順にランキングを行い、ランキングの結果から上位のランクにある１以上のＷｅｂページをリンクオーソリティとして決定する。なお各Ｗｅｂページにあるリンクの数は、リンク切れのリンクの数またはリンク切れではないリンクの数を求める際に一緒に求めてもよいが、これらの数を求める場合とは、別に求めてもよい。一つのＷｅｂページにあるリンクが切れているか否かの確認作業は、時間と手間（アクセス作業、確認作業）を要する。本発明では、リンクオーソリティ候補として予め絞られた複数のＷｅｂページに関してだけ、この確認作業を行うため、収集したＷｅｂページのすべてについて確認作業を行う場合と比べて、リンクオーソリティの決定までの時間と労力を少ないものとすることができる。特に、リンク切れの割合が少ないことを基準にしてランキングすると、より適正なリンクオーソリティ（更新率の高いリンクオーソリティ）を決定できる。この場合、リンクオーソリティ決定ステップでは、リンクの数とリンク切れではないリンクの数の割合を反映した値をキーとしてランキングを行うことが好ましい。このようにすると、ランキングが容易になる上、ランキングの精度を高いものとすることができる。なおキーとする値は、種々の演算法を用いて演算することができる。例えば、リンクの数とリンク切れではないリンクの数の割合の相乗平均により求めた値をキーとしてランキングを行うと、ランキングの精度をより高めることができる。なおリンクオーソリティとして決定する上位のランクにある１以上のＷｅｂページの数は、例えば、リンク数等を参考にして定めればよく、リンク数が少ない場合には、上位の複数のＷｅｂページをリンクオーソリティとして用いればよい。

【００１４】
本発明の方法をコンピュータを用いて実現する場合に用いるプログラムは、監視対象とするＵＲＬへのリンクを持つ複数のＷｅｂページを収集する収集機能と、前記複数のＷｅｂページの中から予め定めた条件を満たす複数のＷｅｂページをリンクオーソリティ候補として定めるリンクオーソリティ候補決定機能と、複数のＷｅｂページの中からリンク切れを修正するために利用可能なリンクオーソリティを決定するリンクオーソリティ決定機能とをコンピュータに実現させるためのプログラムである。特にこのプログラムでは、リンクオーソリティ候補決定機能が、予め定めたリンクオーソリティとなり得るＷｅｂページが有する複数の属性を基準にして、リンクオーソリティ候補に含まれる複数のＷｅｂページをリンクオーソリティとして利用可能性が高いと推測される順にランキングを行う機能と、ランキングの結果から上位のランクにある複数のＷｅｂページを前記リンクオーソリティ候補として定める機能を含む。そしてリンクオーソリティ決定機能が、上位のランクにある複数のＷｅｂページのそれぞれについて、リンク切れのリンクの数またはリンク切れではないリンクの数を求める機能と、この数と各Ｗｅｂページにあるリンクの数とに基づいてリンク切れの割合が少ない順にランキングを行う機能と、このランキングの結果から上位のランクにある１以上のＷｅｂページをリンクオーソリティとして決定する機能とを含む。

【００１５】
また本発明の方法を実施するリンクオーソリティ決定装置は、監視対象とするＵＲＬへのリンクを持つ複数のＷｅｂページを収集するＷｅｂページ収集手段と、収集した複数のＷｅｂページの中から予め定めた条件を満たす複数のＷｅｂページをリンクオーソリティ候補として定めるリンクオーソリティ候補決定手段と、複数のＷｅｂページの中からリンク切れを修正するために利用可能なリンクオーソリティを決定するリンクオーソリティ決定手段とからなる。そして特に、リンクオーソリティ候補決定手段は、予め定めたリンクオーソリティとなり得るＷｅｂページが有する複数の属性を記憶する属性記憶手段と、属性記憶手段に記憶された複数の属性を基準にして、リンクオーソリティ候補に含まれる複数のＷｅｂページをリンクオーソリティとして利用可能性が高いと推測される順にランキングを行うランキング手段と、ランキング手段によるランキングの結果から上位のランク

にある複数のＷｅｂページをリンクオーソリティ候補として定める候補決定手段を含む。またリンクオーソリティ決定手段は、上位のランクにある複数のＷｅｂページのそれぞれについて、リンク切れのリンクの数またはリンク切れではないリンクの数とを求める検索手段と、検索手段により求めた数と各Ｗｅｂページにあるリンクの数とに基づいてリンク切れの割合が少ない順にランキングを行うランキング手段と、ランキング手段によるランキングの結果から上位のランクにある１以上のＷｅｂページをリンクオーソリティとして決定する最終決定手段とを含む。なおＷｅｂページにあるリンクの数は、検索手段により一緒に求めてもよいが、別の手段により求めてもよい。

