

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5379627号
(P5379627)

(45) 発行日 平成25年12月25日(2013.12.25)

(24) 登録日 平成25年10月4日(2013.10.4)

(51) Int.Cl. F I
G 0 6 F 1 7 / 3 0 (2 0 0 6 . 0 1) G 0 6 F 1 7 / 3 0 3 4 0 Z

請求項の数 14 (全 15 頁)

(21) 出願番号	特願2009-225459 (P2009-225459)	(73) 特許権者	399035766 エヌ・ティ・ティ・コミュニケーションズ 株式会社 東京都千代田区内幸町一丁目1番6号
(22) 出願日	平成21年9月29日(2009.9.29)	(74) 代理人	100070150 弁理士 伊東 忠彦
(65) 公開番号	特開2011-76264 (P2011-76264A)	(72) 発明者	阿部 直人 東京都千代田区内幸町一丁目1番6号 エ ヌ・ティ・ティ・コミュニケーションズ株 式会社内
(43) 公開日	平成23年4月14日(2011.4.14)	審査官	伊知地 和之
審査請求日	平成24年1月11日(2012.1.11)		

最終頁に続く

(54) 【発明の名称】 検索制御装置、検索制御方法、及びプログラム

(57) 【特許請求の範囲】

【請求項1】

ウェブ検索サービスが提供されるネットワークに接続される検索制御装置であって、
入力された検索語から生成された検索クエリに基づき、予備検索クエリを複数個生成し、
各予備検索クエリを用いてウェブ検索を行い、当該ウェブ検索の検索結果を、予備検索
結果として予備検索クエリ毎に予備検索結果記憶手段に格納する予備検索手段と、

前記予備検索結果記憶手段から予備検索結果を読み出し、当該予備検索結果の内容の集
中度を評価する処理を予備検索クエリ毎に実行し、内容の集中度が最も高い予備検索結果
に対応する予備検索クエリを、検索クエリとして検索クエリ記憶手段に格納する予備検索
結果評価手段と、

前記検索クエリ記憶手段に格納された前記検索クエリを用いてウェブ検索を行い、当該
ウェブ検索の検索結果が所定の条件を満たす場合に、当該検索結果を出力するウェブ検索
手段と

を備えたことを特徴とする検索制御装置。

【請求項2】

前記予備検索手段は、複数の単語を含む前記検索クエリから1つの単語を削除し、1つ
の単語を削除した当該検索クエリを前記予備検索クエリとすることを特徴とする請求項1
に記載の検索制御装置。

【請求項3】

前記予備検索手段は、前記予備検索クエリに基づく検索結果に含まれる所定数のコンテ

ンツ情報の各々から要約文を抽出し、当該所定数の要約文を前記予備検索結果として前記予備検索結果記憶手段に格納することを特徴とする請求項 1 または 2 に記載の検索制御装置。

【請求項 4】

前記予備検索結果評価手段は、前記所定数の要約文の各々から単語ベクトルを抽出し、当該所定数の単語ベクトルをグループ化し、当該グループ化の結果を用いて前記集中度を評価することを特徴とする請求項 3 に記載の検索制御装置。

【請求項 5】

前記予備検索結果評価手段は、前記所定数に対する各グループ内の単語ベクトルの数の割合を単語ベクトルの出現確率とし、当該出現確率を用いて情報エントロピーを算出することにより前記集中度を評価することを特徴とする請求項 4 に記載の検索制御装置。

10

【請求項 6】

前記ウェブ検索手段において、前記所定の条件は、前記検索結果に含まれるコンテンツ情報の件数が所定の閾値以上であることであり、

前記検索結果が当該条件を満たさない場合に、前記ウェブ検索手段で用いた検索クエリに基づき、前記予備検索手段、前記予備検索結果評価手段、及び前記ウェブ検索手段による処理を再度実行することを特徴とする請求項 1 ないし 5 のうちいずれか 1 項に記載の検索制御装置。

【請求項 7】

ウェブ検索サービスが提供されるネットワークに接続される検索制御装置が実行する検索制御方法であって、

20

入力された検索語から生成された検索クエリに基づき、予備検索クエリを複数個生成し、各予備検索クエリを用いてウェブ検索を行い、当該ウェブ検索の検索結果を、予備検索結果として予備検索クエリ毎に予備検索結果記憶手段に格納する予備検索ステップと、

前記予備検索結果記憶手段から予備検索結果を読み出し、当該予備検索結果の内容の集中度を評価する処理を予備検索クエリ毎に実行し、内容の集中度が最も高い予備検索結果に対応する予備検索クエリを、検索クエリとして検索クエリ記憶手段に格納する予備検索結果評価ステップと、

前記検索クエリ記憶手段に格納された前記検索クエリを用いてウェブ検索を行い、当該ウェブ検索の検索結果が所定の条件を満たす場合に、当該検索結果を出力するウェブ検索ステップと

30

を備えたことを特徴とする検索制御方法。

【請求項 8】

前記予備検索ステップにおいて、前記検索制御装置は、前記予備検索クエリに基づく検索結果に含まれる所定数のコンテンツ情報の各々から要約文を抽出し、当該所定数の要約文を前記予備検索結果として前記予備検索結果記憶手段に格納することを特徴とする請求項 7 に記載の検索制御方法。

【請求項 9】

前記予備検索結果評価ステップにおいて、前記検索制御装置は、前記所定数の要約文の各々から単語ベクトルを抽出し、当該所定数の単語ベクトルをグループ化し、当該グループ化の結果を用いて前記集中度を評価することを特徴とする請求項 8 に記載の検索制御方法。

