

(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2012-146079

(P2012-146079A)

(43) 公開日 平成24年8月2日(2012.8.2)

(51) Int.Cl.
G06F 17/30 (2006.01)

F I
G06F 17/30 320D

テーマコード (参考)
5B075

審査請求 未請求 請求項の数 9 O L (全 12 頁)

(21) 出願番号 特願2011-3242 (P2011-3242)
(22) 出願日 平成23年1月11日 (2011.1.11)

(71) 出願人 000004226
日本電信電話株式会社
東京都千代田区大手町二丁目3番1号
(74) 代理人 100086232
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 佐藤 大祐
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

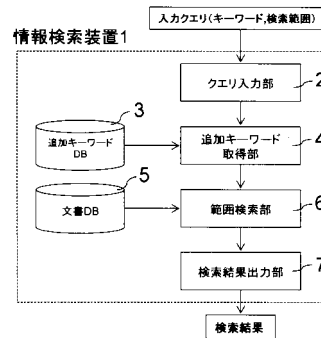
(54) 【発明の名称】 情報検索装置、情報検索方法、及び情報検索プログラム

(57) 【要約】

【課題】略語に基づく情報検索の精度を向上させる。

【解決手段】略語を含んだクエリに適合する文書を検索する情報検索装置1において、追加キーワード取得部4はクエリ入力部2からクエリとして供された略語からなるキーワード及び検索範囲に対応する対応語群を追加キーワードDB3から取得する。そして、追加キーワード取得部4は対応語の主要度に基づき前記対応語群から取得した対応語を前記クエリの追加キーワードとする。範囲検索部6は前記キーワード、前記検索範囲及び前記追加キーワードに基づき文書DB5から情報検索を行う。この検索により得られた検索結果は検索結果出力部7から出力される。

【選択図】図1



【特許請求の範囲】**【請求項 1】**

略語を含んだクエリに適合する文書を検索する情報検索装置であって、

前記クエリとして供された略語からなるキーワード及び検索範囲に対応する対応語群から各対応語の主要度に基づき取得した対応語を当該クエリの追加キーワードとする追加キーワード取得手段と、

前記キーワード、前記検索範囲及び前記追加キーワードに基づき情報検索を行う範囲検索手段と

を備えたこと

を特徴とする情報検索装置。

10

【請求項 2】

前記追加キーワード取得手段は、前記キーワード及び検索範囲に対応する対応語群の各対応語について当該検索範囲における主要度を示す指定範囲中主要度を算出し、この主要度が閾値よりも高い対応語を前記追加キーワードとして取得すること

を特徴とする請求項 1 に記載の情報検索装置。

【請求項 3】

前記検索範囲が複数供された場合、前記追加キーワード取得手段は、前記キーワード及び当該複数の検索範囲に対応する対応語群の各対応語の各検索範囲における主要度を示す指定範囲中主要度を掛け合わせたものを前記複数の検索範囲からなる指定範囲集合における当該対応語の主要度を示す指定範囲中主要度とし、この主要度が閾値よりも高い対応語を前記追加キーワードとして取得すること

を特徴とする請求項 2 に記載の情報検索装置。

20

【請求項 4】

略語で示されるキーワードと前記略語の開示を規定する範囲と前記略語に対応する対応語とこの対応語の主要度とを予め格納したキーワードデータベースをさらに備え、

前記追加キーワード取得手段は前記供されたキーワード及び単数または複数の検索範囲に基づき前記キーワードデータベースから当該キーワード及び検索範囲に対応する対応語群を引き出すこと

を特徴とする請求項 1 から 3 のいずれか 1 項に記載の情報検索装置。

【請求項 5】

略語を含んだクエリに適合する文書を検索する情報検索方法であって、

追加キーワード取得手段が前記クエリとして供された略語からなるキーワード及び検索範囲に対応する対応語群から各対応語の主要度に基づき取得した対応語を当該クエリの追加キーワードとするステップと、

範囲検索手段が前記キーワード、前記検索範囲及び前記追加キーワードに基づき情報検索を行うステップと

を有すること

を特徴とする情報検索方法。

30

【請求項 6】

前記追加キーワードを取得するステップでは、前記キーワード及び検索範囲に対応する対応語群の各対応語について当該検索範囲における主要度を示す指定範囲中主要度を算出し、この主要度が閾値よりも高い対応語を前記追加キーワードとして取得すること