【発明の効果】

【００１６】

本発明によれば、リンクオーソリティ候補として予め絞られた複数のＷｅｂページに関してだけ、リンク切れの確認作業を行うため、収集したＷｅｂページのすべてについて確認作業を行う場合と比べて、リンクオーソリティの決定までの時間を短くして、しかも決定に要する労力を少ないものとすることができる利点が得られる。

【発明を実施するための最良の形態】

【００１７】

以下図面を参照して本発明のリンクオーソリティの決定方法及び装置の実施の形態の一例を詳細に説明する。実施の形態を説明する前に、図２を用いて本発明のリンクオーソリティ決定装置１の主要部を構成するＬＡサーバ（リンクオーソリティ・サーバ）を用いたリンク切れの自動修正について説明する。ここでＬＡサーバとは、あるＷｅｂページ（監視対象となるＷｅｂページ）のＵＲＬを「ｕ」とした場合に、このｕのリンクオーソリティと考えられる候補を決定するサーバである。実際には、リンクオーソリティを一気に求めることは困難であるため、ＬＡサーバは複数のリンクオーソリティ候補を収集し、リンクオーソリティである可能性が高いと考えられる順にランキングしたページのＵＲＬのリストＶ＝[v_1，v_2，．．．]を、結果として出力する。

【００１８】

図３は、自動修正システムの構成（アーキテクチャ）を示す図である。リンクオーソリティ決定装置１として用いられるＬＡサーバ以外の構成は、既に発明者等が前述の論文で発表しているシステムである。簡単化のため、ここでは、システムが監視対象とするＷｅｂページ（リンク）はＵＲＬ「ｕ」で表されるただ一つのリンクに限定する。このシステムは監視対象としてのｕがリンク切れであることを発見すると、新しいリンク先u_{new}を発見し、u_{new}に自動修正することを試みる。本システムの主要な構成要素は、対象となるリンクを監視するＬＩＭ（Ｌｉｎｋ Ｉｎｔｅｎｇｒｉｔｙ Ｍａｉｎｔｅｎａｎｃｅ）サーバ、移動先のページのＵＲＬであるu_{new}の候補集合Ｕを収集するチェーサー（Ｃｈａｓｅｒ）、Ｕに含まれる発見された各候補に対して「移動先らしさ」を表すスコア$score_i$を計算するマーカー（Ｍａｒｋｅｒ）である。これらの動作をまとめた抽象アルゴリズムを図４に示す。このアルゴリズムは、簡単に説明すると次のようになる。（１）ＬＩＭサーバはｕを監視する。リンク切れを発見すると次のようにチェーサー（Ｃｈａｓｅｒ）とマーカー（Ｍａｋｅｒ）を呼び出す。（２）チェーサー（Ｃｈａｓｅｒ）は移動前のＷｅｂページｕのコンテンツとＵＲＬの情報ｗ用いて、Ｗｅｂサーチエンジンによる候補収集やロボットによるサイト内検索を用いた候補収集を行い、Ｕを作成する（４行目）。（３）Ｍａｒｋｅｒは各$u_i \in U$に対し、主に移動前のＷｅｂページと候補先のＷｅｂページとの類似度やＵＲＬの関係などに基づいてスコア$score_i$を計算する（５～７行目）。（４）ＬＩＭサーバは、$score_i$を用いてＵの中のu_iをランキングし、リストＵを計算する（８行目）。

【００１９】

ＬＡサーバを追加した場合の処理は次のようになる。チェーサー（Ｃｈａｓｅｒ）はリンクオーソリティを含むリンク群を、移動先リンクu_{new}の候補として新たにＵを追加する。マーカー（Ｍａｒｋｅｒ）はＵ中の候補ページのランキングの際に、もしそのページがリンクオーソリティとされているならば高いスコアを割り当てる。具体的には、候補

ページが、ＬＡサーバが求めたランキング上位５位までのリンクオーソリティ候補からリンクされている場合、元のｓｃｏｒｅ$_i$にある定数を掛けることにより、スコアを高くする。

【００２０】

ＬＩＭサーバは、監視対象の各リンクに対して、利用者が（１）リンクオーソリティを明示的に指摘する手段と、（２）自動修正を行うためのスコアの閾値を明示的に指定する手段をそれぞれ備えている。（１）では、利用者によって明示的に指定されたリンクオーソリティを参照することにより、ＬＩＭサーバはそのリンクオーソリティに従ってリンクの自動修正を行うための閾値を指定することができる。ＬＩＭサーバは、この閾値を見て、もし指定された閾値より候補ページのスコアが大きい場合、発見されたリンクへの自動修正を行う。また、もし指定された閾値よりも候補ページのスコアが小さい場合、移動先候補の一覧ページを生成し、その生成されたページのリンクへ自動修正を行う。