40

【請求項 10】

前記予備検索結果評価ステップにおいて、前記検索制御装置は、前記所定数に対する各グループ内の単語ベクトルの数の割合を単語ベクトルの出現確率とし、当該出現確率を用いて情報エントロピーを算出することにより前記集中度を評価することを特徴とする請求項 9 に記載の検索制御方法。

【請求項 11】

コンピュータを、ウェブ検索サービスが提供されるネットワークに接続される検索制御装置として機能させるプログラムであって、コンピュータを、

50

入力された検索語から生成された検索クエリに基づき、予備検索クエリを複数個生成し、各予備検索クエリを用いてウェブ検索を行い、当該ウェブ検索の検索結果を、予備検索結果として予備検索クエリ毎に予備検索結果記憶手段に格納する予備検索手段、

前記予備検索結果記憶手段から予備検索結果を読み出し、当該予備検索結果の内容の集中度を評価する処理を予備検索クエリ毎に実行し、内容の集中度が最も高い予備検索結果に対応する予備検索クエリを、検索クエリとして検索クエリ記憶手段に格納する予備検索結果評価手段、

前記検索クエリ記憶手段に格納された前記検索クエリを用いてウェブ検索を行い、当該ウェブ検索の検索結果が所定の条件を満たす場合に、当該検索結果を出力するウェブ検索手段、

として機能させるプログラム。

【請求項 1 2】

前記予備検索手段は、前記予備検索クエリに基づく検索結果に含まれる所定数のコンテンツ情報の各々から要約文を抽出し、当該所定数の要約文を前記予備検索結果として前記予備検索結果記憶手段に格納することを特徴とする請求項 1 1 に記載のプログラム。

【請求項 1 3】

前記予備検索結果評価手段は、前記所定数の要約文の各々から単語ベクトルを抽出し、当該所定数の単語ベクトルをグループ化し、当該グループ化の結果を用いて前記集中度を評価することを特徴とする請求項 1 2 に記載のプログラム。

【請求項 1 4】

前記予備検索結果評価手段は、前記所定数に対する各グループ内の単語ベクトルの数の割合を単語ベクトルの出現確率とし、当該出現確率を用いて情報エントロピーを算出することにより前記集中度を評価することを特徴とする請求項 1 3 に記載のプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、検索語を用いてウェブ検索を行う技術に関し、特に、検索語に関連した内容を多く含む適切な検索結果を自動的に得るための技術に関するものである。

【背景技術】

【0002】

新聞や番組、映像、写真など様々なコンテンツをネットワーク上で提供するサービスが広く普及している。このようなサービスでは、コンテンツに、その詳細が記載されたメタデータが付与されている。メタデータにより、サービス利用者は容易に所望のコンテンツを取得できる。

【0003】

このようなサービスでは、一般に、メタデータはサービス提供者により作成され、コンテンツと共に管理される。しかし、コンテンツの数は膨大になってきており、各コンテンツに対して詳細なメタデータを付与し、管理することを人手で行うことは困難になりつつある。そこで、人手でごく僅かなメタデータを作成し、それを自動的に拡張する技術が提案されている。

【0004】

メタデータを自動的に拡張する技術として、ウェブ検索を利用する技術がある。例えば、特許文献 1 に記載された技術では、ファイアウォールにおけるアクセスログを利用して、メタデータの情報源（ウェブサーバ等）を限定し、当該情報源に対して検索語（基となるメタデータ）に基づき検索を行って、検索語に関連するコンテンツを取得し、当該コンテンツから拡張に用いる新たなメタデータを抽出している。

【先行技術文献】

【特許文献】

10

20

30

40

50

【 0 0 0 5 】

【特許文献 1】特開 2 0 0 7 - 1 4 9 0 3 6 号公報

【発明の概要】

【発明が解決しようとする課題】

【 0 0 0 6 】

さて、一般的にネットワーク上で提供されているウェブ検索サービスにおけるウェブ検索は、入力された検索語を含む情報を、検索語に関連する内容を有する情報としてウェブ上から収集するものである。しかし、現状のウェブ検索は、使用する検索語の選び方によっては、検索語と関連性が低い多くの様々なコンテンツ情報が検索結果に含まれることがある。このような場合、特許文献 1 に記載された技術では、適切な拡張用メタデータを得ることができない。

10

【 0 0 0 7 】

また、ウェブブラウザを備えた PC 端末からウェブ検索を行うなどの一般のウェブ検索においては、1 回の検索で適切な検索結果を得られない場合が少なからず発生し、このような場合、利用者は、目的とする内容を多く含む検索結果を得るために、様々な検索語を用いて検索を試行しなければならず、時間と手間がかかるという問題がある。

【 0 0 0 8 】

ここで、ウェブ検索は、基本的に検索語に関連する情報を収集するものであるから、ウェブ検索結果の内容の散らばり度合いが小さく、ある内容に集中している場合、そのウェブ検索結果は、入力した検索語に関する内容に集中した適切な検索結果である蓋然性が極めて高い。このような適切な検索結果を、人手による再検索を行うことなく自動的に取得することはウェブ検索における課題である。

20

【 0 0 0 9 】

本発明は上記の点に鑑みてなされたものであり、入力された検索語を用いてウェブ検索を実行し、検索結果を出力する技術において、内容の集中度の高い検索結果を自動的に取得する技術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】

