

(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2010-61178

(P2010-61178A)

(43) 公開日 平成22年3月18日(2010.3.18)

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

審査請求 未請求 請求項の数 18 O L (全 29 頁)

(21) 出願番号 特願2008-217635 (P2008-217635)
 (22) 出願日 平成20年8月27日 (2008.8.27)
 (31) 優先権主張番号 特願2008-201639 (P2008-201639)
 (32) 優先日 平成20年8月5日 (2008.8.5)
 (33) 優先権主張国 日本国 (JP)

(71) 出願人 000006747
 株式会社リコー
 東京都大田区中馬込1丁目3番6号
 (74) 代理人 100110607
 弁理士 間山 進也
 (72) 発明者 荒木 禎史
 東京都大田区中馬込1丁目3番6号 株式
 会社リコー内
 (72) 発明者 徐 盈輝
 東京都大田区中馬込1丁目3番6号 株式
 会社リコー内
 (72) 発明者 池田 哲也
 東京都大田区中馬込1丁目3番6号 株式
 会社リコー内
 Fターム(参考) 5B075 PP26 PR03

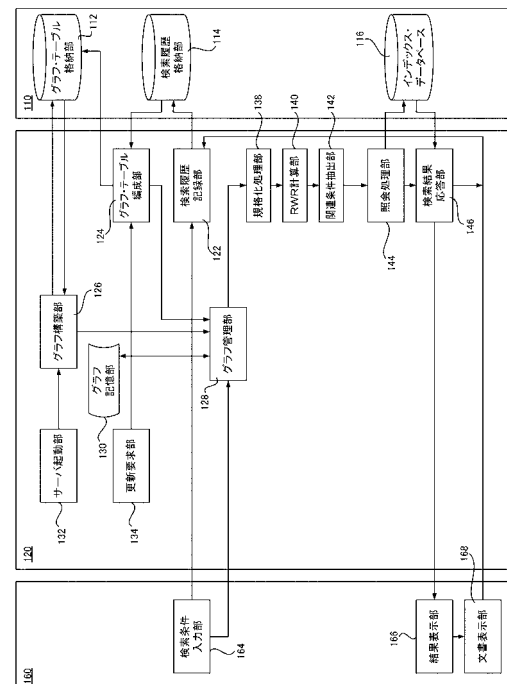
(54) 【発明の名称】 情報処理装置、情報検索システム、情報処理方法およびプログラム

(57) 【要約】

【課題】 入力された検索条件に関連性の高い過去の検索条件を検索する情報処理装置、情報検索システム、情報処理方法およびプログラムを提供すること。

【解決手段】 本発明の情報処理装置120は、検索条件と、該検索条件の検索結果に関連する関連情報とを対応付けて、検索履歴として記録する記録手段122と、検索履歴を読み出して、検索条件と関連情報とをノードとする状態遷移グラフを構成する構成手段(124, 126, 128)と、新たに入力された新規検索条件を含む検索要求にตอบสนองして、状態遷移グラフを用いて、新規検索条件と過去の検索条件との間の関連度を計算する計算手段140と、計算された関連度に従って、新規検索条件に適合する関連検索条件を抽出する抽出手段142と、抽出された関連検索条件を用いて検索要求に対する応答データを生成する応答手段146とを含み、情報検索のための検索条件を処理する。

【選択図】 図3



【特許請求の範囲】**【請求項 1】**

情報検索のための検索条件を処理する情報処理装置であって、
検索条件と、該検索条件の検索結果に関連する関連情報とを対応付けて、検索履歴として記録する記録手段と、

前記検索履歴を読み出して、前記検索条件と前記関連情報とをノードとする状態遷移グラフを構成する構成手段と、

新たに入力された新規検索条件を含む検索要求に応答して、前記状態遷移グラフを用いて、前記新規検索条件と過去の検索条件との間の関連度を計算する計算手段と、

計算された前記関連度に従って、前記新規検索条件に適合する関連検索条件を抽出する抽出手段と、

抽出された前記関連検索条件を用いて前記検索要求に対する応答データを生成する応答手段と

を含む、情報処理装置。

【請求項 2】

前記計算手段は、前記過去の検索条件間での検索結果に含まれる前記関連情報の重複度、および該過去の検索条件間でのそれぞれが含有する単語の類似度の少なくとも一方から算出される該過去の検索条件のノード間の重み付けを用いて、前記関連度を計算する、請求項 1 に記載の情報処理装置。

【請求項 3】

前記計算手段は、さらに、前記新規検索条件と前記過去の検索条件との間でのそれぞれが含有する単語の類似度から算出されるノード間の重み付けを用いて、前記関連度を計算する、請求項 1 または 2 に記載の情報処理装置。

【請求項 4】

前記記録手段は、検索者ユーザの識別値をさらに前記検索条件に対応付けて記録し、前記計算手段は、前記検索者ユーザのノード間の重み付けを用いて、前記関連度を計算する、請求項 1 ～ 3 のいずれか 1 項に記載の情報処理装置。

【請求項 5】

前記応答手段は、抽出された前記関連検索条件を、前記新規検索条件の検索者ユーザの検索履歴から求められた条件と、他の検索者ユーザのものを含む検索履歴から求められた条件とを区別して表示させるように前記応答データを生成する、請求項 4 に記載の情報処理装置。

【請求項 6】

前記計算手段は、前記関連情報のノードとの間の重み付けを含んで前記新規検索条件のノードが加えられた前記状態遷移グラフを用いて、前記状態遷移グラフを表現する隣接行列にランダムウォーク・ウィズ・リスタート方式を適用して前記関連度を計算する、請求項 1 ～ 5 のいずれか 1 項に記載の情報処理装置。

【請求項 7】

前記計算手段は、検索条件と関連情報との関連性に対する信頼度に対応して区別される種別の相違に応答して求められる該検索条件および該関連情報のノード間の重み付けを用いて、上記関連度を計算する、請求項 1 ～ 6 のいずれか 1 項に記載の情報処理装置。

【請求項 8】

前記関連情報の種別は、検索条件による検索結果から参照された参照情報、該検索結果から適合性フィードバックによる再検索のために指定された適合情報、および該検索結果の詳細情報から明示的に指定された指定情報を含む、請求項 7 に記載の情報処理装置。

【請求項 9】

請求項 1 ～ 8 のいずれか 1 項に記載の情報処理装置と、

前記検索履歴を蓄積するデータベースと、

検索条件に従って情報検索を実行して、前記検索された情報の一覧を含む検索結果を前記情報処理装置へ返す情報検索エンジンと

10

20

30

40

50

を含む、情報検索システム。

【請求項 10】

情報検索のための検索条件を処理する情報処理方法であって、情報処理装置が、検索条件と、該検索条件の検索結果に関連する関連情報とを対応付けて、検索履歴として記録するステップと、

前記検索履歴を読み出して、前記検索条件と前記関連情報とをノードとした状態遷移グラフを構成するステップと、

新たに入力された新規検索条件を含む検索要求に応答して、前記状態遷移グラフを用いて、前記新規検索条件と過去の検索条件との間の関連度を計算するステップと、

計算された前記関連度に従って、前記新規検索条件に適合する関連検索条件を抽出するステップと、

抽出された前記関連検索条件を用いて前記検索要求に対する応答データを生成するステップと

を実行する、情報処理方法。

【請求項 11】

情報処理装置は、前記計算するステップでは、前記過去の検索条件間での検索結果に含まれる前記関連情報の重複度、および該過去の検索条件間でのそれぞれが含有する単語の類似度の少なくとも一方から算出される当該過去の検索条件のノード間の重み付けを用いて、前記関連度を計算する、請求項 10 に記載の情報処理方法。

【請求項 12】

情報処理装置は、前記計算するステップでは、さらに、前記新規検索条件と前記過去の検索条件との間でのそれぞれが含有する単語の類似度から算出されるノード間の重み付けを用いて、前記関連度を計算する、請求項 10 または 11 に記載の情報処理方法。

【請求項 13】

情報処理装置は、前記記録するステップでは、検索者ユーザの識別値をさらに前記検索条件に対応付けて前記検索履歴として記録し、前記計算するステップでは、前記検索者ユーザのノード間の重み付けを用いて、前記関連度を計算する、請求項 10 ~ 12 のいずれか 1 項に記載の情報処理方法。

【請求項 14】

情報処理装置は、前記応答するステップでは、抽出された前記関連検索条件を、前記新規検索条件の検索者ユーザの検索履歴から求められた条件と、他の検索者ユーザのものを含む検索履歴から求められた条件とを区別して表示させるように前記応答データを生成する、請求項 13 に記載の情報処理方法。

【請求項 15】

情報処理装置は、前記計算するステップでは、前記関連情報のノードとの間の重み付けを含んで前記新規検索条件のノードが加えられた前記状態遷移グラフを用いて、前記状態遷移グラフを表現する隣接行列にランダムウォーク・ウィズ・リスタート方式を適用して前記関連度を計算する、請求項 10 ~ 14 のいずれか 1 項に記載の情報処理方法。

【請求項 16】

情報処理装置は、前記計算するステップでは、検索条件と関連情報との関連性に対する信頼度に対応して区別される種別の相違に応答して求められる該検索条件および該関連情報のノード間の重み付けを用いて、上記関連度を計算する、請求項 10 ~ 15 のいずれか 1 項に記載の情報処理方法。

【請求項 17】

前記関連情報の種別は、検索条件による検索結果から参照された参照情報、該検索結果から適合性フィードバックによる再検索のために指定された適合情報、および該検索結果の詳細情報から明示的に指定された指定情報を含む、請求項 16 に記載の情報処理方法。

【請求項 18】

情報処理装置を、請求項 1 ~ 8 のいずれか 1 項に記載の各手段として機能させるための装置実行可能なプログラム。

10

20

30

40

50

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、情報検索技術に関し、より詳細には、入力された検索条件に関連性の高い過去の検索条件を検索する情報処理装置、情報検索システム、情報処理方法およびプログラムに関する。

【背景技術】

【0002】

近年の情報技術の発展に伴い、膨大な情報がデータベースやファイルシステム上に蓄積されるようになり、所望の情報を効率的に検索するための情報検索技術の重要性が高まっている。情報検索において、検索者は、検索条件を入力して検索を要求し、ヒットした情報のリストなどとして得られる検索結果の中から、検索者の目的に適合した情報を見つけ出すこととなる。そして、検索結果の中に適当な情報がない場合は、検索者は、検索条件を変更して再度検索することとなるが、その際に、検索者自身や他の利用者が入力した過去の検索条件を容易に参考にすることができれば、検索効率を向上させることができるものと考えられる。

【0003】

このような観点から、特開2001-109752号公報（特許文献1）は、データベース検索時に過去の検索経験を活かすことで、効率の良い検索ができるようにすることを目的として、情報検索システムの提供するデータベース検索のためのグラフィカル・ユーザ・インタフェース（以下、GUIとして参照する。）にて検索者の検索条件を読み込むことで、検索者が目的とする情報を検索・収集するとともに、この時の検索処理と並列して情報のコンテンツを得るまでの検索情報および得られたコンテンツの表題とリンク情報からなる検索履歴を一時的に保持・記録し、そして一時的に保持しておいた検索履歴が検索者によって確認および削除・整理された後、通常の情報とは別に分類して検索履歴データベースに登録し、この登録した検索履歴を新たに類似した目的の情報検索の際に参照して検索者に提示することを特徴とする情報検索方法を開示している。

【特許文献1】特開2001-109752号公報

【発明の開示】

【発明が解決しようとする課題】

【0004】

しかしながら、上記特許文献1に開示される従来技術では、過去の類似した検索履歴を提示する際に、検索用のユーザ・インタフェースとは異なるユーザ・インタフェースを「検索履歴の検索」のために用いる必要があったため、検索者は、検索条件とは別に検索履歴の検索のための検索条件を入力する必要があるため、ユーザ利便性の観点から十分なものではなかった。

【0005】

また、検索履歴に登録する際も、検索結果に含まれる各コンテンツに対して、検索者が登録の可否を判断して明示的に指定する必要があったため、この観点からも十分なユーザ利便性を提供できるものではなかった。さらに上記従来技術では、上記検索履歴は、検索情報と得られたコンテンツの表題とリンク情報とから構成されているが、これらの情報が検索履歴の検索のための検索条件と一致する内容を含む検索履歴のみが提示されることとなるため、過去の検索履歴を十分に活用できていないという点で、十分なものではなかった。例えば、「図書館」と「library」とは、互いに意味の関連性が強い単語であるが、「検索履歴の検索」のための検索条件として「図書館」のキーワードを指定した場合、上記検索履歴の検索情報やコンテンツの表題の中に「図書館」が含まれていない検索履歴については、例え「library」が含まれていた場合であっても、上記類似した目的の情報検索の際に提示されることはなかった。

