

明 細 書

発明の名称：

インデックスを用いた情報検索システム及び情報検索方法

技術分野

[0001] 本発明は、情報検索技術に関し、特に、収集された情報に基づいてデータベースを検索するためのインデックスを作成するための装置及び方法、並びに該方法を実行するためのプログラムに関する。

背景技術

[0002] 現在、コンピュータシステムを用いた様々な情報検索サービスが提供されている。例えば、いわゆるインターネット検索は、インターネット上に分散する膨大な情報（リソース）を検索するためのWebサイト上の情報検索サービスであり、典型的には、サービス提供者における検索エンジンと呼ばれるサーバプログラムが、ユーザから与えられた検索クエリに基づいてインデックス検索を行って、ユーザにその検索の結果を提供する。このような情報検索サービスでは、典型的には、検索エンジンが、予め、クローラと呼ばれるロボットエージェントによりインターネット上のWebページを巡回訪問させ、訪問したWebページの情報を収集、解析し、これにより、データベースを検索するためのインデックスを作成する。ユーザは、Webブラウザを介して、検索サイトにアクセスし、検索したいテキスト（検索クエリ）を入力すると、検索エンジンは、検索クエリを解析して、インデックスを検索し、所定のランキングアルゴリズムに基づいて抽出された結果をスコアリングして、これらをスコア順にソートしたものを検索結果として、ユーザに提供する。

[0003] このような従来のインデックス型検索エンジンは、予め、クローラによって収集された情報に従ってそのままインデックスを作成しているため、ユーザは、インデックス化された情報のみに基づいた検索結果を得ることになる。このため、例えば下記特許文献1は、インデックス基準に従って1つ又は

それ以上の横断検索にインデックスを付与し、クエリをインデックス基準の1つ又はそれ以上と比較し、クエリをインデックス基準とマッチさせ、クエリに関するインデックス基準と関連付けられた1つ又はそれ以上のインデックス付き横断検索からの結果を提供する技術を提案している。

[0004] 一方、下記特許文献2は、高精度な翻訳文を得るために、原文をフレーズに分割し、パラフレーズ情報に基づいて1以上の各フレーズと対になるパラフレーズを取得し、1以上のフレーズと1以上のパラフレーズを用いてフレーズの集合を含むパラフレーズラティスを取得する機械翻訳装置を開示している。

先行技術文献

特許文献

[0005] 特許文献1：特開2015-133151号公報

特許文献2：特開2011-180823号公報

発明の概要

発明が解決しようとする課題

[0006] 上述のような情報検索サービスは、人々のあらゆる社会活動の中に深く浸透しており、例えば、医療分野においても例外ではない。

[0007] 例えば、新しい治療法や薬の候補が標準治療として認められるためには、臨床試験を実施し、そのような治療法や薬の有効性及び安全性などを確認する必要がある。典型的には、医師などの医療従事者は、自身の専門的知識に基づき、治療の内容や意義、副作用など想定されるあらゆる事柄を考慮して、臨床試験計画を策定する。このため、医療従事者は、臨床試験計画の策定にあたって、被験者の病気に関連する各種の情報（例えば、病名や薬名など）を漏れのないように収集する必要がある。例えば、医療従事者は、関連情報を手掛かりにしてデータベースシステムから他の臨床試験の情報や学術論文などを抽出し、それに基づいて実施手順を決定していく。

[0008] しかしながら、実際問題として、一つの特定の概念に、一つの特定の語な

いし表現（フレーズ）が一意に定められているわけではなく、通常、一つの特定の概念に対して、複数のフレーズが存在する。特に、医療分野のような専門分野においては、一つの特定の概念に対して、複数のフレーズ、とりわけ、それらの類義語や同義語、関連語、略語、及び外国語、さらには、これらの語に基づく派生語など（以下、それぞれを「関連フレーズ」といい、これらをまとめて「関連フレーズ群」という。）が多く存在している。

[0009] したがって、仮に、医療従事者が、偏った関連情報（本人は、通常、該情報が偏っているという認識はない。）に基づいてデータベースシステムを検索した場合、適切な情報がヒットしない可能性があり、これにより、臨床試験計画の策定に影響を及ぼす可能性もある。

[0010] しかしながら、既存のインデックス型検索エンジンは、上述したように、収集された情報から直接的に得られる見出し語のみに基づいてインデックスを作成しているため、このような見出し語のみに基づいた検索結果を得ることができるに過ぎないか、或いは、検索クエリとして与えられた特定の語がインデックスに見出し語として存在しない場合、該特定の語に近いと思われる代替語（例えば誤表記に対する正しい表記）を選択して、それに基づいて検索を実行するに過ぎず、検索クエリに基づいて与えられる語の範囲が制限的であった。

[0011] また、特許文献 1 に示される情報検索技術は、与えられた検索語に基づいて、インデックス基準に関連付けられたインデックスを横断的に検索し、その結果を提供するものであるが、与えられた検索クエリにのみ関連する語がインデックス化されるに過ぎず、やはり、検索クエリに基づいて与えられる語の範囲が制限的であった。

[0012] さらに、特許文献 2 に示される機械翻訳装置は、原文のフレーズをパラフレーズすることで、フレーズの集合を得ているものの、該得られるフレーズの集合は、単に、原文のフレーズに直接対応するパラフレーズを用いているに過ぎない。また、特許文献 2 は、機械翻訳技術に向けられており、検索エンジンに用いられるようなインデックスの作成技術にそのまま応用すること

はできない。

[0013] したがって、医療分野のような専門分野においては、特定のフレーズに基づいて、関連フレーズ群を幅広く効率的に検索する技術が望まれるところ、上記従来の情報検索技術は、十分とはいえなかった。

[0014] さらに、ユーザは、検索結果を提供されるにあたり、該検索結果が抽出されることになった根拠が認識できれば、次の検索に利用することができ、便利である。とりわけ、検索結果が検索クエリのフレーズから直接的にヒットしたものでなく、その関連フレーズによってヒットしたものである場合、ユーザは、どのような関連フレーズでヒットしたかを知ることが必要になる。

[0015] そこで、本発明は、インデックス型検索エンジンにおいて、収集された情報を効率的に検索することができるインデックスを作成する技術を提供することを目的とする。

[0016] 具体的には、本発明の目的の一つは、収集されたリソースに含まれる特定のフレーズのみならず、その類義語や同義語、関連語、略語、及び外国語、さらには、これらの語に基づく派生語などにまで拡張したフレーズを見出し語とするインデックスを作成するための技術を提供することである。

[0017] また、本発明の目的の一つは、ユーザからの検索クエリが、該インデックス作成の基となった収集された情報に含まれるフレーズ以外のフレーズからなるものであっても、該情報が効果的に検索され、抽出されるようにする技術を提供することである。