【 0 0 1 0 】

上記の課題を解決するために、本発明は、ウェブ検索サービスが提供されるネットワークに接続される検索制御装置であって、入力された検索語から生成された検索クエリに基づき、予備検索クエリを生成し、各予備検索クエリを用いてウェブ検索を行い、当該ウェブ検索の検索結果を、予備検索結果として予備検索クエリ毎に予備検索結果記憶手段に格納する予備検索手段と、前記予備検索結果記憶手段から予備検索結果を読み出し、当該予備検索結果の内容の集中度を評価する処理を予備検索クエリ毎に実行し、内容の集中度に応じて抽出した予備検索結果に対応する予備検索クエリを、検索クエリとして検索クエリ記憶手段に格納する予備検索結果評価手段と、前記検索クエリ記憶手段に格納された前記検索クエリを用いてウェブ検索を行い、当該ウェブ検索の検索結果が所定の条件を満たす場合に、当該検索結果を出力するウェブ検索手段とを備えたことを特徴とする検索制御装置として構成される。

30

【発明の効果】

40

【 0 0 1 1 】

本発明によれば、入力された検索語を用いてウェブ検索を実行し、検索結果を出力する技術において、内容の集中度の高い検索結果を自動的に取得することが可能となる。

【図面の簡単な説明】

【 0 0 1 2 】

【図 1】本発明の実施の形態に係る検索システムの構成図である。

【図 2】本発明の実施の形態に係る検索システムの動作を示すフローチャートである。

【図 3】検索クエリ生成部 1 3 の処理を示すフローチャートである。

【図 4】検索語と検索クエリの例を示す図である。

【図 5】ウェブ検索部 1 5 の処理を示すフローチャートである。

50

【図6】予備検索実行部16の処理を示すフローチャートである。

【図7】ウェブ検索の検索結果をブラウザ表示した場合の例を示す図である。

【図8】検索結果であるHTMLテキストの一部を示す図である。

【図9】予備検索結果の例を示す図である。

【図10】予備検索結果評価部18の処理を示すフローチャートである。

【図11】予備検索結果評価部18における評価処理を説明するための図である。

【図12】予備検索結果評価部18による出力結果の例を示す図である。

【発明を実施するための形態】

【0013】

以下、図面を参照して本発明の実施の形態を説明する。

10

【0014】

(システム構成)

図1に、本発明の実施の形態に係る検索システムの構成を示す。図1に示すとおり、本実施の形態に係る検索システムは、検索制御装置1、及び検索利用装置3を有する。検索制御装置1は、ウェブ検索サービスが提供されるインターネット等のネットワーク4に接続される。

【0015】

検索利用装置3と検索制御装置1との間はWANやインターネット等の広域ネットワークで接続してもよいし、LAN等のローカルなネットワークで接続してもよい。また、検索利用装置3の中に、検索制御装置1の機能を含めて構成してもよい。もしくは、検索利用装置3を用いず、検索制御装置1内に、情報の入出力を行うためのユーザインタフェースを備え、利用者が検索制御装置1に対して検索語を入力し、検索制御装置1が検索結果を利用者に対して表示する構成としてもよい。

20

【0016】

検索利用装置3は、検索語を検索制御装置1に送信し、検索結果を検索制御装置1から受信する装置である。検索利用装置3は、例えば、ウェブブラウザを備えたPC端末や、特許文献1に記載されたメタデータ生成装置のような、ウェブ検索結果を利用して、メタデータを拡張するための装置である。

【0017】

図1に示すように、検索制御装置1は、入力部11、検索語記憶部12、検索クエリ生成部13、検索クエリ記憶部14、ウェブ検索部15、予備検索実行部16、予備検索結果記憶部17、予備検索結果評価部18、評価結果記憶部19、出力部20、及び制御部21を備える。

30

【0018】

入力部11は、検索利用装置3から受信した検索語を、制御部21を介して検索語記憶部12に格納する。検索語記憶部12には、入力部11から入力された検索語が格納される。

【0019】

検索クエリ生成部13は、メモリ等の記憶装置に格納された設定ファイルの内容に基づき、検索語から、ウェブ検索に不適な文字や記号を削除し、検索制御装置1からウェブ検索を行う際に使用する検索クエリを生成する。そして、生成した検索クエリを制御部21を介して検索クエリ記憶部14に格納する。検索クエリ記憶部14には、検索クエリ生成部13で作成された検索クエリが格納される。

40

【0020】

ウェブ検索部15は、検索クエリ記憶部14に格納されている検索クエリを使用して、ネットワーク4を介してウェブ検索を行い、検索結果と、検索結果に含まれるコンテンツ情報の件数を取得する。本実施の形態において、検索結果は、HTMLで記述されたテキストとして得られるものである。ここでの検索件数に応じて、制御部21は、検索結果を出力部20に出力するか、もしくは、予備検索実行部16に処理を移すかの制御を行なう。

【0021】

50

予備検索実行部 16 は、検索クエリ記憶部 14 に格納されている検索クエリから、複数の単語からなる予備検索クエリを複数個生成し、各予備検索クエリを用いてネットワーク 4 を介してウェブ検索（予備検索という）を実行し、予備検索結果を得る。後に詳細に説明するように、予備検索結果には、上記ウェブ検索の検索結果から抽出された複数の要約文が含まれる。

【0022】

予備検索実行後、予備検索実行部 16 は、予備検索結果を制御部 21 を介して予備検索結果記憶部 17 に格納する。予備検索結果記憶部 17 には、予備検索実行部 16 で得られた予備検索結果が予備検索クエリ毎に格納される。

【0023】

予備検索結果評価部 18 は、予備検索結果記憶部 17 に格納されている予備検索結果に対し、評価関数を用いて内容の集中度を評価し、集中度の高い予備検索結果を特定し、評価結果を評価結果記憶部 19 に格納する。評価結果記憶部 19 には、予備検索結果評価部 18 で得られた評価結果が格納される。出力部 20 は、検索結果を検索利用装置 3 に対して出力する機能部である。