【0006】

さらに、上記従来技術では、自分の検索履歴と他人の検索履歴とが区別されて蓄積され

10

20

30

40

50

ないため、「検索履歴の検索」のための検索条件に対し、検索者の別にかかわらず共通の結果が与えられることとなるため、必ずしも各検索者の興味および関心に適合した検索履歴を提示できるとは限らなかった。また、上記従来技術では、自分のものと他人のものとを区別して「検索履歴の検索」の結果が提示されないため、検索者は、検索目的に合致した結果を識別するにあたって困難性があった。例えば、検索者にとって、自身の過去の検索行動や思考を思い出したい場合と、自身とは異なる新たな発想を広げたい場合とでは、それぞれ適した情報が異なると考えられるため、自身の過去の検索経験が反映された結果と、全体またはグループ内の他の検索者による過去の検索経験が反映された結果とを区別して提示させたいという要望があった。

【0007】

10

本発明は、上記問題点に鑑みてなされたものであり、検索者自身や他の検索者の過去の検索経験が適切に反映された関連の検索条件を効率的に抽出し、検索者に手間をかけることなく、その興味、関心および検索目的に適合した過去の検索条件を容易に再利用可能とする情報処理装置、情報検索システム、情報処理方法およびプログラムを提供することを目的とする。

【課題を解決するための手段】

【0008】

20

本発明では、上記課題を解決するために、検索条件と、該検索条件の検索結果に関連する関連情報とを対応付けて検索履歴として記録し、検索条件と関連情報とをノードとする状態遷移グラフを構成する。そして、新たに入力された新規検索条件を含む検索要求に

【0009】

30

上記構成により、検索条件に含まれる単語自体の類似性のみならず、各検索条件と関連情報との対応付けから辿られる検索条件間の間接的な関係性が活用されて関連検索条件が抽出されるため、共通の単語を含まないような検索条件であっても、間接的な関係性が高いものであれば関連検索条件として抽出される可能性があり、過去の検索履歴をより有効に活用することが可能となる。例えば、上記の例では、「図書館」と「library」とは、互いに意味の関係性が強いものの表現が相違する単語であるが、上記構成によれば、「図書館」を単語として含まない検索条件であっても、「図書館」を単語として含む検索条件は、「library」を含む検索条件と間接的な関係性を有している可能性があり、これにより、「図書館」を単語として含まなくとも「library」を含む過去の検索条件が抽出される可能性が生まれる。また、一度の検索条件の入力により、当該検索条件による検索とともに、検索履歴から関連検索条件の抽出が実行されるため、検索履歴からの検索条件を抽出するための条件の別途入力が必要とされ、高いユーザ利便性が提供される。

【0010】

40

また本発明では、過去の検索条件間での検索結果に含まれる関連情報の重複度、および該過去の検索条件間でのそれぞれが含有する単語の類似度の少なくとも一方から算出される該過去の検索条件のノード間の重み付けを用いて関連度を計算することができる。上記関連情報は、検索結果にリストされた情報および検索結果から参照された情報、またはユーザから指定された適合情報を用いて拡張した検索条件にヒットした情報とすることができる。この構成により、関連検索条件の抽出の際に、過去の検索条件間の内容の直接的な関係性だけでなく検索結果を介した間接的な関係性も好適に取り込むことが可能となる。また、検索結果から参照された情報を関連情報とする場合には、検索者は、なんら追加の動作を行うことなく関係性の高い情報を登録することができ、関係情報の登録の可否を判断する手間を省きつつも、抽出の精度を高めることができる。また、指定の適合情報を用いて拡張した検索条件にヒットした情報を関連情報とする場合には、検索者に特別な心理的負担を感じさせることなく、抽出の精度を高めることができる。

50

【 0 0 1 1 】

さらに本発明では、さらに、新規検索条件と過去の検索条件との間でのそれぞれが含有する単語の類似度から算出されるノード間の重み付けを用いて、上記関連度を計算することができる。新規検索条件と過去の検索条件との間の単語上の類似性が取り込まれるため、検索者の最大の関心を表現している新規検索条件の内容自体に類似する条件が、より高い関連度として抽出されるため、より検索者の意図が反映された過去の検索条件を容易に抽出可能とする。

【 0 0 1 2 】

さらにまた、本発明では、検索者ユーザの識別値をさらに検索条件に対応付けて記録することができる。検索者ユーザのノード間の重み付けを用いて、上記関連度を計算することができる。この構成により、検索者ユーザの興味や関心を反映した過去の検索条件を容易に抽出可能とする。また本発明では、抽出された関連検索条件を、新規検索条件の検索者ユーザの検索履歴から求められた条件と、他の検索者ユーザのものを含む検索履歴から求められた条件とを区別して表示させるように応答データを生成することができる。この構成により、検索者は、検索目的に適した検索条件を容易に識別することが可能となる。

10

【 0 0 1 3 】

またさらに本発明では、関連情報のノードとの間の重み付けを含んで新規検索条件のノードが加えられた状態遷移グラフを用いて、当該状態遷移グラフを表現する隣接行列にランダムウォーク・ウィズ・リスタート方式を適用して関連度を計算することができる。この構成により、新規検索条件と過去の検索条件との間の関連度を効率的に計算することが可能となる。

20

【 0 0 1 4 】

また本発明では、さらに、検索条件と関連情報との関連性に対する信頼度に対応して区別される種別の相違に応答して求められる検索条件および関連情報のノード間の重み付けを用いて、上記関連度を計算することができる。この関連情報の種別としては、例えば、検索条件による検索結果から参照された参照情報、該検索結果から適合性フィードバックによる再検索のために指定された適合情報、および該検索結果の詳細情報から明示的に指定された指定情報を含むことができる。その他、ユーザから指定された、関連情報に対する重要度の指定を上記関連情報の種別の区分に対応させることもできる。

【 0 0 1 5 】

30

ユーザが単に検索結果から参照した参照情報と、適合性フィードバックにおいてユーザが適合するものと指定した適合情報、適合するものとしてユーザが明示的に指定した指定情報とでは、検索者ユーザの興味や関心の程度が異なり、これらの関連情報と検索条件との間の関連性の程度の観点からは、その信頼度は、相違するものと考えられる。上記構成によれば、このような関連情報の種別による関連性の信頼度の相違を、関連情報および検索条件のノード間の重み付けの値の相違として取り込み、検索者ユーザの興味や関心を、より正しく反映する過去の検索条件を抽出可能とし、抽出の精度を高めることができる。

【 発明を実施するための最良の形態 】

【 0 0 1 6 】

40

以下、本発明の実施形態を説明するが、本発明の実施形態は、以下の実施形態に限定されるものではない。なお、以下の実施形態では、文書を対象とした情報検索を実行する文書検索システムを一例として説明する。

【 0 0 1 7 】

図 1 は、第 1 の実施形態による文書検索システム 1 0 0 を示す概略図である。図 1 に示す文書検索システム 1 0 0 は、検索要求を送信するクライアント・コンピュータ（以下、単にクライアントとして参照する。） 1 6 0 a , b と、クライアント 1 6 0 から受信した上記検索要求を処理する W e b アプリケーション・サーバ（以下、W e b サーバとして参照する。） 1 2 0 と、文書検索のためのインデックスを記憶するデータベース 1 1 0 a を管理し、文書検索エンジンとして機能するデータベース・サーバ（以下、D B サーバとして参照する。） 1 1 0 とを含んで構成されている。上記 D B サーバ 1 1 0 、W e b サーバ

50

120およびクライアント160は、イーサネット（登録商標）およびTCP/IPプロトコルによるローカル・エリア・ネットワーク（以下、LANとして参照する。）などのネットワーク170を介して、相互に接続されている。上記ネットワーク170の構成は、特に限定されるものではなく、クライアント160とWebサーバ120の間には、インターネットなどの公衆ネットワークを含んでいてもよい。

【0018】

本実施形態の文書検索システム100は、文書検索に加え、新たに要求された検索条件に関連する過去の検索条件（以下、関連検索条件として参照する。）を抽出して、クライアント160に提示するよう構成されており、上記DBサーバ110は、関連検索条件を抽出するために、さらに、システムに対し過去に要求された検索の履歴をデータベース110a上に蓄積している。クライアント160は、Webサーバ120に対し、文書検索とともに上記関連検索条件の抽出を要求する。Webサーバ120は、クライアント160から受信した検索要求に応答して、文書検索エンジンとしてのDBサーバ110に照会するとともに、上記関連検索条件の抽出の要求に応答して、DBサーバ110が蓄積する検索履歴を利用した関連検索条件の抽出処理を実行し、検索結果および抽出結果をクライアント160へ返信する。

【0019】

上記DBサーバ110は、より具体的には、DB2（登録商標）、Oracle（登録商標）、MySQL（登録商標）、PostgreSQL（登録商標）などのデータベース管理システム（Database Management System）を実装しており、関係データベース、オブジェクト関係データベースまたはオブジェクト・データベースなどとして構成されるデータベース110aを管理している。DBサーバ110は、検索履歴、文書検索のためのインデックスの他、後述する関連検索条件の抽出処理に用いる状態遷移グラフを構成する構成データを格納している。

【0020】

クライアント160は、ブラウザ・ソフトウェアを実装しており、Webサーバ120から送信されるHTML、XHTML、XMLなどの構造化文書を使用して、文書検索および関連検索条件の抽出の指示、および結果表示を行うためのグラフィカル・ユーザ・インタフェース（以下、GUIとして参照する。）を提供している。本実施形態で使うことができるブラウザ・ソフトウェアとしては、これまで知られたいかなるものでも使用することができ、例えば、Internet Explorer（登録商標）、Netscape Navigator（登録商標）、Firefox（登録商標）、Mosaic（登録商標）、Opera（登録商標）などの適切なブラウザ・ソフトウェアを挙げることができる。

【0021】

Webサーバ120は、IIS（Internet Information Service：登録商標）、Apache（登録商標）などのWebサーバ・ソフトウェアおよび、PHP、Java（登録商標）Servlet、JSPなどの適切なWebアプリケーションを実装しており、HTTP（HyperText Transfer Protocol）に従ったトランザクションを処理している。本実施形態のWebサーバ120は、クライアント160から送信される検索条件を受信し、検索結果、および該検索条件に関連する関連検索条件の抽出結果をクライアント160上に表示させるための応答データを返信する。また、Webサーバ120は、JDBC、ODBC（Open Database Connectivity）などのAPI（Application Programming Interface）を含んで構成され、データベース110aに接続して、検索履歴の取得、状態遷移グラフの構成テーブルの取得および格納、およびクライアントから要求された検索の照会を行っている。

【0022】

以下、第1の実施形態の文書検索システム100で用いられるWebサーバ120の構成について、詳細を説明する。図2は、第1の実施形態によるWebサーバ120のハードウェア構成を示す図である。Webサーバ120は、マイクロプロセッサ・ユニット（MPU）12と、BIOS（Basic Input Output System）を格納する不揮発性メモリ14

10

20

30

40

50

と、M P U 1 2 によるプログラム処理を可能とする実行記憶空間を提供するR A Mなどのメモリ16とを含んで構成されている。起動時にM P U 1 2は、不揮発性メモリ14からB I O Sを読み出して、システム診断を行なうとともに入出力装置26の管理を行っている。

【0023】

M P U 1 2は、内部バス22を介して記憶制御用インタフェース18に接続され、ハードディスク20が、M P U 1 2からの入出力要求に应答してデータの書込または読み出しを実行する。記憶制御用インタフェース18としては、I D E (Integrated Device Electronics)、A T A (AT Attachment)、シリアルA T A、U l t r a A T Aなどの規格により、ハードディスク20の入出力を管理するインタフェースを使用することもできる。M P U 1 2は、内部バス22を介してU S B、I E E E 1 1 6 4などのシリアルまたはパラレル・インタフェース24を制御して、キーボード、マウス、プリンタなどの入出力装置26と通信して、ユーザからの入力を受け取る。

【0024】

W e bサーバ120は、さらにV R A M 28とグラフィック・チップ30とを含んでおり、M P U 1 2からの指令に应答してビデオ信号を処理し、ディスプレイ装置32へと表示させている。また、M P U 1 2は、内部バス22を介してネットワークI / F (N I C ; Network Interface Card) 34と通信し、W e bサーバ120を、ネットワーク170を通して、D Bサーバ110およびクライアント160と通信させている。

【0025】

W e bサーバ120は、不揮発性メモリ14やハードディスク20、その他、N V - R A M (図示せず) やS Dカード (図示せず) などの記憶装置に格納されたプログラム (図示せず) を読み出し、メモリ16のメモリ領域に展開することにより、適切なオペレーティング・システム (O S) のもとで、後述する各機能手段および各処理を実現している。上記O Sとしては、W i n d o w s (登録商標)、U N I X (登録商標) またはA I XまたはL I N U X (登録商標) など、如何なるアーキテクチャを有するO Sを採用することができる。なお、本実施形態のD Bサーバ110およびクライアント160も、図1に示すハードウェア構成と同様の構成とすることができる。

【0026】

以下、文書検索システム100の機能構成について説明する。図3は、第1の実施形態による文書検索システム100の機能ブロックを示し、図3には、D Bサーバ110の機能ブロック、W e bサーバ120の機能ブロック、およびクライアント160の機能ブロックが示されている。

