

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5292250号
(P5292250)

(45) 発行日 平成25年9月18日(2013.9.18)

(24) 登録日 平成25年6月14日(2013.6.14)

(51) Int.Cl.	F 1
G 0 6 F 17/30 (2006.01)	G 0 6 F 17/30 3 4 0 B
	G 0 6 F 17/30 3 8 0 E

請求項の数 9 (全 15 頁)

(21) 出願番号	特願2009-236367 (P2009-236367)	(73) 特許権者	000004226
(22) 出願日	平成21年10月13日(2009.10.13)		日本電信電話株式会社
(65) 公開番号	特開2011-85992 (P2011-85992A)		東京都千代田区大手町二丁目3番1号
(43) 公開日	平成23年4月28日(2011.4.28)	(74) 代理人	100086232
審査請求日	平成24年3月8日(2012.3.8)		弁理士 小林 博通
		(74) 代理人	100104938
			弁理士 鶴澤 英久
		(74) 代理人	100140361
			弁理士 山口 幸二
		(74) 代理人	100096459
			弁理士 橋本 剛
		(72) 発明者	村田 真哉
			東京都千代田区大手町二丁目3番1号 日 本電信電話株式会社内

最終頁に続く

(54) 【発明の名称】 文書検索装置、文書検索方法、文書検索プログラム

(57) 【特許請求の範囲】

【請求項1】

検索ログ中のクエリで表現されたユーザの検索意図を分析し、電子文書群の検索結果に反映させる文書検索装置であって、

前記検索ログに含まれるクエリの入力回数に基づきユーザ毎に各クエリの検索頻度をカウントするクエリ頻度集計手段と、

前記クエリ頻度集計手段のカウントした検索頻度における上位頻度のクエリ集合を分析対象とし、ユーザ毎の検索行動のうち最初の検索行動と最後の検索行動とのペアをクラスタリングし、該各クラスタで表された分析対象クエリの検索意図を取得する検索意図分析手段と、

を備えることを特徴とする文書検索装置。

【請求項2】

前記検索意図分析手段の抽出した前記検索意図の周期性の有無を判定する検索意図解析手段をさらに備え、

前記検索意図に周期性があれば、前記検索結果に反映させることを特徴とする請求項1記載の文書検索装置。

【請求項3】

検索ログ中のクエリで表現されたユーザの検索意図の周期性を判別し、周期性を有する前記検索意図を格納するデータベースを利用して、ユーザ端末に入力されたクエリの入力時期を検索結果に反映させる文書検索装置であって、

前記入力クエリを前記データベースの格納データと照合して、該入力クエリの検索意図の周期性を確認し、周期性が確認できれば該入力クエリを入力時期に応じた格納データを検索意図データとして取得するクエリ照合手段と、

前記入力クエリに基づき電子文書群を検索した検索結果を、前記照合手段の取得した検索意図データに基づきリランキングするランキング処理手段と、

を備えることを特徴とする文書検索装置。

【請求項 4】

ユーザ端末に入力されたクエリに基づき電子文書群を検索し、前記ユーザ端末に応答するための検索結果を取得する文書検索装置であって、

ユーザ毎の検索ログを記録する検索ログデータベースを参照し、該検索ログに含まれるクエリを入力回数に基づき各クエリの検索頻度をカウントするクエリ頻度集計手段と、

前記クエリ頻度集計手段のカウントした検索頻度における上位頻度のクエリ集合を分析対象とし、ユーザ毎の検索行動のうち最初の検索行動と最後の検索行動とのペアをクラスタリングし、該各クラスタで表された分析対象クエリの検索意図を任意期間毎に抽出した検索意図時系列データを取得する検索意図分析手段と、

前記検索意図分析手段の抽出した検索意図時系列データにおける周期性を判定し、周期性を有すれば検索意図周期性データとして周期性データベースに格納する検索意図解析手段と、

前記入力クエリを前記周期性データベースの格納データと照合して、該入力クエリの検索意図の周期性を確認し、周期性が確認できれば該入力クエリを入力時期に応じた検索意図周期性データを検索意図データとして取得するクエリ照合手段と、

前記クエリ照合手段の取得した検索意図データに基づき前記検索結果をリランキングするランキング処理手段と、を備え、

前記投入クエリおよび前記最終結果に対するユーザのクリック情報を、前記検索ログデータベースに随時記録することを特徴とする文書検索装置。

【請求項 5】

検索ログ中のクエリで表現されたユーザの検索意図を分析し、電子文書群の検索結果に反映させる文書検索方法であって、

クエリ頻度集計手段が、前記検索ログに含まれるクエリを入力回数に基づきユーザ毎に各クエリの検索頻度をカウントするクエリ頻度集計ステップと、

検索意図分析手段が、前記クエリ頻度集計ステップのカウントした検索頻度における上位頻度のクエリ集合を分析対象とし、ユーザ毎の検索行動のうち最初の検索行動と最後の検索行動とのペアをクラスタリングし、該各クラスタで表された分析対象クエリの検索意図を取得する検索意図分析ステップと、

を有することを特徴とする文書検索方法。

【請求項 6】

検索意図解析手段が、前記検索意図分析ステップの抽出した前記検索意図の周期性の有無を判定する検索意図解析ステップをさらに有し、

前記検索意図に周期性があれば、前記検索結果に反映させることを特徴とする請求項 5 記載の文書検索方法。

【請求項 7】

検索ログ中のクエリで表現されたユーザの検索意図の周期性を判別し、周期性を有する前記検索意図を格納するデータベースを利用して、ユーザ端末に入力されたクエリを入力時期を検索結果に反映させる文書検索方法であって、

クエリ照合手段が、前記入力クエリを前記データベースの格納データと照合することで該入力クエリの検索意図の周期性を確認し、周期性が確認できれば該入力クエリを入力時期に応じた格納データを検索意図データとして取得するクエリ照合ステップと、

ランキング処理手段が、前記入力クエリに基づき電子文書群を検索した検索結果を、前記照合手段の取得した検索意図データに基づきリランキングするランキング処理ステップと、

