

(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2010-33377

(P2010-33377A)

(43) 公開日 平成22年2月12日(2010.2.12)

(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 有 請求項の数 11 O L (全 40 頁)

(21) 出願番号 特願2008-195379 (P2008-195379)
(22) 出願日 平成20年7月29日 (2008.7.29)(71) 出願人 500257300
ヤフー株式会社
東京都港区赤坂9丁目7番1号
(74) 代理人 100106002
弁理士 正林 真之
(72) 発明者 藤田 澄男
東京都港区六本木六丁目10番1号 ヤフー株式会社内
Fターム(参考) 5B075 KK02 NK02 NK44 PQ02 PQ46
PR03 UU40

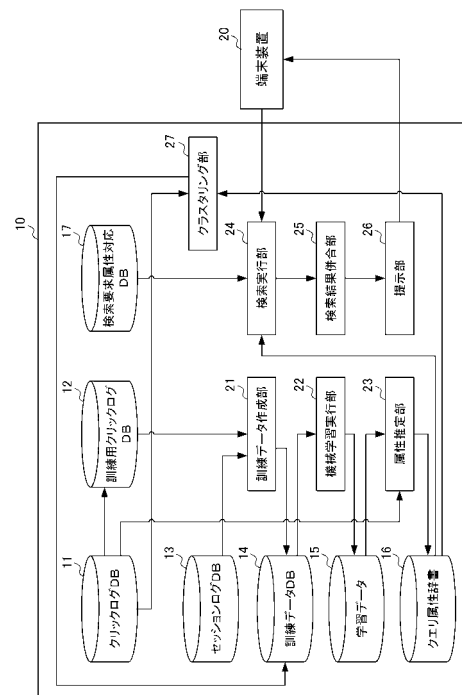
(54) 【発明の名称】 Webクエリに対する検索要求属性を自動推定する検索装置、検索方法及びプログラム

(57) 【要約】

【課題】より一般的な検索要求の属性をクエリに付与することできる検索装置、検索方法及びプログラムを提供すること。

【解決手段】検索サーバ10は、入力クエリと、この入力クエリが検索エンジンにより検索された検索結果のURLのうち、ユーザがクリックしたクリックURLとが対応付けて記憶される訓練用クリックログDB12と、訓練用クリックログDB12に記憶される入力クエリ及びクリックURLに基づいて、機械学習のための訓練データを作成する訓練データ作成部21と、訓練データを用いて学習データの機械学習を実行する機械学習実行部22と、機械学習された学習データに基づいて、訓練用クリックログDB12に記憶され、且つ訓練データには含まれない入力クエリに対し、検索要求属性を推定する属性推定部23と、推定された入力クエリの検索要求属性と、この入力クエリとを記憶するクエリ属性辞書16と、を備える。

【選択図】図3



【特許請求の範囲】**【請求項 1】**

ユーザが操作する端末装置と通信可能に接続され、当該端末装置により入力された入力クエリに応じてウェブページを検索する検索エンジンを備える検索装置において、

前記入力クエリと、当該入力クエリが前記検索エンジンにより検索された検索結果の URL のうち、前記ユーザがクリックしたクリック URL とが対応付けて記憶される検索履歴記憶手段と、

前記検索履歴記憶手段に記憶される前記入力クエリ及び前記クリック URL に基づいて、機械学習のための訓練データを作成する訓練データ作成手段と、

前記訓練データを用いて学習データの機械学習を実行する機械学習実行手段と、

前記機械学習実行手段により機械学習された学習データに基づいて、前記検索履歴記憶手段に記憶され、且つ前記訓練データには含まれない入力クエリに対し、検索要求属性を推定する属性推定手段と、

前記属性推定手段により推定された前記入力クエリの検索要求属性と、当該入力クエリとを記憶するクエリ属性記憶手段と、を更に備えることを特徴とする検索装置。

【請求項 2】

前記訓練データ作成手段は、前記訓練データとして、少なくとも前記検索履歴記憶手段に記憶される前記入力クエリを含む訓練クエリと、前記クリック URL に基づくクリック素性と、前記訓練クエリの検索要求属性とを作成することを特徴とする請求項 1 に記載の検索装置。

【請求項 3】

前記訓練データ作成手段は、前記クリック素性として、ドメイン名、サイト名、ディレクトリ等の前記クリック URL の一部を用いることを特徴とする請求項 2 に記載の検索装置。

【請求項 4】

前記訓練データ作成手段は、前記訓練クエリとして、前記検索履歴記憶手段に記憶された複数の前記入力クエリを含むクエリグループを用いることを特徴とする請求項 2 に記載の検索装置。

【請求項 5】

セッションと、当該セッション内における前記クリック URL とを対応付けて記憶するセッション履歴記憶手段を更に備え、

前記訓練データ作成手段は、前記入力クエリに対する前記検索エンジンによる検索結果において、前記セッション履歴記憶手段に記憶されたセッション内で最後のクリック対象となったクリック URL のファイルの属性を前記検索要求属性として用いることを特徴とする請求項 2 に記載の検索装置。

【請求項 6】

前記訓練データ作成手段は、前記入力クエリに対する前記検索エンジンによる検索結果において、所定期間内又は所定操作数内に予め属性が設定されたファイルに前記端末装置がアクセスした場合に、当該ファイルの属性を前記検索要求属性として用いることを特徴とする請求項 2 に記載の検索装置。

【請求項 7】

前記クエリ属性記憶手段に記憶された入力クエリ及び検索要求属性と当該検索要求属性に応じた所定の条件とに基づいて、前記端末装置により入力されたユーザ入力クエリを前記検索エンジンにより前記検索要求属性に応じた検索処理を実行する検索実行手段と、

前記端末装置により入力されたユーザ入力クエリを前記検索エンジンにより通常検索処理を実行する通常検索実行手段と、

前記検索実行手段により得られた検索結果と、前記通常検索実行手段により得られた検索結果とで重複する検索結果を併合する検索結果併合手段と、

前記検索結果併合手段により併合された検索結果を前記端末装置に提示する提示手段と、を更に備えることを特徴とする請求項 2 に記載の検索装置。

10

20

30

40

50

【請求項 8】

前記クエリ属性記憶手段に記憶された入力クエリ及び検索要求属性に基づいて、前記検索履歴記憶手段に記憶された複数の前記クリック素性にクラスタリングを実行してクリック素性グループを作成するクリック素性クラスタリング手段を更に備えることを特徴とする請求項 3 に記載の検索装置。

【請求項 9】

前記クエリ属性記憶手段に記憶された入力クエリ及び検索要求属性に基づいて、前記検索履歴記憶手段に記憶された複数の入力クエリにクラスタリングを実行してクエリグループを作成するクエリクラスタリング手段を更に備えることを特徴とする請求項 4 に記載の検索装置。

10

【請求項 10】

ユーザが操作する端末装置と通信可能に接続され、当該端末装置により入力された入力クエリに応じてウェブページを検索する検索エンジンを備える検索装置の検索方法において、

前記入力クエリと、当該入力クエリが前記検索エンジンにより検索された検索結果の URL のうち、前記ユーザがクリックしたクリック URL とが対応付けて記憶される検索履歴記憶手段に記憶される前記入力クエリ及び前記クリック URL に基づいて、機械学習のための訓練データを作成する訓練データ作成ステップと、

前記訓練データを用いて学習データの機械学習を実行する機械学習実行ステップと、

前記機械学習実行ステップにより機械学習された学習データに基づいて、前記検索履歴記憶手段に記憶され、且つ前記訓練データには含まれない入力クエリに対し、検索要求属性を推定する推定ステップと、

20

前記推定ステップにより推定された前記入力クエリの検索要求属性と、当該入力クエリとをクエリ属性記憶手段に記憶させる属性記憶制御ステップと、を備えることを特徴とする検索方法。

【請求項 11】

ユーザが操作する端末装置と通信可能に接続され、当該端末装置により入力された入力クエリに応じてウェブページを検索する検索エンジンを備える検索装置のコンピュータにより実行されるコンピュータプログラムにおいて、

前記入力クエリと、当該入力クエリが前記検索エンジンにより検索された検索結果の URL のうち、前記ユーザがクリックしたクリック URL とが対応付けて記憶される検索履歴記憶手段に記憶される前記入力クエリ及び前記クリック URL に基づいて、機械学習のための訓練データを作成する訓練データ作成ステップと、

30

前記訓練データを用いて学習データの機械学習を実行する機械学習実行ステップと、

前記機械学習実行ステップにより機械学習された学習データに基づいて、前記検索履歴記憶手段に記憶され、且つ前記訓練データには含まれない入力クエリに対し、検索要求属性を推定する推定ステップと、

前記推定ステップにより推定された前記入力クエリの検索要求属性と、当該入力クエリとをクエリ属性記憶手段に記憶させる属性記憶制御ステップと、をコンピュータに実行させるコンピュータプログラム。

40

【発明の詳細な説明】**【技術分野】****【0001】**

本発明は、検索装置、検索方法及びプログラムに関する。

【背景技術】**【0002】**

近年、Web 検索サービスの高度化に伴い、Web 検索に用いるクエリの背景となる検索者の検索意図を推定して、検索結果を改良することが重要となってきている。

【0003】

50

Web 検索におけるクエリの検索要求指向としては、ナビゲーション指向、トランザクション指向等が挙げられる。ここで、ナビゲーション指向型の検索要求とは、ある特定の Web ページ（又はある対象物の代表的な Web ページ）に到達することを要求する検索である。例えば、ある特定の大学のホームページに到達することを目的とするような検索は、この検索要求に当たる。

【0004】

トランザクション指向型の検索要求とは、インタラク션을伴うような Web ページ（オンラインショッピング、Web が仲介する様々なサービス等）に到達することを要求する検索である。例えば、東京都に存在するホテルから自分の要求に合ったホテルを探し、予約することを目的としたような検索は、この検索要求に当たる。

10

【0005】

例えば、非特許文献 1 には、機械学習によりナビゲーション指向型の検索要求を有するクエリを判定する技術が開示されている。

【非特許文献 1】Lu, Y. Peng, F. Li, X. and Ahmed, N. (2006) Coupling feature selection and machine learning methods for navigational query identification, In Proceedings of CIKM 2006.

【発明の開示】

【発明が解決しようとする課題】

20

【0006】

しかしながら、特許文献 1 に開示された技術は、特定の検索要求について判定する技術であるため、他の検索要求について判定することはできない。

【0007】

そこで、本発明は、より一般的な検索要求の属性をクエリに付与することできる検索装置、検索方法及びプログラムを提供することを目的とする。

なお、検索要求の属性（以下、検索要求属性という）とは、ナビゲーション指向型やトランザクション指向型等に留まらず、より詳細な利用者の検索意図に関連する属性を指す。

【課題を解決するための手段】

30

【0008】

本発明では、以下のような解決手段を提供する。

【0009】

（１）ユーザが操作する端末装置と通信可能に接続され、当該端末装置により入力された入力クエリに応じてウェブページを検索する検索エンジンを備える検索装置において、前記入力クエリと、当該入力クエリが前記検索エンジンにより検索された検索結果の URL のうち、前記ユーザがクリックしたクリック URL とが対応付けて記憶される検索履歴記憶手段と、前記検索履歴記憶手段に記憶される前記入力クエリ及び前記クリック URL に基づいて、機械学習のための訓練データを作成する訓練データ作成手段と、前記訓練データを用いて学習データの機械学習を実行する機械学習実行手段と、前記機械学習実行手段により機械学習された学習データに基づいて、前記検索履歴記憶手段に記憶され、且つ前記訓練データには含まれない入力クエリに対し、検索要求属性を推定する属性推定手段と、前記属性推定手段により推定された前記入力クエリの検索要求属性と、当該入力クエリとを記憶するクエリ属性記憶手段と、を更に備えることを特徴とする検索装置。

40

【0010】

このように、検索装置は、検索履歴記憶手段に記憶されたクエリと、クリック属性とに基づいて、クエリの検索要求属性を推定する。そして、推定されたクエリの検索要求属性は、例えば検索結果ランキング、広告提示、ユーザに推薦するサービスへのショートカット等の検索サービスに利用することにより、検索サービスの向上を図ることができる。

【0011】

50

(2) 前記訓練データ作成手段は、前記訓練データとして、少なくとも前記検索履歴記憶手段に記憶される前記入力クエリを含む訓練クエリと、前記クリックURLに基づくクリック素性と、前記訓練クエリの検索要求属性とを作成することを特徴とする(1)に記載の検索装置。

【0012】

このように、検索装置は、訓練データとして、訓練クエリと、クリック素性と、訓練クエリの検索要求属性とを作成し、この訓練データを用いて学習データの機械学習を実行する。このため、クエリの検索要求属性を推定するために適した訓練データを用いて学習データの機械学習を実行することができる。したがって、より好適にクエリの検索要求属性を推定することができる。

10

【0013】

(3) 前記訓練データ作成手段は、前記クリック素性として、ドメイン名、サイト名、ディレクトリ等の前記クリックURLの一部を用いることを特徴とする(2)に記載の検索装置。

【0014】

このように、検索装置は、クリックURLの一部は、クエリの検索要求属性を絞り込むための重要な情報となる可能性が高い。したがって、クリックURLの一部を訓練データとして用いることによって、より好適にクエリの検索要求属性を推定することができる。

20

【0015】

(4) 前記訓練データ作成手段は、前記訓練クエリとして、前記検索履歴記憶手段に記憶された複数の前記入力クエリを含むクエリグループを用いることを特徴とする(2)に記載の検索装置。

【0016】

このように、検索装置は、検索要求属性について関連性の高い複数のクエリをクエリグループとする。このため、訓練クエリとしてクエリグループを用いることによって、より好適にクエリの検索要求属性を推定することができる。