【0027】

図3に示すクライアント160は、検索条件の入力を受け付ける検索条件入力部164を含んで構成されている。クライアント160とW e bサーバ120との間のメッセージ交換は、クッキーなどによってセッション管理されている。検索条件が入力され、G U Iを介した検索指示を受けると、検索条件入力部164は、入力された検索条件を含む検索要求をW e bサーバ120へ送信する。

【0028】

ここで、図4を参照して、検索要求のために用いられるG U I画面について説明する。図4は、クライアントのディスプレイ上に表示されるG U I画面を示す。図4 (A) に示すG U I画面300は、第1の実施形態の文書検索システム100において検索要求を発行するためのG U Iを提供し、検索条件を入力するためのテキスト・ボックス302と、検索要求を指示するための検索ボタン304とを含んで構成される。検索者は、G U I画面300を用いて、検索条件の入力および検索要求を指示することができる。

【0029】

再び図3を参照すると、D Bサーバ110は、検索履歴を格納する検索履歴格納部114と、文書検索のための全文テキスト検索インデックスを格納するインデックス・データベース116とを含んで構成される。W e bサーバ120は、検索履歴の記録処理を実行

する検索履歴記録部 1 2 2 と、要求された検索条件による照会処理を行う照会処理部 1 4 4 と、照会結果を得て要求に対する検索結果を含む応答データを生成する検索結果応答部 1 4 6 とを含んで構成される。

【 0 0 3 0 】

上記照会処理部 1 4 4 は、クライアント 1 6 0 から受信した検索条件を、SQL などのデータベース言語に変換してインデックス・データベース 1 1 6 に照会し、検索結果応答部 1 4 6 は、照会結果として検索条件に該当する文書（以下、検索文書として参照する。）の書誌情報およびハイパーリンク情報から構成される文書リストを取得する。照会結果を取得した検索結果応答部 1 4 6 は、検索要求に対する検索結果の応答データを生成して、クライアント 1 6 0 へ返信する。また、検索結果応答部 1 4 6 は、後述する状態遷移グラフのノード間の重み付けを計算するために、検索文書のリストを検索履歴記録部 1 2 2 へ報告することができる。報告を受けた検索履歴記録部 1 2 2 は、検索条件入力部 1 6 4 から送信された検索要求に含まれる検索条件と、報告された該検索条件による検索結果に含まれる検索文書とを相互に関連付けて検索履歴格納部 1 1 4 へ記録する。

10

【 0 0 3 1 】

クライアント 1 6 0 は、さらに、受信した応答データに従って画面表示する結果表示部 1 6 6 と、検索結果中の検索文書へのハイパーリンク情報に従って文書へアクセスし表示するための文書表示部 1 6 8 とを含んで構成される。上記結果表示部 1 6 6 は、GUI を介して検索者からの検索文書の参照の指示を受けて、文書表示部 1 6 8 を呼び出す。呼び出された文書表示部 1 6 8 は、指定された文書へアクセスするとともに、当該検索文書が参照された旨の報告を Web サーバ 1 2 0 へ送信する。報告を受けて検索履歴記録部 1 2 2 は、当該検索条件に対し、さらに参照済みの検索文書（以下、参照文書として参照する。）をさらに関連付けて記録する。

20

【 0 0 3 2 】

つまり、参照された検索文書は、単に検索にヒットしただけではなく、検索者により有為な文書であるとして参照された関連性の高い文書（以下、関連文書として参照する。）であるとして履歴中に記録されることとなる。本実施形態では、参照された参照文書を関連文書として登録するものとして説明するが、他の実施形態では、適合性フィードバックにより、検索者から選択された文書に含まれる単語を用いて拡張した検索条件による二次検索結果に含まれるものを、関連性の高い関連文書として記録することもできる。その場合は、検索結果表示部 1 6 6 は、検索文書のリストとともに適合文書として登録するためのボタンなどの GUI を表示させ、検索者から GUI を介して登録の指示があった文書が検索履歴記録部 1 2 2 に報告され、指示された適合文書を用いて拡張された二次検索結果に含まれる文書が関連文書として記録されることとなる。

30

【 0 0 3 3 】

ここで、図 5 を参照して検索履歴の詳細について説明する。図 5 は、検索履歴のデータ構造を示す。図 5 (A) に示す検索履歴テーブル 2 0 0 は、第 1 の実施形態の文書検索システム 1 0 0 において用いられ、検索条件を識別する検索条件 ID が入力されるフィールド 2 0 0 a と、検索条件の内容が入力されるフィールド 2 0 0 b とを含んで構成され、検索者から要求された検索条件に関する情報が時系列的にエントリされてゆく。また、検索履歴テーブル 2 0 0 は、検索結果としてヒットした検索文書の文書 ID のセットが入力されるフィールド 2 0 0 c と、検索結果から参照された関連文書の文書 ID のセットが入力されるフィールド 2 0 0 d とを含んで構成され、検索条件と、関連文書の集合と、検索文書の集合とをさらに関連付けている。

40

【 0 0 3 4 】

再び図 3 を参照すると、本実施形態の Web サーバ 1 2 0 は、さらに、蓄積した上記検索履歴を利用して検索条件に関連性の高い関連検索条件を抽出するために、グラフ・テーブル編成部 1 2 4 と、グラフ構築部 1 2 6 と、グラフ管理部 1 2 8 とを含んで構成され、後述するランダム・ウォーク・ウィズ・リスタート (Random-walk with Restart ; 以下、RWR として参照する。) 方式により要求された新規検索条件と過去の検索条件との間の

50

関連度の算出に用いるための状態遷移グラフを構成し、管理している。本実施形態のグラフ・テーブル編成部 124、グラフ構築部 126 およびグラフ管理部 128 は、統合的に状態遷移グラフを構成する構成手段として機能する。本実施形態の DB サーバ 110 は、グラフ・テーブル格納部 112 をさらに含んで構成され、グラフ・テーブル格納部 112 は、状態遷移グラフを構成するためのテーブル・セットを格納している。

【0035】

グラフ・テーブル編成部 124 は、時系列データとして構成される検索履歴を検索履歴格納部 114 から読み出して、状態遷移グラフを構成するためのテーブル・セットを作成して、グラフ・テーブル格納部 112 へ格納する。このテーブル・セットは、検索条件毎に履歴情報が整理されて構成される検索条件ノード・テーブルと、関連文書毎に整理されて構成される関連文書ノード・テーブルとを含んで構成される。状態遷移グラフの各ノードは、検索条件 ID および文書 ID により識別される。

10

【0036】

グラフ・テーブル編成部 124 は、同一検索者から短期間に頻発して要求された検索要求などの異常な履歴エントリや、内容が同一の重複する履歴エントリなど、不要なエントリを適宜除去して、状態遷移グラフに追加すべき履歴エントリを取捨選択しつつ、上記テーブル・セットを作成する。Web サーバ 120 は、更新要求部 134 を含んで構成され、更新要求部 134 は、予め設定された間隔やスケジュールなどに従って、グラフ・テーブル編成部 124 に対して上記テーブル・セットの再構成を指令し、検索履歴の経時的な状態変化をテーブル・セットに反映させている。

20

【0037】

Web サーバ 120 は、さらにスタートアップ処理を行うサーバ起動部 132 を含んで構成され、サーバ起動部 132 は、起動時にグラフ構築部 126 を呼び出す。グラフ構築部 126 は、グラフ・テーブル格納部 112 からテーブル・セットを読み出して、状態遷移グラフを構築する。この状態遷移グラフは、検索条件のノードと関連文書のノードとを含んで構成され、さらに、接続されるノード間の関係性を表す重み値を含んで構成される。状態遷移グラフは、より具体的には、各ノード間の重み値からなる隣接行列によって表現される。構築された状態遷移グラフは、グラフ管理部 128 へと渡され、グラフ管理部 128 は、渡された状態遷移グラフをメモリ 16 が提供するグラフ記憶部 130 に記憶させる。

30

【0038】

ここで、図 6 を参照して、状態遷移グラフの構築処理の詳細について説明する。図 6 は、第 1 の実施形態による Web サーバ 120 が実行する状態遷移グラフ構築処理を示すフローチャートである。図 6 に示す処理は、例えば、Web サーバ 120 の起動時にサーバ起動部 132 から呼び出されて、ステップ S100 から開始される。ステップ S101 では、グラフ構築部 126 は、グラフ・テーブル格納部 112 への接続を確立する。ステップ S102 では、グラフ構築部 126 は、検索条件ノード・テーブルおよび関連文書ノード・テーブルを読み出して、検索条件ノード q_i のセットおよび関連文書ノード v_{d_i} のセットからなる状態遷移グラフを構成する。ステップ S102 では、より具体的には、検索条件ノード q_i および関連文書ノード v_{d_i} からなるノード・セットを行および列とした隣接行列が準備される。

40

【0039】

ステップ S103 では、グラフ構築部 126 は、各検索条件ノード間のリンク ($q_i \rightarrow q_j$) の重み値を計算する。このリンク ($q_i \rightarrow q_j$) の重みは、検索条件間の関係性を表し、例えば、対応する検索結果に含まれる検索文書の重複度、対応する関連文書の重複度、および検索条件の内容自体に含まれる単語の類似度、またはこれらの少なくとも 1 つを用いて算出することができる。上記検索文書の重複度 $ObjFeedSim(q_i, q_j)$ は、例えば、検索条件ノード q_i に対応する上位 K の検索文書の集合 $topK(q_i)$ と、検索条件ノード q_j に対応する上位 K の検索文書の集合 $topK(q_j)$ とを用いて、下記式、

50

【 0 0 4 0 】

【 数 1 】

$$ObjFeedSim(q_i, q_j) = \frac{|topK(q_i) \cap topK(q_j)|}{|topK(q_i) \cup topK(q_j)|} \quad \dots\dots(1)$$

により算出することができる。上記関連文書の重複度 $SubFeedSim(q_i, q_j)$ も同様に、検索条件ノード q_i に対応する関連文書の集合 $visDocs(q_i)$ と、検索条件ノード q_j に対応する関連文書の集合 $visDocs(q_j)$ とを用いて、下記式、

10

【 0 0 4 1 】

【 数 2 】

$$SubFeedSim(q_i, q_j) = \frac{|visDocs(q_i) \cap visDocs(q_j)|}{|visDocs(q_i) \cup visDocs(q_j)|} \quad \dots\dots(2)$$

により算出することができる。上記単語の類似度 $ContSim(q_i, q_j)$ は、検索条件ノード q_i の内容に含まれる単語の重み W_s と、検索条件ノード q_j の内容に含まれる単語の重み W_t と、検索条件ノード q_i および検索条件ノード q_j の両方に含まれる単語の重み W_u とを用いて、下記式、

20

【 0 0 4 2 】

【 数 3 】

$$ContSim(q_i, q_j) = \frac{\sum_{w_u \in q_i \cap q_j} w_u}{\text{Max} \left(\sum_{w_s \in q_i} w_s, \sum_{w_t \in q_j} w_t \right)} \quad \dots\dots(3)$$

により算出することができる。ここで、単語の重み W_u, W_s, W_t は、 tf/idf (Term Frequency inverse Document Frequency) 法により求められた単語の重要度を表している。上記検索文書の重複度 $ObjFeedSim(q_i, q_j)$ 、関連文書の重複度 $SubFeedSim(q_i, q_j)$ 、単語の類似度 $ContSim(q_i, q_j)$ を用いるリンクの重み値の具体的な算出方法は、特に限定されるものではないが、これらの値に係数をかけて算出した値を用いることができる。また、同一検索条件ノード間のリンク ($q_i \rightarrow q_j$; $i = j$) の重み値は、「0」に設定することができる。

30

【 0 0 4 3 】

ステップ S 1 0 4 では、グラフ構築部 1 2 6 は、検索条件ノードと関連文書ノードとの間のリンク ($q_i \rightarrow v d_j$) の重み値を計算する。このリンク ($q_i \rightarrow v d_j$) の重み値は、各検索条件と関連文書間の関係性を表し、例えば、各検索条件の検索結果から参照された関連文書と該検索条件との間の値を「1」に設定し、参照関係に無い関連文書と検索条件との間の値を「0」に設定することができる。他の実施形態では、参照関係のある関連文書および検索条件において、検索結果における関連文書のランキングスコアを適宜規格化した値を当該重み値として採用してもよい。

40

【 0 0 4 4 】

ステップ S 1 0 5 では、上記検索条件ノード間のリンク ($q_i \rightarrow q_j$) の重み値、および検索条件ノードおよび関連文書ノード間のリンク ($q_i \rightarrow v d_j$) の重み値を用い、リンク ($v d_i \rightarrow q_j$) の重み値がリンク ($q_i \rightarrow v d_j$) と同一であるものとして、また関連文書ノード間のリンク ($v d_i \rightarrow v d_j$) の重み値が「0」であるものとして、準備された隣接行列の成分にこれら重み値を入力して、状態遷移グラフを表現する隣接行列を生成し、グラフ記憶部 1 3 0 へ格納し、ステップ S 1 0 6 で処理を終了させる。