を特徴とする請求項 5 に記載の情報検索方法。

40

【請求項 7】

前記検索範囲が複数供された場合、前記追加キーワードを取得するステップでは、前記キーワード及び当該複数の検索範囲に対応する対応語群の各対応語の各検索範囲における主要度を示す指定範囲中主要度を掛け合わせたものを前記複数の検索範囲からなる指定範囲集合における当該対応語の主要度を示す指定範囲中主要度とし、この主要度が閾値よりも高い対応語を前記追加キーワードとして取得すること

を特徴とする請求項 6 に記載の情報検索方法。

50

【請求項 8】

前記追加キーワードを取得するステップでは、略語で示されるキーワードと前記略語の開示を規定する範囲と前記略語に対応する対応語とこの対応語の主要度とを予め格納したキーワードデータベースから、前記供されたキーワード及び単数または複数の検索範囲に対応する対応語群を引き出すこと

を特徴とする請求項 5 から 7 のいずれか 1 項に記載の情報検索方法。

【請求項 9】

請求項 1 から 4 のいずれか 1 項に記載の情報検索装置を構成する各手段としてコンピュータを機能させることを特徴とする情報検索プログラム。

【発明の詳細な説明】

10

【技術分野】

【0001】

本発明は入力された検索語に適合する文書を提示する情報検索技術に関する。

【背景技術】

【0002】

情報検索に供される検索キーワードが複数の表記をもつ場合がある。このとき、検索キーワードに加え、同じ意味を持つ他の表記も検索キーワードとして追加するという、いわゆる検索漏れを減らす検索技術が知られている（例えば、特許文献 1、非特許文献 1）。

【0003】

この検索技術によれば、ユーザが入力したキーワードをそのまま含まなくとも、ユーザの要求に合致した内容の文書を提示することができる。特に、検索キーワードが略語であるときには、入力された略語での検索結果に加え、その略語の意味する語での検索結果を追加することにより、より多くの適切な検索結果を得ることができる。例えば、「BBQ」というキーワードを用いて検索を行う場合を考える。「BBQ」は「バーベキュー」の略語である。このとき、「BBQ」というキーワードでの検索に加え、「バーベキュー」というキーワードで検索した結果も提示することにより、ユーザの要求に合致した文書をより多く提示することができる。

20

【0004】

また、指定した範囲に含まれる文書を絞り込んで検索を行う、範囲検索の技術が知られている（例えば、非特許文献 2、3）。指定検索範囲には、日時などの時間範囲や、場所や地域といった地理範囲などがある。例えば、時間範囲を指定した範囲検索で、「2010年8月～9月」と指定した場合、文書に含まれる日時表現や、文書を収集した日付などを利用し、「2010年8月～9月」に含まれる文書を検索することができる。通常、ユーザは検索範囲の指定に加えキーワードによる検索も行い、検索システムは、指定した検索範囲と指定したキーワードとの両方の条件を満たす文書を提示する。

30

【先行技術文献】

【特許文献】

【0005】

【特許文献 1】特開 2008 - 204399 号公報

【非特許文献】

40

【0006】

【非特許文献 1】Akira Terada, Takenobu Tokunaga, Hozumi Tanaka, "Automatic expansion of abbreviations by using context and character information", Information Processing and Management 40, 2004, January 2004, pp.31-45

【非特許文献 2】戸田浩之, 安田宜仁, 松浦由美子, 片岡良治, "地理情報と内容情報の近接性を考慮した地理情報検索", 第 23 回人工知能学会全国大会, 2009 年 6 月

【非特許文献 3】河合英紀, Adam Jatowt, 田中克己, 國枝和雄, 山田敬嗣, "ChronoSeeker Webからの過去・未来情報のオンデマンド検索エンジン", WebDB Forum2009, 2009 年 11 月

【発明の概要】

50

【発明が解決しようとする課題】

【0007】

異なる意味を持つ複数の語が、同じ略語で表記される場合がある。例えば、「ES」という語は、「エントリーシート」という語の略語であると同時に、「従業員満足度 (Employee Satisfaction)」や、「エンジニアリングサービス」といった複数の語の略語でもある。