【００２１】

図５は、主として前述のＬＡサーバーによって構成される本発明のリンクオーソリティ決定装置１の実施の形態の構成の一例を示すブロック図である。そして図６は、リンクオーソリティ決定装置１を、コンピュータを用いて実現する場合に用いるプログラムのアルゴリズムを示すフローチャートである。

【００２２】

リンクオーソリティ決定装置１は、Ｗｅｂページ収集手段３と、リンクオーソリティ候補決定手段５と、リンクオーソリティ決定手段１３とから構成される。Ｗｅｂページ収集手段３は、監視対象とするＷｅｂページのＵＲＬ「ｕ」へのリンクを持つ複数のＷｅｂページを、ネットワークＮＷを介して収集する。Ｗｅｂページ収集手段３は、例えば「ｕ」へのリンクを持つページの集合を計算できるものであればどのような構成でもよい。図３の例では、チェーサー（Ｃｈａｓｅｒ）がこの手段の一部を構成している。収集には、例えば、クローラ（ＷＷＷ巡回プログラム）を用いて収集する方法や、Ｗｅｂアーカイブなどを利用する方法や、Ｗｅｂサーチエンジンを利用する方法などが考えられる。具体的なＬＡサーバの実装では、次の（ａ）及び（ｂ）によって収集を行うことができる。（ａ）Ｇｏｏｇｌｅ Ｗｅｂ Ｓｅｒｖｉｃｅ ＡＰＩ（商標）及びＡｌｅｘｓ Ｗｅｂ ｉｎｆｏｒｍａｔｉｏｎ Ｓｅｒｖｉｃｅ ＡＰＩ（商標）を用いて、ｕへのリンクを持つページを検索する。（ｂ）クローラを用いて、ｕへのリンクを持つページがある可能性が高いと考えられる場所を探索する。すなわち（１）ｕと同じサイト内のページ、（２）ｕの属するサイトをサブドメインとして含むサイトに属するページ、（３）ｕ中のリンクが指す先のページを探索する。したがって、必ずしもｕにリンクを持つ全てのページが収集されるとは限らない。

【００２３】

リンクオーソリティ候補決定手段５は、ランキング手段７と、属性記憶手段９と、候補決定手段１１とを備えて、収集した複数のＷｅｂページの中から予め定めた条件を満たす複数のＷｅｂページをリンクオーソリティ候補として定める。属性記憶手段９は、予め定めたリンクオーソリティとなり得るＷｅｂページが有する複数の属性を記憶する。この属性については、後に詳しく説明する。そしてランキング手段７は、属性記憶手段９に記憶された複数の属性を基準にして、リンクオーソリティ候補に含まれる複数のＷｅｂページをリンクオーソリティとして利用可能性が高いと推測される順にランキングする。候補決定手段１１は、ランキング手段７によるランキングの結果から、上位のランクにある複数のＷｅｂページをリンクオーソリティ候補として定める。

【００２４】

またリンクオーソリティ候補決定手段１３は、検索手段１５と、ランキング手段１７と最終決定手段１９とを備えている。この実施の形態の検索手段１５は、リンクオーソリティ候補中の上位のランクにある複数のＷｅｂページのそれぞれについて、各Ｗｅｂページにあるリンクの数と、リンク切れのリンクの数またはリンク切れではないリンクの数とを求める。なお各Ｗｅｂページにあるリンクの数については、検索手段１５とは別の手段で

求めるようにしてもよいのは勿論である。

【００２５】

そしてランキング手段１７は、検索手段１５により求めた数（リンクの数とリンク切れのリンクの数またはリンク切れではないリンクの数）と各Ｗｅｂページにあるリンクの数に基づいてリンク切れの割合が少ない順にランキングを行う。さらに最終決定手段１９は、ランキング手段１７によるランキングの結果から、上位のランクにある１以上のＷｅｂページをリンクオーソリティとして決定する。

【００２６】

ランキング手段7及びランキング手段１７で行う候補のランキングが、リンクオーソリティ決定装置１の本質的な処理である。図７に示すように、リンクオーソリティ決定装置１が、ｕのリンクオーソリティ候補$v_i \in V$をランキングしＶ及びV^*を求める処理（１）及び（２）は、例えば次のように設計することができる。まず収集した複数のＷｅｂページv_iに対して、それぞれ図８の表に示した属性のうち、「値」の欄に「真偽」と記載した属性を有するかを判定し、ランキング手段７によりランキングを行う。これらの属性は、属性記憶手段９に記憶されている。

【００２７】

次に図８の表に示した属性のうち、下から二つの属性「＃Ｌ」と「Ｂ」に基づいてランキング手段１７がランキングを行い、各v_iの「リンクオーソリティらしさ」を求める。

【００２８】

図８の表に示した属性を決める際に用いるヒューリスティクス（解決法）の選択肢Ｈ１～Ｈ９について、具体的に説明する。以下の選択肢Ｈ１～Ｈ８が、直接的に図８の属性として表現されている場合もあるが、間接的に図８の属性として表現されている場合もある。