10

20

30

40

50

を有することを特徴とする文書検索方法。

【請求項 8】

ユーザ端末に入力されたクエリに基づき電子文書群を検索し、前記ユーザ端末に応答するための検索結果を取得する文書検索方法であって、

クエリ頻度集計手段が、ユーザ毎の検索ログを記録する検索ログデータベースを参照し、該検索ログに含まれるクエリの入力回数に基づき各クエリの検索頻度をカウントするクエリ頻度集計ステップと、

検索意図分析手段が、前記クエリ頻度集計ステップでカウントした検索頻度における上位頻度のクエリ集合を分析対象とし、ユーザ毎の検索行動のうち最初の検索行動と最後の検索行動とのペアをクラスタリングし、該各クラスタで表された分析対象クエリの検索意図を任意期間毎に抽出した検索意図時系列データを取得する検索意図分析ステップと、

検索意図解析手段が、前記検索意図分析ステップで抽出した検索意図時系列データにおける周期性を判定し、周期性を有すれば検索意図周期性データとして周期性データベースに格納する検索意図解析ステップと、

クエリ照合手段が、前記入力クエリを前記周期性データベースの格納データと照合することで該入力クエリの検索意図の周期性を確認し、周期性が確認できれば該入力クエリの入力時期に応じた検索意図周期性データを検索意図データとして取得するクエリ照合ステップと、

ランキング処理手段が、前記クエリ照合ステップで取得した検索意図データに基づき前記検索結果をリランキングするランキング処理ステップと、

前記投入クエリおよび前記最終結果に対するユーザのクリック情報を、前記検索ログデータベースに随時記録する記録ステップと、

を有することを特徴とする文書検索方法。

【請求項 9】

請求項 1 ~ 請求項 4 のいずれか 1 項に記載の文書検索装置を構成する前記各手段としてコンピュータを機能させるための文書検索プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ユーザの投入した検索語（クエリ）に基づき電子文書群を検索し、検索結果を取得してユーザに応答する文書検索の技術に関する。

【背景技術】

【0002】

周知のように、インターネットなどで接続されたクライアントコンピュータを介して、サーバコンピュータに検索語（クエリ）を投入し、検索結果を取得する文書検索システムは現在多くのユーザに用いられている。このサーバコンピュータには時々刻々と検索のログが保存され、該ログには投入されたクエリの情報や、検索結果に対するクリックの情報が残されている。

【0003】

非特許文献 1 の文書検索システムは、検索エンジンに代表される文書検索システムのログを利用することでクエリの検索意図を明確化し、その情報を用いて検索を行い、検索精度の向上を実現させている。この文書検索システムでは検索エンジンのログの内、特に検索結果のクリックに関する情報（クリックログ）を解析し、よくクリックされている検索結果を特定し、そのタイトルとスニペットとから抽出されたキーワードを用いてクエリを拡張している。

【0004】

この文書検索技術は、検索結果をクリックするか否かの判断をそのタイトルとスニペットで行うというユーザの検索行為に基づくもので、拡張されたクエリはそのクエリの検索意図をキーワードで補完する形になっている。このときクリックログにはクエリとそれに対してクリックされた検索結果の URL が時間とユーザ識別子の情報と共に含まれている

10

20

30

40

50

が、非特許文献1の文書検索システムではクエリとクリックされた検索結果のURLの情報のみが使用されている。

【先行技術文献】

【非特許文献】

【0005】

【非特許文献1】Masaya Murata, Hiroyuki Toda, Yumiko Matsuura, Ryoji Kataoka "Improving Mobile Web-IR Using Access Concentration Sites in Search Results" In Proc. of WISE 2008

10

【非特許文献2】Gui-Rong Xue, Hua-jun Zeng, Zheng Chen, Yong Yu, Wei-Ying Ma, WenSi Xi, WeiGuo Fan "Optimizing Web Search Using Web Click-through Date" CIKM 2004

【非特許文献3】Michail Vlachos, Chris Meek, Zografoula Vagena "Identifying Similarities, Periodicities and Bursts for Online Search Queries" SIGMOD 2004.

【発明の概要】

【発明が解決しようとする課題】

20

【0006】

非特許文献1では、クリックログを解析し、よくクリックされている検索結果のタイトルとスニペットから抽出されたキーワードを用いてクエリを拡張し、投入されたクエリの検索意図を該キーワードで補完する手法を提案している。

【0007】

ところが、非特許文献1の検索技術は、クリックログの解析の際にどのユーザがクリックしたのかという情報を利用することなく、単純に各ページに対するクリック回数に注目しているにすぎず、各ユーザのそれぞれの検索行動の分析までは行っていない。

【0008】

その結果、投入されたクエリで表現された各ユーザの検索意図の違いや総数が分からなく、これにより検索精度が不十分となるおそれがあった。すなわち、一般的に同一のクエリが入力されても、入力したそれぞれのユーザによってその検索意図は異なるため、クエリの検索意図をより高精度に把握するためには、クリックログをどのユーザがクリックしたのかという情報を利用して解析し、それぞれの検索行動を分析することが必要である。

30

【0009】

本発明は、上述の実情を解決するためになされたものであって、検索ログ中のクエリに内在する検索意図の相違・総数やそれらの特徴を把握し、高精度の検索結果を提示することを解決課題としている。

【課題を解決するための手段】

【0010】

40

そこで、本発明は、検索ログからユーザの検索行動、即ち最初の検索行動と最後の検索行動とをクラスタリングし、クラスタの特徴を抽出することでクエリに内在する検索意図を把握する。このとき時期(時間)によってクエリで表現された検索意図が変化すれば、検索意図の周期性を判定することが好ましい。ここで把握された検索意図を検索結果に反映させる。

【0011】

本発明の文書検索装置の一態様は、検索ログ中のクエリで表現されたユーザの検索意図を分析し、電子文書群の検索結果に反映させる文書検索装置であって、前記検索ログに含まれるクエリの入力回数に基づきユーザ毎に各クエリの検索頻度をカウントするクエリ頻度集計手段と、前記クエリ頻度集計手段のカウントした検索頻度における上位頻度のクエ