【0024】

制御部 21 は、予め定めた処理手順に従って、各機能部に処理実行を指示する等の処理制御を行う機能部である。つまり、例えば、各機能部が、プログラムのモジュール（関数の実体等）で実現される場合において、制御部 21 は、処理手順に応じた機能部に対応するプログラムを呼び出し、実行する処理を行う。

【0025】

図 1 に示す検索制御装置 1 は、記憶装置及び CPU を備える一般的なコンピュータに、上記各機能部に対応する処理を行うためのプログラムを搭載することにより実現できる。当該プログラムは、メモリやディスク等の記録媒体から上記コンピュータにインストールしてもよいし、ネットワーク上のサーバから上記コンピュータにダウンロードし、インストールすることとしてもよい。

【0026】

また、検索制御装置 1 の上記各機能部は、1 つのコンピュータ内に備える必要はなく、例えば、上記機能部のうちの 1 つ又は複数の機能部を有するコンピュータを、複数個ネットワーク接続した構成を、検索制御装置 1 として用いてもよい。

【0027】

（システムの動作）

次に、本実施の形態に係る検索システム（主に検索制御装置 1）の処理動作を、図 2 に示す処理の流れに沿って詳細に説明する。なお、図 2 には、各ステップを実行する処理の主体と、処理に関わる情報を格納する各記憶部が示されている。

【0028】

まず、入力部 11 が、検索利用装置 3 から検索語を受信し、当該検索語を制御部 21 を介して検索語記憶部 12 に格納する（ステップ 1）。

【0029】

次に、制御部 21 による制御に基づき、検索クエリ生成部 13 が検索クエリの生成を行う（ステップ 2）。検索クエリ生成部 13 の処理を、図 3 のフローチャートを参照して説明する。

【0030】

まず、検索クエリ生成部 13 は、制御部 21 を介して検索語記憶部 12 に格納された検索語を読み出す（図 3 のステップ 21）。そして、検索クエリ生成部 13 は、設定ファイルに設定された区切り文字で指定された表現と一致する箇所を検索語を分割し（ステップ 22）、設定ファイルに不要文字列として指定された表現に一致する文字列を検索語から削除し（ステップ 23）、これらの処理を施した検索語を検索クエリとして出力し、検索クエリ記憶部 14 に格納する。（ステップ 24）。

【0031】

10

20

30

40

50

例えば、設定ファイルとして図3に示す設定ファイルを使用し、検索語として図4(a)に示す検索語が入力されたものとする、検索クエリ生成部13は、図4(b)に示す検索クエリを生成し、検索クエリ記憶部14に格納する。

【0032】

次に、図2のステップ3において、ウェブ検索部15が検索処理を実行する。図5のフローチャートを参照して、ウェブ検索部15の検索処理を説明する。

【0033】

まず、ウェブ検索部15は、制御部21を介して検索クエリ記憶部14から検索クエリを読み出す(ステップ31)。そして、ネットワーク4を介して、検索クエリを用いてウェブ検索を行い、HTMLのテキストである検索結果を取得し、更に、検索されたコンテンツ情報の件数(検索件数)も取得する(ステップ32)。

10

【0034】

図2のステップ3の後、ウェブ検索部15もしくは制御部21は、検索件数が予め定めた閾値以上かどうか判定する(ステップ4)。判定の結果、検索件数が予め定めた閾値未満であればステップ5に進み、検索件数が閾値以上であれば、ステップ7に進む。

【0035】

ステップ5において、予備検索実行部16が、ステップ3で使用した検索クエリから複数の予備検索クエリを生成し、複数の予備検索クエリを用いて予備検索結果を取得する予備検索処理を実行する。図6のフローチャートを参照して、予備検索実行部16の処理を詳細に説明する。

20

【0036】

まず、予備検索実行部16は、制御部21を介して検索クエリ記憶部14から検索クエリを読み出す(ステップ51)。ここで、読み出された検索クエリに含まれる単語の数をKとする。予備検索実行部16は、K個の単語からなる検索クエリから、当該検索クエリの中のK-1個の単語を含む検索クエリである予備検索クエリを複数個生成し、生成した複数の予備検索クエリを予備検索結果記憶部に格納しておく(ステップ52)。

【0037】

例えば、検索クエリが、「ABCde郷土料理」、「ニッポン全国」、「ぎゅう・とり・ぶう〜」の3つの単語を有するものであった場合(K=3)、予備検索実行部16は、「ABCde郷土料理」、「ニッポン全国」、「ニッポン全国」、「ぎゅう・とり・ぶう〜」、及び「ABCde郷土料理」、「ぎゅう・とり・ぶう〜」という3つの予備検索クエリを生成する。

30

【0038】

単語の数をK-1とした予備検索クエリを複数生成した後、予備検索実行部16は、生成した複数の予備検索クエリの中の1つの予備検索クエリを用いてウェブ検索を行い、検索結果を得る(ステップ53)。そして、予備検索実行部16は、HTMLで記述されたテキストからなる検索結果の中のHTMLタグを解析し、予め定めた1以上の数であるN個のコンテンツ情報の各々から要約文を抽出する(ステップ54)。

【0039】

図7に、ウェブ検索の検索結果をブラウザ表示した場合の例を示す。本実施の形態において、要約文とは、図7の例において四角の枠で囲んだ部分に対応するHTMLテキストであり、コンテンツ情報とは、図7の例の中の下線が引かれたタイトルから、要約文の下にあるURLが記載された行までの情報に対応するHTMLテキストのことである。

40

【0040】

図8に、図7の検索結果の最上段にあるコンテンツ情報に対応するHTMLテキストの一部を示す。図8における<div class="abs">から</div>までのテキストが要約文に相当する。予備検索実行部16は、この要約文を、N個のコンテンツ情報から取得することにより、N個の要約文を予備検索結果として取得する。