【0008】

略語の意味する主要な語は、検索範囲によって異なる。例えば、時間を検索範囲にした場合を考える。ある時期においては「ES」といえば「従業員満足度」のことを示しており、ある期間においてはもっぱら「エントリーシート」の意味でつかわれているといた

10

【0009】

しかし、従来の方法では、略語が表す語を抽出する際に、全ての文書集合を用いて行っていた。これにより、検索ユーザが意図した意味以外の語を検索キーワードとして追加することになり、結果として検索の精度が低下するという問題点があった。

【0010】

本発明は、以上の事情に鑑みなされたもので、キーワードとして略語が検索に供された際に、指定された検索範囲における適切な略語の候補を前記検索のための追加キーワードとして取得して、略語に基づく情報検索の精度を向上させることを目的とする。

20

【課題を解決するための手段】

【0011】

そこで、本発明は、略語に基づく情報検索の際、当該略語及び指定した検索範囲に対応した対応語群から主要度に基づき抽出した対応語を追加キーワードとして当該情報検索に供することにより、当該略語に基づく検索の精度を向上させる。

【0012】

本発明の情報検索装置の態様としては、略語を含んだクエリに適合する文書を検索する情報検索装置であって、前記クエリとして供された略語からなるキーワード及び検索範囲に対応する対応語群から各対応語の主要度に基づき取得した対応語を当該クエリの追加キーワードとする追加キーワード取得手段と、前記キーワード、前記検索範囲及び前記追加

30

【0013】

本発明の情報検索方法の態様としては、略語を含んだクエリに適合する文書を検索する情報検索方法であって、追加キーワード取得手段が前記クエリとして供された略語からなるキーワード及び検索範囲に対応する対応語群から各対応語の主要度に基づき取得した対応語を当該クエリの追加キーワードとするステップと、範囲検索手段が前記キーワード、前記検索範囲及び前記追加キーワードに基づき情報検索を行うステップとを有する。

【0014】

前記情報検索装置及び情報検索方法において、前記追加キーワード取得手段は、前記キーワード及び検索範囲に対応する対応語群の各対応語について当該検索範囲における主要度を示す指定範囲中主要度を算出し、この主要度が閾値よりも高い対応語を前記追加キーワードとして取得するようにするとよい。この態様によれば前記指定範囲中主要度の閾値に基づき対応語が抽出されるので所望の検索精度で前記キーワードに基づく情報検索を行

40

【0015】

また、前記検索範囲が複数供された場合、前記追加キーワード取得手段は、前記キーワード及び当該複数の検索範囲に対応する対応語群の各対応語の各検索範囲における主要度を示す指定範囲中主要度を掛け合わせたものを前記複数の検索範囲からなる指定範囲集合における当該対応語の主要度を示す指定範囲中主要度とし、この主要度が閾値よりも高い対応語を前記追加キーワードとして取得するようにするとよい。この態様によれば検索範

50

囲が複数指摘された場合であっても前記キーワードに基づく情報検索を所望の検索精度で行える。

【0016】

さらに、本発明の情報検索装置及び情報検索方法の他の態様としては、前記追加キーワードを取得するにあたり、略語からなるキーワードと前記略語の開示を規定する範囲と前記略語に対応する対応語とこの対応語の主要度とを予め格納したキーワードデータベースから、前記供されたキーワード及び単数または複数の検索範囲に対応する対応語群を引き出すようにするとよい。この態様によれば前記供された略語及び単数または複数の検索範囲に対応する対応語群を効率的に取得できる。

【0017】

尚、本発明は上記の情報検索装置を構成する各手段としてコンピュータを機能させる情報検索プログラムの態様とすることもできる。

【発明の効果】

【0018】

以上の発明によれば、キーワードとして略語が情報検索に供された際に、当該略語と共に指定された検索範囲における適切な略語の候補を前記検索の追加キーワードとして利用するので、当該略語に基づく情報検索の精度が向上する。

【図面の簡単な説明】

【0019】

【図1】 発明の実施形態に係る情報検索装置のブロック構成図。

【図2】 発明の実施形態に係る情報検索装置の処理手順を説明したフロー図。

【図3】 追加キーワード取得部による追加キーワードの取得手順を説明したフロー図。

【図4】 時間的範囲のカラムを有する追加キーワードデータベースの一例。

【図5】 地理的範囲のカラムを有する追加キーワードデータベースの一例。

【発明を実施するための形態】

【0020】

以下、図面を参照しながら本発明の実施の形態を説明するが、本発明は下記の実施形態に限定されるものではない。

【0021】

[概要]