【００２９】

Ｈ１：同一ディレクトリにｕとv_iとが存在すれば、v_iはｕへのリンクを確実に更新する可能性が高いと考えられる。

【００３０】

Ｈ２：v_iがｕに対して論理的に上位の存在である場合であり、この場合には、v_iはリンクオーソリティの可能性が高い。ここで論理的に上位の存在であるとは、例えばv_iが学科のページであるのに対して、ｕが学部のページであるといった場合である。しかし確実に更新される度合いは同一ディレクトリよりはやや劣ると考えられる。

【００３１】

Ｈ３：v_iがｕと同一サイトの上位ディレクトリに配置されている場合、v_iはｕに対して論理的に上位の存在であることが多い。

【００３２】

Ｈ４：v_iがｕに対して論理的に上位の存在である場合、Ｗｅｂサイトの設計の方法によっては、v_iはｕと同一ディレクトリ内のｉｎｄｅｘ．ｈｔｍｌとして配置されることがある。

【００３３】

Ｈ５：Ｈ１～Ｈ４より、v_iがｕと同一ディレクトリに配置され、且つｉｎｄｅｘ．ｈｔｍｌである場合はリンクオーソリティである可能性が非常に高い。

【００３４】

Ｈ６：v_iとｕとが直接的、間接的な相互リンクを持つ場合、v_iはリンクオーソリティである可能性が高い。ここで間接相互リンクとは、v_iとｕとのサイトの間で、異なるページを介し互いにリンクをしているような関係を指す（図９）。

【００３５】

Ｈ７：同一ディレクトリ内でv_1からv_2へリンクがあり、逆方向に存在しない場合、v_2がv_1に対して論理的に上位の存在であるとは考えにくい。

【００３６】

Ｈ８：（Ｈ７と比較して）v_1とv_2とが同一ディレクトリに配置され、お互いに相互

リンクが貼られている場合、これらの間の論理的な上位下位の関係は何ともいえない。なぜなら、「戻る」などのリンクが存在するからである。

【００３７】

Ｈ９：リンク切れが多いページはリンクオーソリティとは考えにくい。このヒューリスティクスは他のヒューリスティクスＨ１～Ｈ８とは独立していると考えられる。

【００３８】

ランキング手段７では、他とは独立していると考えられる上記Ｈ９を除いたヒューリスッティクスの選択肢（Ｈ１～Ｈ８）を考慮して図１０のランキング付けパターン表を作成した。このパターン表の各項目はv_iの属性を現しており、各属性の説明は図８に示してある。図１０において、黒丸はその属性が真であることを表し、空白は偽であることを表す。ハイフンはどちらでもよいことを表す。ランキング手段７におけるランキング処理では、収集された複数のＷｅｂページから選択されたＷｅｂページv_iが与えられると、まずこのパターン表を用いてv_iがどのパターンに属するのかを判定してランキング付けする。そして候補決定手段１１は、ランキング結果に基づいて上位ランク（例えばランク１～ランク５）に属するＷｅｂページをリンクオーソリティ候補として決定する。なお一般に、同じランクを持つv_iは複数存在する。

【００３９】

検索手段１５は、上記Ｈ９を反映させるために、候補決定手段１１が決定した候補となるＷｅｂページのそれぞれについて、そのＷｅｂページに設けられたリンク数とリンク切れではないリンクの数（またはリンク切れのリンクの数）を検索する。そしてランキング手段１７は、リンクの数とリンク切れではないリンクの数の割合を反映した値をキーとしてランキングを行う。

【００４０】

キーとして用いる値の演算方法としては、例えば、Ｗｅｂページに設けられたリンク数とリンク切れではないリンクの数との割合の相乗平均

【数１】

$$\sqrt{\#L \times B}$$

【００４１】

を求めることが考えられる。本実施の形態では、この相乗平均をキーとして降順に並べるランキングを実行する。そしてランキングにより並べられた結果をV^*とする。最終決定手段１９は、この結果V^*の中か上位に位置する１以上のＷｅｂページをリンクオーソリティとして決定する。

【００４２】

なおv_iが与えられたとき、図８に示した相互リンクの属性を求めるためには、ｕとv_iとの間に相互リンクがあるか否かを判断しなければならない。直接的な相互リンクを持つことは簡単に調べられるが、図９のような間接的な相互リンクを発見するためには多くのリンクを探索する必要がある。本来、この処理では図９の$site_u$と$site_v$との両サイトの内部を探索しなければならない。しかし予め処理（１）によってv_iからｕへリンクの存在が保証されていることを利用する。具体的には$site_u$だけの探索を行い、$site_v$中のいずれかのページに対するリンクを発見すると、ｕとv_iとの間に間接的な相互リンクが存在するとみなす。これは厳密にいえば$site_v$のv'とv_i間のリンクの存在を保証しない近似的な処理であるが、これにより探索処理を半分にすることができる。