【0017】

(5) セッションと、当該セッション内における前記クリックURLとを対応付けて記憶するセッション履歴記憶手段を更に備え、前記訓練データ作成手段は、前記入力クエリに対する前記検索エンジンによる検索結果において、前記セッション履歴記憶手段に記憶されたセッション内で最後のクリック対象となったクリックURLのファイルの属性を前記検索要求属性として用いることを特徴とする(2)に記載の検索装置。

30

【0018】

セッション内で最後のクリック対象となったクリックURLのサフィックス(jpg、wmv、mp3、html等)は、クエリの検索要求属性を絞り込むための重要な情報となる可能性が高い。したがって、検索サーバ10は、訓練データとして、セッション内で最後のクリック対象となったクリックURLのファイルの属性を検索要求属性として用いることによって、より好適にクエリの検索要求属性を推定することができる。

【0019】

(6) 前記訓練データ作成手段は、前記入力クエリに対する前記検索エンジンによる検索結果において、所定期間内又は所定操作数内に予め属性が設定されたファイルに前記端末装置がアクセスした場合に、当該ファイルの属性を前記検索要求属性として用いることを特徴とする(2)に記載の検索装置。

40

【0020】

したがって、セッション内で最後のクリック対象が特定できない場合でも、クリックURLのファイルの属性を検索要求属性として用いることができる。

【0021】

(7) 前記クエリ属性記憶手段に記憶された入力クエリ及び検索要求属性と当該検索要求属性に応じた所定の条件とに基づいて、前記端末装置により入力されたユーザ入力クエ

50

りを前記検索エンジンにより前記検索要求属性に応じた検索処理を実行する検索実行手段と、前記端末装置により入力されたユーザ入力クエリを前記検索エンジンにより通常検索処理を実行する通常検索実行手段と、前記検索実行手段により得られた検索結果と、前記通常検索実行手段により得られた検索結果とで重複する検索結果を併合する検索結果併合手段と、前記検索結果併合手段により併合された検索結果を前記端末装置に提示する提示手段と、を更に備えることを特徴とする(2)に記載の検索装置。

【0022】

このように検索装置は、クエリの検索要求属性に応じた検索結果を端末装置へ提示することができる。したがって、検索装置を管理する者にとっては、検索サービスの向上を図ることができる。

10

【0023】

(8) 前記クエリ属性記憶手段に記憶された入力クエリ及び検索要求属性に基づいて、前記検索履歴記憶手段に記憶された複数の前記クリック素性にクラスタリングを実行してクリック素性グループを作成するクリック素性クラスタリング手段を更に備えることを特徴とする(3)に記載の検索装置。

【0024】

このように、検索装置は、クリック素性をクラスタリングして、グループ化する。そして、このグループ化されたクリック素性グループを検索履歴記憶手段に追加することによって、訓練データを用いて機械学習される学習データの精度を向上させることができる。したがって、より好適にクエリの検索要求属性を推定することができる。

20

【0025】

(9) 前記クエリ属性記憶手段に記憶された入力クエリ及び検索要求属性に基づいて、前記検索履歴記憶手段に記憶された複数の前記クリック素性にクラスタリングを実行してクリック素性グループを作成するクリック素性クラスタリング手段を更に備えることを特徴とする(3)に記載の検索装置。

【0026】

このように、検索装置は、クエリをクラスタリングして、グループ化する。そして、このグループ化されたクエリグループを検索履歴記憶手段に追加することによって、訓練データを用いて機械学習される学習データの精度を向上させることができる。したがって、より好適にクエリの検索要求属性を推定することができる。

30

【0027】

(10) ユーザが操作する端末装置と通信可能に接続され、当該端末装置により入力された入力クエリに応じてウェブページを検索する検索エンジンを備える検索装置の検索方法において、前記入力クエリと、当該入力クエリが前記検索エンジンにより検索された検索結果のURLのうち、前記ユーザがクリックしたクリックURLとが対応付けて記憶される検索履歴記憶手段に記憶される前記入力クエリ及び前記クリックURLに基づいて、機械学習のための訓練データを作成する訓練データ作成ステップと、前記訓練データを用いて学習データの機械学習を実行する機械学習実行ステップと、前記機械学習実行ステップにより機械学習された学習データに基づいて、前記検索履歴記憶手段に記憶され、且つ前記訓練データには含まれない入力クエリに対し、検索要求属性を推定する推定ステップと、前記推定ステップにより推定された前記入力クエリの検索要求属性と、当該入力クエリとをクエリ属性記憶手段に記憶させる属性記憶制御ステップと、を備えることを特徴とする検索方法。

40

【0028】

(11) ユーザが操作する端末装置と通信可能に接続され、当該端末装置により入力された入力クエリに応じてウェブページを検索する検索エンジンを備える検索装置のコンピュータにより実行されるコンピュータプログラムにおいて、前記入力クエリと、当該入力クエリが前記検索エンジンにより検索された検索結果のURLのうち、前記ユーザがクリックしたクリックURLとが対応付けて記憶される検索履歴記憶手段に記憶される前記入力クエリ及び前記クリックURLに基づいて、機械学習のための訓練データを作成する訓

50

練データ作成ステップと、前記訓練データを用いて学習データの機械学習を実行する機械学習実行ステップと、前記機械学習実行ステップにより機械学習された学習データに基づいて、前記検索履歴記憶手段に記憶され、且つ前記訓練データには含まれない入力クエリに対し、検索要求属性を推定する推定ステップと、前記推定ステップにより推定された前記入力クエリの検索要求属性と、当該入力クエリとをクエリ属性記憶手段に記憶させる属性記憶制御ステップと、をコンピュータに実行させるコンピュータプログラム。

【発明の効果】

【0029】

本発明によれば、より一般的な検索要求の属性をクエリに付与することできる検索装置、検索方法及びプログラムを提供することができる。

10

【発明を実施するための最良の形態】

【0030】

以下、本発明の実施形態について図面を参照しながら説明する。

【0031】

[検索システム 1 の機能構成]

図 1 は、本発明の検索装置の一実施形態に係る検索システム 1 の概略を示す図である。検索サーバ（検索装置）10、端末装置 20、及び外部サーバ 30 は、通信回線 40 を介して接続されている。端末装置 20 のユーザは、検索サーバ 10 にアクセスし、所定の検索エンジンに対して、所望の Web 文書に辿り着くためのクエリ（検索キーワード）を入力することにより、検索結果を取得する。ユーザは、この検索結果にリストされた URL

20

【0032】

端末装置 20 は、パソコン、携帯電話機、PHS（Personal Handy-phone System：登録商標）、モバイル、ノート PC、PDA（Personal Digital Assistant）等の端末装置を代表とするものであり、有線又は無線通信を行う通信回線 40 を介して、検索サーバ 10 から送信される情報を参照できるものであればどのようなものでもよい。また、検索サーバ 10 は、必要に応じて Web サーバ、DB サーバ、アプリケーションサーバを含んで構成してよく、1 台のサーバで構成しても、それぞれ別のサーバで構成してもよい。

30

【0033】

[検索サーバ 10 のハードウェア構成]

図 2 は、本発明の一実施形態である検索サーバ 10 のハードウェア構成を示す図である。図 2 に示すように、制御部 300 を構成する CPU（Central Processing Unit）310（マルチプロセッサ構成では CPU 320 等複数の CPU が追加されてもよい）、バスライン 200、通信 I/F（I/F：インタフェース）330、メインメモリ 340、BIOS（Basic Input Output System）350、I/O コントローラ 360、ハードディスク 370、光ディスクドライブ 380、並びに半導体メモリ 390 を備える。なお、ハードディスク 370、光ディスクドライブ 380、並びに、半導体メモリ 390 はまとめて記憶装置 410 と呼ばれる。

40

【0034】

制御部 300 は、検索サーバ 10 を統括的に制御する部分であり、ハードディスク 370 に記憶された各種プログラムを適宜読み出して実行することにより、上述したハードウェアと協働し、本発明に係る各種機能を実現している。

【0035】

通信 I/F 330 は、検索サーバ 10 が、ネットワークを介して端末装置 20 等の他の装置と情報を送受信する場合のネットワーク・アダプタである。

【0036】

BIOS 350 は、検索サーバ 10 の起動時に CPU 310 が実行するブートプログラムや、検索サーバ 10 のハードウェアに依存するプログラム等を記録する。

【0037】

50

I/Oコントローラ360には、ハードディスク370、光ディスクドライブ380、及び半導体メモリ390等の記憶装置410を接続することができる。

【0038】

ハードディスク370は、本ハードウェアを検索サーバ10として機能させるための各種プログラム、本発明の機能を実行するプログラム及び後述するデータベース等を記憶する。なお、検索サーバ10は、外部に別途設けたハードディスク(図示せず)を外部記憶装置として利用することもできる。

【0039】

光ディスクドライブ380としては、例えば、DVD-ROMドライブ、CD-ROMドライブ、DVD-RAMドライブ、CD-RAMドライブを使用することができる。この場合は各ドライブに対応した光ディスク400を使用する。光ディスク400から光ディスクドライブ380によりプログラム又はデータを読み取り、I/Oコントローラ360を介してメインメモリ340又はハードディスク370に提供することもできる。

【0040】

なお、本発明でいうコンピュータとは、記憶装置、制御部等を備えた検索装置をいい、検索サーバ10は、記憶装置410、制御部300等を備えた検索装置により構成され、この検索装置は、本発明のコンピュータの概念に含まれる。

【0041】

以上の例は、検索サーバ10について主に説明したが、コンピュータに、プログラムをインストールして、そのコンピュータをサーバ装置として動作させることにより上記で説明した機能を実現することもできる。したがって、本発明において一実施形態として説明した検索サーバ10により実現される機能は、上述の方法を当該コンピュータにより実行することにより、或いは、上述のプログラムを当該コンピュータに導入して実行することによっても実現可能である。

【0042】

図3は、本発明の検索装置の一実施形態である検索サーバ10の機能ブロック図である。検索サーバ10は、クリックログデータベース(以下クリックログDBという)11、訓練用クリックログデータベース(以下訓練用クリックログDBという)12、セッションログデータベース(以下セッションログDBという)13、訓練データデータベース(以下訓練データDBという)14、学習データデータベース(以下学習データDBという)15、クエリ属性辞書16、検索要求属性対応DB17、訓練データ作成部21、機械学習実行部22、属性推定部23、検索実行部24、検索結果併合部25、提示部26、及びクラスタリング部27を備える。

【0043】

クリックログDB11、訓練用クリックログDB12、セッションログDB13、訓練データDB14、学習データDB15、クエリ属性辞書16、及び検索要求属性対応DB17は、主として記憶装置410により構成される。

また、訓練データ作成部21、機械学習実行部22、属性推定部23、検索実行部24、検索結果併合部25、提示部26、及びクラスタリング部27は、主として制御部300により構成される。

【0044】

クリックログDB(検索履歴記憶手段)11は、クエリと、このクエリを検索キーワードとして検索エンジンにより検索された検索結果のURL(Uniform Resource Locator)のうち、ユーザの操作する任意の端末装置20により選択(クリック)されたURL(以下クリックURLともいう)とを対応付けて履歴データとして記憶する。この履歴データをクリックログと呼ぶ。クリックログは、例えば、図4に示すように、クリックデータテーブル111として蓄積的に記憶される。図4に示すクリックデータテーブル111には、クリックURLを選択した端末装置20のIPアドレス、クリックURLが選択された時間、セッション、クエリ、クリックURL、及びクリックURLの検索ランクが対応付けて記憶されている。

10

20

30

40

50

【 0 0 4 5 】

訓練用クリックログDB（検索履歴記憶手段）12は、クリックログDB11から任意に選択したクリックログが記憶され、訓練データの作成に用いられる。クリックログDB11から訓練用クリックログを選択する方法としては、例えば、無作為抽出、頻度の高いデータから順に抽出する、頻度の高いクエリグループから抽出する等が挙げられる。なお、クエリグループについては後述する。

【 0 0 4 6 】

セッションログDB（セッション履歴記憶手段）13は、セッションと、このセッション内における前記クリックURLとを対応付けて履歴データとして記憶する。この履歴データをセッションログと呼ぶ。セッションログは、例えば、図5に示すように、セッションログテーブル131として累積的に記憶される。図5に示すセッションログテーブル131には、クリックURLを選択した端末装置20のIPアドレス、クリックURLが選択された時間、セッション、リファラ、及びクリックURLが対応付けて記憶されている。

10

【 0 0 4 7 】

ここで、セッション及びリファラについて説明する。セッションとは、WWWにおけるアクセス数の単位の一つであり、Webページを訪れたユーザがサイト内で行なう一連の行動をまとめて1セッションという。同一のユーザでも、ある程度間隔が開いた場合は、新しいセッションとしてカウントする。

【 0 0 4 8 】

リファラとは、あるWebページのリンクをクリックして別のWebページに移動したときのリンク元となるWebページであり、本実施形態においては、クエリに応じて検索エンジンにより検索された検索結果である。

20

【 0 0 4 9 】

訓練データ作成部21は、訓練データとして、少なくとも訓練用クリックログDB12に記憶されるクエリを含む訓練クエリと、クリックURLに基づくクリック素性と、訓練クエリの検索要求属性とを作成する。

【 0 0 5 0 】

このように、検索サーバ10は、訓練データとして、訓練クエリと、クリック素性と、訓練クエリの検索要求属性とを作成し、この訓練データを用いて学習データの機械学習を実行する。このため、クエリの検索要求属性を推定するために適した訓練データを用いて学習データの機械学習を実行することができる。したがって、より好適にクエリの検索要求属性を推定することができる。

30

【 0 0 5 1 】

具体的には、訓練データ作成部21は、訓練クエリの検索要求属性を予め付与しておくことができる。

【 0 0 5 2 】

また、訓練データ作成部21は、クリック素性として、ドメイン名、サイト名、ディレクトリ等のクリックURLの一部を用いることができる。

【 0 0 5 3 】

このため、例えば検索要求属性が「自動車」である場合、クリックURLの一部としての「www.carxxxsor.net」のような例が挙げられる。このようにクリックURLの一部は、クエリの検索要求属性を絞り込むための重要な情報となる可能性が高い。したがって、クリックURLの一部を訓練データとして用いることによって、より好適にクエリの検索要求属性を推定することができる。