図1に示された本実施形態の情報検索装置1は、略語からなるキーワードが情報検索に供された際、当該略語と共に指定された検索範囲に対応した対応語群から、当該検索範囲における対応語の主要度に基づき対応語を抽出する。そして、この抽出した対応語を追加キーワードとして情報検索に供する。

【0022】

本実施形態において、情報検索装置1は、略語からなるキーワードと前記略語の開示を規定する範囲と前記略語に対応する対応語とこの対応語の主要度とを予め格納した追加キーワードデータベース3（以下、追加キーワードDB3）を備えている。情報検索装置1では、クエリ入力部2から供されたキーワード及び検索範囲に対応した対応語群を追加キーワードDB3から引き出して、当該対応語群の各対応語について当該検索範囲における主要度を示す指定範囲中主要度を算出する。そして、この主要度が閾値よりも高い対応語を追加キーワードとして範囲検索部6による情報検索に供する。以上のように、略語を検索キーワードとした場合に、指定された検索範囲によって当該略語の表す対応語が絞り込まれ、検索精度が向上する。

【0023】

[情報検索装置1の構成]

情報検索装置1は、図1に示されたように、クエリ入力部2、追加キーワードDB3、追加キーワード取得部4、文書データベース5（以下、文書DB5）、範囲検索部6、検索結果出力部7を備える。

【0024】

10

20

30

40

50

情報検索装置 1 の前記機能部及び DB はサーバ（コンピュータ）のハードウェアリソースによって実現される。すなわち、情報検索装置 1 は CPU、記憶装置（例えば、ハードディスク装置）、I/O デバイス（例えば、通信デバイス）等のコンピュータに係るハードウェアリソースを備える。そして、これらのハードウェアリソースがソフトウェアリソース（OS、アプリケーション等）と協働することにより前記機能部及び DB が実装される。

【0025】

クエリ入力部 2 はネットワークを介してユーザから受け付けたクエリを追加キーワード取得部 4 に供する。クエリは情報検索装置 1 に供される情報要求であり、略語からなるキーワードと、検索範囲とから構成される。前記検索範囲としては時間、地理、言語、ドメイン、分野等が例示される。尚、クエリ入力部 2 は Web インタフェースに例示されるユーザインタフェース上に備えられている。

10

【0026】

追加キーワード DB 3 は、図 4、図 5 に示したように、略語 w を示すキーワードと、略語 w の開示を規定する範囲 t と、略語 w に対応する対応語 e と、対応語 e の主要度 c を予め格納したデータベースである。主要度 c は、略語 w に対する対応語 e の尤もらしさを示すスコアであり、その値が高いものほど対応語としてふさわしいものとする。主要度 c は例えば非特許文献 1 に開示された周知の “Score” の算法によって得られる。

【0027】

図 4 に例示された追加キーワード DB は検索範囲が時間範囲である場合のデータベースの例である。図 5 に例示された追加キーワード DB は検索範囲が地理範囲である場合のデータベースの例である。範囲 t における略語 w に対応する対応語 e の抽出は範囲 t に適合する文書集合を用いることによって行う。範囲 t に適合する文書集合の取得は後述の範囲検索部 6 による指定検索範囲に適合する文書の特定法と同一の方法で行うことができる。

20

【0028】

追加キーワード取得部 4 はクエリ入力部 2 によって指定された前記キーワード及び検索範囲に対応する対応語群を追加キーワード DB 3 から引き出し、各対応語について当該検索範囲における当該対応語の主要度を示す指定範囲中主要度を計算する。そして、この主要度が閾値よりも高い対応語を追加キーワードとして取得する。

【0029】

また、追加キーワード取得部 4 は、クエリ入力部 2 によって複数の検索範囲が指定された場合、クエリ入力部 2 から供されたキーワード及び当該複数の検索範囲に対応する対応語群を追加キーワード DB 3 から引き出す。次いで、前記対応語群の各対応語について前記各検索範囲における当該対応語の主要度を示す指定範囲中主要度を計算する。次いで、この各検索範囲における指定範囲中主要度を掛け合わせたものを前記複数の検索範囲からなる指定範囲集合における前記対応語の主要度を示す指定範囲中主要度として算出する。そして、この主要度が閾値よりも高い対応語を追加キーワードとして取得する。