【００４３】

次に図６に示したフローチャートを参照して本発明のリンクオーソリティの決定方法の実施の形態を、コンピュータを用いて実行する場合について説明する。本発明のリンクオーソリティ決定方法では、収集ステップ（ＳＴ１，ＳＴ２）と、リンクオーソリティ候補

決定ステップ（ＳＴ３，ＳＴ４）と、リンクオーソリティ決定ステップ（ＳＴ５～ＳＴ１０）とから構成される。収集ステップでは、監視対象とする複数のＷｅｂページから１つのＷｅｂページを選択し（ＳＴ１）、このＷｅｂページを監視対象として、このＷｅｂページのＵＲＬへのリンクを持つ複数のＷｅｂページを収集する（ＳＴ２）。次にリンクオーソリティ候補決定ステップでは、収集した複数のＷｅｂページの中から予め定めた条件を満たす複数のＷｅｂページをリンクオーソリティ候補として定める。そのために、まず予めリンクオーソリティとなり得るＷｅｂページが有する複数の属性を定める。実際には事前に定めて記憶した属性を用いる。次に複数の属性を基準にして、複数のＷｅｂページをリンクオーソリティとして利用可能性が高いと推測される順にランキングする（ＳＴ３）。ランキングの手法は任意である。そしてランキングの結果から上位のランクにある複数のＷｅｂページをリンクオーソリティ候補として定める（ＳＴ４）。ここで上位のランクにある複数のＷｅｂページの定め方は任意であり、例えば予め定めた順位（例えば１位及び２位）に属するＷｅｂページを候補と定める方法を採用してもよいし、また上位３０件以内に入るＷｅｂページを候補と定める方法を採用することができる。

【００４４】

次に、リンクオーソリティ決定ステップでは、リンクオーソリティ候補に含まれる複数のＷｅｂページの中からリンク切れを修正するために利用可能なリンクオーソリティを決定する。まず上位のランクにある複数のＷｅｂページのそれぞれについて、リンク切れのリンクの数またはリンク切れではないリンクの数を求める（ＳＴ５，ＳＴ６）。そしてこの数と各Ｗｅｂページにあるリンクの数とに基づいてリンク切れの割合が少ない順にランキングを行い（ＳＴ９）、ランキングの結果から上位のランクにある１以上のＷｅｂページをリンクオーソリティとして決定する（ＳＴ１０）。なお本実施の形態では、各Ｗｅｂページにあるリンクの数は、リンク切れのリンクの数またはリンク切れではないリンクの数を求めるときに一緒に求めている。

【００４５】

ランキングのために、具体的には、各Ｗｅｂページにあるリンクの数とリンク切れではないリンクの数の割合を反映した値（本実施の形態では相乗平均の演算値）を演算する（ＳＴ７）。すべての候補について、演算を実行し、その後この値（相乗平均の演算値）をキーとしてランキングを行う（ＳＴ９）。そしてランキングの上位にある１以上のＷｅｂページをリンクオーソリティとして決定する（ＳＴ１０）。リンクオーソリティとして決定する上位のランクにある１以上のＷｅｂページの数は、例えば、リンク数等を参考にして定めればよく、リンク数が少ない場合には、上位の複数のＷｅｂページをリンクオーソリティとして用いればよい。一つのＷｅｂページにあるリンクが切れているか否かの確認作業は、時間と手間（アクセス作業、確認作業）を要する。そこで本実施の形態では、リンクオーソリティ候補として予め絞られた複数のＷｅｂページに関してだけ、この確認作業を行う。そのため、収集したＷｅｂページのすべてについて確認作業を行う場合と比べて、本発明によれば、リンクオーソリティの決定までの時間と労力を少ないものとすることができる。特に、リンク切れの割合が少ないことを基準にしてランキングすると、より適正なリンクオーソリティ（更新率の高いリンクオーソリティ）を決定できる。図６のアルゴリズムからなるプログラムにおいて、ステップＳＴ２によってＷｅｂページ収集機能が実現され、ステップＳＴ３によりランキング機能が実現され、ステップＳＴ４によってリンクオーソリティ候補を定める機能が実現され、ステップＳＴ５及びＳＴ６によってリンクの数を求める機能が実現され、ステップＳＴ７からステップ法ＳＴ９によってランキング機能が実現され、ステップＳＴ１０でリンクオーソリティを決定する機能が実現されている。

【００４６】

次に、本発明の実施の形態を用いて実験を行った結果について説明する。リンクオーソリティ決定装置によって発見されたリンクオーソリティを利用することにより、リンク切れ自動修正システムの移動先発見精度はどのように変わるか検証を行った。