[0018] さらに、本発明の目的の一つは、該ユーザが、該結果が抽出されることになった根拠を認識できるようにする技術を提供することである。

課題を解決するための手段

[0019] 上記課題を解決するための本発明は、以下に示す発明特定事項乃至は技術的特徴を含んで構成される。

[0020] ある観点に従う本発明は、テキスト文を含むリソースに基づいて、データベースを検索するためのインデックスを作成する装置である。前記装置は、例えば、特定の概念ごとに、相互に関連付けられた複数のフレーズを含む辞

書と、前記リソースの入力を受け付ける入力インターフェース部と、前記受け付けたリソースのテキスト文から複数のフレーズをトークンとして切り出すための処理を行うテキスト処理部と、前記テキスト文に従って、前記切り出した複数のトークンどうしの接続関係を表現した有向グラフ（ラティス）を生成する有向グラフ生成部と、探索対象フレーズに基づいて前記有向グラフ内を探索し、前記探索対象フレーズが前記辞書中に発見される場合に、前記有向グラフ内における前記探索対象タームが顕出した箇所を特定する有向グラフ探索部と、前記有向グラフ内における複数のトークンに基づいて、前記リソースとの関連付けを示すインデックスを作成するインデックス作成部と、を備え得る。前記有向グラフ探索部は、前記有向グラフ内における少なくとも1つのトークンを前記探索対象フレーズに設定し、前記設定した探索対象フレーズに基づいて前記辞書を参照するように構成され得る。また、前記有向グラフ生成部は、前記辞書に基づいて、前記探索対象フレーズに関連付けられた少なくとも1つのフレーズに基づく新たなトークンを前記有向グラフに追加することによって、前記有向グラフを更新するように構成され得る。

[0021] これにより、前記装置は、リソースのテキスト文に含まれるフレーズのみならず、該フレーズに対するパラフレーズを含む関連フレーズ群のそれぞれを該リソースの見出し語とするインデックスを作成することができるようになる。また、前記装置は、テキスト文に基づいて生成した有向グラフを更新しながら探索を進めていくので、該テキスト文に含まれるフレーズに対する直接的なパラフレーズのみならず、更新された時点の有向グラフに含まれるフレーズに対する間接的ないしは派生的なパラフレーズを関連フレーズ群とすることができる。

[0022] 前記テキスト処理部は、前記テキスト文に基づいて所定の構文解析を行って前記複数の語を前記トークンとして切り出すように構成され得る。これにより、前記テキスト文に基づく有向グラフを生成することができるようになる。

- [0023] 前記有向グラフ生成部は、前記複数のトークンのうちの隣接するトークンどうしをノードによって接続した前記有向グラフを生成するように構成され得る。前記有向グラフは、前記トークンをエッジのラベルとして定義し、前記エッジを前記ノードによって接続することにより表現されるデータ構造であり得る。
- [0024] 前記有向グラフ生成部は、前記有向グラフ内における前記顕出した箇所の前後のノードに、前記辞書中に発見された前記探索対象フレーズに関連付けられた前記少なくとも1つのフレーズを接続するように構成され得る。
- [0025] 前記有向グラフ探索部は、前記有向グラフ生成部によって更新された有向グラフを探索するように構成され得る。これにより、探索される有向グラフは、動的に変化していくため、前記テキスト文に含まれるフレーズのみならず、パラフレーズ化された新たなフレーズについても、さらに、パラフレーズ化されることになる。
- [0026] また、前記有向グラフ探索部は、前記更新された有向グラフにおいて既に探索し終えたトークンに対する再探索が回避されるように、前記更新された有向グラフ内を探索し得る。
- [0027] 前記有向グラフは、前記複数のトークンのうちの隣接するトークンどうしを接続する複数のノードを含み得る。そして、前記有向グラフ探索部は、前記探索をし終えたトークンに関連付けられたノードを探索済みノードとして記憶し、前記探索済みノード以外のノードに接続されたトークンに対して再探索し得る。
- [0028] また、前記装置は、前記辞書に含まれる前記複数のフレーズに基づいて構成される木構造を示すデータ配列をさらに備え得る。前記木構造は、前記複数のフレーズのそれぞれに対応する順序付きノードを含むトライ木又はプレフィックス木であり得る。そして、前記有向グラフ探索部は、前記木構造を参照しながら、前記有向グラフ内を探索し得る。
- [0029] また、前記有向グラフ探索部は、前記探索対象フレーズが前記木構造における少なくとも1つの前記順序付きノードに対応する少なくとも1つのフレ

ーズと一致する場合に、前記有向グラフ内に前記探索対象フレーズが顕出したと判断し得る。

[0030] また、前記有向グラフ生成部は、前記一致する探索対象フレーズに関連付けられた、前記辞書における少なくとも1つのフレーズに基づく新たなトークンを前記有向グラフに追加することによって、前記有向グラフを更新し得る。

[0031] 前記インデックス作成部は、更新された前記有向グラフ内における複数のトークンに基づいてN-gramを抽出し、前記抽出したN-gramに基づいて、前記インデックスを作成するように構成され得る。これにより、リソースのテキスト文に含まれるフレーズ以外のフレーズをN-gramとするインデックスが作成されることになる。

[0032] また、前記インデックス作成部は、前記N-gramとして、単語ベースのN-gram（単語N-gram）を抽出するように構成され得る。かかる単語N-gramは、前記有向グラフ内におけるトークンに基づいているため、関連フレーズ群に基づく検索に適したインデックスが作成されることになる。

[0033] また、別の観点に係る本発明は、上記の装置により作成されたインデックスと、前記インデックスに関連付けられた前記リソースに関するリソース情報ファイルと、検索クエリに基づいて前記インデックスを検索し、該検索の結果に基づいて前記リソース情報ファイルから所定のリソースを特定し、前記特定した所定のリソースを示す検索結果を出力するクエリサーバとを備える検索装置ないしは検索エンジンである。前記クエリサーバは、前記特定した所定のリソースから得られる、前記所定のリソースにおけるフレーズに関連する所定の関連フレーズを含む有向グラフに基づいて、前記所定のリソースが特定された根拠となった根拠フレーズ及び前記フレーズに関連付けられた前記所定のリソースにおける元フレーズを特定する。また、前記検索結果は、前記根拠フレーズ及び前記元フレーズを含む。これにより、ユーザは、検索クエリに対する検索結果として、検索されたリソースに加え、該リソー