【0041】

その後、図6のステップ54において、予備検索実行部16は、各要約文に含まれる不

50

要なタグを除去し(ステップ55)、不要なタグを除去した要約文を、予備検索結果として処理対象の予備検索クエリと対応付けて予備検索結果記憶部17に格納する(ステップ56)。予備検索実行部16は、全ての予備検索クエリについての処理が終了するまで、1つずつステップ53からステップ56の処理を実行する(ステップ57)。

【0042】

検索クエリが(「ABCde郷土料理」、「ニッポン全国」、「ぎゅう・とり・ぶう〜」)である場合において、上記の処理の結果、予備検索結果記憶部17に格納される予備検索クエリ毎の予備検索結果の例を図9に示す。

【0043】

次に、図2のステップ6において、制御部21による制御に基づき、予備検索結果評価部18が、予備検索結果評価処理を実行する。図10のフローチャートを参照して、予備検索結果評価部18による予備検索結果評価処理を説明する。

【0044】

まず、予備検索結果評価部18は、制御部21を介して予備検索結果記憶部17から1つの予備検索クエリに対応する予備検索結果を読み出す(ステップ61)。そして、予備検索結果評価部18は、処理対象の予備検索結果におけるN個の要約文のそれぞれを単語ベクトルに変換する(ステップ62)。ここでの単語とは、名詞、形容詞、形容動詞、動詞、副詞のものを指し、活用形の無い単語はそのまま利用し、活用形のある単語は原形のものを利用する。

【0045】

例えば、図11(a)に示すようなN=10個の要約文($T_1 \sim T_{10}$)を有する予備検索結果に対する単語ベクトルとして、図11(b)に示すように10個の単語ベクトルが得られる。

【0046】

本実施の形態における単語ベクトルとは、予備検索結果について定めた特定の複数の単語の各々について、要約文における当該単語の有無を0(含まれない)または1(含む)で表現したベクトルである。

【0047】

例えば、上記特定の複数の単語が、単語ベクトルの要素の並びの順に、「天気」、「日食」、「キャンプ」、「晴れる」...である場合、図11(a)に示す要約文 T_1 は、「天気」、「日食」という単語を含むため、 T_1 に対して $V_1 = (1, 1, 0, 0, \dots)$ という単語ベクトルが得られる。

【0048】

単語ベクトルの生成後、図10のステップ63において、予備検索結果評価部18は、生成された単語ベクトル同士の距離を測り、単語ベクトルのグループ化を行う。つまり、ここでは実質的に、要約文の内容の近さに応じてグループ化が行われる。ここでの処理において、本実施の形態では、2つの単語ベクトル間の距離を以下の式を用いて測る。

【0049】

【数1】

$$Sim(V_i, V_j) = \frac{\sum_t w_i^t w_j^t}{\sqrt{\sum_t (w_i^t)^2 \sum_t (w_j^t)^2}}$$

上記の式において、 $i = 1, 2, \dots, N$, $j = 1, 2, \dots, N$, $i \neq j$ である。 w_i^t は単語tの要素位置における単語ベクトルの値である。例えば、図11の例において、 $i=1$, $j=2$ とすると、 $t = \text{日食}$ に対して、 $w_1^{\text{日食}} = 1$, $w_2^{\text{日食}} = 0$ である。この式は、コサイン距離を求める式で

10

20

30

40

50

あり、0以上1以下の値をとり、2つの単語ベクトルの近さを測る際によく用いられる式である。より詳細には、2つの単語ベクトルの共通性が高い場合には1に近い値をとり、共通性が低い場合は、0に近い値をとる。2つの単語ベクトル間の共通性が高いということは、単語ベクトルの元となった2つの要約文の内容が似ていると解釈できる。図11(c)に、図11(a)の予備検索結果に対してグループ化をした結果の例を示す。

【0050】

なお、要約文をグループ化する手法は上記の手法に限られるわけでない。要約文の内容の近さに応じてグループ化を行うことができる手法であればどのような手法を用いてもよい。

【0051】

続いて、図10のステップ64において、予備検索結果評価部18は、得られたグループ化結果を用いて、予備検索結果における要約文の内容の異なり度合い(すなわち、予備検索結果の内容の集中度)を評価し(ステップ64)、評価値を対象の予備検索結果に対応する予備検索クエリとともに評価結果格納部19に格納する(ステップ65)。

【0052】

そして、予備検索結果評価部18は、全ての予備検索結果に対する処理が終了するまで、ステップ61~ステップ65の処理を繰り返す(ステップ66)。

【0053】

本実施の形態では、以下の評価式を用いて予備検索結果内の要約文の異なり度合い(予備検索結果の内容の集中度)を評価している。

【0054】

【数2】

$$Score(T) = -\frac{1}{\log N} \sum_{i=1}^M \left(\frac{n_i}{N} \right) \log \left(\frac{n_i}{N} \right)$$

上記の評価式において、Mは、対象とする予備検索結果から生成されたグループの個数、 n_i はグループiに含まれる単語ベクトルの数である。例えば、図11(c)に示す例では、10個の単語ベクトルが4つのグループに分けられており、グループの個数は4である。つまり、 $N=10$ 、 $M=4$ である。また、 $n_1=2$ 、 $n_2=3$ 、 $n_3=1$ 、 $n_4=4$ である。

【0055】

上記の評価式は、情報エントロピーを求めるための式であり、ある事象が起こる確率を用いて事象全体の偏りを評価するための式である。全ての事象が均等に起こる場合に最大値をとり、特定の事象だけが起こる場合に最小値をとる。この性質を利用して、本実施の形態では予備検索結果の内容の偏りを判定している。すなわち、各グループにおける単語ベクトル数のNに対する割合(n_i/N ($i=1,2,\dots,M$))を、事象の出現確率とみなし、上記の評価式に代入して、予備検索結果の内容の偏りを評価している。