50

り集合を分析対象とし、ユーザ毎の検索行動のうち最初の検索行動と最後の検索行動とのペアをクラスタリングし、該各クラスタで表された分析対象クエリの検索意図を取得する検索意図分析手段と、を備える。

【 0 0 1 2 】

前記文書検索装置は、前記検索意図分析手段の抽出した前記検索意図の周期性の有無を判定する検索意図解析手段をさらに備え、前記検索意図に周期性があれば、前記検索結果に反映させてもよい。

【 0 0 1 3 】

前記文書検索装置の他の態様は、検索ログ中のクエリで表現されたユーザの検索意図の周期性を判別し、周期性を有する前記検索意図を格納するデータベースを利用して、ユーザ端末に入力されたクエリの入力時期を検索結果に反映させる文書検索装置であって、前記入力クエリを前記データベースの格納データと照合して、該入力クエリの検索意図の周期性を確認し、周期性が確認できれば該入力クエリの入力時期に応じた格納データを検索意図データとして取得するクエリ照合手段と、前記入力クエリに基づき電子文書群を検索した検索結果を、前記照合手段の取得した検索意図データに基づきリランキングするランキング処理手段と、を備える。

【 0 0 1 4 】

前記文書検索装置のさらに他の態様は、ユーザ端末に入力されたクエリに基づき電子文書群を検索し、前記ユーザ端末に回答するための検索結果を取得する文書検索装置であって、ユーザ毎の検索ログを記録する検索ログデータベースを参照し、該検索ログに含まれるクエリの入力回数に基づき各クエリの検索頻度をカウントするクエリ頻度集計手段と、前記クエリ頻度集計手段のカウントした検索頻度における上位頻度のクエリ集合を分析対象とし、ユーザ毎の検索行動のうち最初の検索行動と最後の検索行動とのペアをクラスタリングし、該各クラスタで表された分析対象クエリの検索意図を任意期間毎に抽出した検索意図時系列データを取得する検索意図分析手段と、前記検索意図分析手段の抽出した検索意図時系列データにおける周期性を判定し、周期性を有すれば検索意図周期性データとして周期性データベースに格納する検索意図解析手段と、前記入力クエリを前記周期性データベースの格納データと照合して、該入力クエリの検索意図の周期性を確認し、周期性が確認できれば該入力クエリの入力時期に応じた検索意図周期性データを検索意図データとして取得するクエリ照合手段と、前記クエリ照合手段の取得した検索意図データに基づき前記検索結果をリランキングするランキング処理手段と、を備え、前記投入クエリおよび前記最終結果に対するユーザのクリック情報を、前記検索ログデータベースに随時記録する。

【 0 0 1 5 】

本発明の文書検索方法の一態様は、検索ログ中のクエリで表現されたユーザの検索意図を分析し、電子文書群の検索結果に反映させる文書検索方法であって、クエリ頻度集計手段が、前記検索ログに含まれるクエリの入力回数に基づきユーザ毎に各クエリの検索頻度をカウントするクエリ頻度集計ステップと、検索意図分析手段が、前記クエリ頻度集計ステップのカウントした検索頻度における上位頻度のクエリ集合を分析対象とし、ユーザ毎の検索行動のうち最初の検索行動と最後の検索行動とのペアをクラスタリングし、該各クラスタで表された分析対象クエリの検索意図を取得する検索意図分析ステップと、を有する。

【 0 0 1 6 】

前記文書検索方法は、検索意図解析手段が、前記検索意図分析ステップの抽出した前記検索意図の周期性の有無を判定する検索意図解析ステップをさらに有し、前記検索意図に周期性があれば、前記検索結果に反映させてもよい。

【 0 0 1 7 】

前記文書検索方法の他の態様は、検索ログ中のクエリで表現されたユーザの検索意図の周期性を判別し、周期性を有する前記検索意図を格納するデータベースを利用して、ユーザ端末に入力されたクエリの入力時期を検索結果に反映させる文書検索方法であって、ク

10

20

30

40

50

エリ照合手段が、前記入力クエリを前記データベースの格納データと照合することで該入力クエリの検索意図の周期性を確認し、周期性が確認できれば該入力クエリを入力時期に応じた格納データを検索意図データとして取得するクエリ照合ステップと、ランキング処理手段が、前記入力クエリに基づき電子文書群を検索した検索結果を、前記照合手段の取得した検索意図データに基づきリランキングするランキング処理ステップと、を有する。

【0018】

前記文書検索方法のさらに他の態様は、ユーザ端末に入力されたクエリに基づき電子文書群を検索し、前記ユーザ端末に回答するための検索結果を取得する文書検索方法であって、クエリ頻度集計手段が、ユーザ毎の検索ログを記録する検索ログデータベースを参照し、該検索ログに含まれるクエリを入力回数に基づき各クエリの検索頻度をカウントするクエリ頻度集計ステップと、検索意図分析手段が、前記クエリ頻度集計ステップでカウントした検索頻度における上位頻度のクエリ集合を分析対象とし、ユーザ毎の検索行動のうち最初の検索行動と最後の検索行動とのペアをクラスタリングし、該各クラスタで表された分析対象クエリの検索意図を任意期間毎に抽出した検索意図時系列データを取得する検索意図分析ステップと、検索意図解析手段が、前記検索意図分析ステップで抽出した検索意図時系列データにおける周期性を判定し、周期性を有すれば検索意図周期性データとして周期性データベースに格納する検索意図解析ステップと、クエリ照合手段が、前記入力クエリを前記周期性データベースの格納データと照合することで該入力クエリの検索意図の周期性を確認し、周期性が確認できれば該入力クエリを入力時期に応じた検索意図周期性データを検索意図データとして取得するクエリ照合ステップと、ランキング処理手段が、前記クエリ照合ステップで取得した検索意図データに基づき前記検索結果をリランキングするランキング処理ステップと、前記投入クエリおよび前記最終結果に対するユーザのクリック情報を、前記検索ログデータベースに随時記録する記録ステップと、を有する。