スが検索された根拠となった根拠フレーズ及びこれに関連する元フレーズを認識することができるようになる。

[0034] さらに別の観点に係る本発明は、テキスト文を含むリソースに基づいて、データベースを検索するためのインデックスを作成する装置の作動方法である。前記方法は、例えば、特定の概念ごとに、相互に関連付けられた複数のフレーズを含む辞書を、前記装置のプロセッサによってアクセスされるメモリに展開することと、入力インターフェース部を介して、前記リソースの入力を受け付けることと、前記受け付けたリソースのテキスト文から複数の語をトークンとして切り出すためのテキスト処理を行うことと、前記テキスト文に従って、前記切り出した複数のトークンどうしの接続関係を表現した有向グラフを生成することと、探索対象フレーズに基づいて前記有向グラフ内を探索することと、前記有向グラフ内における複数のトークンに基づいて、前記リソースとの関連付けを示すインデックスを作成することと、を含み得る。前記有向グラフ内を探索することは、前記有向グラフ内における少なくとも1つのトークンを前記探索対象フレーズに設定し、前記設定した探索対象フレーズに基づいて前記辞書を参照し、前記探索対象フレーズが前記辞書中に発見される場合に、前記有向グラフ内における前記探索対象フレーズが顕出した箇所を特定することを含み得る。また、前記有向グラフを生成することは、前記辞書に基づいて、前記探索対象フレーズに関連付けられた少なくとも1つのフレーズに基づく新たなトークンを前記有向グラフに追加することによって、前記有向グラフを更新することを含み得る。

[0035] さらに別の観点に係る本発明は、検索クエリに基づいてインデックスを検索し、該検索の結果に基づいてリソース情報ファイルから少なくとも1つのリソースを特定する検索装置の作動方法である。前記作動方法は、前記特定した少なくとも1つのリソースから得られる、前記少なくとも1つのリソースにおけるフレーズに関連する所定の関連フレーズを含む有向グラフに基づいて、前記少なくとも1つのリソースが特定された根拠となった根拠フレーズ及び前記フレーズに関連付けられた前記少なくとも1つのリソースにおけ

る元フレーズを特定することと、前記所定のリソース並びに前記根拠フレーズ及び前記元フレーズを示す情報を検索結果として出力することと、を含み得る。

[0036] さらに別の観点に従う本発明は、コンピューティングデバイスに、各前記方法を実現させるためのコンピュータプログラム又はこれを非一時的に記録したコンピュータ可読記録媒体であり得る。

[0037] なお、本明細書等において、手段とは、単に物理的手段を意味するものではなく、その手段が有する機能をソフトウェアによって実現する場合も含む。また、1つの手段が有する機能が2つ以上の物理的手段により実現されても、2つ以上の手段の機能が1つの物理的手段により実現されても良い。

発明の効果

[0038] 本発明によれば、インデックス型検索エンジンにおいて、収集された情報を効率的に検索することができるインデックスが作成される。したがって、インデックス作成の基となった収集された情報に含まれるフレーズ以外のフレーズによっても、該情報を検索することができるようになる。

[0039] また、本発明によれば、ユーザは、検索結果の根拠を認識できるようになる。とりわけ、検索クエリに対してヒットしたリソースが、検索クエリのフレーズから直接的にヒットしたリソースではなく、その関連フレーズによってヒットしたリソースである場合であっても、ユーザは、検索によりヒットした根拠となった関連フレーズを認識できるようになる。

[0040] 本発明の他の技術的特徴、目的、及び作用効果乃至は利点は、添付した図面を参照して説明される以下の実施形態により明らかにされる。

図面の簡単な説明

[0041] [図1]図1は、本発明の一実施形態に係る情報検索システムの概略構成を示すブロックダイアグラムである。

[図2]図2は、本発明の一実施形態に係る検索エンジンにおけるデータベースの一例を説明するための図である。

[図3]図3は、本発明の一実施形態に係る検索エンジンにおける辞書の一例を

説明するための図である。

[図4]図4は、本発明の一実施形態に係る検索エンジンのインデクサの概略的構成を示すブロックダイアグラムである。

[図5]図5は、本発明の一実施形態に係る検索エンジンのテキスト処理部に入力されるテキスト文の一例を示す図である。

[図6]図6は、本発明の一実施形態に係る検索エンジンによるインデックスの作成処理を説明するためのフローチャートである。

[図7]図7は、本発明の一実施形態に係る検索エンジンにより生成された初期有向グラフの一部の例を示している

[図8]図8は、本発明の一実施形態に係る検索エンジンによる有向グラフの探索・更新処理を説明するためのフローチャートである。

[図9]図9は、本発明の一実施形態に係る検索エンジンにより有向グラフの一部が展開される処理過程の一例を示す図である。

[図10]図10は、本発明の一実施形態に係る検索エンジンにより有向グラフが展開される処理過程の一例を説明するための図である。

[図11]図11は、本発明の一実施形態に係る検索エンジンによる検索処理を説明するためのフローチャートである。

[図12A]図12Aは、本発明の一実施形態に係る検索エンジンにより検索されたりソースの有向グラフから検索根拠を特定するための処理過程の一例を説明するための図である。

[図12B]図12Bは、本発明の一実施形態に係る検索エンジンにより検索されたりソースの有向グラフから検索根拠を特定するための処理過程の一例を説明するための図である。

[図13A]図13Aは、本発明の一実施形態に係る情報検索システムにおけるクライアント上に表示される検索結果画面の一例を示す図である。

[図13B]図13Bは、本発明の一実施形態に係る情報検索システムにおけるクライアント上に表示される検索結果画面の一例を示す図である。

[図13C]図13Cは、本発明の一実施形態に係る情報検索システムにおけるク

ライアント上に表示される検索結果画面の一例を示す図である。

[図14]図14は、本発明の一実施形態に係る検索エンジンにおける辞書の他の例を説明するための図である。

[図15]図15は、本発明の一実施形態に係る検索エンジンにおけるトライ木の一例を説明するための図である。

[図16A]図16Aは、本発明の一実施形態に係る検索エンジンによる有向グラフの探索・更新処理の他の例を説明するためのフローチャートである。

[図16B]図16Bは、本発明の一実施形態に係る検索エンジンによる有向グラフの探索・更新処理の他の例を説明するためのフローチャートである。

[図17]図17は、本発明の一実施形態に係る検索エンジンにより有向グラフの一部が展開される処理過程の他の例を示す図である。

[図18]図18は、本発明の一実施形態に係る検索エンジンのハードウェア構成の一例を示すブロックダイアグラムである。

発明を実施するための形態

[0042] 以下、図面を参照して本発明の実施の形態を説明する。ただし、以下に説明する実施形態は、あくまでも例示であり、以下に明示しない種々の変形や技術の適用を排除する意図はない。本発明は、その趣旨を逸脱しない範囲で種々変形（例えば各実施形態を組み合わせる等）して実施することができる。また、以下の図面の記載において、同一又は類似の部分には同一又は類似の符号を付して表している。図面は模式的なものであり、必ずしも実際の寸法や比率等とは一致しない。図面相互間においても互いの寸法の関係や比率が異なる部分が含まれていることがある。