50

【 0 0 4 5 】

上述したように、本実施形態のWebサーバ120は、新たに要求された検索条件に関連する関連検索条件を抽出するために、要求された検索条件と、該検索条件による検索結果に含まれる検索文書と、参照あるいは指定された関連文書とを対応付けて検索履歴として記録し、この検索条件に対応するノード、および関連文書に対応するノードからなる状態遷移グラフを予め構成しておく。新たな検索要求があった場合には、Webサーバ120は、以下説明するように、この状態遷移グラフを用いて関連検索条件の抽出処理を実行する。

【 0 0 4 6 】

再び図3を参照すると、本実施形態のWebサーバ120は、さらに、規格化処理部138と、RWR計算部140と、関連条件抽出部142とを含んで構成され、RWR方式を適用して、蓄積した上記検索履歴を利用した関連検索条件の抽出処理を実行する。上記グラフ管理部128は、クライアント160から検索要求を受信すると、関連検索条件を抽出するために、この新たな検索条件のノードを状態遷移グラフに追加して新規検索要求に適合させて更新し、この更新済みの状態遷移グラフの隣接行列を規格化処理部138へと渡す。

【 0 0 4 7 】

図7は、第1の実施形態による隣接行列のデータ構造を示す図である。図7に示す隣接行列は、検索条件ノードおよび関連文書ノードを、行および列に配したデータ構造となっており、リンク($q_i \rightarrow q_j$)の重み値、リンク($q_i \rightarrow v_{d_j}$)の重み値、リンク($v_{d_i} \rightarrow q_j$)の重み値、リンク($v_{d_i} \rightarrow v_{d_j}$)の重み値(=0)を成分として構成されている。また、最後の行および列には、新たに要求された新規検索条件に対応する行ベクトルおよび列ベクトルがそれぞれ追加されている。図7に示す実施形態では、新規検索条件ノードと、過去の検索条件ノードとのリンク($q_{new} \rightarrow q_j$; $q_i \rightarrow q_{new}$)の重み値が「0」であり、一方、新規検索条件ノードと履歴に含まれる関連文書ノードとのリンク($q_{new} \rightarrow v_{d_j}$; $v_{d_i} \rightarrow q_{new}$)には、ある重み値が与えられている。リンク($q_{new} \rightarrow v_{d_j}$; $v_{d_i} \rightarrow q_{new}$)の重み値は、例えば、検索履歴に含まれる関連文書の中から新規検索条件にて全文検索した場合に得られるランキングスコアを利用することができ、また所定の閾値以下のスコアを有する関連文書との重み値を「0」に設定することもできる。

【 0 0 4 8 】

再び図3を参照すると、規格化処理部138は、隣接行列の各行の総和を1となるよう規格化し、規格化済みの隣接行列をRWR計算部140へ渡す。RWR計算部140は、規格化済みの隣接行列にRWR方式を適用して、新たに要求された検索条件と、過去の検索条件との間の関連度を計算する。RWR方式では、隣接行列Aに対し、リスタート・ベクトルVを繰り返し適用して、得られる定常状態のベクトル U_{ss} を求める。このベクトル U_{ss} には、新規検索条件ノード q_{new} と、過去の各検索条件ノード q_i との関連度を成分として含まれることとなる。

【 0 0 4 9 】

以下、図8および図9を参照して、RWR方式による関連度計算処理の詳細について説明する。図8は、第1の実施形態による隣接行列A、リスタート・ベクトルV、および定常状態のベクトル U_{ss} を模式的に例示している。リスタート・ベクトルVは、新規検索条件に対応する成分が1である他、他の成分が0であるベクトルである。また隣接行列Aの行および列には、検索条件ノード q_i に対応する検索条件 Q_i 、および関連文書ノード v_{d_i} に対応する関連文書 V_i が示されている。図9は、第1の実施形態のWebサーバ120が実行するRWR方式による関連度計算処理を示すフローチャートである。図9に示す処理は、規格化処理部138から規格化済みの隣接行列Aを渡されて、ステップS200から開始される。

【 0 0 5 0 】

ステップS201では、RWR計算部140は、まず、初期値のベクトル U_0 としてリ

スタート・ベクトル V を与える。ステップ $S202$ では、 RWR 計算部 140 は、所与の定数 c 、隣接行列 A 、リスタート・ベクトル V 、次のベクトル U_{i+1} 、および現在のベクトル U_i を用いて表される下記式、

【0051】

【数4】

$$U_{i+1} = c \times A \times U_i + (1 - c) \times V \quad \dots\dots(4)$$

に従って、現在の現在のベクトル U_i から次のベクトル U_{i+1} を算出する。ステップ $S203$ では、ベクトル U_i とベクトル U_{i+1} との差のノルム ($= |U_{i+1} - U_i|$)、より具体的には距離を算出し、ノルムが所定の閾値以下となり、収束したか否かを判定する。

【0052】

ステップ $S203$ で、収束していないと判定された場合 (NO) には、ステップ $S204$ へ処理を進める。ステップ $S204$ では、イタレーションの回数 i が、所定の打ち切り限度を越えたか否かを判定する。ステップ $S204$ で、未だ打ち切り限度を越えていないと判定された場合 (NO) には、ステップ $S205$ へ処理を進めて、 i を加算し、ステップ $S202$ へと再びループさせる。

【0053】

一方、ステップ $S203$ で、収束したと判定された場合 (YES) には、ステップ $S206$ へ処理を進めて、処理を終了させる。同様に、ステップ $S204$ で、打ち切り限度を超えたと判定された場合 (YES) には、ステップ $S206$ へ処理を進めて、処理を終了させる。ステップ $S204$ の判定によって、所定回数以上のイタレーションが行われても十分に収束しない場合にでも処理を終了させることができる。そして、ステップ $S203$ または $S204$ からループを出たとき、最後に算出されたベクトル U_{i+1} が、定常状態のベクトル U_{ss} として与えられる。図8に示すように、ベクトル U_{ss} は、新規検索条件 Q_{new} に対する、検索履歴中の過去の検索条件 Q_i および過去の関連文書 V_i の関連度を成分として含んでいる。

【0054】

再び図3を参照すると、 RWR 計算部 140 は、得られた定常状態のベクトル U_{ss} を関連条件抽出部 142 へ渡す。関連条件抽出部 142 は、得られたベクトル U_{ss} の成分である関連度に従って、検索条件 Q_i をソーティングし、関連度が上位の関連検索条件のリスト (以下、関連検索条件リストとして参照する。) を、その内容および関連度とともに照会処理部 144 を経由して検索結果応答部 146 へ渡す。

【0055】

関連検索条件リストを得て検索結果応答部 146 は、取得した文書リストを含む照会結果に従って当該検索要求に対する検索結果の応答データを生成する際に、関連条件抽出部 142 から取得した関連検索条件の表示および該関連検索条件による検索要求を指令するハイパーリンク情報を検索結果の応答データに含める。そして、検索要求に対する検索結果と関連検索条件の抽出結果との両方を含む応答データがクライアント 160 へ返信されることとなる。なお、本実施形態では、関連検索条件による検索要求を指令するハイパーリンク情報を応答データに含める構成として説明したが、他の実施形態では、関連検索条件による照会を実際実施し、この検索結果をも応答データに含める構成を採用することもできる。

【0056】

図10は、第1の実施形態においてWebサーバ 120 より返信された応答データに従ってクライアント 160 のディスプレイ上に表示されるGUI画面を示す。図10に示すGUI画面 310 は、図4(A)に示したGUI画面 300 で入力された検索条件を示すテキスト 312 と、検索履歴から抽出された過去の関連検索条件を示すテキスト 314 と

10

20

30

40

50

、上記検索条件の検索結果に含まれる上位検索文書をサマリとともに一覧表示するテキスト 316 と、GUI 画面 300 に戻るためのボタン 318 とを含んで構成される。テキスト 314 の各関連検索条件には、ハイパーリンクが付されており、該ハイパーリンクをクリックすることにより、関連検索条件を条件とした検索要求が容易に行えるように構成されている。図 10 に示す例では、GUI 画面 310 には、オリジナルの検索条件として「電子図書館」が表示され、「電子図書館 マイライブラリ」「専門分野 文献検索」など、過去の関連する複数の検索条件が提示されている。なお、上記クリックに替えてリターンキーの押下など、別の動作により関連検索条件を条件とした検索要求が行えるよう構成することができることは、言うまでもない。

【0057】

10

上述したように、上記第 1 の実施形態の文書検索システム 100 における関連検索条件の抽出処理により、検索条件に含まれる単語自体の類似性のみならず、各検索条件と関連文書との対応付けから辿られる検索条件間の間接的な関係性が活用されて関連検索条件が抽出されることとなる。したがって、共通の単語を含まないような検索条件であっても、間接的な関係性が高いものであれば関連検索条件として抽出される可能性があり、過去の検索履歴をより有効に活用できるものと言える。また、第 1 の実施形態では、一度検索条件を入力することにより、当該検索条件による検索結果と当該検索条件に関係性の高い関連検索条件の抽出結果との両方が得られるため、検索履歴からの検索条件を抽出するための条件の別途入力が必要とされ、高いユーザ利便性が提供される。

【0058】

20

また、検索結果から参照された文書を関連文書として履歴に記録するよう構成可能であるため、検索者が検索文書について登録の要否を判断し明示的に指定する手間を省くことが可能であり、ユーザ利便性高く関連度の精度を高めることができる。また、適合性フィードバックにより対話的に指定された関連文書を記録する他の実施形態でも、検索者に特別の心理的負担をかけることなく、容易に関連文書の登録を行うことが可能とされる。

【0059】

しかしながら、第 1 の実施形態の隣接行列では、新規検索条件ノードと履歴中の関連文書ノードとのリンク ($q_{new} \rightarrow v_{dj}$; $v_{di} \rightarrow q_{new}$) には、ある重み値が与えられている一方で、新規検索条件ノードと、過去の検索条件ノードとのリンク ($q_{new} \rightarrow q_j$; $q_i \rightarrow q_{new}$) の重み値は、「0」に設定されていた。そのため、新規検索条件と過去の検索条件とが、例え共通の単語を多く含んでいた場合でも、RWR 法による計算結果として得られる両者の関連度が低く計算され、この過去の検索条件が表示されない、または下位になってしまう蓋然性がある。以下、RWR 法の最終的な計算結果において、入力した検索条件と共通の単語をより多く含む過去の検索条件との間の関連度が高くなるよう構成される、第 2 の実施形態の文書検索システム 100 について説明する。

30

【0060】

第 2 の実施形態による文書検索システムは、隣接行列のデータ構造以外の構成は、第 1 の実施形態の文書検索システムと同一であるため、説明の便宜上、隣接行列以外の構成についての説明は割愛する。図 11 は、第 2 の実施形態における隣接行列のデータ構造を示す図である。図 12 は、第 2 の実施形態による隣接行列 A、リスタート・ベクトル V、および定常状態のベクトル Uss を模式的に例示する図である。図 11 および図 12 に示す隣接行列は、第 1 の実施形態と同様に、検索条件ノードおよび関連文書ノードを行および列に配したデータ構造となっており、最後の行および列には、新規検索条件に対応する行ベクトルおよび列ベクトルがそれぞれ追加されている。第 2 の実施形態のデータ構造では、新規検索条件ノードと履歴中の関連文書ノードとのリンク ($q_{new} \rightarrow v_{dj}$; $v_{di} \rightarrow q_{new}$) に加えて、新規検索条件ノードと、過去の検索条件ノードとのリンク ($q_{new} \rightarrow q_j$; $q_i \rightarrow q_{new}$) の重み値が与えられている。リンク ($q_{new} \rightarrow q_j$; $q_i \rightarrow q_{new}$) の重み値は、過去の検索条件間のリンク ($q_i \rightarrow q_j$) と同様に、例えば、上記式 (3) の q_i を q_{new} に読み替えて表現されるような検索文書の含有する単語の類似度 $ContSim(q_{new}, q_j)$ を用いて算出することができる。

40

50

【 0 0 6 1 】

図 1 1 に示す第 2 の実施形態の隣接行列のデータ構造を用いた関連条件の抽出処理により、検索者の検索目的に応じて自然なランク付けの関連検索条件の抽出結果を提示することが可能となる。例えば、新規検索条件として「電子図書館」が入力された場合、関連する過去の検索条件が、含有する単語の類似度の高い順、図 1 0 に示す例では、「電子図書館 マイライブラリ」、「図書館評価」、「電子ジャーナル」、「専門分野 文献検索」、「図書館ポータル 利用状況」といった順に表示されることとなり、「電子図書館」をそのまま含む「電子図書館 マイライブラリ」が上位に表示されることとなる。つまり、利用者の最大の関心対象である「電子図書館」をそのまま含む検索条件が上位にランクされるようになる。

10

【 0 0 6 2 】

以下、検索者に関する情報を含む検索履歴を活用して関連検索条件を抽出する第 3 の実施形態による文書検索システムについて説明する。第 3 の実施形態の文書検索システムの概略構成および各装置のハードウェア構成は、第 1 の実施形態の文書検索システムと同一であるため、説明は割愛する。