30

【0030】

「単一の検索範囲が指定された場合」「複数の検索範囲が指定された場合」の追加キーワード取得部 4 による対応語の指定範囲中主要度の具体的な計算についてそれぞれ説明する。

40

【0031】

（単一の検索範囲が指定された場合の指定範囲中主要度の計算）

ユーザが指定する検索範囲は、データベース中の範囲の単位と異なる場合がある。例えば、図 4 のデータベース例では、時間単位を 1 年としているが、ユーザは 2008 年 7 月から 2009 年 12 月までを指定している場合などである。このような場合に対応するため、ユーザの指定する検索範囲中における対応語の主要度である、指定範囲中主要度を計算する。

【0032】

略語 w に対する、指定範囲（検索範囲） s における対応語 e の指定範囲中主要度 s_c （

50

w, s, e) は、以下の式 (1) によって算出できる。

【0033】

【数1】

$$sc(w, s, e) = \frac{\sum_{t \in T} \lambda(t) c(w, t, e)}{\sum_{t \in T} \lambda(t)} \quad (1)$$

【0034】

10

ここで、 t は範囲を表し、 T は指定範囲 s を含んでいる範囲 t の集合を表す。指定範囲 s が 2008 年 7 月から 2009 年 12 月である場合、 T は 2008 年と 2009 年である。 $\lambda(t)$ は、範囲 t のうち指定範囲 s が含まれる割合を表す。先ほどの例で考えると、 t が 2008 年である場合、指定時間に含まれるのは 7 月から 12 月までの半年であるため、割合 $\lambda(t)$ は 0.5 となる。一方、範囲 t が 2009 年である場合、1 月から 12 月まで全て指定範囲が含まれるため、割合 $\lambda(t)$ は 1 となる。 $c(w, t, e)$ はキーワード w に対する、範囲 t における対応語 e の主要度を表す。そして、指定範囲中主要度 sc が予め設定された閾値より高い対応語 e が追加キーワードとして決定される。

【0035】

20

前記閾値は任意に設定される。すなわち、より多くのキーワードを追加してより多くの検索結果を提示したい場合には閾値は低く設定され、関係のない検索結果を減らして精度を高めたい場合には閾値は高く設定される。

【0036】

(複数の検索範囲が指定された場合の指定範囲中主要度の計算)

複数の検索範囲が検索条件として指定された場合、各検索範囲で算出された指定範囲中主要度を掛け合わせることで、前記複数の検索範囲からなる指定範囲集合における対応語の指定範囲中主要度を算出する。

【0037】

30

すなわち、 n 個の検索範囲が指定されたとき、それぞれの指定範囲 (検索範囲) を S_i ($i = 1, 2, 3, \dots, n$) で表し、指定範囲 S_i の集合を S とする。略語 w に対する、指定範囲集合 S における対応語 e の指定範囲中主要度 $sc(w, S, e)$ は以下の式 (2) によって算出できる。そして、この指定範囲中主要度 sc の値が予め設定された閾値より高い対応語 e が追加キーワードとして決定される。前記閾値も前述の趣旨により任意に設定される。

【0038】

【数2】

$$sc(w, S, e) = \prod_i^n sc(w, s_i, e) \quad (2)$$

40

【0039】

(追加キーワード取得部 4 による追加キーワードの取得手順)

図 3 を参照しながら追加キーワード取得部 4 による追加キーワードの取得手順 S41 ~ S47 について説明する。ここでは検索範囲として指定範囲 s が指定された事例での追加キーワードの取得手順について説明する。

【0040】

(S41) クエリ入力部 2 からキーワード (略語) と指定範囲 s が入力される。

【0041】

(S42) 追加キーワード DB 3 において前記キーワード、指定範囲 s を含む範囲 t を有する行を検索する。

50

【 0 0 4 2 】

(S 4 3) 前記行が存在する場合、 S 4 4 に移行する。前記行が存在しない場合、処理を終了する。

【 0 0 4 3 】

(S 4 4) S 4 2 で取得された行から対応語 e の指定範囲中主要度 s c を式 (1) による演算によって算出する。

【 0 0 4 4 】

(S 4 5) 前記算出された指定範囲中主要度 s c が閾値以上である場合、 S 4 6 に移行する。前記指定範囲中主要度 s c が閾値以上でない場合、 S 4 7 に移行する。