【0056】

例えば、N個の要約文が、ある1つの内容に関するものであれば、グループ化を行うとN個の単語ベクトルは全て1つのグループにまとまる。このとき、 $M=1$ 、 $n_1/N=1$ となり、評価式は最小値0をとる。

【0057】

逆に、N個の要約文が全て異なる内容であれば、1つの単語ベクトルが1つのグループとなり、N個のグループができる。このとき、 $M=N$ 、 $n_1/N=1/N$ 、 $n_2/N=1/N$ 、 \dots 、 $n_M/N=1/N$ となり、評価式の値は最大値1をとる。

【0058】

より具体的に説明すると、図9に示すような予備検索結果が得られた場合、予備検索結

10

20

30

40

50

果評価部 18 は、図 12 に示す結果を算出し、これを評価結果記憶部 19 に格納する。図 12 に示すように、予備検索クエリ：（「ABCde郷土料理」、「ニッポン全国」）に対する評価値が最小である。

【0059】

実際に、図 9 に示すように、予備検索クエリ：（「ABCde郷土料理」、「ニッポン全国」）に対する予備検索結果が、他の予備検索結果に比べて、クエリに関連する内容が多く含まれていることがわかる。予備検索クエリ：（「ニッポン全国」、「ぎゅう・とり・ぶう〜」）については、複数の要約文が予備検索結果として得られているものの、要約文にはクエリとは異なる内容が多く含まれており、内容の集中度が低いため、評価値が高くなっている。

10

【0060】

以上の結果から、複数の予備検索結果において、評価式の値が最小となる予備検索結果が、内容の集中度が最も高い予備検索結果であるといえる。なお、内容の集中度を評価する手法は上記の手法に限られるわけでない。内容の集中度を適切に評価できる手法であればどのような手法を用いてもよい。

【0061】

全ての予備検索結果について図 10 に示すステップ 61～ステップ 65 の処理が終了すると、予備検索結果評価部 18 は、評価結果記憶部 19 から、評価値が最小となった予備検索クエリを読み出し、当該予備検索クエリを、制御部 21 を介して検索クエリとして検索クエリ記憶部 14 に格納する（ステップ 67）。図 12 の例では、予備検索クエリ：（「ABCde郷土料理」、「ニッポン全国」）が読み出され、検索クエリ記憶部 14 に格納される。

20

【0062】

その後、図 2 のステップ 3 において、ウェブ検索部 15 は、予備検索結果評価部 18 により検索クエリ記憶部 14 に格納された検索クエリを読み出し、ウェブ検索を行う。

【0063】

このウェブ検索の結果、ステップ 4 において、Yes の判定結果が得られれば、出力部 20 が、検索結果を検索利用装置 3 に対して出力する（図 2 のステップ 7）。判定結果が No であれば、今回の検索クエリに基づいて、再びステップ 5 からの処理が行われる。この処理は、ステップ 4 において、Yes の判定結果が得られるまで繰り返し行われる。

30

【0064】

ステップ 7 の後、検索利用装置 3 は、例えば、検索結果を利用してメタデータの拡張を行う。また、検索利用装置 3 が一般的なウェブ検索を行うブラウザを備えた PC 端末である場合、当該検索利用装置 3 は、検索結果の表示を行う。

【0065】

（実施の形態の効果）

本実施の形態で説明した検索制御装置 1 によれば、ある検索語から、内容の集中度の高い検索結果を自動的に取得することができる。

【0066】

例えば、検索利用装置 3 として、ウェブ検索を用いたメタデータ拡張用の装置を適用した場合、従来であればウェブ検索に失敗した場合には、コンテンツの管理者が検索語の再入力等の人手による操作を行うことが必要であったところ、本実施の形態の技術では、自動的に再検索を行うので、人手による操作を一切行うことなくメタデータの拡張を適切に行うことができる。

40

【0067】

また、例えば、検索利用装置 3 として、一般的なウェブ検索を行う PC 端末を適用した場合、本実施の形態によれば、利用者は検索語の再入力を行うことなく適切な検索結果を迅速に得ることができる。

【0068】

本発明は、上記の実施の形態に限定されることなく、特許請求の範囲内において、種々

50

変更・応用が可能である。

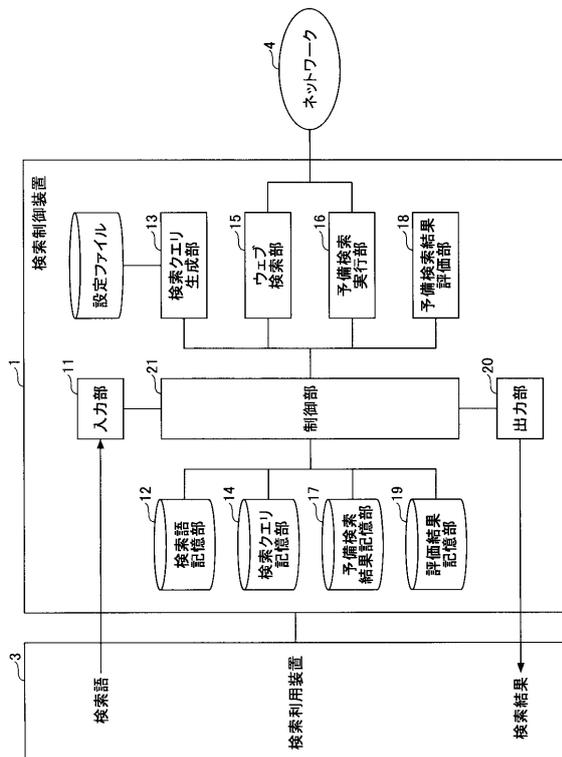
【符号の説明】

【0069】

- 1 検索制御装置
- 3 検索利用装置
- 4 ネットワーク
- 11 入力部
- 12 検索語記憶部
- 13 検索クエリ生成部
- 14 検索クエリ記憶部
- 15 ウェブ検索部
- 16 予備検索実行部
- 17 予備検索結果記憶部
- 18 予備検索結果評価部
- 19 評価結果記憶部
- 20 出力部
- 21 制御部