40

【 0 0 5 4 】

また、訓練データ作成部21は、訓練クエリとして、訓練用クリックログDB12に記憶された複数のクエリを含むクエリグループを用いることができる。クエリグループとは、後述するクラスタリング部27により作成される。クエリグループは、例えば[カローラ, トヨタカローラ, カローラトヨタ]のように表される。

50

【 0 0 5 5 】

このように、検索サーバ 10 は、検索要求属性について関連性の高い複数のクエリをクエリグループとする。このため、訓練クエリとしてクエリグループを用いることによって、より好適にクエリの検索要求属性を推定することができる。

【 0 0 5 6 】

また、訓練データ作成部 21 は、クエリに対する検索エンジンによる検索結果において、セッションログ DB 13 に記憶されたセッション内で最後のクリック対象となったクリック URL のファイルの属性を検索要求属性として用いることができる。

【 0 0 5 7 】

セッション内で最後のクリック対象となったクリック URL のサフィックス (j p g 、 w m v 、 m p 3 、 h t m l 等) は、クエリの検索要求属性を絞り込むための重要な情報となる可能性が高い。例えば、「 j p g 」は静止画、「 w m v 」は動画、「 m p 3 」は音楽、「 h t m l 」は Web ページにそれぞれ関連性が強いと考えられる。したがって、検索サーバ 10 は、訓練データとして、セッション内で最後のクリック対象となったクリック URL のファイルの属性を検索要求属性として用いることによって、より好適にクエリの検索要求属性を推定することができる。

【 0 0 5 8 】

また、訓練データ作成部 21 は、入力クエリに対する検索エンジンによる検索結果において、所定期間内又は所定操作数内に予め属性が設定されたファイルに端末装置 20 がアクセスした場合に、このファイルの属性を検索要求属性として用いることができる。

【 0 0 5 9 】

したがって、セッション内で最後のクリック対象が特定できない場合でも、クリック URL のファイルの属性を検索要求属性として用いることができる。

【 0 0 6 0 】

このような訓練データとしての訓練クエリ、クリック素性、及び検索要求属性を訓練データ作成部 21 は、三つ組みに変換する。具体的には、訓練データ作成部 21 は、クリック素性が「カ○ーラ」、クリック素性が「 w w w . c a r x x x s o r . n e t 」、検索要求属性が「自動車」の場合には、(カ○ーラ、 w w w . c a r x x x s o r . n e t 、自動車) のような三つ組みに変換する。

【 0 0 6 1 】

そして、この三つ組み変換された訓練データを機械学習実行部 22 により機械学習する。具体的には、機械学習実行部 22 によりクエリの検索要求属性を判別するための判別関数を学習する場合には、学習する各々の検索要求属性をクリック素性のベクター表現で以下のように表す。

(数 1)

$$a_1 = (s_1 = 1, s_2 = 0, s_3 = 0, s_4 = 1, s_5 = 0, \dots, s_m = 0)$$

$$a_2 = (s_1 = 0, s_2 = 1, s_3 = 0, s_4 = 1, s_5 = 0, \dots, s_m = 0)$$

...

$$a_n = (s_1 = 0, s_2 = 0, s_3 = 0, s_4 = 0, s_5 = 1, \dots, s_m = 0) \cdot$$

・ ・ 式 (1)

【 0 0 6 2 】

ここで、 a_n は検索要求属性であり、 s_m はクリック素性である。具体的には、検索要求属性 a_1 において、 $s_1 = 1$ である。これは、クリック素性 s_1 が検索要求属性 a_1 に関連することを示している。すなわち、検索要求属性 a_1 を有するクエリは、クリック素性がサイト名である場合には、このサイト名を含むクリック URL が選択されたことを意味する。また、検索要求属性 a_1 において、 $s_2 = 0$ である。これは、クリック素性 s_2 が検索要求属性 a_1 に関連しないことを示している。

【 0 0 6 3 】

訓練データ DB 14 は、訓練データ作成部 21 により作成された訓練データを記憶する。具体的には、上述したように検索要求属性を予め付与した訓練クエリやクリック素性の

10

20

30

40

50

ベクター表現等を記憶する。

【0064】

機械学習実行部22は、訓練データを用いて学習データの機械学習を実行する。訓練データは、上述したように判別関数を学習する場合には、ベクター表現によって表現するため、SVM(Support Vector Machine)、kNN(k Nearest Neighbor)、Naive bayes classifier、ニューラルネットワーク、判別木等の様々な公知の機械学習手法を用いて、判別関数を学習することができる。

【0065】

学習データDB15は、機械学習実行部22により機械学習が実行された学習データを記憶する。具体的には、学習データDB15には、機械学習実行部22により機械学習が実行された判別関数が記憶される。

【0066】

属性推定部23は、機械学習実行部22により機械学習された学習データに基づいて、クリックログDB11に記憶され、且つ訓練用クリックログDB12には含まれないクエリに対し、検索要求属性を推定する。ここで、学習データには、クリックURLと、クリック頻度が含まれる。なお、クリック頻度とは、同一のセッションにおける、クリックURLの頻度であり、訓練用クリックログDB12に記憶されるクリックURLと、入力クエリとに基づいて、制御部300により算出される値である。

そして、属性推定部23は、推定した検索要求属性と、クエリとをクエリ属性辞書16に記憶する。

【0067】

クエリ属性辞書16は、属性推定部23により推定されたクエリの検索要求属性と、このクエリとを対応付けて記憶する。

図6は、クエリ属性辞書16の一例を示す図である。図6では、クエリの検索要求属性が関連性の高いグループに分けられている。例えば、クエリ「カローラ」に対応する検索要求属性1(図6中では属性1と表記)「自動車」、検索要求属性2(図6中では属性2と表記)「中古車」、及び検索要求属性3(図6中では属性3と表記)「国産車」が検索要求属性グループ1として記憶されている。

【0068】

検索実行部24は、クエリ属性辞書16に記憶されたクエリ及び検索要求属性と当該検索要求属性に応じた所定の条件とに基づいて、端末装置20により入力されたユーザ入力クエリを検索エンジンにより検索要求属性に応じた検索処理を実行する。ここで、検索実行部24は、所定の条件として、検索要求属性対応DB17を適用することができる。

【0069】

図7は、検索要求属性対応DB17に記憶される検索要求属性対応テーブルの一例を示す図である。検索要求属性対応テーブルは、検索要求属性と、この検索供給属性に対応する処理とが対応付けて記憶される。例えば、検索要求属性が「中古車」のときには、対応する処理は「該当するクリックURLのランクを2上げる」となる。また、検索要求属性が「自動車」のときには、対応する処理は「該当する検索要求属性を有する広告を表示する」となる。また、検索要求属性が「音楽ダウンロード」のときには、対応する処理は「該当するクリックURLで最も上位のものをランクの1位にする」となる。また、検索要求属性が「音楽ダウンロード」のときには、対応する処理は「該当するクリックURLのスコアを5%上げる」となる。また、検索要求属性が「音楽ダウンロード」のときには、対応する処理は「該当する検索要求属性を有する広告を表示する」となる。

【0070】

このように、本実施形態の検索サーバ10は、クリックログDB11に記憶されたクエリと、クリック素性とに基づいて、クエリの検索要求属性を推定する。そして、推定されたクエリの検索要求属性は、例えば検索結果ランキング、広告提示、ユーザに推薦するサービスへのショートカット等の検索サービスに利用することにより、検索サービスの向上

10

20

30

40

50

を図ることができる。

【0071】

また、検索実行部24は、端末装置20により入力されたクエリを検索エンジンにより通常検索処理を実行する。

【0072】

検索結果併合部25は、検索実行部24の検索要求属性に応じた検索処理により得られた検索結果と、検索実行部24の通常検索処理により得られた検索結果とで重複する検索結果を併合する。

【0073】

提示部26は、検索結果併合部25により併合された検索結果を端末装置20に提示する。

10

【0074】

このように検索サーバ10は、クエリの検索要求属性に応じた検索結果を端末装置20へ提示することができる。したがって、検索サーバ10を管理する者にとっては、検索サービスの向上を図ることができる。また、端末装置20を操作するユーザにとっては、クエリの検索要求属性に応じた入力クエリの検索結果を得ることができる。

【0075】

なお、上述した判別関数を機械学習する場合に代えて、クエリの検索要求属性を機械学習により直接、学習することもできる。

この場合には、訓練データ作成部21は、クリックログDB11に記憶されたクエリの頻度の大きいものから100万件ほどについて訓練用クリックログDB12に記憶する。

20

【0076】

そして、訓練データ作成部21は、訓練用クリックログDB12に記憶されたクエリのうち、1万件ほどを訓練データとして人手により検索要求属性を付与する。

機械学習実行部22は、この訓練データを用いて検索要求属性を有するか否かを判別する2値分類器を上述した公知の機械学習手法を用いて機械学習を実行する。

【0077】

制御部300は、100万件ほどのクエリのうち、残る99万件のクエリを学習した2値分類器を用いて自動分類する。そして、検索要求属性に対応するクエリのリストを作成する。

30

属性推定部23は、端末装置20から受け付けたユーザ入力クエリと、作成されたリストに含まれるクエリとが一致する場合には、このユーザ入力クエリの検索要求属性を作成されたクエリに対応する検索要求属性であると推定する。

【0078】

次に、クラスタリング部27について説明する。クラスタリング部27は、クエリ属性辞書16に記憶されたクエリ及び検索要求属性に基づいて、訓練用クリックログDB12に記憶された複数のクリック素性にクラスタリングを実行してクリック素性グループを作成する。具体的には、クラスタリング部27は、クリック素性としてのドメイン名、サイト名、ディレクトリ等のクリックURLの一部分をクラスタリングを実行してクリック素性グループを作成する。

40

【0079】

また、クラスタリング部27は、クエリ属性辞書16に記憶されたクエリ及び検索要求属性に基づいて、訓練用クリックログDB12に記憶された複数のクエリにクラスタリングを実行してクエリグループを作成する。具体的には、クラスタリング部27は、訓練用クリックログDB12に記憶されたクリックURL全てにクラスタリングを実行し、クエリグループを作成する。

【0080】

ここで、クラスタリングを実行する際のクラスタリング方式としては、例えば各々のクエリをクリックURLの集合で表現して、クリックURLの集合間での一致度によるワンパスクラスタリングや階層クラスタリング等を適用することができる。

50

【 0 0 8 1 】

クラスタリング部 2 7 によりクラスタリングを実行した場合には、三つ組み変換された訓練データは、例えば（「カローラ、トヨタカローラ、カローラトヨタ」、www.carxxxsor.net、自動車）のような構成となる。ここで、クラスタリングされたクエリグループは、「カローラ、トヨタカローラ、カローラトヨタ」、クリック素性は、「www.carxxxsor.net」、検索要求属性が「自動車」である。

【 0 0 8 2 】

そして、クラスタリング部 2 7 は、クラスタリングを実行したことにより得られたクリック素性グループ及びクエリグループを訓練データ DB 1 4 に記憶される訓練データに追加する。すなわち、訓練データにおけるクリック素性としてクリック素性グループを追加する。また、訓練データにおける訓練クエリとしてクエリグループを追加する。

10

【 0 0 8 3 】

このように、検索サーバ 1 0 は、クエリ及びクリック素性をクラスタリングして、グループ化する。そして、このグループ化されたクエリグループ又はクリック素性グループを訓練データ DB 1 4 に追加することによって、訓練データを用いて機械学習される学習データの精度を向上させることができる。したがって、より好適にクエリの検索要求属性を推定することができる。

【 0 0 8 4 】

ここで、クラスタリング部 2 7 による処理について詳細に説明する。

上述した式（ 1 ）では、クリック素性についてベクターで表現した。ところで、このベクターは、クリックログ DB 1 1 に記憶されるクエリについて、

20

（ 数 2 ）

$$q\ 1 = (s\ 1 = 1, s\ 2 = 0, s\ 3 = 0, s\ 4 = 1, s\ 5 = 0, \dots, s\ m = 0)$$

$$q\ 2 = (s\ 1 = 0, s\ 2 = 1, s\ 3 = 0, s\ 4 = 1, s\ 5 = 0, \dots, s\ m = 0)$$

...

$$q\ n = (s\ 1 = 0, s\ 2 = 0, s\ 3 = 0, s\ 4 = 0, s\ 5 = 1, \dots, s\ m = 0) \cdot$$

・ 式（ 2 ）

と表現することができる。式（ 2 ）は、n 個のクエリと m 個のクリック素性という n × m の 2 次元の行列で表現される。これにより、クラスタリング部 2 7 は、式（ 2 ）をクエリごとに分けたベクターを用いてクエリのクラスタリングを実行することができる。また、クラスタリング部 2 7 は、式（ 2 ）をクリック素性ごとに分けたベクターを用いてクリック素性のクラスタリングを実行することができる。

30

【 0 0 8 5 】

この場合、特にクエリの次元数は、数百万・数千万にも増大する可能性があるため、例えば同一のクラスタ内のクエリは、全て同一視して一つの次元として扱うことが必要となる。

【 0 0 8 6 】

[処理の流れ]

図 8 は、本実施形態の検索サーバ 1 0 における検索要求属性の推定処理の流れを示すフローチャートである。

40

ステップ S 1 : 訓練データ作成部 2 1 は、訓練用クリックログ DB 1 2 に記憶された入力クエリに属性を付与する。具体的には、訓練データ作成部 2 1 は、入力クエリに対して予め用意されたクエリを人手により付与される。