【 0 0 4 5 】

(S 4 6) 指定範囲中主要度 s c が閾値以上である対応語句 e を追加キーワードとして取得する。

【 0 0 4 6 】

(S 4 7) 追加キーワード D B 3 における前記キーワード、指定範囲 s を含む範囲 t を有する行中の全ての対応語 e について S 4 4 にて指定範囲中主要度 s c を算出する。

【 0 0 4 7 】

以上の手順により取得された追加キーワードは範囲検索部 6 に供される。

【 0 0 4 8 】

文書 D B 5 は予め収集した文書を格納したデータベースである。例えば、 W e b 上に公開されている文書をクローラ等の周知の情報収集手段によって収集した文書を記録することで構築されたデータベースである。そして、インターネット上の W e b サイトのように文書の内容が逐次更新される情報源である場合、当該文書が更新されると、新たな文書が作成されたとみなされ、前記情報収集手段によって収集されて文書 D B に格納される。尚、既存の W e b ページの検索エンジン群を文書 D B 5 の代わりとして用いるようにしてもよい。

【 0 0 4 9 】

範囲検索部 6 は、前記入力されたキーワードに適合する文書、前記入力された検索範囲に適合する文書、及び追加キーワード取得部 4 で得られた追加キーワードに適合する文書を文書 D B 5 から検索する。前記文書の検索には周知の検索技術が適用される。

【 0 0 5 0 】

検索結果出力部 7 は範囲検索部 6 による検索によって得られた文書を検索結果として出力する。検索結果出力部 7 はクエリ入力部 2 と同様に前記ユーザインタフェース上に備えられている。

【 0 0 5 1 】

[情報検索装置 1 による検索の処理手順の説明]

図 2 を参照しながら情報検索装置 1 による文書検索の手順について具体的に説明する。

【 0 0 5 2 】

S 1 : クエリ入力部 2 はユーザからクエリとして受け付けたキーワード (略語) 及び検索範囲を追加キーワード取得部 4 に供する。

【 0 0 5 3 】

検索範囲としては例えば時間、地理、言語、ドメイン、分野などである。検索範囲が時間である場合には、検索範囲としては「 2 0 0 8 年 7 月 ~ 2 0 0 9 年 1 2 月 」が例示される。検索範囲が地理である場合には、検索範囲としては「 北東座標 緯度 3 5 度 0 0 分 0 0 秒 , 経度 1 3 9 度 0 0 分 0 0 秒 ~ 南西座標 緯度 3 4 度 5 0 分 0 0 秒 , 経度 1 3 8 度 5 0 分 0 0 秒 」が例示される。検索範囲が言語である場合には、「 日本語 」といった検索範囲クエリとしてもよい。他にもドメイン、分野などが検索範囲として例示される。

【 0 0 5 4 】

S 2 : 追加キーワード取得部 4 はクエリ入力部 2 から供された検索範囲が追加キーワード D B 3 に含まれているかを判断する。

【 0 0 5 5 】

10

20

30

40

50

S 2では具体的には図3を参照して説明した前述のステップS 4 1～S 4 3が実行される。例えば、図4, 図5に示された追加キーワードDB 3に前記検索範囲が含まれているかを判断する。前記検索範囲が追加キーワードDB 3に含まれている場合、S 3に移行する。一方、前記検索範囲が追加キーワードDB 3に含まれていない場合、S 7に移行する。

【0056】

S 3：追加キーワード取得部4はクエリ入力部2から供された検索範囲に対応する範囲における対応語群を追加キーワードDB 3から引き出す。

【0057】

S 4：追加キーワード取得部4はS 3で取得した対応語群の各対応語の指定範囲中主要度を計算し、これによって算出された主要度が閾値よりも高い対応語を追加キーワードとして取得する。

10

【0058】

S 4では具体的には図3を参照して説明した前述のステップS 4 4～S 4 7が実行される。クエリ入力部2から供された検索範囲が単一である場合、式(1)に基づく演算によって対応語の指定範囲中主要度が算出される。一方、クエリ入力部2から供された検索範囲が複数である場合、式(2)に基づく演算によって対応語の指定範囲中主要度が算出される。そして、指定範囲中主要度が閾値以上である対応語句が追加キーワードとして取得される。