[0043] 図1は、本発明の一実施形態に係る情報検索システムの概略構成を示すブロックダイアグラムである。同図に示すように、本実施形態の情報検索システム1は、コンピュータネットワーク10を介して通信可能に接続される、少なくとも1つの検索エンジン20と少なくとも1つのクライアント30とを含み構成される。

[0044] コンピュータネットワーク10は、典型的には、IPベースのコンピュー

タネットワークを含むが、これに限られない。例えば、コンピュータネットワーク 10 は、ノード間通信を可能とするあらゆるプロトコルのネットワークを適用することができる。

[0045] 検索エンジン 20 は、クライアント 30 に情報検索サービスを提供するコンピュータシステムであり、例えば、1 以上の汎用のコンピューティングデバイスにより実現され得る。すなわち、検索エンジン 20 は、コンピューティングデバイス上の 1 つ又は複数の CPU（プロセッサ）が、例えば、所定のオペレーティングシステム（OS）上で本実施形態の検索エンジンプログラムを実行することにより、他のハードウェア要素と相俟って、具現化される仮想マシンである。検索エンジン 20 は、仮想化技術により実現されても良い。検索エンジン 20 が実現されるコンピューティングデバイスのハードウェア構成は、図 18 に例示されるが、既知であるため、ここではその詳細を省略する。本実施形態では、検索エンジン 20 は、インデックス型検索エンジンであり、例えば、インデックスの作成及び更新の機能と、該インデックスに基づくデータベース検索の機能を有する。図 1 に示すように、検索エンジン 20 は、例えば、クローラ 21 と、辞書 22 と、インデクサ 23 と、データベース 24 と、クエリサーバ 25 とを含み、構成される。なお、情報検索をしようとするユーザからみると、検索エンジン 20 は、インデックスに基づくデータベース検索機能を実現する仮想マシン（検索装置ないしは狭義の検索エンジン）である。

[0046] クローラ 21 は、Web 上の多種多様なリソース（例えば、Web ページ、文書、画像、プログラムなど）を自動的に収集するための自動巡回エージェントプログラムである。具体的には、クローラ 21 は、Web ページ内のリンクを辿りながら、Web 上のサイトを周期的に訪問して、サイト内にあるリソースを収集する。例示として、クローラ 21 は、特定の専門分野に属するサイト内のリソースを収集するものであっても良い。クローラ 21 は、例えば、収集したリソースをストレージ（図示せず）に一時的に記憶させても良い。本実施形態の情報検索システム 1 においては、既知のクローラ 21

を適用することができ、したがって、ここではその詳細は省略する。

[0047] 辞書 2 2 は、ある特定の概念についての複数のフレーズを相互に関連付けて記憶した 1 つ又は複数のファイルないしはテーブルであるが、これに限られるものではない。例えば、辞書 2 2 は、静的なファイルないしはテーブルとして構成され、メモリ上に常駐するのではなく、所定のアルゴリズムに従って、フレーズ群の関係が構築される辞書モデル（例えばニューラルネットワークモデルなど）として動的に構成されても良い。本開示では、フレーズは、1 つの単語及び 2 つ以上の単語の並びからなる語ないし表現をいうものとする。また、相互に関連付けられる複数のフレーズを、関連フレーズ群と称するものとし、一例として、類義語、同義語、関連語、略語、及び対応外国語、さらには、これらの語に基づく派生語などを含む。辞書 2 2 は、例えば、入手可能な情報ソースに基づいて、手動で、半自動で、又は全自動で構築される。例えば、医療分野であれば、アメリカ国立がん研究所（National Cancer Institute : N C I）が提供する様々な辞書は、辞書 2 2 の構築のために用いられる。辞書 2 2 は、後述するインデクサ 2 3 がインデックス 2 4 1 を作成する際に参照される。

[0048] インデクサ 2 3 は、クローラ 2 1 によって収集されたリソース（例えば、リソースに関する情報）をデータベース 2 4 に登録するとともに、該リソースの内容（典型的には、テキスト文）を解析し、該解析の結果に従ってインデックス 2 4 1 を作成し、更新する。本実施形態のインデクサ 2 3 は、概略的には、リソースに含まれるテキスト文をいくつかのトークンに分解し、これらに基づいて有向グラフ（ラティス）を作成し、所定のフレーズに従って、辞書 2 2 を参照しながら有向グラフ内の探索し、所定の条件の下、有向グラフを展開し、更新していく。更新された有向グラフは、新たな探索対象の有向グラフとなる。その後、インデクサ 2 3 は、展開された有向グラフに基づいて、所定の手法（例えば N - g r a m 手法など）で見出し語を切り出して、これを収集したリソースと関連付けて（索引付けて）インデックス 2 4 1 を作成及び／又は更新する。また、インデクサ 2 3 は、収集したリソース

と作成した有向グラフとを関連付けて、データベース24に登録する。

[0049] データベース24は、クローラ21によって収集されたリソースに関する情報を格納し、そのインデックス241を含み構成される。また、データベース24は、インデックス241を作成するために用いた有向グラフを格納する。図2は、本発明の一実施形態に係る検索エンジンにおけるデータベースの一例を説明するための図である。図2(a)は、リソースに関する情報の一例を示す。リソースに関する情報は、例えば、リソースを識別するリソースID、リソース名称、URL、及び有向グラフを識別するための有向グラフIDなどを含む構造ファイルである。本開示では、リソースに関する情報を格納した構造ファイルをリソース情報ファイルと称するものとする。インデックス241は、インデクサ23によって作成され、更新される索引情報を含む構造ファイルである。図2(b)は、インデックス241の一例を示す。典型的には、インデックス241は、転置インデックスと呼ばれるデータ構造を有する。索引情報は、例えば、N-gram手法により有向グラフ中の語の並びから切り出された見出し語である。インデックス241内の見出し語は、リソース（すなわち、リソースID）と関連付けられ、これにより、データベース24のリソース情報ファイルの個々のデータセットと関連付けられる。なお、本開示では、インデックス241は、データベース24の一部として構成されているが、これに限られるものではなく、それぞれ別体として構成されても良い。

[0050] 図1に戻り、クエリサーバ25は、クライアント30から与えられる検索クエリを受け付けて、インデックス241を検索し、その検索結果に基づいてデータベース24から該当するリソースに関連する情報を抽出し、これを検索結果としてクライアント30に提供する。クエリサーバ25は、例えば、検索クエリと抽出したリソースとの間の適合度を所定の手法により算出し、その算出結果に応じて抽出したリソースを順位付けした検索結果を提供しても良い。本実施形態では、クエリサーバ25は、クライアント30からの検索クエリに対する検索結果を、クライアント30に直接返しているが、こ

れに限られるものではない。例えば、クエリサーバ25は、検索結果を、所定のAPI (Application Program Interface) を介して、他のアプリケーションプログラムに渡し、他のアプリケーションプログラムにより該検索結果に対する所定の情報処理を行った結果をクライアント30に返すようにしても良い。本実施形態の情報検索システム1においては、既知のクエリサーバ25を適用することができる。