【 0 0 8 7 】

ステップ S 2 : 訓練データ作成部 2 1 は、ステップ S 1 において属性が付与された入力クエリを含む訓練クエリと、訓練用クリックログ DB 1 2 に記憶されるクリック素性及び検索要求属性とを三つ組みに変換する。具体的には、訓練データ作成部 2 1 は、訓練クエリが「カローラ」、クリック素性が「www.carxxxsor.net」、検索要求属性が「自動車」の場合には、（カローラ、www.carxxxsor.net、自動車）のような三つ組みに変換する。そして、三つ組みに変換された訓練データは、訓練デ

50

ータ作成部 2 1 により訓練データ DB 1 4 に記憶される。

【 0 0 8 8 】

ステップ S 3 : 訓練データ作成部 2 1 は、ステップ S 3 において三つ組みに変換された訓練データの検索要求属性をクリック素性のベクター表現で表現する

【 0 0 8 9 】

ステップ S 4 : 機械学習実行部 2 2 は、ステップ 3 において得られた訓練データを用いて、学習データの機械学習を実行する。具体的には、機械学習実行部 2 2 は、ステップ S においてベクター表現された訓練データを用いて入力クエリの検索要求属性を判別するための判別関数を学習する。そして、機械学習実行部 2 2 は、学習した判別関数を学習データ DB 1 5 に記憶させる。

【 0 0 9 0 】

ステップ S 5 : 属性推定部 2 3 は、機械学習実行部 2 2 により機械学習された判別関数によって、クリックログ DB 1 1 に記憶され、且つ訓練用クリックログ DB 1 2 には含まれないクエリに対し、検索要求属性を推定する。

【 0 0 9 1 】

ステップ S 6 : 属性推定部 2 3 は、ステップ S 5 において推定された各々の検索要求属性と、この検索要求属性が推定されたクエリとを対応付けてクエリ属性辞書 1 6 に記憶させる。

【 0 0 9 2 】

図 9 は、本実施形態の検索サーバ 1 0 におけるクエリ属性辞書 1 6 を用いた検索処理の流れを示すフローチャートである。

ステップ S 1 1 : 検索実行部 2 4 は、端末装置 2 0 からのユーザ入力クエリを受け付ける。

ステップ S 1 2 : 検索実行部 2 4 は、クエリ属性辞書 1 6 に記憶されたクエリ及び検索要求属性と検索要求属性対応 DB 1 7 とに基づいて、ステップ S 1 1 において受け付けたユーザ入力クエリを検索エンジンにより検索要求属性に応じた検索処理を実行する。

【 0 0 9 3 】

ステップ S 1 3 : 検索実行部 2 4 は、ステップ S 1 1 において受け付けたユーザ入力クエリを検索エンジンにより通常検索処理を実行する。

ステップ S 1 4 : 検索結果併合部 2 5 は、ステップ S 1 2 において得られた検索結果と、ステップ S 1 3 において得られた検索結果とで重複する検索結果を併合する。

【 0 0 9 4 】

ステップ S 1 5 : 提示部 2 6 は、ステップ S 1 4 において併合された検索結果を端末装置 2 0 に提示する。

【 0 0 9 5 】

図 1 0 は、本実施形態の検索サーバ 1 0 におけるクラスタリング処理の流れを示すフローチャートである。

ステップ S 2 1 : クラスタリング部 2 7 は、クリックログ DB 1 1 に記憶されたクエリ又は / 及びクリックログ DB 1 1 に記憶されたクリック URL のクラスタリングを実行する。

【 0 0 9 6 】

ステップ S 2 2 : クラスタリング部 2 7 は、ステップ S 2 1 においてクラスタリングされた複数のクエリをグループ化する。これにより、上述したように

ステップ S 2 3 : クラスタリング部 2 7 は、ステップ S 2 1 においてクラスタリングされた複数のクリック URL をグループ化する。これにより、上述したように

【 0 0 9 7 】

ステップ S 2 3 : クラスタリング部 2 7 は、ステップ S 2 2 又は / 及びステップ S 2 3 においてグループ化されたクエリ又は / 及びクリック URL を訓練クエリ又は / 及びクリック素性として、訓練データ DB 1 4 に追加して記憶させる。

【 0 0 9 8 】

10

20

30

40

50

以上、本発明の実施形態について説明したが、本発明は上述した実施形態に限るものではない。また、本発明の実施形態に記載された効果は、本発明から生じる最も好適な効果を列挙したに過ぎず、本発明による効果は、本発明の実施例に記載されたものに限定されるものではない。

【図面の簡単な説明】

【0099】

【図1】本発明の検索装置の一実施形態に係る検索システム1の概略を示す図である。

【図2】本発明の一実施形態である検索サーバ10のハードウェア構成を示す図である。

【図3】本発明の検索装置の一実施形態である検索サーバ10の機能ブロック図である。

【図4】クリックデータテーブル111の一例を示す図である。

10

【図5】セッションログテーブル131の一例を示す図である。

【図6】クエリ属性辞書16の一例を示す図である。

【図7】検索要求属性対応DB17の一例を示す図である。

【図8】本実施形態の検索サーバ10における検索要求属性の推定処理の流れを示すフローチャートである。

【図9】本実施形態の検索サーバ10におけるクエリ属性辞書16を用いた検索処理の流れを示すフローチャートである。

【図10】本実施形態の検索サーバ10におけるクラスタリング処理の流れを示すフローチャートである。

【符号の説明】

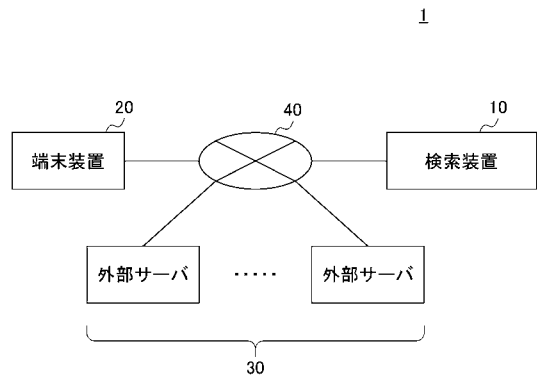
20

【0100】

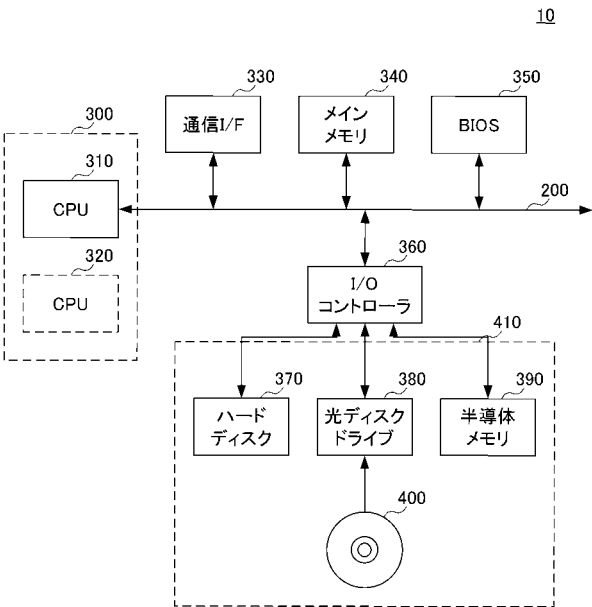
- 10 検索サーバ
- 11 クリックログDB
- 12 訓練用クリックログDB
- 13 セッションログDB
- 14 訓練データDB
- 15 学習データDB
- 16 クエリ属性辞書
- 17 検索要求属性対応動作DB
- 20 端末装置
- 21 訓練データ作成部
- 22 機械学習実行部
- 23 属性推定部
- 24 検索実行部
- 25 検索結果併合部
- 26 提示部
- 27 クラスタリング部

30

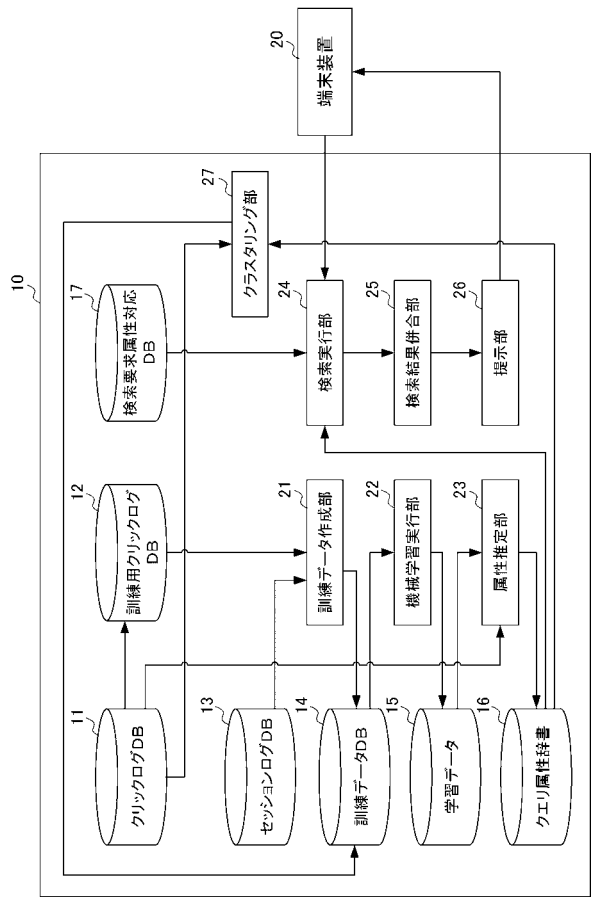
【 図 1 】



【 図 2 】



【 図 3 】



【 図 4 】

クリックログ				クエリ	URL
IP アドレス	時間	セッション	ランク	クエリ	URL
110.149.145.1	2006/10/6 21:45:53	Bp4ed6917	3	自動車	http://xxx.yyy.zzz
110.149.145.1	2006/10/6 21:49:01	Bp4ed6917	2	トヨタ	http://xxx.yyy.aaa

【 図 5 】

セッションログ		時間	セッション	リファラ	URL
IP	アドレス				
110.149.145.1		2006/10/6 21:45:53	Bp4ed6917	http://search.yahoo.co.jp/search/?p=tomaruta	http://xxx.yyy.zzz
110.149.145.1		2006/10/6 21:49:01	Bp4ed6917	http://search.yahoo.co.jp/search/?p=tomaruta	http://xxx.yyy.zzz

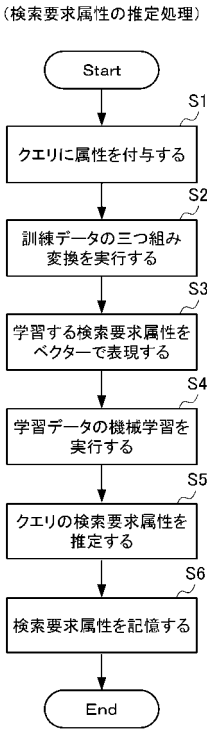
【 図 6 】

クエリ属性辞書	検索要求属性グループ1					検索要求属性グループ2	
	属性1	属性2	属性3	属性4	属性5	属性1	属性2
	自動車	中古車	国産車			保険	車検
	カローラ						

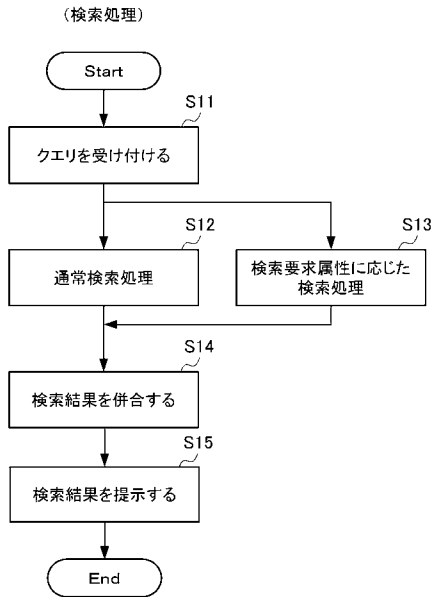
【 図 7 】

検索要求属性対応テーブル	
検索要求属性	対応する処理
中古車	該当するクリックURLのランクを2上げる
自動車	該当する検索要求属性を有する広告を表示する
音楽ダウンロード	該当するクリックURLで最も上位のものをランクの1位にする
音楽ダウンロード	該当するクリックURLのスコアを5%上げる
音楽ダウンロード	該当する検索要求属性を有する広告を表示する
...	...