【0019】

なお、本発明は、前記各文書検索装置としてコンピュータを機能させるプログラムの態様としてもよく、このプログラムは記録媒体に記録した態様でもよい。

【発明の効果】

【0020】

本発明によれば、検索ログ中のクエリに内在する異なった意図の総数や、それらの特徴が把握でき、検索精度が向上する。特に、検索意図の周期的変化を発見することにより、時期（時間）によって検索意図が変化するクエリに対しても高精度な検索結果が提示される。

【図面の簡単な説明】

【0021】

【図1】本発明の実施形態に係る文書検索装置の構成図。

【図2】同 分析処理ステージの処理動作図。

【図3】同 分析処理ステージおよび解析処理ステージの処理データ構成図。

【図4】(a)はユーザの検索行動例を示す図、(b)は2分グラフの一例を示す図。

【図5】解析処理ステージおよびクエリ照合部の処理動作を示す図。

【図6】検索意図周期性データとクエリの検索意図データと入力クエリとクエリ入力情報のデータ構成図。

【図7】検索エンジンおよび時間連動型ランキング処理部の処理動作を示す図。

【図8】検索エンジンの検索結果とリランキングされた最終検索結果とクリック情報のデータ構成図。

【発明を実施するための形態】

【0022】

以下、本発明の実施形態に係る文書検索装置を説明する。この文書検索装置によれば、ユーザのクエリ投入時期（時間）に応じて、ユーザの検索意図を把握する時間連動型の文書検索を実施する。すなわち、検索エンジンの検索ログをどのユーザがクリックしたのかという情報を解析し、各ユーザのクエリを入力から検索結果のクリック、そして該検索の

終了までの一連の検索セクションを取得する。

【 0 0 2 3 】

つぎに各セクションの最初の検索行動と最後の検索行動に注目し、クエリに内在する異なった検索意図の違いや総数を把握し、検索意図を詳細に分析する。ここで分析されたクエリの検索意図の時間変化を長期間の検索ログを用いて解析し、周期性を発見する。周期性の発見されたクエリは、時間（時期）によってそのクエリで検索されるユーザの検索意図が変化することを意味する。

【 0 0 2 4 】

したがって、同一クエリが再度投入されれば、その入力時間に応じて検索結果のランキングを適切に調整し、ユーザに提示される。このとき検索結果の上位ランクには、ユーザがクエリ入力時に真に捜し求めていた情報が提示され、この点で検索精度が向上する。

【 0 0 2 5 】

装置構成例

図 1 に基づき前記文書検索装置の構成例を説明する。ここでは前記文書検索装置 1 は、ユーザ端末 2 とネットワーク経由でデータ送受信可能に接続され、ユーザ毎の検索ログに含まれたクエリで表現された検索意図の時間的な変化を捉え、前記ユーザ端末 2 のクエリ入力時期（時間）に応じて検索結果を適切にランキングする時間連動型の文書検索システムを構成している。

【 0 0 2 6 】

前記ユーザ端末 2 は、クエリの入力、検索結果の表示、そして一連のユーザの検索行動を記録するユーザインターフェース 1 4 0 を有している。このユーザインターフェース 1 4 0 は、前記文書検索装置 1 に送信するクエリを入力するクエリ入力部 1 4 1 と、前記文書検索装置 1 から返信された検索結果を表示する検索結果表示部 1 4 2 と、ユーザの検索行動を監視・記録する検索ログ記録部 1 4 3 を備えていればよく、例えばパーソナルコンピュータ（PC）や携帯電話・PDAなどのモバイル端末のブラウザにより構成される。なお、ユーザのクエリ入力作業は、キーボードなどの入力手段を通じて実施される。

【 0 0 2 7 】

前記文書検索装置 1 は、Web 検索エンジン（いわゆる検索エンジン）のシステムサーバにより構成され、通常のコンピュータのハードウェアリソース、例えば CPU、メモリ（RAM）、ハードディスクドライブ装置、通信デバイスなどを備える。このハードウェアリソースとソフトウェアリソースとの協働の結果、前記文書検索装置 1 は、検索エンジンの検索ログに含まれるクエリの検索意図を分析する分析処理ステージと、該分析処理ステージを通じて分析されたクエリの検索意図の時間的な変化を解析する解析処理ステージと、解析処理ステージで解析された検索意図の時間的な変化を検索結果に反映させた時間連動型の検索ランキングを前記ユーザ端末 2 に返信する検索処理ステージと、を実施する。なお、前記文書検索装置 1 は、必ずしも単一のコンピュータにより構成される必要は無く、複数のコンピュータにて構成してもよいものとする。

【 0 0 2 8 】

分析処理ステージは、検索ログ DB 1 1 0、クエリ頻度集計部 1 1 2、クエリ頻度ランキング DB 1 1 4、検索意図分析部 1 1 6 を通じて実施され、解析処理ステージは、検索意図時系列解析部 1 2 0、検索意図周期性 DB 1 2 2 を通じて実施される。この両ステージは、いずれも前記ユーザ端末 2 からのクエリ送信前のオフライン段階で実施される。一方、検索処理ステージは、前記ユーザ端末 2 からクエリ送信後のオンライン段階にてクエリ照合部 1 3 0、検索エンジン 1 3 1（インデックス DB 1 3 3 を含む）、時間連動型検索結果ランキング処理部 1 3 5 を通じて実施される。このうち前記各 DB 1 1 0、1 1 4、1 2 2、1 3 3 は、前記ハードディスクドライブ装置に構築されているものとする。また、前記ユーザ端末 2 とのデータ送受信は、前記通信デバイスを通じて実施されるものとする。