[0051] クライアント30は、典型的には、ユーザが所有するコンピューティングデバイスであり、例えば、パーソナルコンピュータであり得る。クライアント30は、例えば、コンピュータネットワーク10を介して、検索エンジン20にアクセスするための、ユーザインターフェースとして機能するビューワを含む。ビューワは、例えば、Webブラウザであり得るが、これに限られるものではない。クライアント30は、例えば、SSLなどのセキュアな通信技術を利用することによって、コンピュータネットワーク10を介して、検索エンジン20との通信セッションを構築し得る。クライアント30は、1つ又は複数のCPU (プロセッサ) 及びメモリなどを含み構成されるが、そのハードウェア構成は既知であるため、ここではその詳細は省略する。

[0052] 図3は、本発明の一実施形態に係る検索エンジンにおける辞書の一例を説明するための図である。図3に示すように、辞書22は、特定の概念ごとに、1つ又は複数の語からなる複数のフレーズのそれぞれを相互に関連付けて記憶するテーブルとして構成されている。例示的に、図3においては、フレーズ 'p a r p'、フレーズ 'p o l y | a d p | r i b o s e | p o l y | m e r a s e' 及びフレーズ 'ポリ | a d p | リボース | ポリメラーゼ' が、それぞれ相互に関連付けられて登録されている。記号 '|' は、フレーズ中の個々の単語を区切るために追加されたデリミタである。このように、フレーズを構成する単語をデリミタで区切ることにより、後述するように、フレーズを構成する個々の単語を有向グラフ中の個々のトークンとして扱えるようになる。なお、ここでは、3つのフレーズが相互に関連付けられた1つのデータセットとして示されているが、これに限られるものでなく、その

数に制限はない。

- [0053] 図4は、本発明の一実施形態に係る検索エンジンのインデクサの概略的構成を示すブロックダイアグラムである。同図に示すように、本実施形態のインデクサ23は、例えば、入力インターフェース部231と、テキスト処理部232と、有向グラフ生成部233と、有向グラフ探索部234と、インデックス作成部235とを含み構成される。
- [0054] 入力インターフェース部231は、例えば、クローラ21とテキスト処理部232との間のインターフェースとして機能する。入力インターフェース部231は、クローラ21と通信して、クローラ21が収集したリソースに含まれる例えば図5に示すようなテキスト文をテキスト処理部232に入力する。
- [0055] テキスト処理部232は、入力されたテキスト文について例えば形態素解析を行って、複数のトークンを切り出す。テキスト処理部232は、典型的には、日本語や英語などの各国の言語に対応した言語解析エンジン（図示せず）を含み得る。言語解析エンジンは、既知のものを用いることができる。
- [0056] 有向グラフ生成部233は、テキスト処理部232によって切り出された複数のトークンに基づく有向グラフを生成する。具体的には、有向グラフ生成部233は、テキスト処理部232によって切り出された複数のトークンのそれぞれについて、隣接するトークンをノードによって接続し、これによって、複数のトークンどうしの接続関係を表現した有向グラフ（例えば図7（a））を生成する。つまり、一般的に、有向グラフは、エッジ及びノードにより表現されるところ、本実施形態の有向グラフは、個々のトークンをエッジのラベルとして定義し、かかるラベル付きエッジ及びノードにより表現されたものということができる。このように、有向グラフ生成部233によってテキスト文から直接生成された有向グラフは、先頭ノードから最終ノードまで1つのパスのみを含む。なお、本開示で図示される有向グラフは、理解容易のために描かれたもので、実際上は、コンピューティングデバイス内部では、典型的には、プロセッサにより解釈可能なある種のデータ構造とし

て扱われる。有向グラフ生成部233は、後述する有向グラフ探索部234による探索の結果に従って、有向グラフの接続関係を動的に変更することにより、有向グラフを展開し、更新していく。

[0057] 有向グラフ探索部234は、探索の対象となるフレーズ（以下「探索対象フレーズ」という。）を設定し、これに基づいて、辞書22を参照しながら有向グラフ内を探索する。探索対象フレーズは、有向グラフ中の任意の1つ又は複数のトークンであり、例えば、先頭のトークンから順に選択され、設定される。より具体的には、本実施形態では、有向グラフ探索部234は、設定された探索対象フレーズに基づいて、有向グラフ内をエッジ及びノードに従って順番に辿りながら、該探索対象フレーズが辞書22に登録されているか否かを判定する。有向グラフ探索部234は、該探索対象フレーズが辞書22に登録されていると判定する場合、すなわち、該探索対象フレーズを辞書22中に発見した場合、有向グラフ探索部234は、有向グラフ内における該探索対象フレーズが顕出した箇所を特定し、特定した箇所を有向グラフ生成部233に通知する。これを受けて、有向グラフ生成部233は、有向グラフ内における特定された箇所の前後のノードに、辞書22における該探索対象フレーズに関連付けられたフレーズを追加的に接続することにより、有向グラフを更新する。したがって、新たにフレーズが追加され、更新された有向グラフは、例えば図9（b）などに示すように、テキスト文の先頭ノードから最終ノードまでの間に、複数のパスが存在し得ることになる。

[0058] インデックス作成部235は、有向グラフ探索部234が探索を終えることにより得られる最終的な有向グラフに基づいて、収集されたりソースとの関連付けを示すインデックスを作成する。具体的には、インデックス作成部235は、更新された最終的な有向グラフに従い、そこから単語（トークン）ベースのN-gram（以下「単語N-gram」という。）（すなわち、単語の並び）を抽出し、該抽出した単語N-gramを見出し語としてリソースに関連付けることにより、インデックス241を作成する。本実施形態では、 $N \leq 3$ （ただし、Nは正数）、すなわち、1～3個の単語の並び（

すなわち、1-gram、2-gram及び3-gram)がインデックス241の見出し語として用いられる。また、本実施形態は、単語N-gramを用いているが、これに限られるものではなく、従前のN-gramのように、文字単位の並び(すなわち、1文字、2文字、…)を用いることを排除する趣旨ではない。

[0059] 図6は、本発明の一実施形態に係る検索エンジンによるインデックスの作成処理を説明するためのフローチャートである。かかる処理は、例えば、コンピューティングデバイス上のプロセッサが、本実施形態に係る所定の検索エンジンプログラムを実行することにより、実現される。

[0060] 図6に示すように、検索エンジン20は、クローラ21によって収集されたリソースに含まれるテキスト文を受信する(S601)。例えば、入力インターフェース部231は、クローラ21によって収集されたリソースを受信すると、リソースに含まれるテキスト文をテキスト処理部232に入力する。検索エンジン20は、テキスト文を受信すると、テキスト文に対して形態素解析を行って、テキスト文をトークン化、すなわち、テキスト文から複数のトークンを切り出す(S602)。例えば、テキスト処理部232は、テキスト文の言語を判定し、該言語に応じた形態素解析処理により、テキスト文から複数のトークンを切り出す。