この実験ではまず、筑波大学、芝浦工業大学、北海道大学、東北大学、東京大学、名古

屋大学、京都大学、大阪大学、九州大学の計９大学のドメインに属するサイトの中に含まれるリンクを収集した。本実験で監視対象とするリンクは、これらの収集したリンクのうち、リンク元とリンク先が異なるサイトであるようなリンク（合計４９７５０個）である。したがって、リンクが指している先は学内のサイトとは限らない。監視対象をこれらのリンクに絞った理由は、リンク先とリンク元とが異なるサイトであればリンク切れが発生する確率が高いと考えたためである。

【００４７】

これらのリンクを対象に２００５年１月９日より実験を行った。２００５年２月４日時点で、ＬＩＭサーバは監視対象のうち１４６個のリンク切れを発見した（図１１）。このうちＷｅｂページ移動によって生じたと考えられるリンク切れは４７個存在した。ここで、ページ移動によるリンク切れとの判断は次のように行った。つまり、ＬＩＭサーバ探索ログなどを基に、様々な方法で移動先の探索を行い、移動先と考えられるページが発見できたものを移動ページとした。ここで移動先発見が成功した場合とは、ＬＩＭサーバが出力した結果の上位３位以内に正しい移動先が含まれている場合とした。

【００４８】

まず、移動先の探索結果を図１２に示す。図における「ＬＡサーバなし」の行はＬＡサーバを利用せず、「ＬＡサーバあり」の行はＬＡサーバを利用して探索を行った結果である。ＬＡサーバを利用せずに移動先の発見に成功したものは２８個、失敗したものは１９個、成功率は５９．６％であった。それに対してＬＡサーバを利用した場合は、移動先の発見に成功したものは３４個、失敗したものは１３個、成功率は７２．３％となった。この結果から見て分かるとおり、ＬＡサーバを利用することにより、ページの移動先発見数が２１％増加した。

【００４９】

一方で、ＬＡサーバを利用しても移動先の発見ができなかったものが１３個存在した。そこで、実験において発見されたＷｅｂページ移動に伴うリンク切れ４７個についてＬＡサーバが出力するログなどからＬＡサーバの探索結果についての分析を行った（図１３）。

【００５０】

その結果、ＬＡサーバがリンクオーソリティと考えられるページを発見でき、正しく更新されているページは４７個中１２個（図１３ｅ）であった。残る３５個については図１３ａ～ｄのように分析した。以下に分類毎の原因の検証を行う。

【００５１】

分類ａ：この分類に当てはまるものは１８個存在した。これは今回の実験で利用したリンクオーソリティ候補収集の手法が原因であると考えられる。本来、リンクオーソリティ候補群Ｖとしては、ｕに対してリンクを貼っている全てのページが収集されるべきである。しかし、今回実験で用いたリンクオーソリティ候補の収集手法ではすべてのページを収集していないためである。

【００５２】

分類ｂ：この分類に当てはまるものは１４個存在した。これは上述のａと同様の理由も考えられる。しかし、全てのＷｅｂページにリンクオーソリティが存在するわけではないとも考えられる。つまり、リンクオーソリティ候補の収集手法の追加や改良を行ってもこのパターンに該当するものを限りなく減少させることは不可能であると考えられる。

【００５３】

分類ｃ：実験ではこのパターンは存在していない。本実験の結果より、前述のリンクオーソリティ候補ランキングのヒューリスティクスが適切だったと考えることができる。この分類に当てはまるものが増加するようであれば、リンクオーソリティ候補のランキングヒューリスチクスの見直しが必要である。

【００５４】

分類ｄ：この分類に当てはまるものは３個存在した。これは、Ｗｅｂページの移動が発生してから、リンクオーソリティが持つリンクを参照するまでの時間が短かったためと考

えられる。正しいリンクオーソリティは、Ｗｅｂページの移動に伴いリンクを正しく修正すると考えられるが、それがページ移動が起こった直後に行われるとは限らない。この分類に当てはまるものは、ページ移動後、ある程度時間が経過した後にリンクオーソリティが有効に機能するのではないかと考えられる。

【００５５】

本実験では、特に分類ａと分類ｂとに該当するものが多く存在した。この原因は、本実験で利用したＬＡサーバのリンクオーソリティ候補Ｖの収集手法に偏りがあったものと考えられる。リンクオーソリティ候補ｖを効率的に収集するために、候補収集の手法を次に示す手順で行うことが好ましい。