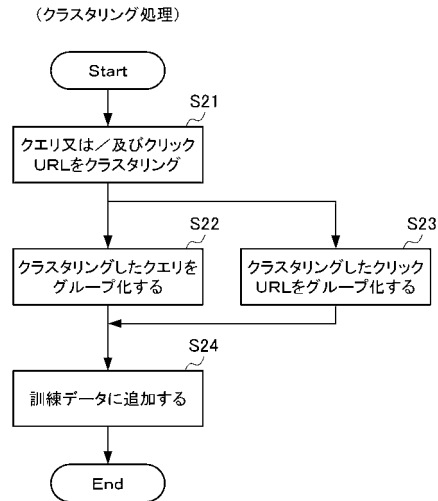
【 図 8 】



【図 9】



【図 10】



【手続補正書】

【提出日】平成21年9月15日(2009.9.15)

【手続補正 1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項 1】

ユーザが操作する端末装置と通信可能に接続され、当該端末装置により入力された入力クエリに応じてウェブページを検索する検索エンジンを備える検索装置において、

前記入力クエリと、当該入力クエリが前記検索エンジンにより検索された検索結果のURLのうち、前記ユーザがクリックしたクリックURLとが対応付けて記憶される検索履歴記憶手段と、

前記検索履歴記憶手段に記憶される前記入力クエリ及び前記クリックURLに基づいて、機械学習のための訓練データを作成する訓練データ作成手段と、

前記訓練データを用いて学習データの機械学習を実行する機械学習実行手段と、

前記機械学習実行手段により機械学習された前記学習データに基づいて、前記検索履歴記憶手段に記憶され、かつ、前記訓練データには含まれない前記入力クエリに対し、検索要求属性を推定する属性推定手段と、

前記属性推定手段により推定された前記入力クエリの検索要求属性と、当該入力クエリとを記憶するクエリ属性記憶手段と、を更に備えることを特徴とする検索装置。

【請求項 2】

前記訓練データ作成手段は、前記訓練データとして、少なくとも前記検索履歴記憶手段に記憶される前記入力クエリを含む訓練クエリと、前記クリックURLに基づくクリック

素性と、前記訓練クエリの前記検索要求属性とを作成することを特徴とする請求項 1 に記載の検索装置。

【請求項 3】

前記訓練データ作成手段は、前記クリック素性として、ドメイン名、サイト名、ディレクトリ等の前記クリック URL の一部分を用いることを特徴とする請求項 2 に記載の検索装置。

【請求項 4】

前記訓練データ作成手段は、前記訓練クエリとして、前記検索履歴記憶手段に記憶された複数の前記入力クエリを含むクエリグループを用いることを特徴とする請求項 2 に記載の検索装置。

【請求項 5】

セッションと、当該セッション内における前記クリック URL とを対応付けて記憶するセッション履歴記憶手段を更に備え、

前記訓練データ作成手段は、前記入力クエリに対する前記検索エンジンによる検索結果において、前記セッション履歴記憶手段に記憶された前記セッション内で最後のクリック対象となった前記クリック URL のファイルの属性を前記検索要求属性として用いることを特徴とする請求項 2 に記載の検索装置。

【請求項 6】

前記訓練データ作成手段は、前記入力クエリに対する前記検索エンジンによる検索結果において、所定期間内又は所定操作数内に予め属性が設定されたファイルに前記端末装置がアクセスした場合に、当該ファイルの属性を前記検索要求属性として用いることを特徴とする請求項 2 に記載の検索装置。

【請求項 7】

前記クエリ属性記憶手段に記憶された前記入力クエリ及び前記検索要求属性と当該検索要求属性に応じた所定の条件とに基づいて、前記端末装置により入力されたユーザ入力クエリを用いて前記検索エンジンにより前記検索要求属性に応じた検索処理を実行する検索実行手段と、

前記端末装置により入力された前記ユーザ入力クエリを用いて前記検索エンジンにより通常検索処理を実行する通常検索実行手段と、

前記検索実行手段により得られた検索結果と、前記通常検索実行手段により得られた検索結果とで重複する検索結果を併合する検索結果併合手段と、

前記検索結果併合手段により併合された前記検索結果を前記端末装置に提示する提示手段と、を更に備えることを特徴とする請求項 2 に記載の検索装置。

【請求項 8】

前記クエリ属性記憶手段に記憶された前記入力クエリ及び前記検索要求属性に基づいて、前記検索履歴記憶手段に記憶された複数の前記クリック素性にクラスタリングを実行してクリック素性グループを作成するクリック素性クラスタリング手段を更に備えることを特徴とする請求項 3 に記載の検索装置。

【請求項 9】

前記クエリ属性記憶手段に記憶された前記入力クエリ及び前記検索要求属性に基づいて、前記検索履歴記憶手段に記憶された複数の前記入力クエリにクラスタリングを実行してクエリグループを作成するクエリクラスタリング手段を更に備えることを特徴とする請求項 4 に記載の検索装置。

【請求項 10】

ユーザが操作する端末装置と通信可能に接続され、当該端末装置により入力された入力クエリに応じてウェブページを検索する検索エンジンを備える検索装置の検索方法において、

前記入力クエリと、当該入力クエリが前記検索エンジンにより検索された検索結果の URL のうち、前記ユーザがクリックしたクリック URL とが対応付けて記憶される検索履歴記憶手段に記憶される前記入力クエリ及び前記クリック URL に基づいて、機械学習の

ための訓練データを作成する訓練データ作成ステップと、

前記訓練データを用いて学習データの機械学習を実行する機械学習実行ステップと、

前記機械学習実行ステップにより機械学習された前記学習データに基づいて、前記検索履歴記憶手段に記憶され、かつ、前記訓練データには含まれない前記入力クエリに対し、検索要求属性を推定する推定ステップと、

前記推定ステップにより推定された前記入力クエリの検索要求属性と、当該入力クエリとをクエリ属性記憶手段に記憶させる属性記憶制御ステップと、を備えることを特徴とする検索方法。

【請求項 11】

ユーザが操作する端末装置と通信可能に接続され、当該端末装置により入力された入力クエリに応じてウェブページを検索する検索エンジンを備える検索装置のコンピュータにより実行されるコンピュータプログラムにおいて、

前記入力クエリと、当該入力クエリが前記検索エンジンにより検索された検索結果のURLのうち、前記ユーザがクリックしたクリックURLとが対応付けて記憶される検索履歴記憶手段に記憶される前記入力クエリ及び前記クリックURLに基づいて、機械学習のための訓練データを作成する訓練データ作成ステップと、

前記訓練データを用いて学習データの機械学習を実行する機械学習実行ステップと、

前記機械学習実行ステップにより機械学習された前記学習データに基づいて、前記検索履歴記憶手段に記憶され、かつ、前記訓練データには含まれない前記入力クエリに対し、検索要求属性を推定する推定ステップと、

前記推定ステップにより推定された前記入力クエリの検索要求属性と、当該入力クエリとをクエリ属性記憶手段に記憶させる属性記憶制御ステップと、をコンピュータに実行させるコンピュータプログラム。

【手続補正2】

【補正対象書類名】明細書

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、検索装置、検索方法及びプログラムに関する。

【背景技術】

【0002】

近年、Web検索サービスの高度化に伴い、Web検索に用いるクエリの背景となる検索者の検索意図を推定して、検索結果を改良することが重要となってきている。

【0003】

Web検索におけるクエリの検索要求指向としては、ナビゲーション指向、トランザクション指向等が挙げられる。ここで、ナビゲーション指向型の検索要求とは、ある特定のWebページ（又はある対象物の代表的なWebページ）に到達することを要求する検索である。例えば、ある特定の大学のホームページに到達することを目的とするような検索は、この検索要求に当たる。

【0004】

トランザクション指向型の検索要求とは、インタラク션을を伴うようなWebページ（オンラインショッピング、Webが仲介する様々なサービス等）に到達することを要求する検索である。例えば、東京都に存在するホテルから自分の要求に合ったホテルを探し、予約することを目的としたような検索は、この検索要求に当たる。

【0005】

例えば、非特許文献1には、機械学習によりナビゲーション指向型の検索要求を有するクエリを判定する技術が開示されている。

【非特許文献 1】Lu, Y. Peng, F. Li, X. and Ahmed, N. (2006) Coupling feature selection and machine learning methods for navigational query identification, In Proceedings of ICKM 2006.

【発明の開示】

【発明が解決しようとする課題】

【0006】

しかしながら、非特許文献 1 開示された技術は、特定の検索要求について判定する技術であるため、他の検索要求について判定することはできない。

【0007】

そこで、本発明は、より一般的な検索要求の属性をクエリに付与することのできる検索装置、検索方法及びプログラムを提供することを目的とする。

なお、検索要求の属性（以下、検索要求属性という）とは、ナビゲーション指向型やトランザクション指向型等に留まらず、より詳細な利用者の検索意図に関連する属性を指す。

【課題を解決するための手段】

【0008】

本発明では、以下のような解決手段を提供する。

【0009】

（１）ユーザが操作する端末装置と通信可能に接続され、当該端末装置により入力された入力クエリに応じてウェブページを検索する検索エンジンを備える検索装置において、前記入力クエリと、当該入力クエリが前記検索エンジンにより検索された検索結果の URL のうち、前記ユーザがクリックしたクリック URL とが対応付けて記憶される検索履歴記憶手段と、前記検索履歴記憶手段に記憶される前記入力クエリ及び前記クリック URL に基づいて、機械学習のための訓練データを作成する訓練データ作成手段と、前記訓練データを用いて学習データの機械学習を実行する機械学習実行手段と、前記機械学習実行手段により機械学習された前記学習データに基づいて、前記検索履歴記憶手段に記憶され、かつ、前記訓練データには含まれない前記入力クエリに対し、検索要求属性を推定する属性推定手段と、前記属性推定手段により推定された前記入力クエリの検索要求属性と、当該入力クエリとを記憶するクエリ属性記憶手段と、を更に備えることを特徴とする検索装置。

【0010】

このように、検索装置は、検索履歴記憶手段に記憶され、かつ、訓練データには含まれない入力クエリに対し、クエリの検索要求属性を推定する。そして、推定されたクエリの検索要求属性は、例えば検索結果ランキング、広告提示、ユーザに推薦するサービスへのショートカット等の検索サービスに利用することにより、検索サービスの向上を図ることができる。

【0011】

（２）前記訓練データ作成手段は、前記訓練データとして、少なくとも前記検索履歴記憶手段に記憶される前記入力クエリを含む訓練クエリと、前記クリック URL に基づくクリック素性と、前記訓練クエリの前記検索要求属性とを作成することを特徴とする（１）に記載の検索装置。

【0012】

このように、検索装置は、訓練データとして、訓練クエリと、クリック素性と、訓練クエリの検索要求属性とを作成し、この訓練データを用いて学習データの機械学習を実行する。このため、クエリの検索要求属性を推定するために適した訓練データを用いて学習データの機械学習を実行することができる。したがって、より好適にクエリの検索要求属性を推定することができる。

【0013】

(3) 前記訓練データ作成手段は、前記クリック素性として、ドメイン名、サイト名、ディレクトリ等の前記クリックURLの一部分を用いることを特徴とする(2)に記載の検索装置。

【0014】

このように、検索装置は、クリックURLの一部分は、クエリの検索要求属性を絞り込むための重要な情報となる可能性が高い。したがって、クリックURLの一部分を訓練データとして用いることによって、より好適にクエリの検索要求属性を推定することができる。

【0015】

(4) 前記訓練データ作成手段は、前記訓練クエリとして、前記検索履歴記憶手段に記憶された複数の前記入力クエリを含むクエリグループを用いることを特徴とする(2)に記載の検索装置。

【0016】

このように、検索装置は、検索要求属性について関連性の高い複数のクエリをクエリグループとする。このため、訓練クエリとしてクエリグループを用いることによって、より好適にクエリの検索要求属性を推定することができる。

【0017】

(5) セッションと、当該セッション内における前記クリックURLとを対応付けて記憶するセッション履歴記憶手段を更に備え、前記訓練データ作成手段は、前記入力クエリに対する前記検索エンジンによる検索結果において、前記セッション履歴記憶手段に記憶された前記セッション内で最後のクリック対象となった前記クリックURLのファイルの属性を前記検索要求属性として用いることを特徴とする(2)に記載の検索装置。

【0018】

セッション内で最後のクリック対象となったクリックURLのサフィックス(jpg、wmv、mp3、html等)は、クエリの検索要求属性を絞り込むための重要な情報となる可能性が高い。したがって、検索サーバは、訓練データとして、セッション内で最後のクリック対象となったクリックURLのファイルの属性を検索要求属性として用いることによって、より好適にクエリの検索要求属性を推定することができる。

【0019】

(6) 前記訓練データ作成手段は、前記入力クエリに対する前記検索エンジンによる検索結果において、所定期間内又は所定操作数内に予め属性が設定されたファイルに前記端末装置がアクセスした場合に、当該ファイルの属性を前記検索要求属性として用いることを特徴とする(2)に記載の検索装置。

【0020】

したがって、セッション内で最後のクリック対象が特定できない場合でも、クリックURLのファイルの属性を検索要求属性として用いることができる。

【0021】

(7) 前記クエリ属性記憶手段に記憶された前記入力クエリ及び前記検索要求属性と当該検索要求属性に応じた所定の条件とに基づいて、前記端末装置により入力されたユーザ入力クエリを用いて前記検索エンジンにより前記検索要求属性に応じた検索処理を実行する検索実行手段と、前記端末装置により入力された前記ユーザ入力クエリを用いて前記検索エンジンにより通常検索処理を実行する通常検索実行手段と、前記検索実行手段により得られた検索結果と、前記通常検索実行手段により得られた検索結果とで重複する検索結果を併合する検索結果併合手段と、前記検索結果併合手段により併合された前記検索結果を前記端末装置に提示する提示手段と、を更に備えることを特徴とする(2)に記載の検索装置。

【0022】

このように検索装置は、クエリの検索要求属性に応じた検索結果を端末装置へ提示することができる。したがって、検索装置を管理する者にとっては、検索サービスの向上を図ることができる。

【 0 0 2 3 】

(8) 前記クエリ属性記憶手段に記憶された前記入力クエリ及び前記検索要求属性に基づいて、前記検索履歴記憶手段に記憶された複数の前記クリック素性にクラスタリングを実行してクリック素性グループを作成するクリック素性クラスタリング手段を更に備えることを特徴とする (3) に記載の検索装置。

【 0 0 2 4 】

このように、検索装置は、クリック素性をクラスタリングして、グループ化する。そして、このグループ化されたクリック素性グループを検索履歴記憶手段に追加することによって、訓練データを用いて機械学習される学習データの精度を向上させることができる。したがって、より好適にクエリの検索要求属性を推定することができる。

【 0 0 2 5 】

(9) 前記クエリ属性記憶手段に記憶された前記入力クエリ及び前記検索要求属性に基づいて、前記検索履歴記憶手段に記憶された複数の前記入力クエリにクラスタリングを実行してクエリグループを作成するクエリクラスタリング手段を更に備えることを特徴とする (4) に記載の検索装置。

【 0 0 2 6 】

このように、検索装置は、クエリをクラスタリングして、グループ化する。そして、このグループ化されたクエリグループを検索履歴記憶手段に追加することによって、訓練データを用いて機械学習される学習データの精度を向上させることができる。したがって、より好適にクエリの検索要求属性を推定することができる。