【 0 0 6 3 】

図 1 3 は、第 3 の実施形態による文書検索システム 1 0 0 の機能ブロックを示す。図 1 3 に示す文書検索システム 1 0 0 は、クライアント 1 6 0 のログイン手続部 1 6 2 と、W e b サーバ 1 2 0 の新規ユーザ判定部 1 5 0、検索条件ノード付加部 1 5 2 およびユーザ・ノード付加部 1 5 4 からなる機能構成 1 4 8 を、図 3 と相違する部分として含んで構成されている。なお、図 1 3 に示す機能部のうち、図 3 に示したものと同様の機能を有するものについては、同一符号を付し、詳細な説明は省略する。

20

【 0 0 6 4 】

ログイン手続部 1 6 2 は、ユーザ名およびパスワードの入力を求めてシステムに対しログイン手続処理を実行する。クライアント 1 6 0 と W e b サーバ 1 2 0 との間のメッセージ交換は、クッキーなどによってセッション管理されており、第 3 の実施形態の W e b サーバ 1 2 0 は、検索者であるユーザを識別可能とされている。なお、第 3 の実施形態では、パスワードを伴わないゲストユーザのログインを許可するよう構成することもできる。検索条件が入力され、G U I を介した検索指示を受けると、検索条件入力部 1 6 4 は、入力された検索条件と、ユーザを識別する情報とを含む検索要求を W e b サーバ 1 2 0 へ送信する。

30

【 0 0 6 5 】

ここで、図 4 を参照して、第 3 の実施形態において検索要求のために用いられる G U I 画面について説明する。図 4 (B) に示す G U I 画面 3 3 0 は、第 3 の実施形態の文書検索システムにおける検索要求を発行するための G U I を提供し、ログイン手続後に表示され、検索条件を入力するためのテキスト・ボックス 3 3 2 と、検索要求を指示するための検索ボタン 3 3 4 と、ログアウトするためのログアウト・ボタン 3 3 6 とを含んで構成される。

【 0 0 6 6 】

再び図 1 3 を参照すると、第 3 の実施形態の検索履歴記録部 1 2 2 は、検索条件入力部 1 6 4 から送信された検索要求に含まれる検索条件と、ユーザ識別値と、該検索条件による検索結果に含まれる検索文書とを相互に関連付けて検索履歴格納部 1 1 4 へ記録する。ここで図 5 を参照して、第 3 の実施形態の検索履歴の詳細について説明する。図 5 (B) に示す検索履歴テーブル 2 1 0 は、第 1 の実施形態と同様に、検索条件 I D が入力されるフィールド 2 1 0 a と、検索条件の内容が入力されるフィールド 2 1 0 c と、検索文書の文書 I D のセットが入力されるフィールド 2 1 0 d と、関連文書の文書 I D のセットが入力されるフィールド 2 1 0 e とを含む。第 3 の実施形態の検索履歴テーブル 2 1 0 は、さらに、該検索条件の検索者を識別するユーザ I D が入力されるフィールド 2 1 0 b を含んで構成され、検索条件と、ユーザと、関連文書の集合と、検索文書の集合とを関連付けている。

40

50

【 0 0 6 7 】

再び図 1 3 を参照すると、第 3 の実施形態のグラフ・テーブル編成部 1 2 4 は、時系列データとして構成される検索履歴テーブル 2 1 0 を検索履歴格納部 1 1 4 から読み出して、検索条件ノード・テーブルおよび関連文書ノード・テーブルに加え、検索者ユーザ毎に整理されて構成されるユーザ・ノード・テーブルを作成し、グラフ・テーブル格納部 1 1 2 へ格納する。第 3 の実施形態では、状態遷移グラフの各ノードは、検索条件 ID、文書 ID、およびユーザ ID により識別されることとなる。そして、第 3 の実施形態のグラフ・テーブル編成部 1 2 4 は、同一ユーザから短期間に頻発して要求された検索要求などの異常な履歴エントリや、同一ユーザからの内容が同一の重複する履歴エントリなど、不要なエントリを適宜除去して、状態遷移グラフに追加すべき履歴エントリを取捨選択しつつ、上記テーブル・セットを作成する。

10

【 0 0 6 8 】

以下、第 3 の実施形態の文書検索システム 1 0 0 におけるグラフ構築部 1 2 6 の機能について説明する。グラフ構築部 1 2 6 は、第 3 の実施形態では、検索条件のノード、関連文書のノードに加えて、ユーザのノードを含む状態遷移グラフを構築する。以下、図 1 4 を参照して、第 3 の実施形態における状態遷移グラフの構築処理の詳細について説明する。

【 0 0 6 9 】

図 1 4 は、第 3 の実施形態による Web サーバ 1 2 0 が実行する状態遷移グラフ構築処理を示すフローチャートである。図 1 4 に示す処理は、例えば起動時にサーバ起動部 1 3 2 から呼び出されて、ステップ S 3 0 0 から開始される。ステップ S 3 0 1 では、グラフ構築部 1 2 6 は、グラフ・テーブル格納部 1 1 2 への接続を確立する。ステップ S 3 0 2 では、グラフ構築部 1 2 6 は、検索条件ノード・テーブル、ユーザ・ノード・テーブルおよび関連文書ノード・テーブルを読み出して、検索条件ノード q_i のセット、ユーザ・ノード u_i のセット、および関連文書ノード $v d_i$ のセットからなる状態遷移グラフを構成する。ステップ S 3 0 2 では、より具体的には、検索条件ノード q_i 、ユーザ・ノード u_i 、および関連文書ノード $v d_i$ からなるノード・セットを行および列に配した隣接行列が準備される。

20

【 0 0 7 0 】

ステップ S 3 0 3 では、グラフ構築部 1 2 6 は、各ユーザ・ノード間のリンク ($u_i \rightarrow u_j$) の重み値を計算する。このリンク ($u_i \rightarrow u_j$) の重み値は、検索者ユーザ間の関係性を表し、検索者により入力された過去の検索条件が検索者の興味を好適に反映しているものとして、上記式 (3) と同様に算出される各検索者ユーザの過去の検索条件の内容が含有する単語の類似度を用いて算出することができる。その他、システムがユーザ・プロフィールとして保持する検索者ユーザの所属、ブックマーク情報、RSS フィード情報などを用いて、上記重み値を算出することができる。例えば、検索者が同じ研究室に属していた場合に、対応するユーザ・ノード間の重みを「0.9」とし、研究室が相違していても学科が同一であれば「0.6」とし、学科が相違していても学部が同一であれば 0.3 とし、学部も相違すれば 0 とする、というように設定することができる。

30

【 0 0 7 1 】

ステップ S 3 0 4 では、グラフ構築部 1 2 6 は、第 1 の実施形態と同様に、対応する検索結果に含まれる検索文書の重複度 $ObjFeedSim(q_i, q_j)$ 、対応する関連文書の重複度 $SubFeedSim(q_i, q_j)$ 、および検索条件の内容に含まれる単語の類似度 $ContSim(q_i, q_j)$ 、またはこれらの少なくとも 1 つを用いて、各検索条件ノード間のリンク ($q_i \rightarrow q_j$) の重み値を計算する。

40

【 0 0 7 2 】

ステップ S 3 0 5 では、グラフ構築部 1 2 6 は、検索条件ノードとユーザ・ノードとの間のリンク ($q_i \rightarrow u_j$) の重み値を計算する。このリンク ($q_i \rightarrow u_j$) の重み値は、各検索条件と検索者ユーザとの間の関係性を表し、例えば、検索条件と該検索条件による検索要求を行った検索者ユーザとの間の値を「1」に設定し、それ以外の場合に「0」に

50

設定することができる。

【0073】

ステップS306では、グラフ構築部126は、ユーザ・ノードと関連文書ノードとの間のリンク($u_i \rightarrow v_{dj}$)の重み値を計算する。このリンク($u_i \rightarrow v_{dj}$)の重み値は、各検索者ユーザと関連文書との間の関係性を表し、例えば、各検索者により検索結果から参照された関連文書と該検索者ユーザとの間の値を「1」に設定し、参照関係に無い関連文書と検索者ユーザとの間の値を「0」に設定することができる。

【0074】

ステップS307では、グラフ構築部126は、第1の実施形態と同様に、関連文書と検索条件との参照関係の有無から、検索条件ノードと関連文書ノードとの間のリンク($q_i \rightarrow v_{dj}$)の重み値を計算する。その他、参照関係のある関連文書および検索条件において、検索結果における関連文書のランキングスコアを適宜規格化した値を当該重み値として採用してもよい。

【0075】

ステップS308では、上記ユーザ・ノード間のリンク($u_i \rightarrow u_j$)の重み値、上記検索条件ノード間のリンク($q_i \rightarrow q_j$)の重み値、検索条件ノードおよびユーザ・ノード間のリンク($u_i \rightarrow q_j$)の重み値、ユーザ・ノードおよび関連文書ノード間のリンク($u_i \rightarrow v_{dj}$)の重み値、および検索条件ノードおよび関連文書ノード間のリンク($q_i \rightarrow v_{dj}$)の重み値を用い、逆向きのリンクの重み値が同一であるものとして、また関連文書ノード間のリンク($v_{di} \rightarrow v_{dj}$)の重み値が「0」であるものとして、準備された隣接行列の成分にこれら重み値を入力して、状態遷移グラフを表現する隣接行列を生成し、ステップS309で処理を終了させる。

【0076】

再び図13を参照すると、第3の実施形態の文書検索システム100では、入力された検索条件と、検索者ユーザを識別する情報とを含む検索要求を検索条件入力部164から受信すると、上記新規ユーザ判定部150は、当該検索者ユーザが既に状態遷移グラフの中に組み込まれているか否かを判定する。

【0077】

検索者ユーザが既に状態遷移グラフ中に存在していた場合には、検索条件ノード付加部152のみが呼び出される。この場合、検索条件ノード付加部152は、第1の実施形態の文書検索システム100と同様に、上記グラフ管理部128に対し、この新たな検索条件のノードを状態遷移グラフに追加させて更新する。一方、検索者ユーザが状態遷移グラフ中に存在していない場合には、検索条件ノード付加部152とともにユーザ・ノード付加部154が呼び出される。この場合、検索条件ノード付加部152は、第1の実施形態の文書検索システム100と同様に、上記グラフ管理部128に対し、新たな検索条件のノードを追加させるとともに、ユーザ・ノード付加部154は、新たな検索者ユーザのノードを状態遷移グラフに追加させる。

【0078】

図15は、第3の実施形態による隣接行列のデータ構造を示す図である。図15に示す隣接行列は、検索条件ノード、ユーザ・ノードおよび関連文書ノードを行および列に配したデータ構造となっており、リンク($q_i \rightarrow q_j$)の重み値、リンク($q_i \rightarrow v_{dj} : v_{di} \rightarrow q_j$)の重み値に加えて、リンク($u_i \rightarrow u_j$)の重み値、リンク($q_i \rightarrow u_j : u_i \rightarrow q_j$)の重み値、リンク($u_i \rightarrow v_{dj} : v_{di} \rightarrow u_j$)の重み値を成分として構成されている。また、最後の行および列のそれぞれには、新規検索条件および新規ユーザに対応する行ベクトルおよび列ベクトルのセットが追加されている。

【0079】

図15に示すデータ構造では、新規検索条件ノードと履歴中の関連文書ノードとのリンク($q_{new} \rightarrow v_{dj} ; v_{di} \rightarrow q_{new}$)、新規検索条件ノードおよび過去の検索条件ノードのリンク($q_{new} \rightarrow q_j ; q_i \rightarrow q_{new}$)に加えて、新規ユーザ・ノードと既存のユーザ・ノードとのリンク($u_{new} \rightarrow u_j ; u_i \rightarrow u_{new}$)の重み値が与えられ

10

20

30

40

50

ている。リンク ($u_{new} \rightarrow u_j ; u_i \rightarrow u_{new}$) の重み値は、既存のユーザ・ノード間のリンク ($u_i \rightarrow u_j$) と同様に、上記式 (3) と同様に算出される新規ユーザの当該検索条件と、既存のユーザの過去の検索条件の内容自体に含まれる単語の類似度、検索者ユーザの所属、ブックマーク情報、RSSフィード情報などのユーザ・プロフィールを用いて算出することができる。また、適当な情報が無い場合には、リンク ($u_{new} \rightarrow u_j ; u_i \rightarrow u_{new}$) の重み値をすべて「0」に設定することもできる。

【0080】

図16は、第3の実施形態による隣接行列A、リスタート・ベクトルV、および定常状態のベクトルU_sを模式的に例示している。リスタート・ベクトルVは、第1の実施形態のデータ構造と同様に、新規検索条件に対応する成分が1である他、他の成分が0であるベクトルである。また隣接行列Aの行および列には、検索条件ノード q_i に対応する検索条件 Q_i 、ユーザ・ノード u_i に対応するユーザ U_i 、および関連文書ノード vd_i に対応する関連文書 V_i が示されている。第3の実施形態の文書検索システムでは、図16に示す隣接行列Aが規格化処理部138へ渡されて規格化され、さらにRWR計算部140へ渡される。RWR計算部140は、規格化済みの隣接行列AにRWR方式を適用して、図9に示すフローチャートと同様の処理によって、新規検索条件と、過去の検索条件との間の関連度を計算する。RWR方式を適用して得られたベクトルU_sは、図16に示すように、新規検索条件 Q_{new} に対する過去の検索条件 Q_i 、既存のユーザ U_i および過去の関連文書 V_i の関連度を成分として含んでいる。