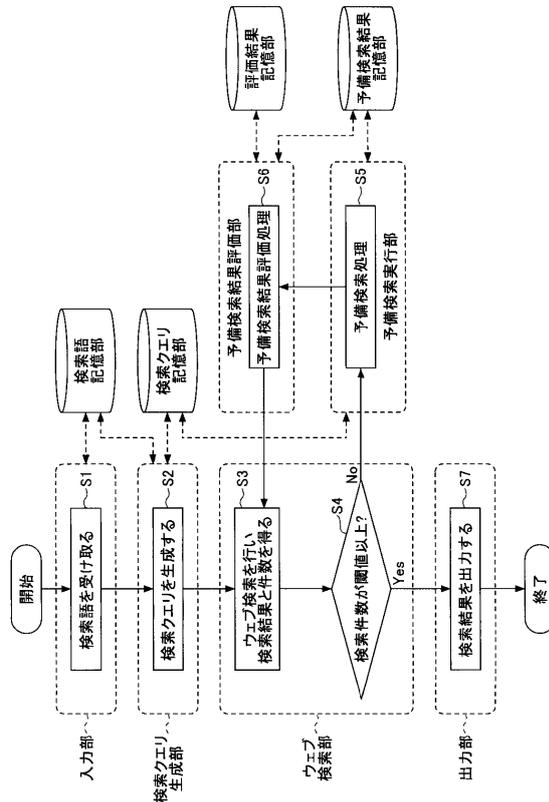
【図1】

本発明の実施の形態に係る検索システムの構成図

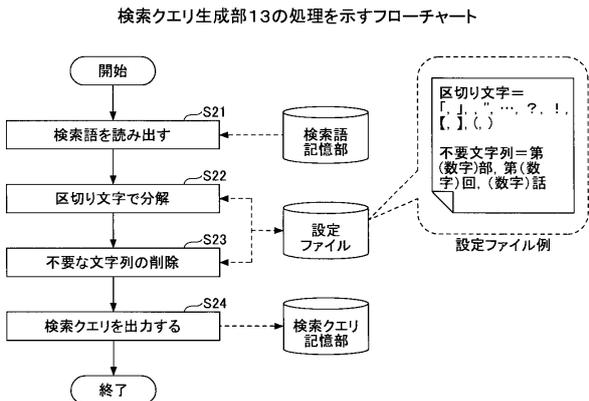


【図2】

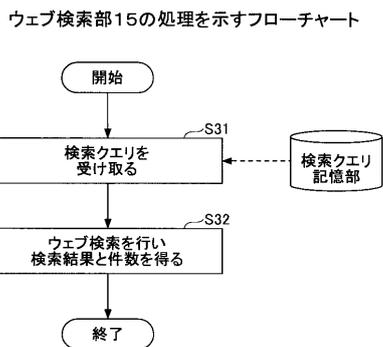
本発明の実施の形態に係る検索システムの動作を示すフローチャート



【図3】



【図5】



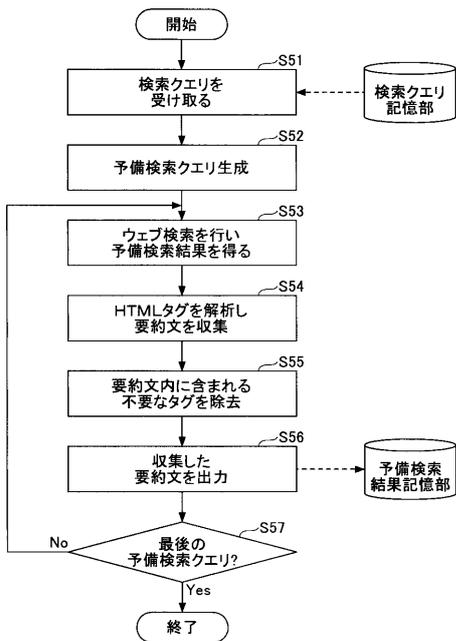
【図4】

検索語と検索クエリの例を示す図

(a) 検索語	ola! セレソン密着 12話 パロンドール XYZの全て特集
(b) 検索クエリ	ola セレソン密着 パロンドール XYZの全て特集

【図6】

予備検索実行部16の処理を示すフローチャート



【図7】

ウェブ検索の検索結果をブラウザ表示した場合の例を示す図

ウェブ検索結果 皆既日食 天気 で検索した結果 1~10件目 / 約21,500,000件-0.03秒

要約文

皆既日食 - 日本気象協会 otenki.jp
 日本気象協会による。各地の皆既日食。部分日食情報。部分日食の始め/終り。皆既日食(太陽がすべて隠れる)が始まる/終わる時刻です...otenki.jp 皆既日食は、日本天文委員会公認イベントです。天気サイト...

URL1

2009年7月22日 皆既日食・部分日食 goo 天気
 皆既日食が見られる地点の天気 鹿児島 奄美 - あやまる 奄美 - 大浜海浜公園
 皆既日食が起きると、あることが大きく変化する?~西洋占星術から見た日食の影響
 ~ (goo 占い)...