【００５６】

手順１：全ての監視対象のＵＲＬに対してＡｌｅｘａ Ｗｅｂ ｉｎｆｏｒｍａｔｉｏｎ Ｓｅｖｅｉｃｅ ＡＰＩ（商標）を用いて、ｕへのリンクを持つページを収集。

【００５７】

手順２：全ての監視対象のＵＲＬに対してＧｏｏｇｌｅ Ｗｅｂ Ｓｅｒｖｉｃｅ ＡＰＩ（商標）を用いて、ｕへのリンクを持つページを収集。

【００５８】

手順３：全ての監視対象のＵＲＬに対してクローラを用いてｕへのリンクを持つページがある可能性が高いと考えられる場所からの収集。

【００５９】

上記実験結果では、手順１を用いた収集しか行っていない。そのため、候補の収集が必ずしも十分なものとならなかったものと考えられる。

【００６０】

また、ＬＡサーバによってリンクオーソリティと考えられるページが発見されていた１２個については、ページ移動先の発見に及ぼした影響について検討した（図１４）。その結果、「リンクオーソリティが発見されていなければ、正しい移動先を上位３位以内にランキングできなかった」というものが２個存在した。つまり、ＬＡサーバを利用することで、これまでのＬＩＭサーバだけでは発見できない移動先を発見することができた。その一方で、「リンクオーソリティは発見できたが、正しい移動先を上位３位以内評価できなかった」というものが１個存在した。しかし、これは、ＬＩＭサーバによる移動先候補の評価手法の問題であり、ＬＡサーバの問題ではなかった。

【００６１】

上記の実験結果と検討から分かるように本発明の方法及び装置の実施の形態によれば、従来と比べて、短時間の内に高い精度でリンクオーソリティを決定できることが分かる。

【図面の簡単な説明】

【００６２】

【図１】リンクオーソリティの考え方を説明するための図である。

【図２】リンクオーソリティ決定装置の概要を説明するために用いる図である。

【図３】自動修正システムの構成（アーキテクチャ）を示す図である。

【図４】ＬＩＭサーバの基本動作を説明するためのアルゴリズムを示す図である。

【図５】主として前述のＬＡサーバーによって構成される本発明のリンクオーソリティ決定装置の実施の形態の構成の一例を示すブロック図である。

【図６】リンクオーソリティ決定装置をコンピュータを用いて実現する場合に用いるプログラムのアルゴリズムを示すフローチャートである。

【図７】リンクオーソリティ決定装置の処理を説明するために用いる図である。

【図８】属性の内容を示す図である。

【図９】間接相互リンクを説明するための図である。

【図１０】ランキング付けパターンの表を示す図である。

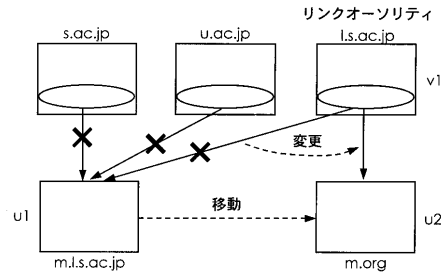
【図１１】実験期間内に発生したリンク切れの数を示す図である。

【図１２】移動先の探索結果を示す図である。

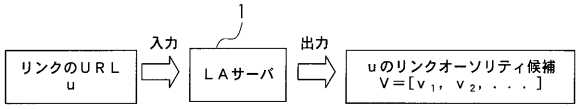
【図１３】リンクオーソリティの探索結果の分類を示す図である。

【図１４】移送先探索におけるリンクオーソリティを利用した影響を示す図である。
【符号の説明】
【００６３】
１　リンクオーソリティ決定装置
３　Ｗｅｂページ収集手段
５　リンクオーソリティ候補決定手段
７　ランキング手段
９　属性記憶手段
１１　候補決定手段
１３　リンクオーソリティ決定手段
１５　検索手段
１７　ランキング手段
１９　最終決定手段