[0061] 次に、検索エンジン20は、切り出した複数のトークンに基づいて有向グラフを生成する(S603)。例えば、有向グラフ生成部233は、テキスト処理部232によって切り出された複数のトークンのそれぞれについて、該トークンをエッジのラベルとして、隣接するエッジをノードによって接続し、これによって、複数のトークンどうしの接続関係を表現した有向グラフを生成する。テキスト文から最初に生成された有向グラフを、ここでは、初期有向グラフと呼ぶものとする。図7は、図5に示した入力されたテキスト文から生成された初期有向グラフの一部を示している。すなわち、図7(a)に示すように、本実施形態の有向グラフは、個々のトークンをエッジのラベルとして定義し、かかるラベル付きエッジ及びノードにより表現されてい

る。初期有向グラフは、図7に示すように、先頭ノードから続くパスは1つのみである。

[0062] なお、本実施形態では、上述したように、個々のトークンを、有向グラフにおけるエッジのラベルとして定義した有向グラフが示されているが、これに限られるものではない。例えば、図7(b)に示すように、トークンもまたノードと定義し（つまり、2種類のノードが存在する。）、これら2種類のノードを単なるエッジで接続した有向グラフであっても良い。なお、このようなグラフは、自然言語処理の分野ではラティスと称されることもあり、本開示では、これを区別するものではない。

[0063] 図6に戻り、検索エンジン20は、辞書22を参照しながら有向グラフを探索し、探索の結果に従って該有向グラフを展開し、更新する(S604)。すなわち、検索エンジン20は、有向グラフの中から所定の語を探索の対象となる探索対象フレーズとして順番に設定し、これに基づいて、有向グラフ内をエッジ及びノードに従って探索しながら、該探索対象フレーズを辞書22中に発見するごとに、該発見した探索対象フレーズに関連付けられたフレーズに基づいて、有向グラフを展開し、更新する。なお、有向グラフの探索及び更新処理の詳細は、図8を用いて説明される。

[0064] 上述した有向グラフの探索及び更新処理の後、検索エンジン20は、更新された有向グラフに基づいて、収集されたリソースとの関連付けを示すインデックス241を作成する(S605)。例えば、インデックス作成部235は、更新された最終的な有向グラフに従い、そこから1-gramから3-gramまでの単語の並びをそれぞれ抽出し、該抽出した語の並びを見出し語として、収集されたリソースに関連付けることにより、例えば図2(b)に示したようなインデックス241を作成する。インデックス作成部235はまた、作成したインデックス241に対応するように、例えば図2(a)に示したようなリソース情報ファイルをデータベース24に登録する。これにより、後の検索クエリに対して、検索クエリの関連フレーズの範囲にまで拡張された見出し語からリソースが検索されるようになる。さらに、本実

施形態では、インデックス作成部 235 は、インデックス 241 の作成に際し、収集されたリソースに更新された最終的な有向グラフを関連付け、リソースに関連付けた有向グラフをデータベース 24 に格納する。リソースに関連付けられた有向グラフは、検索クエリに基づいて検索が実行され、該リソースがヒットし、抽出された場合に、該検索によりヒットした根拠（例えば関連フレーズ）を提示するために用いられる。

[0065] 図 8 は、本発明の一実施形態に係る検索エンジンによるインデックスの作成処理を説明するためのフローチャートであり、より具体的には、図 6 に示した有向グラフの探索及び更新の処理の詳細を説明するためのフローチャートである。

[0066] すなわち、図 8 に示すように、例えば、有向グラフ探索部 234 は、まず、現在の有向グラフを探索対象の有向グラフに設定し（S801）、続いて、有向グラフ中の現時点の探索箇所にある所定のトークンを探索対象フレーズとして設定する（S802）。探索の開始時点では、初期有向グラフが探索対象の有向グラフであり、初期有向グラフの先頭ノードに接続されたトークンが探索対象フレーズとして設定される。また、有向グラフの探索に伴い、有向グラフの中の所定のトークンが順次に設定される。探索対象フレーズは、1 つ又はそれ以上のトークンである。

[0067] 有向グラフ探索部 234 は、次に、設定された探索対象フレーズに基づいて、辞書 22 を参照しながら有向グラフ内を探索する（S803）。例えば、有向グラフ探索部 234 は、設定された探索対象フレーズに基づいて、有向グラフ内をエッジ及びノードに従って順番に辿りながら、該探索対象フレーズが辞書 22 に登録されているか否かを判定する（S804）。有向グラフ内の探索は、例えば、KMP（クヌースーモリスープラット）法に従って行われ得るが、これに限られるものではない。

[0068] 有向グラフ探索部 234 は、該探索対象フレーズが辞書 22 に登録されていると判定する場合（S804 の Yes）、すなわち、該探索対象フレーズを辞書 22 中で発見した場合、有向グラフ探索部 234 は、該有向グラフ内

における該探索対象フレーズが顕出した箇所を特定し、該特定した箇所を有向グラフ生成部233に通知する(S805)。有向グラフ生成部233は、これを受けて、該有向グラフ内における該特定された箇所の前後のノードに、辞書22における該探索対象フレーズに関連付けられたフレーズを追加的に接続することにより、該有向グラフを更新する(S806)。フレーズは、典型的には、1つ以上のトークンとして切り出されて、有向グラフに接続される。その後、有向グラフ探索部234は、探索を続行するため、S801の処理に戻る。すなわち、有向グラフ探索部234は、現在の有向グラフを探索対象の有向グラフとして(S801)、その探索箇所を次に進め、そのトークンを新たな探索対象フレーズに設定して、探索処理を進める(S802)。

[0069] 一方、有向グラフ探索部234は、該探索対象フレーズが辞書22中で発見されない場合(S804のNo)、有向グラフの終端ノードに到達したか否かを判定する(S807)。有向グラフ探索部234は、有向グラフの終端ノードに到達していないと判定する場合(S807のNo)、有向グラフ内の探索を続行するため、S802の処理に戻る。一方、有向グラフ探索部234は、有向グラフの最終ノードに到達したと判定する場合(S807のYes)、探索処理を終了する。

[0070] 以上のような処理により、検索エンジン20は、入力されたテキスト文に基づいて検索のためのインデックスを作成するに際して、相互に関連付けられた複数のフレーズが登録された辞書22を用いて、元のテキスト文に含まれるフレーズのみならず、関連フレーズ群にまで拡張した個々のフレーズに基づいてインデックスを作成することができるようになる。