【 0 0 2 9 】

概略を説明すれば、分析処理ステージは、前記検索ログ DB 1 1 0 に記録された検索ロ

10

20

30

40

50

グを利用する。ここでは前記検索ログDB110には、前記ユーザ端末2から送信された前記検索ログ記録部143の記録データが蓄積されているものとする。この記録データを前記頻度集計部112が参照し、検索ログ中の各クエリの出現頻度をカウントする。カウントされた集計結果は、前記ランキングDB114に転送され、蓄積される。つぎに前記分析部116は、前記ランキングDB114に検索意図の分析対象となる上位頻度のクエリ集合を要求する。要求に対する回答を受け取ると、前記分析部116は、クエリ集合に属するそれぞれのクエリの検索意図を分析する。ここではユーザの検索行動からクエリに内在する検索意図を特徴語(キーワード)として抽出し、該特徴語を任意期間毎に時系列に抽出した検索意図時系列データを前記解析部120に転送する。

【0030】

解析処理ステージは、検索結果時系列データに基づきクエリの時間的な変化を捉える。ここでは前記解析部120は、前記分析部116から転送されたクエリの検索意図時系列データを解析し、特徴語に周期性があれば周期性を示すキーワードとみなす。この解析結果、即ち周期を示すキーワードおよび時期(時間)を検索意図周期性データとして、前記周期性DB122に登録され、検索処理ステージにて検索結果のランキングに利用される。

【0031】

すなわち、検索処理ステージは、前記周期性DB122の登録データに基づき前記ユーザ端末2からのクエリ入力時期(時間)に適した検索ランキングを生成し、前記ユーザ端末2に返信する。ここではユーザが前記クエリ入力部141にクエリを入力・送信すれば、前記文書検索装置1では受信したクエリを前記照合部130に転送する。このとき前記照合部130は、前記周期性DB122にアクセスして入力クエリ情報の周期性を照合する。照合の結果、入力クエリ情報に周期性があれば、検索意図周期性データを該クエリ情報の検索意図情報として取得する。

【0032】

取得した検索意図情報と入力クエリ情報とは、前記検索エンジン131に転送される。前記検索エンジン131は、事前にWWW(World Wide Web)132からWebページ集合を収集し、インデクスDB133に保存しているため、転送された入力クエリ情報を基に検索を実施し、通常の結果を前記ランキング処理部135に転送する。その際に検索意図情報も併せて転送するものとする。ここで前記ランキング処理部135は、前記検索エンジン131から検索意図情報が転送されれば、該検索意図情報を基に検索結果をリランキングし、クエリ入力時期(時間)に適合する検索結果ランキングを生成する。生成した検索結果は、前記ユーザ端末2に返信され、検索結果表示部142に表示される。

【0033】

これによりユーザは、検索結果表示部142に表示された検索結果を閲覧し、検索結果のクリックなどの検索行動を続けることができる。このとき前記検索ログ記録部143は、ユーザのクエリ入力や検索結果に対するクリックなどの検索行動を監視し、その情報は前記文書検索装置1に送信され、前記検索ログDB110に記録・蓄積される。以下、オフライン段階とオンライン段階とに大別して各処理ステージの処理動作を個別具体的に説明する。

【0034】

オフライン段階

(1)図2は、分析処理ステージにおけるクエリの検索意図の分析処理に関する動作処理を示している。ここでは前記検索ログDB110には、図3中のログ111が保存されているものとする。具体的には、ログ111には、ユーザのクエリ入力時における「入力時間・ユーザID・入力クエリ」と、検索結果に対するクリック時における「入力時間・ユーザID・入力クエリ・検索結果URL」とが保存されている。この検索結果URLは、検索結果に対してクリックされた電子文書(Webページ)のURLを示している。また、ユーザIDは、ユーザの識別子であり、これを固定してログ111を分析することでユ

10

20

30

40

50

ーザの検索行動、即ち検索セッションを取得する。

【0035】

前記集計部112および前記分析部116は、前記検索ログDB110に対してログの要求を行い、ログ111を取得する。このうち前記集計部112は、取得したログ111をクエリの入力回数に注目して解析し、クエリ頻度113を算出する(クエリ頻度集計ステップ)。このクエリ頻度113には、図3に示すように、クエリ毎に入力頻度がデータ格納されているものとする。算出したクエリ頻度113は、前記ランキングDB114に転送され、保存される。

【0036】

前記分析部116は、ログ111の要求と併せて前記ランキングDB114に検索意図の分析処理を実施するクエリ集合を要求する。ここでは前記分析部116は、図3に示す処理対象クエリ115、即ち入力回数が多い上位頻度のクエリ集合を取得する。このログ111および分析対象クエリ115の取得後にクエリで表された検索意図を分析する(検索意図分析ステップ)。

【0037】

詳細を説明すれば、処理対象クエリ115に属する各クエリに対してユーザIDを固定してログ111を解析し、各ユーザの検索行動を抽出する。その際、クエリを入力してから最初の検索行動(検索結果のクリック、クエリの変更、もしくは検索終了)と最後の検索行動に注目する。

【0038】

最後の検索行動とは、他のキーワードの入力や検索終了の一つ前の検索行動を意味する。ここでは最初の検索行動を「First Action」、最後の検索行動を「Last Action」と呼ぶものとする。「First Action」に注目する理由は、例えば図4(a)に示す「UserID A」のユーザのように最初に検索結果1位をクリックするユーザと、「UserID B」や「ユーザID C」のユーザのように検索結果1位を飛ばして最初に検索結果2位あるいは検索結果3位をクリックするユーザとが存在し、ユーザ毎の検索意図の違いが明確に表れ易い検索行動と考えられるためである。

【0039】

また、「Last Action」に注目する理由は、その検索行動で自分の検索意図が満たされ、検索を辞めたと考えることができるからである。ここではクエリで表現されたそれぞれのユーザの検索意図の違いを明確化し、それらを満足させることができた検索行動に注目するために、「First Action」と「Last Action」との間のつながりを考慮して解析し、その結果、図4(b)に示す2分グラフを得る。