【 0 0 2 7 】

(1 0) ユーザが操作する端末装置と通信可能に接続され、当該端末装置により入力された入力クエリに応じてウェブページを検索する検索エンジンを備える検索装置の検索方法において、前記入力クエリと、当該入力クエリが前記検索エンジンにより検索された検索結果のURLのうち、前記ユーザがクリックしたクリックURLとが対応付けて記憶される検索履歴記憶手段に記憶される前記入力クエリ及び前記クリックURLに基づいて、機械学習のための訓練データを作成する訓練データ作成ステップと、前記訓練データを用いて学習データの機械学習を実行する機械学習実行ステップと、前記機械学習実行ステップにより機械学習された前記学習データに基づいて、前記検索履歴記憶手段に記憶され、かつ、前記訓練データには含まれない前記入力クエリに対し、検索要求属性を推定する推定ステップと、前記推定ステップにより推定された前記入力クエリの検索要求属性と、当該入力クエリとをクエリ属性記憶手段に記憶させる属性記憶制御ステップと、を備えることを特徴とする検索方法。

【 0 0 2 8 】

(1 1) ユーザが操作する端末装置と通信可能に接続され、当該端末装置により入力された入力クエリに応じてウェブページを検索する検索エンジンを備える検索装置のコンピュータにより実行されるコンピュータプログラムにおいて、前記入力クエリと、当該入力クエリが前記検索エンジンにより検索された検索結果のURLのうち、前記ユーザがクリックしたクリックURLとが対応付けて記憶される検索履歴記憶手段に記憶される前記入力クエリ及び前記クリックURLに基づいて、機械学習のための訓練データを作成する訓練データ作成ステップと、前記訓練データを用いて学習データの機械学習を実行する機械学習実行ステップと、前記機械学習実行ステップにより機械学習された前記学習データに基づいて、前記検索履歴記憶手段に記憶され、かつ、前記訓練データには含まれない前記入力クエリに対し、検索要求属性を推定する推定ステップと、前記推定ステップにより推定された前記入力クエリの検索要求属性と、当該入力クエリとをクエリ属性記憶手段に記憶させる属性記憶制御ステップと、をコンピュータに実行させるコンピュータプログラム。

【 発明の効果 】

【 0 0 2 9 】

本発明によれば、より一般的な検索要求の属性をクエリに付与することのできる検索装

置、検索方法及びプログラムを提供することができる。

【発明を実施するための最良の形態】

【0030】

以下、本発明の実施形態について図面を参照しながら説明する。

【0031】

[検索システム 1 の機能構成]

図 1 は、本発明の検索装置の一実施形態に係る検索システム 1 の概略を示す図である。検索サーバ（検索装置）10、端末装置 20 及び外部サーバ 30 は、通信回線 40 を介して接続されている。端末装置 20 のユーザは、検索サーバ 10 にアクセスし、所定の検索エンジンに対して、所望の Web 文書に辿り着くためのクエリ（検索キーワード）を入力することにより、検索結果を取得する。ユーザは、この検索結果にリストされた URL を選択し、外部サーバ 30 にて管理される Web ページを閲覧する。

【0032】

端末装置 20 は、パソコン、携帯電話機、PHS（Personal Handy-phone System：登録商標）、モバイル、ノート PC、PDA（Personal Digital Assistant）等の端末装置を代表とするものであり、有線又は無線通信を行う通信回線 40 を介して、検索サーバ 10 から送信される情報を参照できるものであればどのようなものでもよい。また、検索サーバ 10 は、必要に応じて Web サーバ、DB サーバ及びアプリケーションサーバを含んで構成してよく、1 台のサーバで構成しても、それぞれ別のサーバで構成してもよい。

【0033】

[検索サーバ 10 のハードウェア構成]

図 2 は、本発明の一実施形態である検索サーバ 10 のハードウェア構成を示す図である。図 2 に示すように、制御部 300 を構成する CPU（Central Processing Unit）310（マルチプロセッサ構成では CPU 320 等複数の CPU が追加されてもよい）、バスライン 200、通信 I/F（I/F：インタフェース）330、メインメモリ 340、BIOS（Basic Input Output System）350、I/O コントローラ 360、ハードディスク 370、光ディスクドライブ 380 及び半導体メモリ 390 を備える。なお、ハードディスク 370、光ディスクドライブ 380 及び半導体メモリ 390 はまとめて記憶装置 410 と呼ばれる。

【0034】

制御部 300 は、検索サーバ 10 を統括的に制御する部分であり、ハードディスク 370 に記憶された各種プログラムを適宜読み出して実行することにより、上述したハードウェアと協働し、本発明に係る各種機能を実現している。

【0035】

通信 I/F 330 は、検索サーバ 10 が、ネットワークを介して端末装置 20 等の他の装置と情報を送受信する場合のネットワーク・アダプタである。

【0036】

BIOS 350 は、検索サーバ 10 の起動時に CPU 310 が実行するブートプログラムや、検索サーバ 10 のハードウェアに依存するプログラム等を記録する。

【0037】

I/O コントローラ 360 には、ハードディスク 370、光ディスクドライブ 380、半導体メモリ 390 等の記憶装置 410 を接続することができる。

【0038】

ハードディスク 370 は、本ハードウェアを検索サーバ 10 として機能させるための各種プログラム、本発明の機能を実行するプログラム及び後述するデータベース等を記憶する。なお、検索サーバ 10 は、外部に別途設けたハードディスク（図示せず）を外部記憶装置として利用することもできる。

【0039】

光ディスクドライブ 380 としては、例えば、DVD-ROM ドライブ、CD-ROM

ドライブ、DVD-RAMドライブ及びCD-RAMドライブを使用することができる。この場合は各ドライブに対応した光ディスク400を使用する。光ディスク400から光ディスクドライブ380によりプログラム又はデータを読み取り、I/Oコントローラ360を介してメインメモリ340又はハードディスク370に提供することもできる。

【0040】

なお、本発明でいうコンピュータとは、記憶装置、制御部等を備えた検索装置をいい、検索サーバ10は、記憶装置410、制御部300等を備えた検索装置により構成され、この検索装置は、本発明のコンピュータの概念に含まれる。

【0041】

以上の例は、検索サーバ10について主に説明したが、コンピュータに、プログラムをインストールして、そのコンピュータをサーバ装置として動作させることにより上記で説明した機能を実現することもできる。したがって、本発明において一実施形態として説明した検索サーバ10により実現される機能は、上述の方法を当該コンピュータにより実行することにより、或いは、上述のプログラムを当該コンピュータに導入して実行することによっても実現可能である。

【0042】

図3は、本発明の検索装置の一実施形態である検索サーバ10の機能ブロック図である。検索サーバ10は、クリックログデータベース（以下クリックログDBという）11、訓練用クリックログデータベース（以下訓練用クリックログDBという）12、セッションログデータベース（以下セッションログDBという）13、訓練データデータベース（以下訓練データDBという）14、学習データデータベース（以下学習データDBという）15、クエリ属性辞書16、検索要求属性対応DB17、訓練データ作成部21、機械学習実行部22、属性推定部23、検索実行部24、検索結果併合部25、提示部26及びクラスタリング部27を備える。

【0043】

クリックログDB11、訓練用クリックログDB12、セッションログDB13、訓練データDB14、学習データDB15、クエリ属性辞書16及び検索要求属性対応DB17は、主として記憶装置410により構成される。

また、訓練データ作成部21、機械学習実行部22、属性推定部23、検索実行部24、検索結果併合部25、提示部26及びクラスタリング部27は、主として制御部300により構成される。

【0044】

クリックログDB（検索履歴記憶手段）11は、クエリと、このクエリを検索キーワードとして検索エンジンにより検索された検索結果のURL（Uniform Resource Locator）のうち、ユーザの操作する任意の端末装置20により選択（クリック）されたURL（以下クリックURLともいう）とを対応付けて履歴データとして記憶する。この履歴データをクリックログと呼ぶ。クリックログは、例えば、図4に示すように、クリックデータテーブル111として蓄積的に記憶される。図4に示すクリックデータテーブル111には、クリックURLを選択した端末装置20のIPアドレス、クリックURLが選択された時間、セッション、クエリ、クリックURL及びクリックURLの検索ランクが対応付けて記憶されている。

【0045】

訓練用クリックログDB（検索履歴記憶手段）12は、クリックログDB11から任意に選択したクリックログが記憶され、訓練データの作成に用いられる。クリックログDB11から訓練用クリックログを選択する方法としては、例えば、無作為抽出、頻度の高いデータから順に抽出する、頻度の高いクエリグループから抽出する等が挙げられる。なお、クエリグループについては後述する。

【0046】

セッションログDB（セッション履歴記憶手段）13は、セッションと、このセッション内における前記クリックURLとを対応付けて履歴データとして記憶する。この履歴デ

ータをセッションログと呼ぶ。セッションログは、例えば、図5に示すように、セッションログテーブル131として累積的に記憶される。図5に示すセッションログテーブル131には、クリックURLを選択した端末装置20のIPアドレス、クリックURLが選択された時間、セッション、リファラ及びクリックURLが対応付けて記憶されている。

【0047】

ここで、セッション及びリファラについて説明する。セッションとは、WWWにおけるアクセス数の単位の一つであり、Webページを訪れたユーザがサイト内で行なう一連の行動をまとめて1セッションという。同一のユーザでも、ある程度間隔が開いた場合は、新しいセッションとしてカウントする。

【0048】

リファラとは、あるWebページのリンクをクリックして別のWebページに移動したときのリンク元となるWebページであり、本実施形態においては、クエリに応じて検索エンジンにより検索された検索結果である。

【0049】

訓練データ作成部21は、訓練データとして、少なくとも訓練用クリックログDB12に記憶されるクエリを含む訓練クエリと、クリックURLに基づくクリック素性と、訓練クエリの検索要求属性とを作成する。

【0050】

このように、検索サーバ10は、訓練データとして、訓練クエリと、クリック素性と、訓練クエリの検索要求属性とを作成し、この訓練データを用いて学習データの機械学習を実行する。このため、クエリの検索要求属性を推定するために適した訓練データを用いて学習データの機械学習を実行することができる。したがって、より好適にクエリの検索要求属性を推定することができる。

【0051】

具体的には、訓練データ作成部21は、訓練クエリの検索要求属性を予め付与しておくことができる。

【0052】

また、訓練データ作成部21は、クリック素性として、ドメイン名、サイト名、ディレクトリ等のクリックURLの一部を用いることができる。

【0053】

このため、例えば検索要求属性が「自動車」である場合、クリックURLの一部としての「www.carxxxxsor.net」のような例が挙げられる。このようにクリックURLの一部は、クエリの検索要求属性を絞り込むための重要な情報となる可能性が高い。したがって、クリックURLの一部を訓練データとして用いることによって、より好適にクエリの検索要求属性を推定することができる。

【0054】

また、訓練データ作成部21は、訓練クエリとして、訓練用クリックログDB12に記憶された複数のクエリを含むクエリグループを用いることができる。クエリグループとは、後述するクラスタリング部27により作成される。クエリグループは、例えば[カ。ーラ, ト タカ。ーラ, カ。ーラト タ]のように表される。

【0055】

このように、検索サーバ10は、検索要求属性について関連性の高い複数のクエリをクエリグループとする。このため、訓練クエリとしてクエリグループを用いることによって、より好適にクエリの検索要求属性を推定することができる。

【0056】

また、訓練データ作成部21は、クエリに対する検索エンジンによる検索結果において、セッションログDB13に記憶されたセッション内で最後のクリック対象となったクリックURLのファイルの属性を検索要求属性として用いることができる。

【0057】

セッション内で最後のクリック対象となったクリックURLのサフィックス(jpg、

w m v、m p 3、h t m l等)は、クエリの検索要求属性を絞り込むための重要な情報となる可能性が高い。例えば、「j p g」は静止画、「w m v」は動画、「m p 3」は音楽、「h t m l」はWebページにそれぞれ関連性が強いと考えられる。したがって、検索サーバ10は、訓練データとして、セッション内で最後のクリック対象となったクリックURLのファイルの属性を検索要求属性として用いることによって、より好適にクエリの検索要求属性を推定することができる。

【0058】

また、訓練データ作成部21は、入力クエリに対する検索エンジンによる検索結果において、所定期間内又は所定操作数内に予め属性が設定されたファイルに端末装置20がアクセスした場合に、このファイルの属性を検索要求属性として用いることができる。

【0059】

したがって、セッション内で最後のクリック対象が特定できない場合でも、クリックURLのファイルの属性を検索要求属性として用いることができる。

【0060】

このような訓練データとしての訓練クエリ、クリック素性及び検索要求属性を訓練データ作成部21は、三つ組みに変換する。具体的には、訓練データ作成部21は、クリック素性が「カローラ」、クリック素性が「www.carxxxsor.net」、検索要求属性が「自動車」の場合には、(カローラ、www.carxxxsor.net、自動車)のような三つ組みに変換する。

【0061】

そして、この三つ組み変換された訓練データを機械学習実行部22により機械学習する。具体的には、機械学習実行部22によりクエリの検索要求属性を判別するための判別関数を学習する場合には、学習する各々の検索要求属性をクリック素性のベクター表現で以下のように表す。

(数1)

$$a_1 = (s_1 = 1, s_2 = 0, s_3 = 0, s_4 = 1, s_5 = 0, \dots, s_m = 0)$$

$$a_2 = (s_1 = 0, s_2 = 1, s_3 = 0, s_4 = 1, s_5 = 0, \dots, s_m = 0)$$

...

$$a_n = (s_1 = 0, s_2 = 0, s_3 = 0, s_4 = 0, s_5 = 1, \dots, s_m = 0) \cdot$$

・ ・ 式(1)

【0062】

ここで、 a_n は検索要求属性であり、 s_m はクリック素性である。具体的には、検索要求属性 a_1 において、 $s_1 = 1$ である。これは、クリック素性 s_1 が検索要求属性 a_1 に関連することを示している。すなわち、検索要求属性 a_1 を有するクエリは、クリック素性がサイト名である場合には、このサイト名を含むクリックURLが選択されたことを意味する。また、検索要求属性 a_1 において、 $s_2 = 0$ である。これは、クリック素性 s_2 が検索要求属性 a_1 に関連しないことを示している。