[0071] 次に、検索エンジン20による有向グラフの展開処理の一例を説明する。図9は、本発明の一実施形態に係る検索エンジンにより有向グラフが展開される処理過程の一例を説明するための図である。本例では、検索エンジン20は、図4に示した辞書22に基づいて、図5に示したテキスト文を含むリソースに対してインデックス241を作成する際の有向グラフの展開の様子

を示している。

[0072] まず、検索エンジン20は、上述したように、入力されたテキスト文に基づいて、例えば図7(a)に示したような初期有向グラフを生成する。次に、検索エンジン20は、有向グラフの探索を開始し、その過程で、設定された探索対象フレーズ‘poly|adp|ribose|polymerase’を、辞書22中に発見したとする。検索エンジン20は、有向グラフ内における該探索対象フレーズ‘poly|adp|ribose|polymerase’の箇所前後にある始端ノード及び終端ノードに、辞書22中のフレーズ‘poly|adp|ribose|polymerase’に関連付けられたフレーズ‘parp’及びフレーズ‘ポリ|adp|リボース|ポリメラーゼ’のそれぞれを並列に接続することにより、有向グラフを更新する(図9(a))。

[0073] さらに、検索エンジン20は、有向グラフの探索を進め、有向グラフ内における‘adp|リボース’が探索対象フレーズとして設定されたときに、これを辞書22中に発見する。探索対象フレーズ‘adp|リボース’は、上述した探索の過程で、検索エンジン20によって有向グラフに追加された語である。検索エンジン20は、同様に、有向グラフ内における該探索対象フレーズ‘adp|リボース’の箇所前後にある始端ノード及び終端ノードに、辞書22中のフレーズ‘adp|リボース’に関連付けられたフレーズ‘adenosine|diphosphate|ribose’を並列に接続することにより、有向グラフを更新する(図9(b))。このように、追加されたフレーズを含むパスについても、さらにパラフレーズが追加され得る。

[0074] さらに、検索エンジン20は、有向グラフの探索を進め、有向グラフ内における‘parp|inhibitor’が探索対象フレーズとして設定されたときに、これを辞書22中に発見する。探索対象フレーズ‘parp|inhibitor’は、探索の過程で、検索エンジン20によって‘parp’が有向グラフに追加された結果、隣接する語‘inhibitor’

との結合により得られた語である。検索エンジン20は、同様に、有向グラフ内における該探索対象フレーズ‘p a r p | i n h i b i t o r’の箇所
の前後にある始端ノード及び終端ノードに、辞書22中のフレーズ‘p a r
p | i n h i b i t o r’に関連付けられたフレーズ‘p a r p | 阻害 | 剤’
’を並列に接続することにより、有向グラフを更新する(図9(c))。す
なわち、フレーズ‘p a r p | 阻害 | 剤’は、初期有向グラフにおける元の
パス(p o l y → a d p → r i b o s e → p o l y m e r a s e → i n h i
b i t o r)に並列的に接続されたことになる。

[0075] 検索エンジン20は、以上のような、有向グラフに対する探索動作を行う
ことにより、有向グラフを展開し、更新していく。この結果、更新された有
向グラフは、初期有向グラフにおける元のパスに対して、いくつかの並列的
なパスが形成されることになる。したがってこのような有向グラフに基づい
て見出し語を抽出することにより、元のテキスト文に含まれるフレーズのみ
ならず、関連フレーズ群にまで拡張した個々のフレーズ(パラフレーズ)に
基づいてインデックスを作成することができるようになる。

[0076] 図10は、本発明の一実施形態に係る検索エンジンによる検索処理を説明
するためのフローチャートである。かかる処理は、例えば、コンピューティ
ングデバイス上のプロセッサが、所定の検索エンジンプログラムを実行する
ことにより、実現される。

[0077] 図10に示すように、検索エンジン20は、クライアント30から検索ク
エリを受信する(S1001)。例えば、クライアント30は、例えば、ユ
ーザが、ウェブブラウザ(図示せず)上に表示された検索画面に検索クエリ
を入力することにより、検索エンジン20に検索クエリを送信する。検索エ
ンジンは、受信した検索クエリに対して所定の解析処理を行う(S1002
)。例えば、クエリサーバ25は、クライアント30から受信した検索クエ
リに対して、N-g r a mの検索式を作成する。本実施形態では、N=3を
最大値とするが、これに限られない。したがって、検索クエリが4つ以上の
単語を含む場合には、クエリサーバ25は、3-g r a mの単語を抽出して

、これらのAND検索を行う。例えば、検索クエリが ‘p l a t i n u m
a n d p a r p i n h i b i t o r’ である場合、検索式は、 ‘(p l
a t i n u m a n d p a r p) A N D (a n d p a r p i n h
i b i t o r)’ となる。

[0078] 次に、検索エンジン20は、作成した検索式に基づいて、インデックス241を検索する(S1003)。例えば、クエリサーバ25は、作成した検索式に基づいて、インデックス241を検索し、インデックス241内の見出し語がヒットするか否かをチェックする。見出し語がヒットした場合、クエリサーバ25は、該見出し語に従ってデータベース24を参照し、見出し語に関連付けられたリソースの名称及びその位置情報などの情報を抽出し、該抽出した結果をクライアント30に返答する。例えば、クエリサーバ25は、検索式 ‘(p l a t i n u m a n d p a r p) A N D (a n d p a r p i n h i b i t o r)’ に基づいてインデックス241を検索することにより、図5に示したようなテキスト文を含むリソースがヒットし、これを抽出することになる。クエリサーバ25は、抽出される結果が複数ある場合、これらを所定のスコア計算に基づくそのスコアに従ってランキングしても良い。また、本実施形態では、クエリサーバ25は、見出し語に関連付けられたリソースを抽出するに際し、該リソースに関連付けられた有向グラフに基づいて、該リソースがヒットした根拠となった関連フレーズ及び該リソースのテキスト文のフレーズを特定する(S1004)。検索根拠の特定処理の詳細については、図11を参照して説明される。

[0079] そして、検索エンジン20は、検索根拠となったフレーズを特定した後、該検索根拠となったフレーズを含む検索結果をクライアントに送信する(S1005)。クライアント30は、これを受けて、検索クエリを入力した検索画面に検索結果を表示する。

[0080] 従来の検索エンジンは、収集したリソースのテキスト文に含まれるフレーズのみに基づく見出し語をインデックスに登録していたため、検索により該リソースが抽出されるためには、検索クエリが、該リソースに含まれていた