URL2

皆既日食とは?(連載)回目X(2009年6月1日) - 日直予報士...
 天気予報、世界天気、最新予報、日直予報士、過去天気、皆既日食...日食は、太陽が月によって隠される現象のことです。見え方によって「皆既日食」、「金環日食」、「部分日食」に分かれます。...日本で皆既日食が見られるのは実に46年ぶりとなります。...

URL3

皆既日食 天気に関するニュース
 今世紀最後のショーへ準備万端 きょう46年ぶり皆既日食
 日本の機上で46年ぶりとなる皆既日食が22日、鹿児島・トカラ列島や奄美大島などで起きる。皆既継続時間が8分5秒と有人観測で最も長いカラ列島、鹿児島では21日、観測客が望遠鏡やカメラを致し、太平洋上...

【図 8】

検索結果であるHTMLテキストの一部を示す図

```

<div class="web"><div class="hd"><h3><a
class="yschurl" href="http://wrs.search.yahoo.co.jp/S=2114736003/K=%E7%9A%
86%E6%97%A2%E6%97%A5%E9%A3%9F+%E5%A4%A9%E6%B0%97/v=2/SID=w/TID=jpc
007_jpc007/l=WS1/R=1/VCA=1/wdm=0/IPC=jp/ln=ja/H=0/:_yit=A3yTKiOrVmZK
HKsAq2aDTwx...yilu=X3oDMTFjYXAwbXQzBGnVbG8DdwRsA1dTMQRwb3MDMQR
zZWMDc3EdnRpZANqcGMwMDfanBjMDA3/SIG=11kujasbr/EXP=1248307243/*
-http%3A//otenki.jp/forecast/eclipse"><b>皆既日食</b></div><div class="bd"><div class="abs">日本気協会
otenki.jp</a></h3></div><div class="bd"><div class="abs">日本気協会による。
各地の<b>皆既日食</b>、<b>部分日食情報... 皆既食の始め/ 終り</b>、<b>皆既日食
</b>(太陽がすべて隠れる)が始まる/ 終わる時刻です...otenki.jp<b>皆既日食</b>は、
日本天文委員会公認イベントです。<b>天気</b>/<b>ガイド</b></div><div class="snif">
<em class="yschurl">...

```

<div class="abs">から</div>までを抽出し要約文を取得

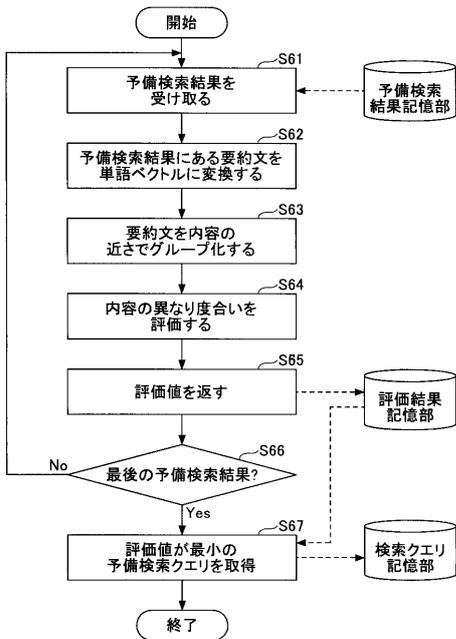
【図 9】

予備検索結果の例を示す図

予備検索クエリ	予備検索結果
ABC de 郷土料理 ニッポン全国	ABC de 郷土料理 ニッポン全国 ぎゅう・とり・ぶう? 山形県 ABC de 郷土料理 ニッポン全国 ぎゅう・とり・ぶう? 青森県 70. ABC de 郷土料理 ニッポン全国 ぎゅう・とり・ぶう - qoo ...
ニッポン全国 ぎゅう・とり・ぶう~	『x x x 市』の施行範囲の推移は「日本全国が非難まつり?」 「いま、かアちゃんとお、お湯(ぶう)に入ってます。 「たんまぎゅう」ではないですよ。「...」と諦めます。 ...
ABC de 郷土料理 ぎゅう・とり・ぶう~	(検索結果無し)

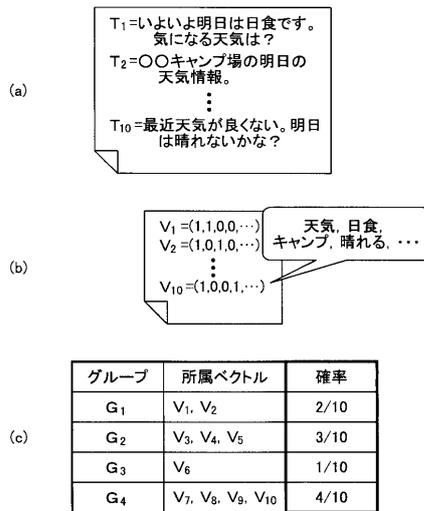
【図 10】

予備検索結果評価部18の処理を示すフローチャート



【図 11】

予備検索結果評価部18における評価処理を説明するための図



【図 12】

予備検索結果評価部18による出力結果の例を示す図

予備検索クエリ	評価値
ABC de 郷土料理 ニッポン全国	0.86992
ニッポン全国 ぎゅう・とり・ぶう～	0.96023
ABC de 郷土料理 ぎゅう・とり・ぶう～	(値無し)

フロントページの続き

(56)参考文献 特開2009-015796(JP,A)
特開2007-172179(JP,A)
特開2003-281183(JP,A)
特開平08-006970(JP,A)
特開2005-173847(JP,A)
国際公開第2009/019860(WO,A1)

(58)調査した分野(Int.Cl., DB名)
G06F 17/30
CSDB(日本国特許庁)