【0081】

関連検索条件リストを得て検索結果応答部146は、上述したように、関連検索条件の表示および該関連検索条件による検索要求を指令するハイパーリンク情報を検索結果の応答データに含めて生成する。第3の実施形態では、検索結果応答部146は、その際さらに、本検索要求の検索者ユーザと関連付けられている条件をユーザ自身の過去の関連検索条件とし、本検索要求のユーザに関連付けられていない条件を、他の人の過去の関連検索条件として、区別して表示されるように上記応答データを構成する。そして、検索要求に対する検索結果と、自他で区分された関連検索条件の抽出結果の両方を含む応答データがクライアント160へ返信されることとなる。

【0082】

図17は、第3の実施形態においてWebサーバ120より返信された応答データに従ってクライアント160のディスプレイ上に表示されるGUI画面を示す。図17に示すGUI画面320は、図4に示したGUI画面300で入力された検索条件を示すテキスト322と、検索履歴から抽出された当該検索者の過去の関連検索条件を示すテキスト324と、他の検索者の過去の関連検索条件を示すテキスト326と、検索結果に含まれる上位検索文書をサマリとともに一覧表示するテキスト328と、GUI画面300に戻るためのボタン329とを含んで構成される。図17に示す例では、GUI画面320には、オリジナルの検索条件として「電子図書館」が表示され、検索者自身の過去の検索条件として「電子図書館 マイライブラリ」「専門分野 文献検索」が提示され、他の人の過去の関連検索条件として、「図書館評価」や「図書館ポータル 利用状況」「電子ジャーナル」が提示されている。なお、他の実施形態では、さらに、ユーザが属するグループ情報などに従って、グループの他メンバの過去の関連検索条件を、さらに区別して提示することもできる。なお、第3の実施形態の文書検索システム100では、ゲストユーザにてログインされていた場合には、図17に示すGUI画面320において、テキスト324が表示されず、他の人の過去の関連検索条件を示すテキスト326のみが表示されることとなる。

【0083】

上述したように、第3の実施形態の文書検索システムにおける関連検索条件の抽出処理により、検索要求を行った検索者に関する情報を活用して、新規検索条件と過去の検索条件との関連度を算出するため、その検索者の興味や関心を反映した関連検索条件を抽出することが可能となる。例えば、過去において異なる検索者が同一の検索条件を用いて検索

10

20

30

40

50

した場合であっても、それぞれ異なる文書を関連文書として参照または指定する可能性があるため、当該抽出処理によれば、検索要求した検索者の相違も応答して関連度が算出され、得られる関連検索条件も異なることとなる。したがって、その検索者の興味や関心を反映した関連検索条件を抽出することが可能となる。また、第3の実施形態では、当該検索者の過去の検索条件と、他の検索者の過去の検索条件とが識別可能に提示されるため、検索者は、自身の検索目的に合致した関連検索条件を容易に認識することができる。例えば、自身の過去の検索行動や思考を想起したい場合には、自身の過去の検索条件を見ればよく、自身の発想を広げたい場合には、他の人の関連検索条件を参照すればよい。

【0084】

上述までの実施形態では、検索条件ノードと関連文書ノードとの間のリンク ($q_i \rightarrow v d_j$) の重み値は、各検索条件の検索結果から参照された関連文書と該検索条件との間の値を「1」に設定し、参照関係に無い関連文書と検索条件との間の値を「0」に設定するものとして説明してきた。しかしながら、検索条件と関連文書との関連性としては、参照関係の有無のみならず、種々の信頼性の程度を有する文書を定義付けることができる。例えば、ユーザが単に検索結果から参照した参照文書と、適合性フィードバックにおいてユーザが適合するものと指定した適合文書、詳細な書誌情報を参照した後に適合するものとしてユーザが明示的に指定した指定文書とでは、検索者ユーザの興味や関心の程度が異なり、これらの関連情報と検索条件との間の関連性に対する信頼性は、相違するものと考えられる。

【0085】

以下、参照関係の有無のみならず、上記の参照文書、適合文書および指定文書の別に応じて求められるリンク ($q_i \rightarrow v d_j$) の重み値を用いる第4の実施形態の文書検索システム100について説明する。第4の実施形態の文書検索システムの概略構成および各装置のハードウェア構成は、第1の実施形態の文書検索システムと同一であるため、説明は割愛する。また、第4の実施形態の文書検索システムの機能ブロックは、第1の実施形態と同様であるため、図3を再び参照して、第4の実施形態による文書検索システム100の機能ブロックについて説明する。なお、図3に示す機能部のうち、第1および第2の実施形態と同様の機能を有するものについては、詳細な説明は省略し、相違点を中心として説明する。

【0086】

第4の実施形態の検索結果応答部146は、検索要求に応答して、その検索結果の応答データを生成して、クライアント160へ返信する。このとき、第4の実施形態の応答データには、例えばチェックボックスなど、検索文書の中から適合性フィードバックによる適合文書を指定するためのGUI部品、適合文書を用いた再検索を指示するためのGUI部品の表示指令が含まれている。クライアント160の結果表示部166は、受信した応答データに従って、適合文書を指定するためのGUI部品、および適合文書を用いた再検索を指令するためのGUI部品を含む検索結果を画面表示する。

【0087】

図18は、第4の実施形態においてWebサーバ120より返信された応答データに従ってクライアント160のディスプレイ上に表示されるGUI画面を示す。図18に示すGUI画面340は、例えば図4(A)に示したGUI画面300で入力された検索条件を示すテキスト342と、検索履歴から抽出された過去の関連検索条件を示すテキスト344と、上記検索条件の検索結果に含まれる上位検索文書を、適合文書として指定するためのチェックボックス352a~eおよびサマリとともに一覧表示する結果リスト346と、GUI画面300に戻るためのボタン348と、指定された適合文書を用いた再検索を指令するための再検索ボタン350とを含んで構成される。再検索ボタン350がクリックされると、結果表示部166は、適合文書の指定を含めてWebサーバ120へ適合性フィードバックによる再検索の要求を発行する。

【0088】

適合性フィードバックによる再検索の要求を受領したWebサーバ120では、照会処

理部 1 4 4 が、指定された適合文書に含まれるキーワードで検索条件を拡張してインデックス・データベース 1 1 6 に照会するとともに、指定された適合文書のリストを検索履歴記録部 1 2 2 へ報告する。検索履歴記録部 1 2 2 は、先の検索要求に含まれる検索条件と、報告された適合文書とを相互に関連付けて検索履歴格納部 1 1 4 へ記録する。一方、検索結果応答部 1 4 6 は、二次検索結果として拡張された検索条件に該当する検索文書の文書のリストを取得し、二次検索結果の文書リストを検索履歴記録部 1 2 2 へ報告する。検索履歴記録部 1 2 2 は、先の検索要求に含まれる検索条件と、報告された二次検索結果に含まれる文書とを相互に関連付けて検索履歴格納部 1 1 4 へ記録することができる。

【 0 0 8 9 】

図 1 8 に示す G U I 画面 3 4 0 の結果リスト 3 4 6 にリストアップされる各上位検索文書には、その詳細な書誌情報を表示させるためのハイパーリンクが付されている。結果表示部 1 6 6 は、G U I を介して検索者からの検索文書の詳細な書誌情報の参照の指示を受けた場合、文書表示部 1 6 8 を呼び出す。呼び出された文書表示部 1 6 8 は、指定された検索文書の詳細な書誌情報へアクセスするとともに、当該検索文書が参照された旨の報告を W e b サーバ 1 2 0 へ送信する。報告を受けて検索履歴記録部 1 2 2 は、検索条件に対し、さらに参照文書をさらに関連付けて記録する。

【 0 0 9 0 】

図 1 9 は、第 4 の実施形態において、クライアント 1 6 0 のディスプレイ上に表示される指定文書を指定するための G U I 画面を示す。図 1 9 に示す G U I 画面 3 6 0 は、詳細な書誌情報を示すテキスト 3 6 6、図 1 8 に示す G U I 画面 3 4 0 に戻るための戻るボタン 3 6 4 と、さらに、適合する文書であるとして明示的に指定するためのチェックボックス 3 6 2 とを含んで構成される。戻るボタン 3 6 4 がクリックされると、結果表示部 1 6 6 は、チェックボックス 3 6 2 に対するチェックの有無を参照文書に関連付けて、W e b サーバ 1 6 8 へ報告する。報告を受けて W e b サーバ 1 2 0 の検索履歴記録部 1 2 2 は、チェックが有りの場合には、当該参照文書を明示的に適合すると指定された指定文書として、当該検索条件に対しさらに関連付けて記録する。

【 0 0 9 1 】

つまり、第 4 の実施形態の W e b サーバ 1 2 0 は、新たに要求された検索条件に対して、ユーザが検索結果から参照した参照文書と、適合性フィードバックにおいてユーザが適合するものと指定した適合文書、詳細な書誌情報の表示から適合するものとしてユーザが明示的に指定した指定文書とを区別して関連付ける。その他、指示された適合文書を用いて拡張された二次検索結果に含まれる文書をさらに区別して関連付けても良い。

【 0 0 9 2 】

ここで、図 2 0 を参照して、第 4 の実施形態における検索履歴の詳細について説明する。図 2 0 は、第 4 の実施形態による検索履歴のデータ構造を示す。図 2 0 に示す検索履歴テーブル 2 2 0 は、第 1 の実施形態と同様に、検索条件を識別する検索条件 I D が入力されるフィールド 2 2 0 a と、検索条件の内容が入力されるフィールド 2 2 0 b とを含んで構成される。また、検索履歴テーブル 2 2 0 は、検索結果として上位にヒットした検索文書の文書 I D のセットが入力されるフィールド 2 2 0 c と、検索結果から参照された参照文書の文書 I D のセットが入力されるフィールド 2 2 0 d と、適合性フィードバックによる再検索のために指定された適合文書の文書 I D のセットが入力されるフィールド 2 2 0 e と、詳細な書誌情報の表示から明示的に指定された指定文書の文書 I D のセットが入力されるフィールド 2 2 0 f とを含んで構成され、検索条件と、検索文書の集合と、関連文書の集合とを、該関連情報の種別を区別してさらに関連付けている。

【 0 0 9 3 】

第 4 の実施形態のグラフ・テーブル編成部 1 2 4 は、時系列データとして構成される上記検索履歴を検索履歴格納部 1 1 4 から読み出して、状態遷移グラフを構成するためのテーブル・セットを作成して、グラフ・テーブル格納部 1 1 2 へ格納する。ここで、再び図 6 を参照して、第 4 の実施形態における状態遷移グラフの構築処理の詳細について説明する。なお、ステップ S 1 0 4 までの処理については、第 1 の実施形態と同様であるため、

説明を割愛する。

【0094】

第4の実施形態において、ステップS104では、グラフ構築部126は、検索条件ノードと関連文書ノードとの間のリンク($q_i \rightarrow v d_j$)の重み値を計算する。このリンク($q_i \rightarrow v d_j$)の重み値は、各検索条件と関連文書との間の関係性を表し、第4の実施形態においては、関連情報が、参照文書であるか、適合文書であるか、指定文書であるかの別に応じて求められる。

【0095】

参照文書は、上述したように、検索結果の一覧に含まれる文書のタイトルあるいはサマリのみから判断されて、利用者が興味を持って詳細情報などを参照した文書である。しかしながら、利用者に与えられる情報が限定されているため、単に参照したというだけでは、関連度の程度という観点からは信頼性が低いものとなる。一方、適合文書は、適合性フィードバックのためにチェックボックスなどにより指定された文書である。少なくとも利用者が適合するものとして判断した文書であることから、参照文書と比較して、関連度の程度という観点からは信頼性が高いものとなる。しかしながら、より適合性の高い文書を検索するために、手頃な文書を取りあえず選択された場合もあるため、利用者の興味および関心には完全には適合していない可能性がある。これに対して、指定文書は、詳細な書誌情報を表示させた上で、チェックボックスなどにより明示的に適合するものとして指定された文書である。この指定文書は、利用者が適合するものとして判断し、陽に指定した文書であることから、適合文書と比較しても、関連度の程度という観点から信頼性が高いものといえる。