【0059】

20

S 5：追加キーワード取得部4はS 1で供されたキーワード及び検索範囲に対してS 4で取得した追加キーワードを追加したクエリを範囲検索部6に供する。

【0060】

S 6：範囲検索部6は、S 1で入力されたキーワードに適合する文書、S 1で入力された検索範囲に適合する文書、及びS 5で得られた追加キーワードに適合する文書を文書DB 5から検索する。

【0061】

S 7：追加キーワード取得部4はS 1で入力された検索範囲を非特許文献2, 3に例示される従来技術に係る範囲検索を実行する検索エンジンに供してキーワードを取得し、これを追加キーワードとする。そして、S 1で供されたキーワード及び検索範囲に対して本ステップS 7で取得された追加キーワードを追加したクエリを範囲検索部6に供する。

30

【0062】

S 8：範囲検索部6は、S 1で入力されたキーワードに適合する文書、S 1で入力された検索範囲に適合する文書、及びS 7で得られた追加キーワードに適合する文書を文書DB 5から検索する。

【0063】

S 9：検索結果出力部7はS 6またはS 8での検索で取得された文書を検索結果として出力する。

【0064】

[本実施形態の効果]

40

以上のように本実施形態の情報検索装置1によれば、略語及び検索範囲を検索クエリとした際に、検索漏れが低減する共に検索精度が向上する。また、検索クエリとして略語を指定した際に指定した範囲によって当該略語の表す対応語が絞り込まれるので、当該略語及び当該範囲に関係のない文書が除外される。

【0065】

特に、本実施形態においては追加キーワードDB 3が具備されているので、前記略語及び単数または複数の検索範囲に対応する対応語群を効率的に取得できる。

【0066】

また、前記略語及び検索範囲に対応する対応語群の各対応語について当該検索範囲における主要度を示す指定範囲中主要度が算出され、この主要度が閾値よりも高い対応語が前

50

記検索クエリへの追加キーワードとして取得される。このように前記指定範囲中主要度の閾値に基づき対応語が抽出されるので、前記略語に基づく情報検索を所望の検索精度で行える。

【0067】

さらに、前記検索範囲が複数指定された場合、前記略語及び当該複数の検索範囲に対応する対応語群の各対応語の各検索範囲における主要度を示す指定範囲中主要度を掛け合わせたものが前記複数の検索範囲からなる指定範囲集合における当該対応語の主要度を示す指定範囲中主要度とされる。そして、この主要度が閾値よりも高い対応語が前記検索クエリへの追加キーワードとして取得される。したがって、検索範囲が複数指摘された場合であっても、前記略語に基づく情報検索を所望の検索精度で行える。

10

【0068】

また、本実施形態においては予め収集された情報を格納した文書DB5が具備されているので、前記略語、前記検索範囲及び前記追加キーワードに適合した文書を迅速に取得することができる。

【0069】

尚、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変更・応用が可能である。

【0070】

[本発明のプログラムとしての態様]

本発明は上記の実施形態の情報検索装置1の各機能部の一部もしくは全部の機能をコンピュータのプログラムで構成し、そのプログラムをコンピュータによって実行して本発明を実現することができる。また、コンピュータで前記機能部を実現するためのプログラムをそのコンピュータが読み取り可能な記録媒体、例えば、FD(Floppy(登録商標)Disk)や、MO(Magneto-Optical disk)、ROM(Read Only Memory)、メモリカード、CD(Compact Disk)-ROM、DVD(Digital Versatile Disk)-ROM、CD-R、CD-RW、HDD、SSD、リムーバブルディスクなどに記録して、保存したり、配布したりすることが可能である。また、上記のプログラムをインターネットや電子メールなど、ネットワークを通して提供することも可能である。