【0063】

訓練データDB14は、訓練データ作成部21により作成された訓練データを記憶する。具体的には、上述したように検索要求属性を予め付与した訓練クエリやクリック素性のベクター表現等を記憶する。

【0064】

機械学習実行部22は、訓練データを用いて学習データの機械学習を実行する。訓練データは、上述したように判別関数を学習する場合には、ベクター表現によって表現するため、SVM(Support Vector Machine)、kNN(k Nearest Neighbor)、Naive bayes classifier、ニューラルネットワーク、判別木等の様々な公知の機械学習手法を用いて、判別関数を学習することができる。

【0065】

学習データDB15は、機械学習実行部22により機械学習が実行された学習データを

記憶する。具体的には、学習データDB15には、機械学習実行部22により機械学習が実行された判別関数が記憶される。

【0066】

属性推定部23は、機械学習実行部22により機械学習された学習データに基づいて、クリックログDB11に記憶され、かつ、訓練用クリックログDB12には含まれないクエリに対し、検索要求属性を推定する。ここで、学習データには、クリックURLと、クリック頻度とが含まれる。なお、クリック頻度とは、同一のセッションにおける、クリックURLの頻度であり、訓練用クリックログDB12に記憶されるクリックURLと、入力クエリとに基づいて、制御部300により算出される値である。

そして、属性推定部23は、推定した検索要求属性と、クエリとをクエリ属性辞書16に記憶する。

【0067】

クエリ属性辞書16は、属性推定部23により推定されたクエリの検索要求属性と、このクエリとを対応付けて記憶する。

図6は、クエリ属性辞書16の一例を示す図である。図6では、クエリの検索要求属性が関連性の高いグループに分けられている。例えば、クエリ「カオラ」に対応する検索要求属性1（図6中では属性1と表記）「自動車」、検索要求属性2（図6中では属性2と表記）「中古車」、及び検索要求属性3（図6中では属性3と表記）「国産車」が検索要求属性グループ1として記憶されている。

【0068】

検索実行部24は、クエリ属性辞書16に記憶されたクエリ及び検索要求属性と当該検索要求属性に応じた所定の条件とに基づいて、端末装置20により入力されたユーザ入力クエリを用いて検索エンジンにより検索要求属性に応じた検索処理を実行する。ここで、検索実行部24は、所定の条件として、検索要求属性対応DB17を適用することができる。

【0069】

図7は、検索要求属性対応DB17に記憶される検索要求属性対応テーブルの一例を示す図である。検索要求属性対応テーブルは、検索要求属性と、この検索供給属性に対応する処理とが対応付けて記憶される。例えば、検索要求属性が「中古車」のときには、対応する処理は「該当するクリックURLのランクを2上げる」となる。また、検索要求属性が「自動車」のときには、対応する処理は「該当する検索要求属性を有する広告を表示する」となる。また、検索要求属性が「音楽ダウンロード」のときには、対応する処理は「該当するクリックURLで最も上位のものをランクの1位にする」となる。また、検索要求属性が「音楽ダウンロード」のときには、対応する処理は「該当するクリックURLのスコアを5%上げる」となる。また、検索要求属性が「音楽ダウンロード」のときには、対応する処理は「該当する検索要求属性を有する広告を表示する」となる。

【0070】

このように、本実施形態の検索サーバ10は、クリックログDB11に記憶されたクエリと、クリック素性とに基づいて、クエリの検索要求属性を推定する。そして、推定されたクエリの検索要求属性は、例えば検索結果ランキング、広告提示、ユーザに推薦するサービスへのショートカット等の検索サービスに利用することにより、検索サービスの向上を図ることができる。

【0071】

また、検索実行部24は、端末装置20により入力されたクエリを検索エンジンにより通常検索処理を実行する。

【0072】

検索結果併合部25は、検索実行部24の検索要求属性に応じた検索処理により得られた検索結果と、検索実行部24の通常検索処理により得られた検索結果とで重複する検索結果を併合する。

【0073】

提示部 26 は、検索結果併合部 25 により併合された検索結果を端末装置 20 に提示する。

【0074】

このように検索サーバ 10 は、クエリの検索要求属性に応じた検索結果を端末装置 20 へ提示することができる。したがって、検索サーバ 10 を管理する者にとっては、検索サービスの向上を図ることができる。また、端末装置 20 を操作するユーザにとっては、クエリの検索要求属性に応じた入力クエリの検索結果を得ることができる。

【0075】

なお、上述した判別関数を機械学習する場合に代えて、クエリの検索要求属性を機械学習により直接、学習することもできる。

この場合には、訓練データ作成部 21 は、クリックログ DB 11 に記憶されたクエリの頻度の大きいものから 100 万件ほどについて訓練用クリックログ DB 12 に記憶する。

【0076】

そして、訓練データ作成部 21 は、訓練用クリックログ DB 12 に記憶されたクエリのうち、1 万件ほどを訓練データとして用いて、人手により検索要求属性を付与する。

機械学習実行部 22 は、この訓練データを用いて検索要求属性を有するか否かを判別する 2 値分類器を上述した公知の機械学習手法を用いて機械学習を実行する。

【0077】

制御部 300 は、100 万件ほどのクエリのうち、残る 99 万件のクエリを学習した 2 値分類器を用いて自動分類する。そして、検索要求属性に対応するクエリのリストを作成する。

属性推定部 23 は、端末装置 20 から受け付けたユーザ入力クエリと、作成されたリストに含まれるクエリとが一致する場合には、このユーザ入力クエリの検索要求属性を作成されたクエリに対応する検索要求属性であると推定する。

【0078】

次に、クラスタリング部 27 について説明する。クラスタリング部 27 は、クエリ属性辞書 16 に記憶されたクエリ及び検索要求属性に基づいて、訓練用クリックログ DB 12 に記憶された複数のクリック素性にクラスタリングを実行してクリック素性グループを作成する。具体的には、クラスタリング部 27 は、クリック素性としてのドメイン名、サイト名、ディレクトリ等のクリック URL の一部分を用いてクラスタリングを実行してクリック素性グループを作成する。

【0079】

また、クラスタリング部 27 は、クエリ属性辞書 16 に記憶されたクエリ及び検索要求属性に基づいて、訓練用クリックログ DB 12 に記憶された複数のクエリにクラスタリングを実行してクエリグループを作成する。具体的には、クラスタリング部 27 は、訓練用クリックログ DB 12 に記憶されたクリック URL 全てにクラスタリングを実行し、クエリグループを作成する。

【0080】

ここで、クラスタリングを実行する際のクラスタリング方式としては、例えば各々のクエリをクリック URL の集合で表現して、クリック URL の集合間での一致度によるワンパスクラスタリングや階層クラスタリング等を適用することができる。

【0081】

クラスタリング部 27 によりクラスタリングを実行した場合には、三つ組み変換された訓練データは、例えば（[カ。ーラ，ト タカ。ーラ，カ。ーラト タ]、www.carxxxxsor.net、自動車）のような構成となる。ここで、クラスタリングされたクエリグループは、「カ。ーラ，ト タカ。ーラ，カ。ーラト タ」、クリック素性は、「www.carxxxxsor.net」、検索要求属性は、「自動車」である。

【0082】

そして、クラスタリング部 27 は、クラスタリングを実行したことにより得られたクリック素性グループ及びクエリグループを訓練データ DB 14 に記憶される訓練データに追

加する。すなわち、訓練データにおけるクリック素性としてクリック素性グループを追加する。また、訓練データにおける訓練クエリとしてクエリグループを追加する。

【0083】

このように、検索サーバ10は、クエリ及びクリック素性をクラスタリングして、グループ化する。そして、このグループ化されたクエリグループ又はクリック素性グループを訓練データDB14に追加することによって、訓練データを用いて機械学習される学習データの精度を向上させることができる。したがって、より好適にクエリの検索要求属性を推定することができる。

【0084】

ここで、クラスタリング部27による処理について詳細に説明する。

上述した式(1)では、クリック素性についてベクターで表現した。ところで、このベクターは、クリックログDB11に記憶されるクエリについて、

(数2)

$$q_1 = (s_1 = 1, s_2 = 0, s_3 = 0, s_4 = 1, s_5 = 0, \dots, s_m = 0)$$

$$q_2 = (s_1 = 0, s_2 = 1, s_3 = 0, s_4 = 1, s_5 = 0, \dots, s_m = 0)$$

...

$$q_n = (s_1 = 0, s_2 = 0, s_3 = 0, s_4 = 0, s_5 = 1, \dots, s_m = 0) \cdot$$

・式(2)

と表現することができる。式(2)は、n個のクエリとm個のクリック素性という $n \times m$ の2次元の行列で表現される。これにより、クラスタリング部27は、式(2)をクエリごとに分けたベクターを用いてクエリのクラスタリングを実行することができる。また、クラスタリング部27は、式(2)をクリック素性ごとに分けたベクターを用いてクリック素性のクラスタリングを実行することができる。

【0085】

この場合、特にクエリの次元数は、数百万・数千万にも増大する可能性があるため、例えば同一のクラスタ内のクエリは、全て同一視して一つの次元として扱うことが必要となる。

【0086】

[処理の流れ]

図8は、本実施形態の検索サーバ10における検索要求属性の推定処理の流れを示すフローチャートである。

ステップS1：訓練データ作成部21は、訓練用クリックログDB12に記憶された入力クエリに属性を付与する。具体的には、訓練データ作成部21は、入力クエリに対して予め用意されたクエリを人手により付与される。

【0087】

ステップS2：訓練データ作成部21は、ステップS1において属性が付与された入力クエリを含む訓練クエリと、訓練用クリックログDB12に記憶されるクリック素性と検索要求属性とを三つ組みに変換する。具体的には、訓練データ作成部21は、訓練クエリが「カローラ」、クリック素性が「www.carxxxsor.net」、検索要求属性が「自動車」の場合には、(カローラ、www.carxxxsor.net、自動車)のような三つ組みに変換する。そして、三つ組みに変換された訓練データは、訓練データ作成部21により訓練データDB14に記憶される。

【0088】

ステップS3：訓練データ作成部21は、ステップS2において三つ組みに変換された訓練データの検索要求属性をクリック素性のベクター表現で表現する。

【0089】

ステップS4：機械学習実行部22は、ステップ3において得られた訓練データを用いて、学習データの機械学習を実行する。具体的には、機械学習実行部22は、ステップS3においてベクター表現された訓練データを用いて入力クエリの検索要求属性を判別するための判別関数を学習する。そして、機械学習実行部22は、学習した判別関数を学習デ

ータDB15に記憶させる。

【0090】

ステップS5：属性推定部23は、機械学習実行部22により機械学習された判別関数によって、クリックログDB11に記憶され、かつ、訓練用クリックログDB12には含まれないクエリに対し、検索要求属性を推定する。

【0091】

ステップS6：属性推定部23は、ステップS5において推定された各々の検索要求属性と、この検索要求属性が推定されたクエリとを対応付けてクエリ属性辞書16に記憶させる。

【0092】

図9は、本実施形態の検索サーバ10におけるクエリ属性辞書16を用いた検索処理の流れを示すフローチャートである。

ステップS11：検索実行部24は、端末装置20からのユーザ入力クエリを受け付ける。

ステップS12：検索実行部24は、クエリ属性辞書16に記憶されたクエリ及び検索要求属性と検索要求属性対応DB17とに基づいて、ステップS11において受け付けたユーザ入力クエリを検索エンジンにより検索要求属性に応じた検索処理を実行する。

【0093】

ステップS13：検索実行部24は、ステップS11において受け付けたユーザ入力クエリを用いて検索エンジンにより通常検索処理を実行する。

ステップS14：検索結果併合部25は、ステップS12において得られた検索結果と、ステップS13において得られた検索結果とで重複する検索結果を併合する。

【0094】

ステップS15：提示部26は、ステップS14において併合された検索結果を端末装置20に提示する。

【0095】

図10は、本実施形態の検索サーバ10におけるクラスタリング処理の流れを示すフローチャートである。

ステップS21：クラスタリング部27は、クリックログDB11に記憶されたクエリ又は／及びクリックログDB11に記憶されたクリックURLのクラスタリングを実行する。

【0096】

ステップS22：クラスタリング部27は、ステップS21においてクラスタリングされた複数のクエリをグループ化する。

ステップS23：クラスタリング部27は、ステップS21においてクラスタリングされた複数のクリックURLをグループ化する。

【0097】

ステップS23：クラスタリング部27は、ステップS22 及び／又はステップS23においてグループ化されたクエリ 及び／又はクリックURLを訓練クエリ 及び／又はクリック素性として、訓練データDB14に追加して記憶させる。

【0098】

以上、本発明の実施形態について説明したが、本発明は上述した実施形態に限るものではない。また、本発明の実施形態に記載された効果は、本発明から生じる最も好適な効果を列挙したに過ぎず、本発明による効果は、本発明の実施例に記載されたものに限定されるものではない。

【図面の簡単な説明】

【0099】

【図1】本発明の検索装置の一実施形態に係る検索システム1の概略を示す図である。

【図2】本発明の一実施形態である検索サーバ10のハードウェア構成を示す図である。

【図3】本発明の検索装置の一実施形態である検索サーバ10の機能ブロック図である。

【図４】クリックデータテーブル１１１の一例を示す図である。

【図５】セッションログテーブル１３１の一例を示す図である。

【図６】クエリ属性辞書１６の一例を示す図である。

【図７】検索要求属性対応ＤＢ１７の一例を示す図である。

【図８】本実施形態の検索サーバ１０における検索要求属性の推定処理の流れを示すフローチャートである。

【図９】本実施形態の検索サーバ１０におけるクエリ属性辞書１６を用いた検索処理の流れを示すフローチャートである。

【図１０】本実施形態の検索サーバ１０におけるクラスタリング処理の流れを示すフローチャートである。

【符号の説明】

【０１００】

- １０ 検索サーバ
- １１ クリックログＤＢ
- １２ 訓練用クリックログＤＢ
- １３ セッションログＤＢ
- １４ 訓練データＤＢ
- １５ 学習データＤＢ
- １６ クエリ属性辞書
- １７ 検索要求属性対応ＤＢ
- ２０ 端末装置
- ２１ 訓練データ作成部
- ２２ 機械学習実行部
- ２３ 属性推定部
- ２４ 検索実行部
- ２５ 検索結果併合部
- ２６ 提示部
- ２７ クラスタリング部