【0040】

そして2分グラフを分析し、結び付きの強い「First Action」と「Last Action」とのペアを算出していく。ここでは一例として、非特許文献2に示されたような手法(「3.3 Co-Visited Method(CVM)」、3.4 Iterative Algorithm(IA))で、それぞれのノード間(「First Action」と「Last Action」)の類似度行列を生成し、凝集法のような既存のクラスタリング手法で各ユーザの検索行動をクラスタリングするものとする。各クラスタからクエリに内在する検索意図を表す特徴語、例えばクラスタに属する検索結果を得るために頻繁に利用(クリック)されているクエリを特徴語として付与する。なお、非特許文献2の手法を用いる際には、「Queries」・「Web Pages」を「First Action」・「Last Action」に読み替える。

【0041】

このような検索意図の分析を1週間毎や1ヶ月毎に実施し、図3に示すように、「クエリ入力時期・クエリ・検索意図番号・特徴語」がペアにデータ格納された検索意図時系列データ117を取得する。ここではクエリ入力時期は、春夏秋冬などの季節、あるいは午前午後や朝昼晩の時間帯などでもよく、その入力時期に応じた特徴語毎に検索意図番

10

20

30

40

50

号が付与されている。例えば、図3中の検索意図時系列データ117のクエリ1は「豆腐」のフレーズからなり、入力時期は「夏」と「冬」とする。この場合に入力時期「夏」の特徴語は「冷奴」が得られ、入力時期「冬」の特徴語は「湯豆腐」が得られれば、それぞれ異なる検索意図番号が付与される。

(2) つぎに図5に基づき解析処理ステージの処理動作を説明すれば、前記解析部120は、検索意図時系列データ117の転送後に解析処理を実施する。ここでは非特許文献3に示された高速フーリエ変換の手法を通じて、検索意図時系列データ117中の特徴語の周期性の有無を検索意図番号に基づき判別する(検索意図解析ステップ)。

【0042】

解析の結果、周期性があれば、その周期で代表的に表れるクラスタの特徴語を、周期を特徴付けるキーワードとみなす。この解析結果を検索意図周期性データ121として前記周期性DB122に登録する。この処理動作によりオフライン処理を終了する。なお、検索意図周期性データ121には、図6に示すように、クエリ毎に周期を示す時期(季節や時間帯など)と前記キーワードを構成する特徴語とがペアにデータ格納されている。

10

【0043】

オンライン段階

検索処理ステージにおける文書検索段階の処理動作を説明すれば、まずユーザは検索意図を表現するクエリを前記クエリ入力部141に対して入力する。このクエリ144は、図6に示すように、入力された時間(時期)とキーワード集合により構成され、前記ユーザ端末2から前記文書検索装置1に送信される。このときクエリ144は、ユーザIDの情報を付与したクエリ入力情報145のデータ形式で前記検索ログ記録部143に転送され、蓄積される。

20

【0044】

このクエリ144は、図6に示すように、入力された時間(時期)とキーワード集合により構成され、前記ユーザ端末2から前記文書検索装置1に送信される。このときクエリ144は、ユーザIDの情報を付与したクエリ入力情報145のデータ形式で前記検索ログ記録部143に転送され、蓄積される。

【0045】

前記文書検索装置1は、クエリ144を受信すると、前記照合部130に転送する。前記照合部130は、前記周期性DB122にアクセスして、クエリ144を検索意図周期性データ121のクエリ(クエリ1.2...)と照合し、周期性を確認する(クエリ照合ステップ)。具体的には、検索意図周期性データ121にクエリ144と一致するクエリがあれば周期性が肯定され、一致するクエリなければ周期性が否定される。ここではクエリ144に含まれるキーワードの一致性を照合するものとする。

30

【0046】

照合の結果、クエリ144の周期性が肯定されれば、入力された時間(時期)に応じた検索意図周期性データ121を、クエリの検索意図データ123として取得する。この検索意図データ123には、「クエリ・時期(時間)・特徴語」が格納されている。ただし、クエリ144に周期性が否定された場合、あるいは入力時間(時期)に応じた検索意図周期性データ121が無い場合には空のデータが検索意図データ123として構成される。このクエリの照合後に前記照合部130は、クエリ144とクエリの検索意図データ123とを前記検索エンジン131に転送する。

40

【0047】

前記検索エンジン131は、事前にWWW(World Wide Web)132からWebページ集合(電子文書群)を収集し、インデクスDB133に保存している。したがって、検索エンジン131は、図7に示すように、前記照合部130から転送されたクエリ144に基づきWebページ集合を検索し(検索処理ステップ)、これにより検索結果134が取得される。ここで取得された検索結果134とクエリの検索意図データ123は、前記ランキング処理部135に転送される。

【0048】

50

前記ランキング処理部 135 は、転送された検索結果 134、検索意図 123 に基づき時間連動型の検索結果ランキングを生成する（ランキング処理ステップ）。すなわち、クエリの検索意図 123 に基づき検索結果 134 をリランキングする。例えば検索結果 134 のうち、検索意図データ 123 の特徴語を含むものを上位にランキングする。したがって、入力されたクエリの時間（時期）の検索意図に適合した検索結果ランキングが生成され、最終検索結果 136 が前記ユーザ端末 2 に返信され、前記表示部 142 に表示される。これによりユーザは、最終検索結果 136 を確認し、閲覧した検索結果に対してクリックなどの検索行動を行うことができる。

【0049】

この検索結果のクリックの情報に対して、クリックされた時間とユーザIDを付与したクリック情報 146 が検索ログ記録部 143 に転送され、記録される。また、最終検索結果 136 の閲覧に続くユーザの検索行動のうち、クエリの変更（再入力）はクエリ入力部 141 を通じて入力され、該クエリ入力情報 145 も前記検索ログ記録部 143 に転送され、記録される。このとき前記検索ログ記録部 143 の記録データ 145、146 は、前記文書検索装置 1 に送信され、前記検索ログDB 110 に記録され、ログ 111 としてフィールドバックされる（記録ステップ）。したがって、前記検索ログDB 110 のログ 111 が更新され、更新された検索ログDB 110 は次に同一ユーザから入力されるクエリの時間連動型検索結果ランキングの生成の際に使用される。このような前記文書検索装置 1 によれば、以下の効果を得ることができる。