20

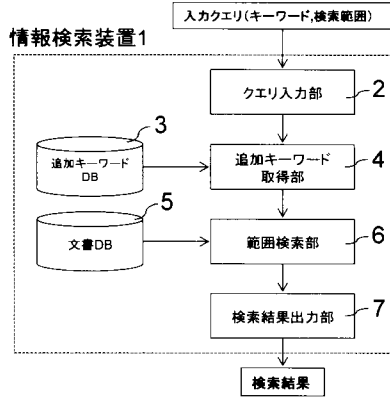
【符号の説明】

30

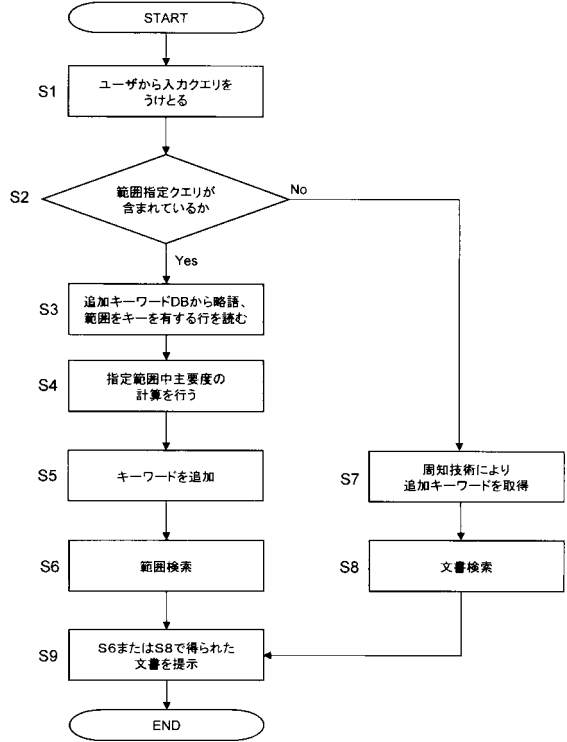
【0071】

- 1 ... 情報検索装置
- 3 ... 追加キーワードDB(キーワードデータベース)
- 4 ... 追加キーワード取得部(追加キーワード取得手段)
- 6 ... 範囲検索部(範囲検索手段)

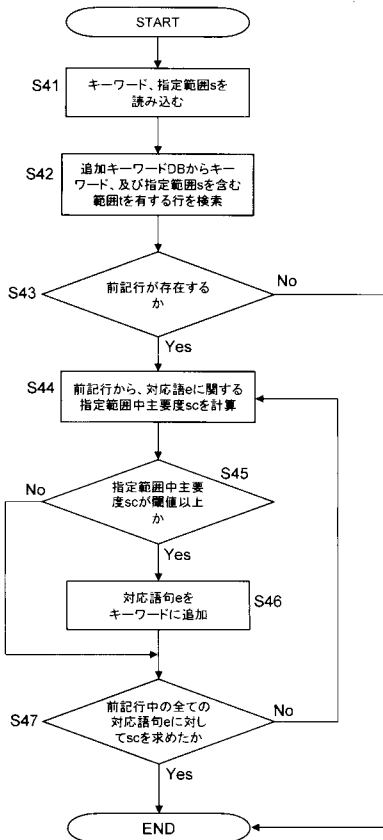
【 図 1 】



【 図 2 】



【 図 3 】



【 図 4 】

略語w	範囲 t(時間)	対応語e	主要度 c
ES	2010年1月~12月	エントリーシート	0.8
ES	2010年1月~12月	従業員満足度	0.3
⋮	⋮	⋮	⋮
ES	2009年1月~12月	エントリーシート	0.7
⋮	⋮	⋮	⋮
CC	1995年1月~12月	キャッチコピー	0.7

【 図 5 】

略語w	範囲 t(地理)		対応語e	主要度 c
	北東座標	南西座標		
ES	緯度35度00分00秒 経度139度00分00秒	緯度34度50分00秒 経度138度50分00秒	エントリーシート	0.8
ES	緯度35度00分00秒 経度139度00分00秒	緯度34度50分00秒 経度138度50分00秒	従業員満足度	0.3
⋮	⋮	⋮	⋮	⋮
ES	緯度29度00分00秒 経度135度00分00秒	緯度28度50分00秒 経度134度50分00秒	エントリーシート	0.7
⋮	⋮	⋮	⋮	⋮
CC	緯度35度00分00秒 経度139度00分00秒	緯度34度50分00秒 経度138度50分00秒	キャッチコピー	0.7

フロントページの続き

- (72)発明者 安田 宜仁
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 小池 義昌
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
- Fターム(参考) 5B075 NR05 QP03 QP05