10

20

【0096】

第4の実施形態において、ステップS104では、上述した参照文書、適合文書および指定文書の順に、重み値を大きく設定する。より具体的には、例えば、参照文書、適合文書、指定文書に対して、それぞれ「0.3」、「0.7」、「1.0」などの多値の重み値を与えるよう構成することができる。なお、参照文書、適合文書、指定文書に対して設定されるこれらの数値はあくまで一例であり、その他の値を設定することもできる。例えば、単に参照された文書は関連性がないものとして、参照文書にも「0」を設定してもよい。また、関連文書が複数の上記種別に属する場合には、値の大きい方の重み値を採用すればよい。ステップS105では、上記検索条件ノード間のリンク($q_i \rightarrow q_j$)の重み値、および検索条件ノードおよび関連文書ノード間のリンク($q_i \rightarrow v d_j$)の重み値を用い、リンク($v d_i \rightarrow q_j$)の重み値がリンク($q_i \rightarrow v d_j$)と同一であるものとして、また関連文書ノード間のリンク($v d_i \rightarrow v d_j$)の重み値が「0」であるものとして、準備された隣接行列の成分にこれら重み値を入力して、状態遷移グラフを表現する隣接行列を生成し、グラフ記憶部130へ格納し、ステップS106で処理を終了させる。

30

【0097】

上述したように、第4の実施形態のWebサーバ120は、新たに要求された検索条件に関連する関連検索条件を抽出するために、要求された検索条件と、該検索条件による検索文書と、参照文書と、適合文書と、指定文書とを対応付けて検索履歴として記録し、この検索条件に対応するノード、および関連文書に対応するノードからなる状態遷移グラフを予め構成しておく。新たな検索要求があった場合には、Webサーバ120は、第1の実施形態について説明したように、状態遷移グラフを用いて関連検索条件の抽出処理を実行する。

40

【0098】

再び図3を参照すると、第4の実施形態のWebサーバ120は、作成された状態遷移グラフを用いて、RWR方式を適用して、蓄積した上記検索履歴を利用した関連検索条件の抽出処理を実行する。上記グラフ管理部128は、クライアント160から検索要求を受信すると、関連検索条件を抽出するために、この新たな検索条件のノードを状態遷移グラフに追加して新規検索要求に適合させて更新し、この更新済みの状態遷移グラフの隣接行列を規格化処理部138へと渡す。

50

【 0 0 9 9 】

図 2 1 は、第 4 の実施形態における隣接行列のデータ構造を示す図である。図 2 2 は、第 4 の実施形態による隣接行列 A、リスタート・ベクトル V、および定常状態のベクトル U_{ss} を模式的に例示する図である。図 2 1 および図 2 2 に示す隣接行列は、第 1 の実施形態と同様に、検索条件ノードおよび関連文書ノードを行および列に配したデータ構造となっており、最後の行および列には、新規検索条件に対応する行ベクトルおよび列ベクトルがそれぞれ追加されている。第 4 の実施形態のデータ構造では、リンク ($v d_i \rightarrow q_j$) およびリンク ($q_i \rightarrow v d_j$) に、「0」および「1」以外の多値の重みが入力されている様子が示されている。つまり、第 4 の実施形態では、単に参照された参照文書よりも適合文書の関連性が強化され、適合文書よりも指定文書の関連性が強化されて、状態遷移グラフが構成され、関連検索条件の検索に用いられることとなる。

10

【 0 1 0 0 】

再び図 3 を参照すると、規格化処理部 1 3 8 は、隣接行列の各行の総和を 1 となるよう規格化し、規格化済みの隣接行列を RWR 計算部 1 4 0 へ渡し、RWR 計算部 1 4 0 は、規格化済みの隣接行列に RWR 方式を適用して、新たに要求された検索条件と、過去の検索条件との間の関連度を計算する。RWR 方式による関連度計算処理の詳細については、第 1 の実施形態と同様であるため、詳細な説明は割愛する。

【 0 1 0 1 】

図 2 1 に示す第 4 の実施形態の隣接行列のデータ構造を用いた関連条件の抽出処理により、参照文書であるか、適合文書であるか、または指定文書であるかといった、関連文書の種別による関連性の信頼度の相違を、関連文書および検索条件のノード間の重み付けの値の相違として取り込み、検索者ユーザの興味や関心を、より反映した過去の検索条件を容易に抽出可能とし、抽出の精度を高めることができる。

20

【 0 1 0 2 】

なお、関連文書の種別としては、上述した参照文書、適合文書、指定文書に特に限定されるものではない。他の実施形態では、適合文書から再検索される文書を種別を含むことができ、またユーザに対して文書に対する重要度を指定させ、該重要度に対応させて関連文書の種別を区別してもよい。また、同様に第 3 の実施形態におけるリンク ($q_i \rightarrow v d_j$: $v d_i \rightarrow q_j$) の重み値を求める際に、第 4 の実施形態の構成を採用することもできる。

30

【 0 1 0 3 】

以上説明したように、上述までの実施形態によれば、検索者自身や他の検索者の過去の検索経験が適切に反映された関連の検索条件を効率的に抽出し、検索者に手間をかけることなく、その興味、関心および検索目的に適合した過去の検索条件を容易に再利用可能とする情報処理装置、情報検索システム、情報処理方法およびプログラムを提供することが可能となる。なお、上述までの実施形態では、文書を検索対象とする文書検索システム 1 0 0 を用いて説明してきたが、検索対象は、文書に限られるものではなく、画像データ、マルチメディア・データ等の種々のコンテンツを対象とすることができる。その際には、これらの検索対象のデータは、好適には、テキスト情報をメタ情報として含むことができる。

40

【 0 1 0 4 】

また上記機能は、アセンブラ、C、C++、C#、Java（登録商標）、などのレガシープログラミング言語やオブジェクト指向プログラミング言語などで記述されたコンピュータ実行可能なプログラムにより実現でき、ROM、EEPROM、EPROM、フラッシュメモリ、フレキシブルディスク、CD-ROM、CD-RW、DVD、SDカード、MOなど装置可読な記録媒体に格納して頒布することができる。

【 0 1 0 5 】

これまで本発明の実施形態について説明してきたが、本発明の実施形態は上述した実施形態に限定されるものではなく、他の実施形態、追加、変更、削除など、当業者が想到することができる範囲内で変更することができ、いずれの態様においても本発明の作用・効

50

果を奏する限り、本発明の範囲に含まれるものである。

【図面の簡単な説明】

【0106】

【図1】第1の実施形態による文書検索システムを示す概略図。

【図2】第1の実施形態によるWebサーバ120のハードウェア構成を示す図。

【図3】第1の実施形態による文書検索システム100の機能ブロック図。

【図4】クライアントのディスプレイ上に表示されるGUI画面を示す図。

【図5】検索履歴のデータ構造を示す図。

【図6】第1の実施形態による状態遷移グラフ構築処理を示すフローチャート。

【図7】第1の実施形態による隣接行列のデータ構造を示す図。

10

【図8】第1の実施形態による隣接行列A、リスタート・ベクトルV、および定常状態のベクトルUsを模式的に例示する図。

【図9】第1の実施形態によるRWR方式による関連度計算処理を示すフローチャート。

【図10】第1の実施形態においてクライアントのディスプレイ上に表示されるGUI画面を示す。

【図11】第2の実施形態による隣接行列のデータ構造を示す図。

【図12】第2の実施形態による隣接行列A、リスタート・ベクトルV、および定常状態のベクトルUsを模式的に例示する図。

【図13】第3の実施形態による文書検索システム100の機能ブロック図。

【図14】第3の実施形態による状態遷移グラフ構築処理を示すフローチャート。

20

【図15】第3の実施形態による隣接行列のデータ構造を示す図。

【図16】第3の実施形態による隣接行列A、リスタート・ベクトルV、および定常状態のベクトルUsを模式的に例示する図。

【図17】第3の実施形態においてクライアントのディスプレイ上に表示されるGUI画面を示す図。

【図18】第4の実施形態においてクライアントのディスプレイ上に表示される検索結果のGUI画面を示す。

【図19】第4の実施形態においてクライアントのディスプレイ上に表示される指定文書を指定するためのGUI画面を示す。

【図20】第4の実施形態による検索履歴のデータ構造を示す図。

30

【図21】第4の実施形態による隣接行列のデータ構造を示す図。

【図22】第4の実施形態による隣接行列A、リスタート・ベクトルV、および定常状態のベクトルUsを模式的に例示する図。

【符号の説明】

【0107】

12...MPU、14...不揮発メモリ、16...メモリ、18...記憶制御用インタフェース、20...ハードディスク、22...内部バス、24...インタフェース、26...入出力装置、28...VRAM、30...グラフィック・チップ、32...ディスプレイ装置、34...ネットワークI/F(NIC; Network Interface Card)、34...NIC、100...文書検索システム、110...DBサーバ、110a...データベース、112...グラフ・テーブル格納部、114...検索履歴格納部、116...インデックス・データベース、120...Webサーバ、122...検索履歴記録部、124...グラフ・テーブル編成部、126...グラフ構築部、128...グラフ管理部、130...グラフ記憶部、132...サーバ起動部、134...更新要求部、138...規格化処理部、140...RWR計算部、142...関連条件抽出部、144...照会処理部、146...検索結果応答部、148...機能構成、150...新規ユーザ判定部、152...検索条件ノード付加部、154...ユーザ・ノード付加部、160...クライアント、162...ログイン手続部、164...検索条件入力部、166...結果表示部、168...文書表示部、170...ネットワーク、200, 210, 220...検索履歴テーブル、300, 330...GUI画面、302, 332...テキスト・ボックス、304, 334...検索ボタン、336...ログアウト・ボタン、310...GUI画面、312...テキスト、31

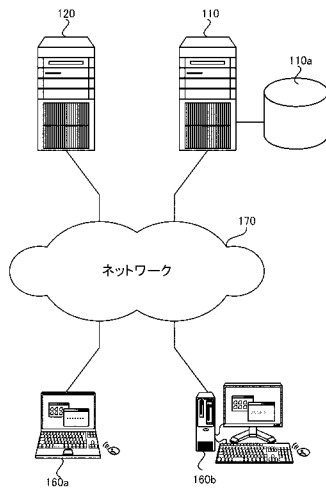
40

50

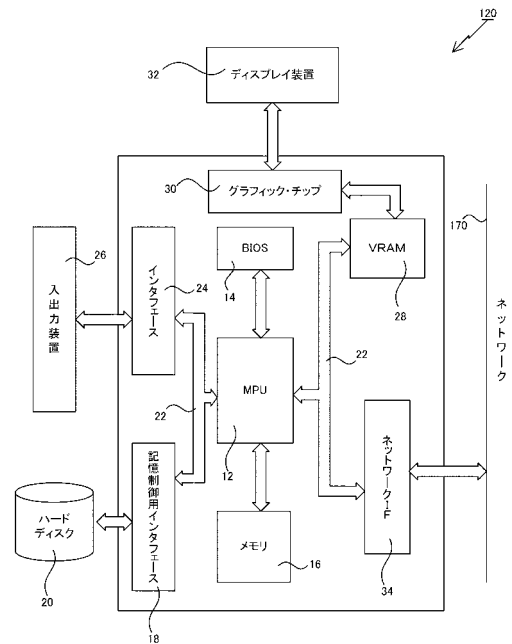
4 ... テキスト、316 ... テキスト、318 ... ボタン、320 ... G U I 画面、322 ... テキスト、324 ... テキスト、326 ... テキスト、328 ... テキスト、329 ... ボタン、340 ... G U I 画面、342 ... テキスト、344 ... テキスト、346 ... 結果リスト、348 ... ボタン、350 ... ボタン、352 ... チェックボックス、360 ... G U I 画面、362 ... チェックボックス、364 ... ボタン、366 ... テキスト

【 図 1 】

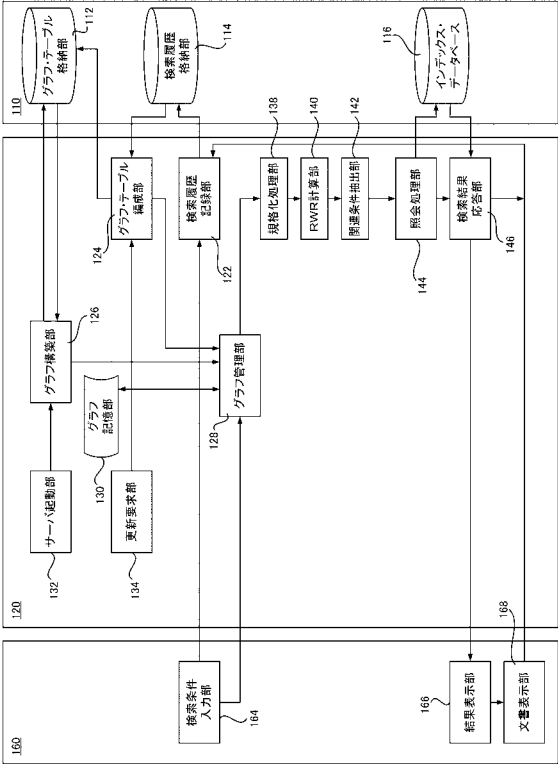
100



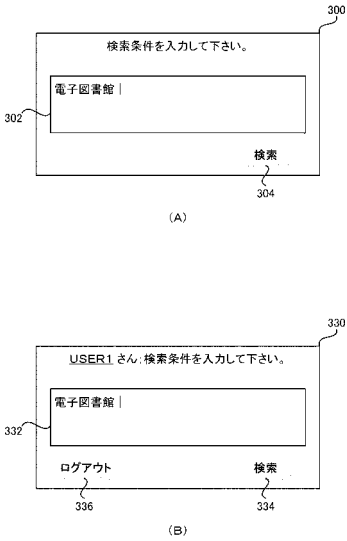
【 図 2 】



【図 3】



【図 4】



【図 5】

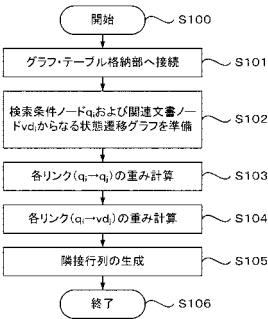
検索条件ID				検索条件	検索文書の文書IDセット	関連文書の文書IDセット
0001				電子図書館	40976.34771.0083...	40976.10083...
0002				電子ジャーナル	68930.57954.28071...	68930.28071...
0003				専門分野 AND 文献検索	17330.52166	52166
...				...	...	...