【手続補正３】

【補正対象書類名】図面

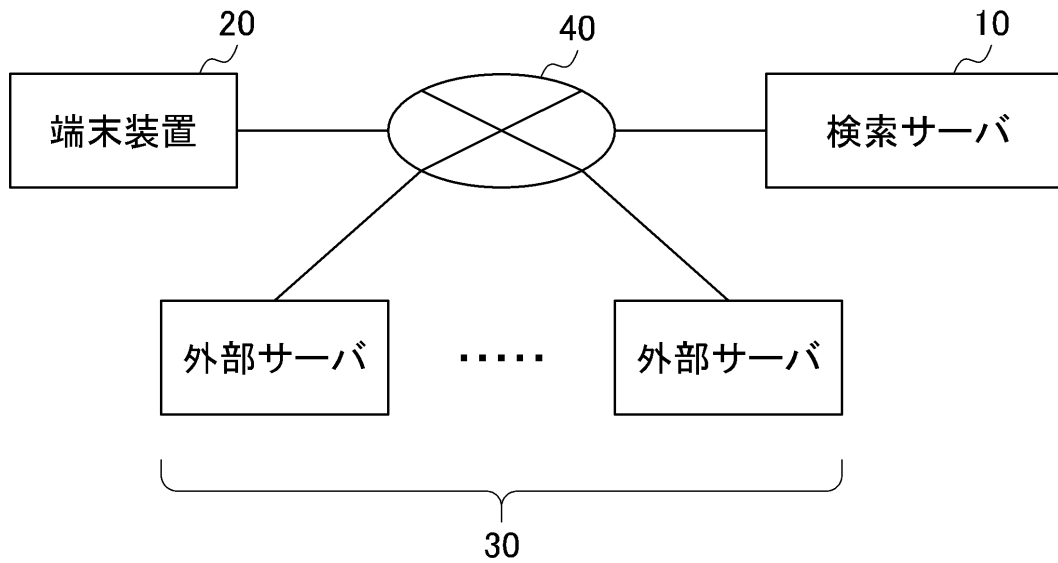
【補正対象項目名】図１

【補正方法】変更

【補正の内容】

【図 1】

1



【手続補正 4】

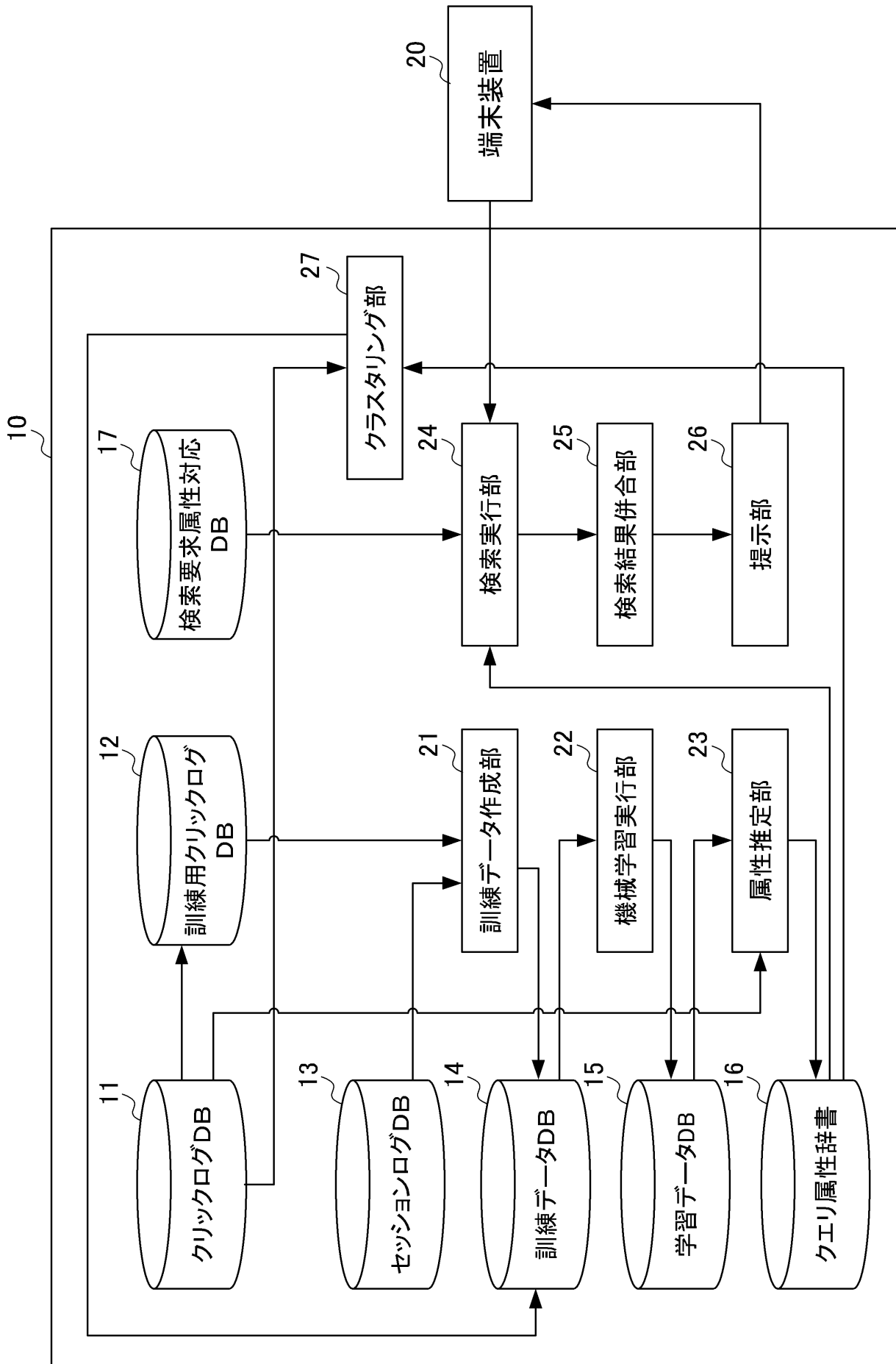
【補正対象書類名】図面

【補正対象項目名】図 3

【補正方法】変更

【補正の内容】

【図 3】



【 手 続 補 正 5 】

【 補 正 対 象 書 類 名 】 図 面

【 補 正 対 象 項 目 名 】 図 4

【 補 正 方 法 】 変 更

【 補 正 の 内 容 】

【 図 4 】

クリックデータテーブル111

IP アドレス	時間	セッション	ランク	クエリ	URL
110.149.145.1	2006/10/6 21:45:53	Bp4ed6917	3	自動車	http://xxx.yyy.zzz
110.149.145.1	2006/10/6 21:49:01	Bp4ed6917	2	トヨタ	http://xxx.yyy.aaa

【 手 続 補 正 6 】

【補正対象書類名】図面
 【補正対象項目名】図 5
 【補正方法】変更
 【補正の内容】
 【図 5】

セッションログテーブル131

IP アドレス	時間	セッション	リファラ	URL
110.149.145.1	2006/10/6 21:45:53	Bp4ed6917	http://search.yahoo.co.jp /search/?p=tomaruta	http://xxx.yyy.zzz
110.149.145.1	2006/10/6 21:49:01	Bp4ed6917	http://search.yahoo.co.jp /search/?p=tomaruta	http://xxx.yyy.zzz

【手続補正 7】
 【補正対象書類名】図面
 【補正対象項目名】図 6

【補正方法】変更

【補正の内容】

【図 6】

クエリ属性辞書16

	検索要求属性グループ1			検索要求属性グループ2	
	属性1	属性2	属性3	属性4	属性5
クエリ	自動車	中古車	国産車	保険	車検
カテゴリー					

【手続補正 8】

【補正対象書類名】図面

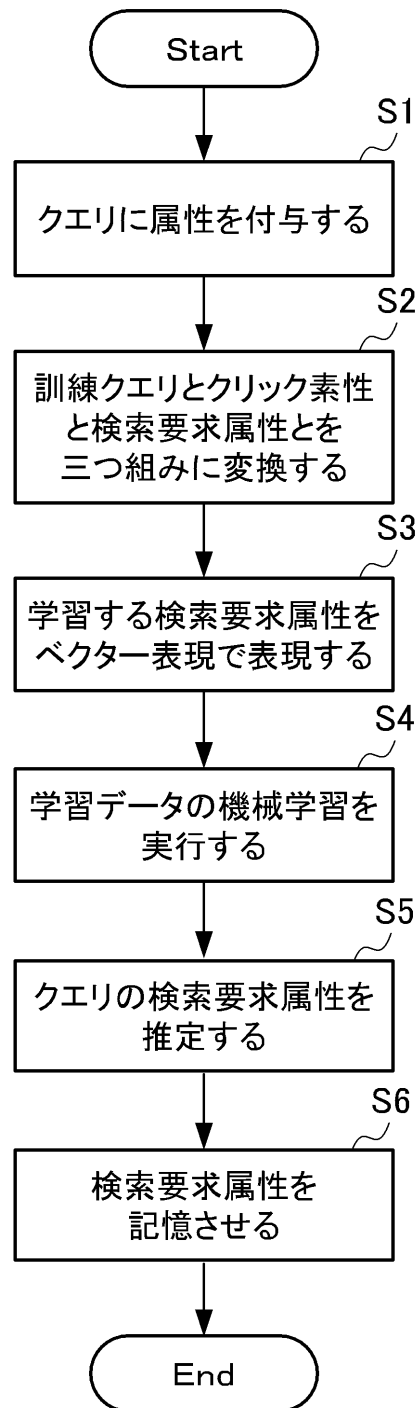
【補正対象項目名】図 8

【補正方法】変更

【補正の内容】

【図 8】

(検索要求属性の推定処理)



【手続補正 9】

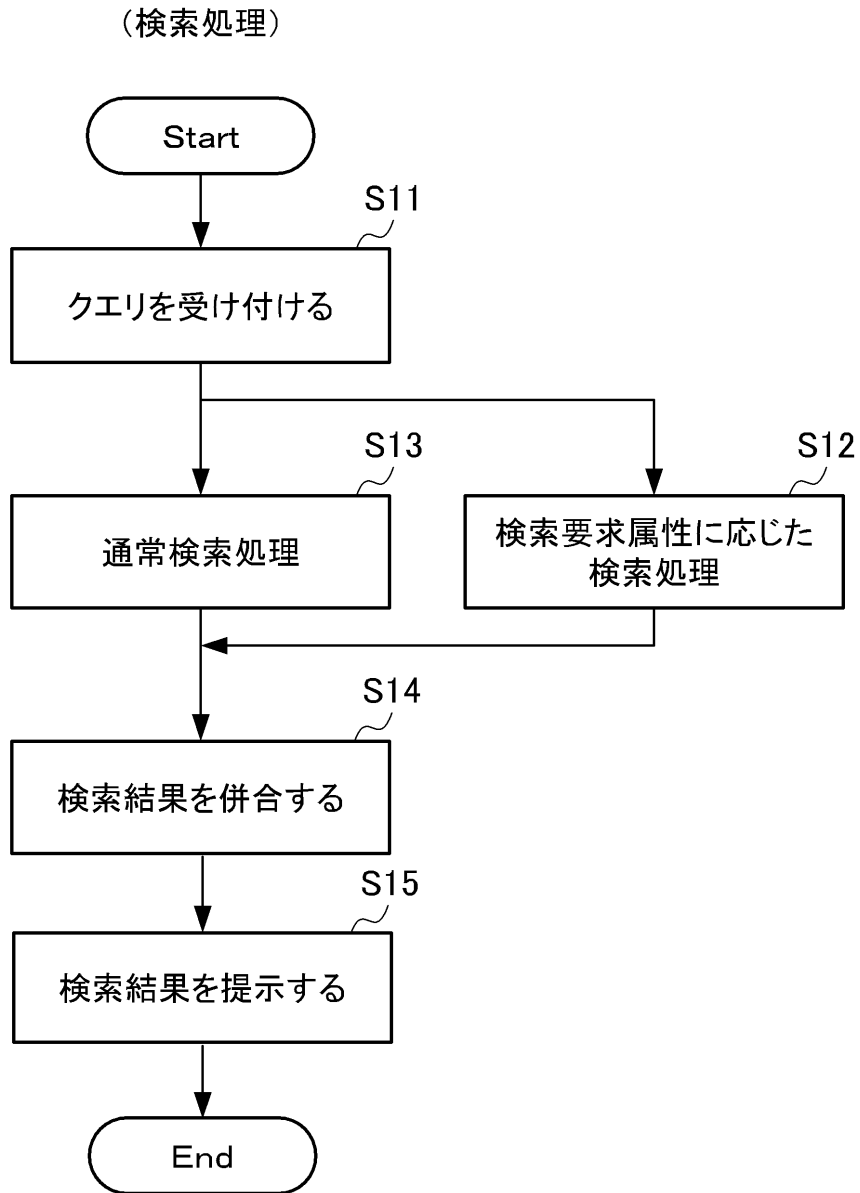
【補正対象書類名】図面

【補正対象項目名】図 9

【補正方法】変更

【補正の内容】

【図 9】



【手続補正 10】

【補正対象書類名】図面

【補正対象項目名】図 10

【補正方法】変更

【補正の内容】

【図 10】

(クラスタリング処理)