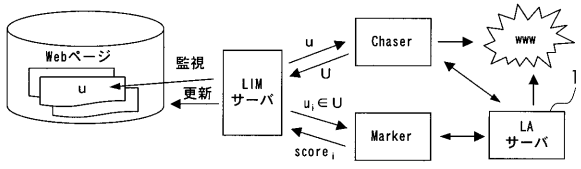
【図１】



【図２】



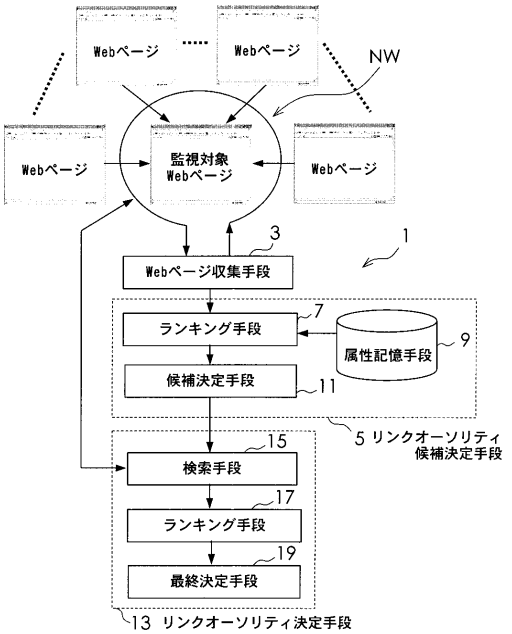
【図３】



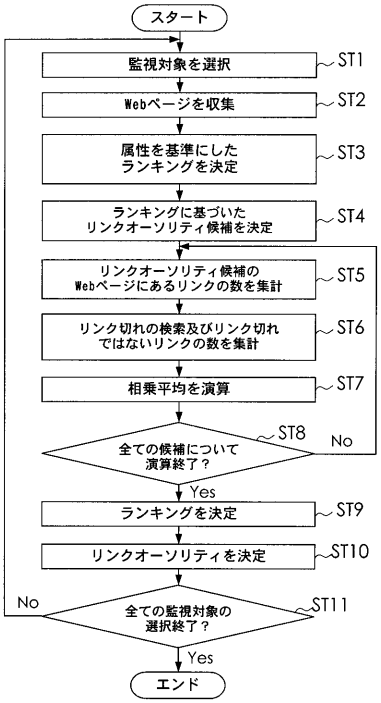
【図４】

```
1.LIM- Server(u){
2.  while(true){
3.    if(isBroken(u)){//link is broken
4.      Set U=Chaser(u);
5.      for each u i in U{
6.        score i=Maker(u i);
7.    }
8.    ListU=the result of scoring U by score i;
9.  }
10.  interval( ):
11. }
12.}
```

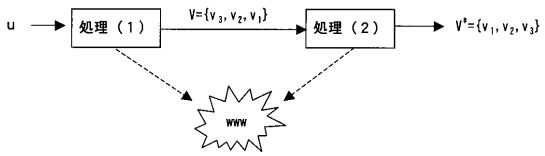
【図５】



【図６】



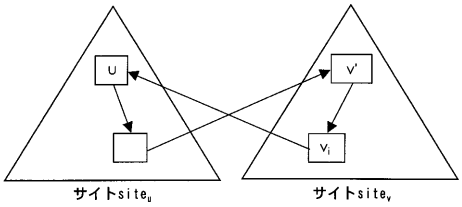
【図７】



【図８】

v_iの属性	値	意味
同一同一	真偽	uと同一サイトの同一ディレクトリにv_iが存在
同一上位	真偽	uと同一サイトの上位ディレクトリにv_iが存在
同一下位	真偽	uと同一サイト且つ下位ディレクトリにv_iが存在
同一その他	真偽	uと同一サイト且つその他のディレクトリにv_iが存在
上位サイト	真偽	uが属するサイトをサブドメインとして含むサイトにv_iが存在
外部サイト	真偽	uの上位サイト及び同一サイト以外のサイトにv_iが存在
相互リンク	真偽	uとv_iとの間に直接的若しくは間接的な相互リンクが存在
index	真偽	v_iのファイル名がデフォルトファイル名 （典型的にはindex.html）である
#L	自然数	v_iのページに含まれているリンクの数
B	[0,1]	v_iのページに含まれているリンクのうち、 リンク切れではないものの割合

【図９】



【図１０】

パターン	ランク	同一サイト				上位サイト	下位サイト	相互リンク	index
		上位	同一	下位	その他				
a	1		●					-	●
b	2	●						●	-
c	2					●		●	-
d	2						●	●	-
e	3		●						-
f	4	●							-
g	5		●					●	-
h	6				●			●	-
i	7				●				-
j	7					●			-
k	8			●				-	-
l	8						●		-

【図１１】

監視対象ページ数	49750
リンク切れ総数	146
内、移動に伴うリンク切れ	47

【図１２】

	成功	失敗	成功率
LAサーバなし	28	19	59.6%
LAサーバあり	34	13	72.3%

【図１３】

分類	詳細	個数
a	V＝φ	18
b	V≠φだが、LAらしきページは含まれていない	14
c	VにLAと思われるページが含まれているが、V'の上位ではない	0
d	V'の上位３位以内で発見されたが、まだリンクが更新されていない	3
e	V'の上位５位以内で発見され、正しくリンクが変更された	12

【図１４】

効果	詳細	個数
あり	リンクオーソリティが発見されていなければ、正しい移動先を発見できなかった	4
	リンクオーソリティが発見されていなければ、正しい移動先を上位３位以内に評価できなかった	2
なし	リンクオーソリティは発見されたが、正しい移動先を上位３位以内に評価できなかった	1
	リンクオーソリティは発見されたが、移動先の発見に特に影響を与えなかった	5

フロントページの続き

(72)発明者　北川　博之
茨城県つくば市天王台一丁目１番１　国立大学法人筑波大学内
(72)発明者　中溝　昌佳
神奈川県川崎市川崎区貝塚２－７－９

Ｆターム(参考) 5B075 NK44
5B082 HA08