(A)

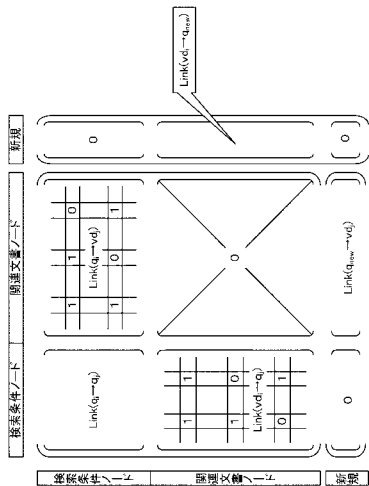
検索条件ID	ユーザID	検索条件	検索文書の文書IDセット	関連文書の文書IDセット
0001	00634	電子図書館	40976.34771.0083...	40976.10083...
0002	00289	電子ジャーナル	68930.57954.28071...	68930.28071...
0003	00093	専門分野 AND 文献検索	17330.52166	52166
...	...	...	...	...

(B)

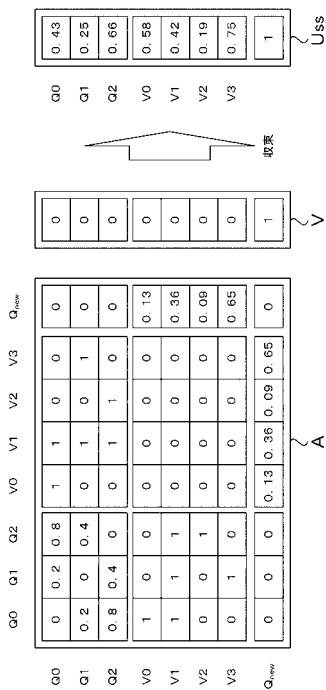
【図 6】



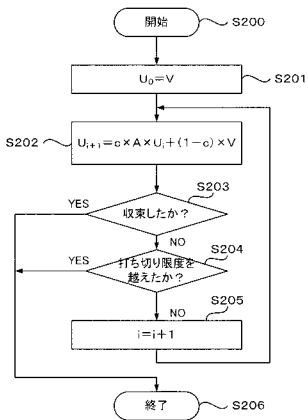
【図 7】



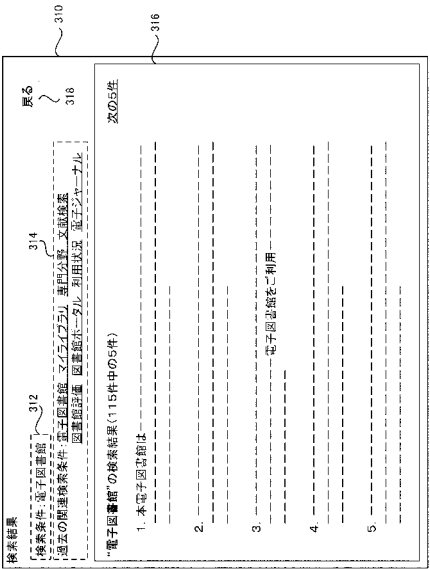
【図 8】



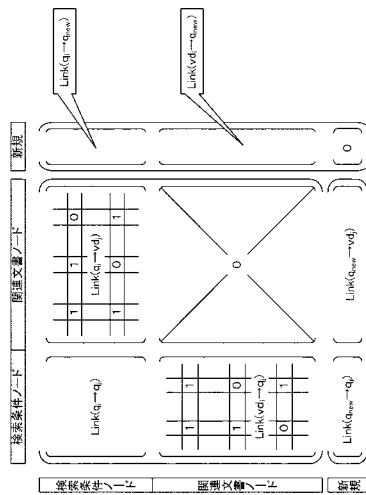
【図 9】



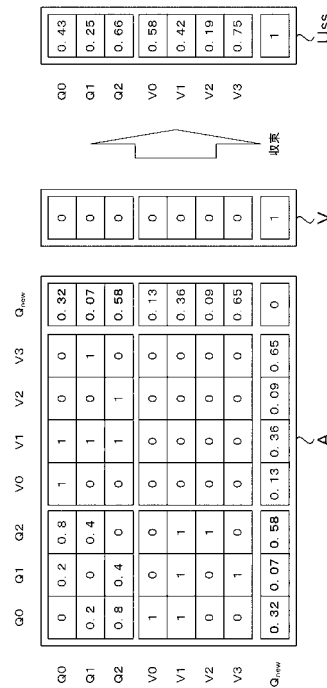
【図 10】



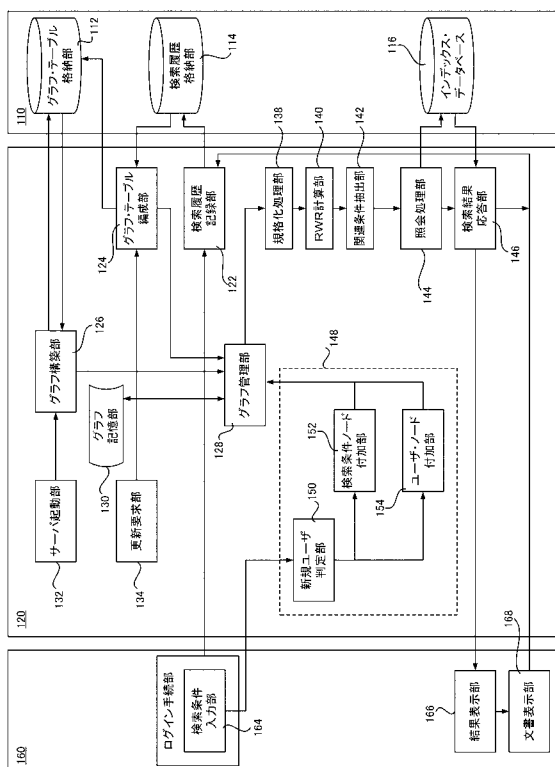
【図 1 1】



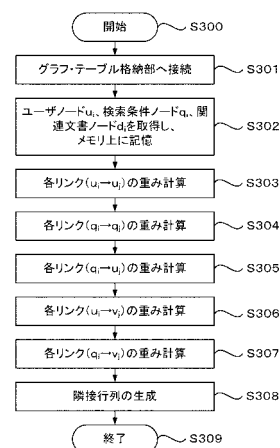
【図 1 2】



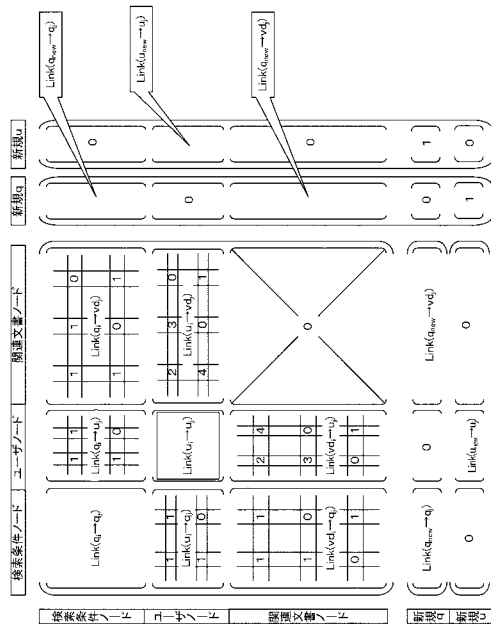
【図 1 3】



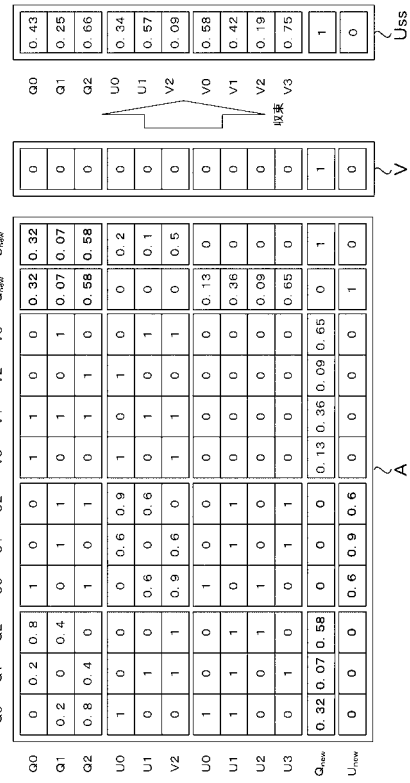
【図 1 4】



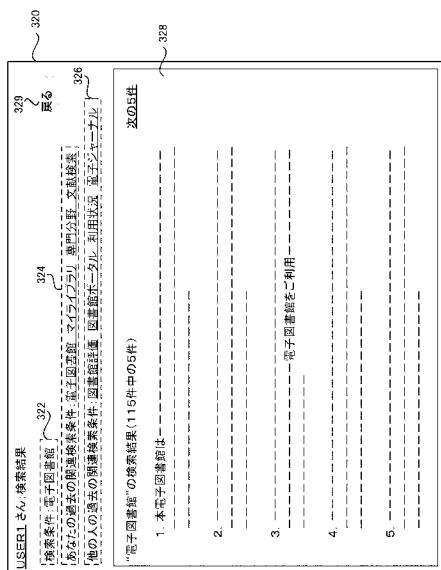
【 図 1 5 】



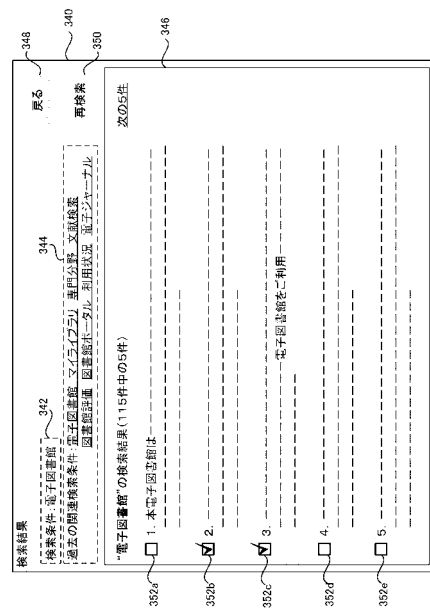
【 図 1 6 】



【 図 1 7 】



【 図 1 8 】



【図 19】

この文庫があなただの興味・関心に合っている場合には、ここをチェックしてください →

もよう

364

362

360

文庫タイトル

著者名

出版社名

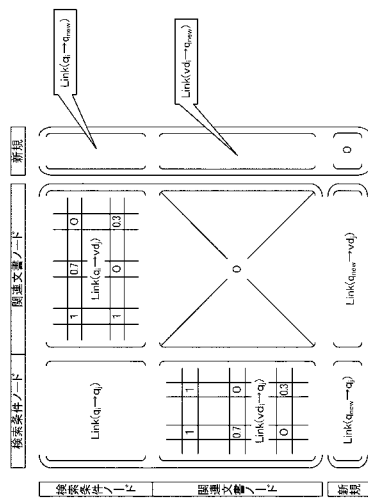
出版年

種別

【図 20】

検索条件ID	検索条件	検索文庫の文庫IDセット	参照文庫IDセット	適合文庫IDセット	指定文庫IDセット
0001	電子図書館	40976, 3477, 10083...	40976, 10083...	10083...	...
0002	電子ジャーナル	68930, 57954, 28071...	68930, 28071...	28071...	...
0003	専門分野 AND 文献検索	17330, 32166	52166	52166	...
...	...	...	...	...	...

【図 21】



【図 22】

Q0	Q1	Q2	V0	V1	V2	V3	Q _{max}
0	0.2	0.8	1	0.3	0	0	0.32
0.2	0	0.4	0	0.7	0	0.3	0.07
0.8	0.4	0	0	1	0.7	0	0.58
1	0	0	0	0	0	0	0.13
0.3	0.7	1	0	0	0	0	0.36
0	0	0.7	0	0	0	0	0.09
0	0.3	0	0	0	0	0	0.65
0.32	0.07	0.58	0.13	0.36	0.09	0.65	0

Q0	Q1	Q2	V0	V1	V2	V3
0.43	0.25	0.66	0.58	0.42	0.19	0.75
1						