(1) すなわち、前記文書検索装置 1 によれば、前記検索ログDB 110 にはユーザIDを固定して検索エンジン 131 のログ 111 が保存されているため、前記分析部 116 を通じて各ユーザのクエリの入力から検索結果のクリック、そしてその検索の終了までの一連の検索セッションが取得される。

【0050】

ここでは各セッションの最初の行動（First Action）と最後の行動（Last Action）とに注目し、両者間の二分グラフ「Bipartite Graph」を構築することで、「First Action」と「Last Action」のペアをクラスタリング処理する。これらクラスタはクエリに内在する異なった検索意図を表しており、特徴語を抜き出すことで、検索意図の把握や総数を詳細に分析することができ、検索精度の向上に貢献する。

(2) また、前記分析部 116 は、クエリの分析意図の時間的変化を長期間の検索ログに基づき時系列データ化し、前記解析部 120 を通じて周期性の変化を発見する。ここで検索意図の周期性変化が発見されたクエリは、時間（時期）によってそのクエリで検索されるユーザの検索意図が変化することを意味するため、かかるクエリに対しては、検索処理ステージにて入力された時期（時間）に応じて検索結果ランキングを適切にリランキングしてユーザに提示される。

【0051】

すなわち、前記照合部 130 にて前記周期性DB 122 の検索意図周期性データに基づき入力クエリの周期性が確認され、入力時期（時刻）に適切な検索意図データ 123 が取得され、前記ランキング処理部 135 にて時間連動型ランキングが生成される。

【0052】

したがって、前記表示部 142 に表示される検索結果 136 の上位ランクには、ユーザがクエリ入力時期（時間）に真に探し求めていた情報が提示される。これにより時間（時期）によって検索意図が変化するような難しいクエリに対しても、高精度な検索結果を提示でき、時間（時期）を考慮した時間連動型の文書検索装置が実現され、この点で文書検索システムの検索精度が改善される。

【0053】

プログラムなど

本発明は、前記文書検索装置 1 の各部 110、112、114、116、120、122、130、131、133、135 の一部もしくは全部として、コンピュータを機能さ

10

20

30

40

50

せるプログラムとして構成することもできる。このプログラムによれば、前記各ステップの全てあるいは、その一部をコンピュータに実行させることができる。

【0054】

このプログラムは、Webサイトや電子メールなどネットワークを通じて提供することができる。また、前記プログラムは、CD-ROM、DVD-ROM、CD-R、CD-RW、DVD-R、DVD-RW、MO、HDD、Blu-ray Disk（登録商標）などの記録媒体に記録して、保存・配布することも可能である。この記録媒体は、記録媒体駆動装置を利用して読み出され、そのプログラムコード自体が前記実施形態の処理を実現するので、該記録媒体も本発明を構成する。

【符号の説明】

10

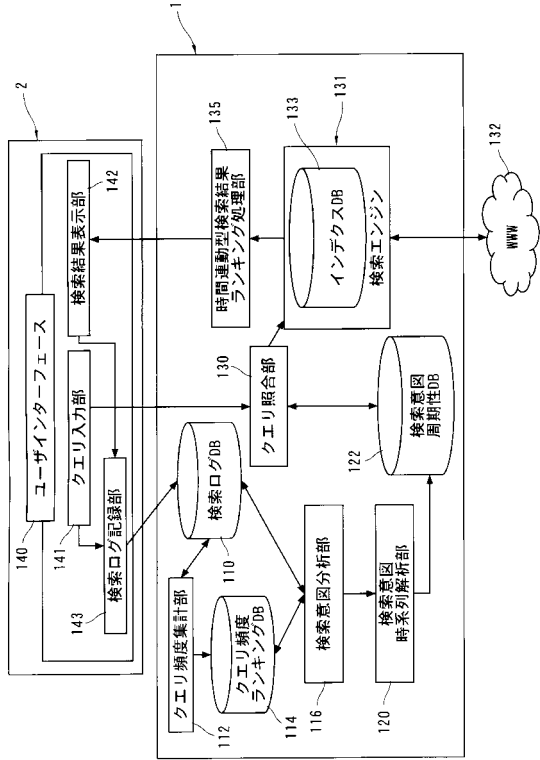
【0055】

- 1 ... 文書検索装置
- 2 ... ユーザ端末
- 110 ... 検索ログDB
- 111 ... ログ
- 112 ... クエリ頻度集計部（クエリ頻度集計手段）
- 113 ... クエリ頻度
- 114 ... クエリ頻度ランキングデータベース
- 115 ... 処理対象クエリ（分析対象クエリ）
- 116 ... 検索意図分析部（検索意図分析手段）
- 117 ... 検索意図時系列データ（検索意図）
- 120 ... 検索意図時系列解析部（検索意図解析手段）
- 121 ... 検索意図周期性データ
- 122 ... 検索意図周期性DB（データベース）
- 123 ... クエリの検索意図データ
- 130 ... クエリ照合部（クエリ照合手段）
- 131 ... 検索エンジン
- 132 ... WWW（World Wide Web）
- 133 ... インデックスDB
- 134 ... 検索結果
- 135 ... 時間連動型検索結果ランキング処理部（ランキング処理手段）
- 136 ... 最終検索結果
- 140 ... ユーザインターフェース
- 141 ... クエリ入力部
- 142 ... 検索結果表示部
- 143 ... 検索ログ記録部
- 144 ... クエリ
- 145 ... クエリ入力情報
- 146 ... クリック情報

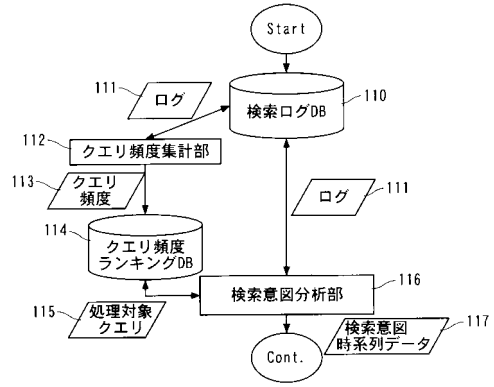
20

30

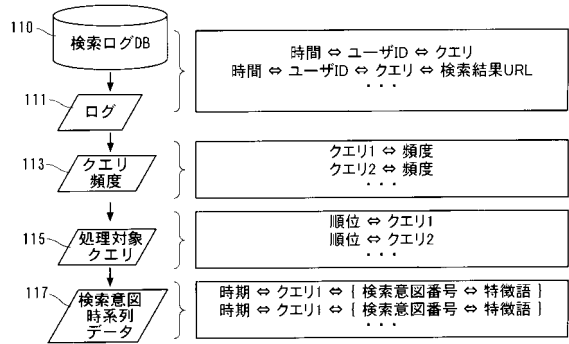
【図1】



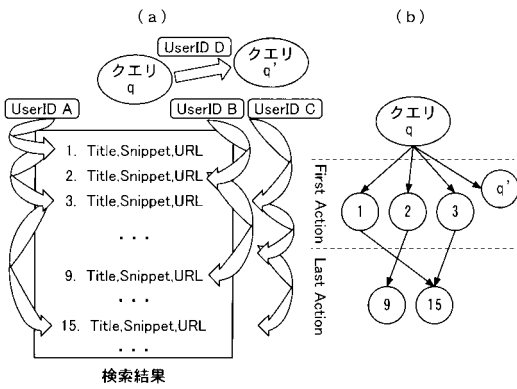
【図2】



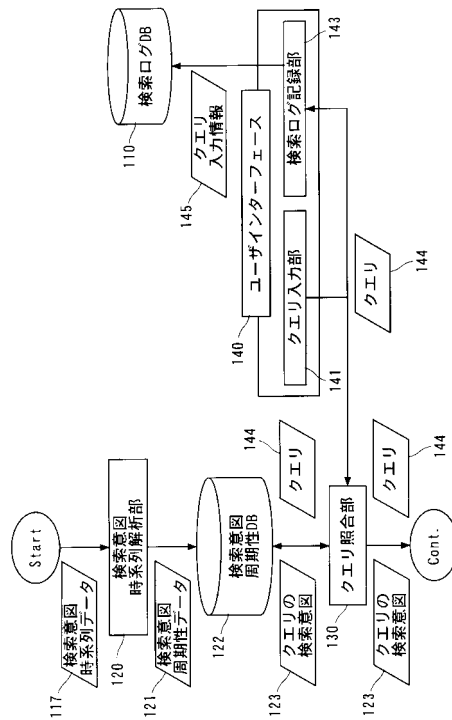
【図3】



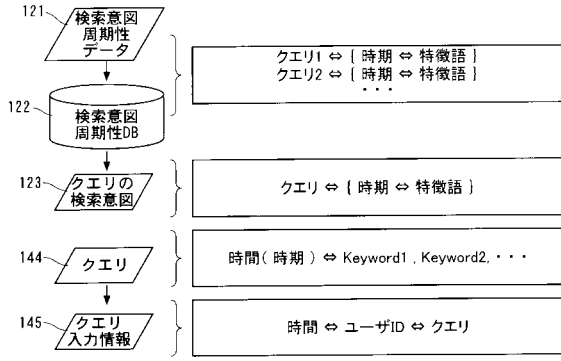
【図4】



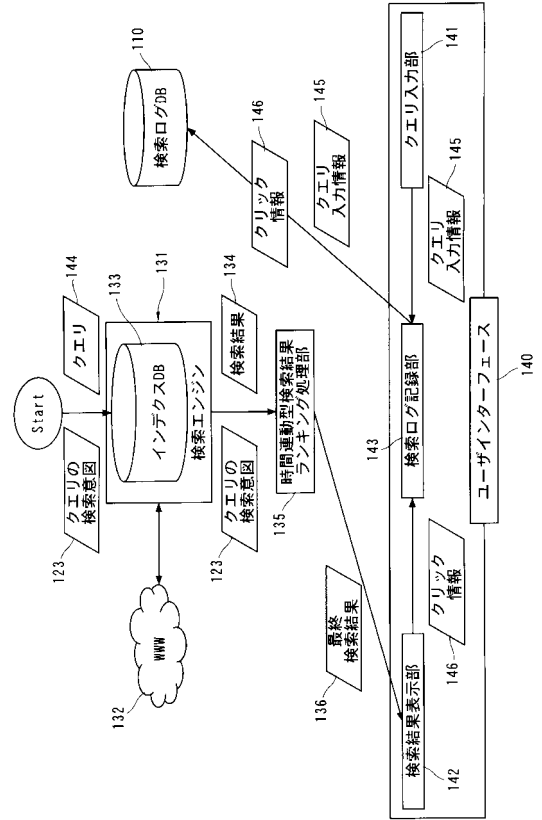
【図5】



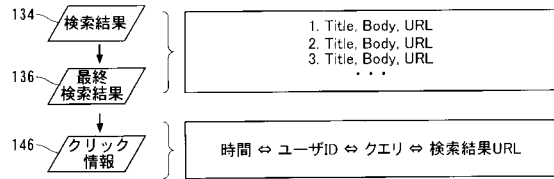
【図 6】



【図 7】



【図 8】



フロントページの続き

- (72)発明者 戸田 浩之
東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内
- (72)発明者 松浦 由美子
東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内
- (72)発明者 片岡 良治
東京都千代田区大手町二丁目3番1号 日本電信電話株式会社内

審査官 野崎 大進

- (56)参考文献 特開2003-173352(JP,A)
特開2006-107467(JP,A)
特表2008-541228(JP,A)
特開2009-110231(JP,A)
米国特許出願公開第2006/0074902(US,A1)
米国特許出願公開第2006/0248068(US,A1)
米国特許出願公開第2005/0278317(US,A1)

- (58)調査した分野(Int.Cl., DB名)
G06F 17/30
JSTPlus(JDreamIII)