見出し語を含む必要がある。例えば、従来の検索エンジンは、図5に示したようなテキスト文を含むリソースに対して、フレーズ ‘p a r p i n h i b i t o r’ を見出し語とするインデックスを作成することはない。したがって、クライアント30からの検索クエリが ‘p l a t i n u m a n d p a r p i n h i b i t o r’ である場合、 ‘p a r p i n h i b i t o r’ を含まないリソースは検索結果として提示されないか、仮に他の要因で提示される場合であっても、ランキングが低いものとして提示されるにすぎない。これに対して、本実施形態の検索エンジン20は、テキスト文に含まれるフレーズのみならず関連フレーズ群にまで拡張した個々のフレーズに基づく見出し語をインデックス241に登録することができる。したがって、クライアント30からの検索クエリが ‘p l a t i n u m a n d p a r p i n h i b i t o r’ である場合であっても、 ‘p a r p i n h i b i t o r’ を含まないリソースが検索結果として提示されるようになる。

[0081] 図11は、本発明の一実施形態に係る検索エンジンによる検索処理を説明するためのフローチャートであり、より具体的には、図10に示した検索根拠の特定処理の詳細を説明するためのフローチャートである。

[0082] 図11に示すように、まず、検索エンジン20は、検索によりヒットしたリソースの有向グラフを特定する(S1101)。例えば、クエリサーバ25は、図2に示したようなデータベース24のリソース情報ファイルを参照し、ヒットしたリソースのテキスト文に関連付けられた有向グラフIDを特定し、該特定した有向グラフIDに基づいて、データベース24に格納されている有向グラフを読み出す。なお、本実施形態では、検索エンジン20は、ヒットしたリソースに関連付けられて格納された有向グラフIDを用いているが、これに限られず、ヒットしたリソースについて、図8に示した探索及び更新処理に従って、有向グラフを再度生成するようにしても良い。

[0083] 次に、検索エンジン20は、検索クエリに基づいて、特定した有向グラフ内を探索し(S1102)、検索クエリが顕出した、有向グラフにおける箇所を特定する(S1103)。例えば、クエリサーバ25は、有向グラフに

おける各パスをエッジ及びノードに従って順番に辿りながら、検索クエリに一致する箇所を特定する。特定される箇所は、1つのパスの全部又は一部であり、したがって、複数のエッジ及びノードを含み得る。探索は、例えば、KMP法に従って行われる。或いは、トライ木Tを用いて行われる。

[0084] 検索エンジン20は、次に、特定された箇所のパスが接続されている、有向グラフにおける元のパス上の分岐ノードを特定するとともに（S1104）、合流ノードを特定する（S1105）。元のパスとは、リソースのテキスト文から直接得られた有向グラフのパスであり、上述した初期有向グラフ（図7（a）参照）のパスに等しい。例えば、クエリサーバ25は、特定された箇所の前端ノードからテキスト文に基づく元のパスのノードに到達するまで、特定された箇所のパスを後ろ向きに（すなわち、エッジの向きとは逆方向に）辿り、到達したノードを分岐ノードとして特定する。また、クエリサーバ25は、特定された箇所の後端ノードから元のパスのノードに達するまで、特定された箇所のパスを前向きに（すなわち、エッジの向きの方向に）辿り、到達したノードを合流ノードとして特定する。なお、本例では、クエリサーバ25は、分岐ノードを特定した後に合流ノードを特定しているが、合流ノードを特定した後に分岐ノードを特定しても良いし、或いは、同時並行的にこれらを特定しても良い。

[0085] 検索エンジン20は、分岐ノード及び合流ノードを特定すると、特定された箇所のパス上で、分岐ノードから合流ノードまでフレーズを特定する（S1106）。例えば、クエリサーバ25は、ステップS1104においてパスを後ろ向きに辿る過程及びステップS1105において前向きに辿る過程で、エッジのラベルを抽出し、バッファに記憶しておき、該抽出し記憶したラベルに従ってフレーズを特定する。このステップS1106で特定されたフレーズが、検索によりヒットした根拠となった関連フレーズである。

[0086] 続いて、検索エンジン20は、同様に、元のパス上で、分岐ノードから合流ノードまでのフレーズを抽出する（S1106）。例えば、クエリサーバ25は、分岐ノードからエッジの向きに従って合流ノードまで、エッジのラ

ベルを抽出し、バッファに記憶しておき、該抽出し記憶したラベルに従ってフレーズを特定する。このステップS 1 1 0 7で特定されたフレーズが、関連フレーズが生成される基となったテキスト文のフレーズである。

[0087] なお、検索エンジン20は、有向グラフにおいて検索クエリが顕出した箇所ごとに、上記の検索根拠の特定処理を行い得る。

[0088] 以上のような処理により、検索エンジン20は、検索クエリに対してヒットしたリソースが、検索クエリのフレーズから直接的にヒットしたリソースではなく、その関連フレーズによってヒットしたリソースである場合であっても、検索によりヒットした根拠となった関連フレーズをユーザに提供することができるようになる。

[0089] また、例えば複数のフレーズのOR検索を含む検索クエリに基づいて本実施形態のインデックス241に対して検索が行われた場合であっても、ユーザに検索根拠が提示されることで、ユーザは、複数のフレーズの中から、リソースのヒットに貢献したフレーズを容易に把握することができるようになる。また、検索クエリからユーザが予期しないリソースがヒットされた場合であっても、検索根拠が提示されるため、ユーザは、検索エンジン20が正常に機能していることを認識できるようになる。

[0090] 次に、検索エンジン20による検索根拠の特定処理の一例を説明する。図12A及び12Bは、本発明の一実施形態に係る検索エンジンにより検索されたリソースの有向グラフから検索根拠を特定するための処理過程の一例を説明するための図である。本例では、検索エンジン20は、図9に示した有向グラフに基づいて、検索根拠を特定する様子を示している。

[0091] まず、検索エンジン20は、上述したように、データベース24のリソース情報ファイルからヒットしたリソースの有向グラフIDを特定し、例えば図12A(a)に示すような有向グラフをデータベース24から読み出して、検索クエリに含まれるフレーズ‘d i p h o s p h a t e r i b o s e’に基づいて、有向グラフ内の探索を開始する。かかる探索過程において、検索エンジン20は、有向グラフにおけるある1つのパスに、検索クエリの

フレーズに一致するフレーズを発見すると、有向グラフにおける該フレーズが顕出した箇所を特定する（図12A（b））。以下では、特定した箇所の前端のノードを前端ノードAとし、その後端のノードを後端ノードBとする。

[0092] 次に、検索エンジン20は、前端ノードAからテキスト文に基づく元のパスのノードに達するまで、特定された箇所のパスを逆方向に辿り、到達したノードを分岐ノードA'として特定するとともに、後端ノードBから元のパスのノードに達するまで、特定された箇所のパスを順方向に辿り、到達したノードを合流ノードB'として特定する（図12B（c））。

[0093] そして、検索エンジン20は、分岐ノードA'から、特定した箇所を含むパスを経由した、合流ノードB'までのパスを辿って、フレーズ‘ポリ | adenosine | diphosphate | ribose | ポリメラーゼ’を特定する。該特定した箇所のフレーズが、検索クエリに関連するフレーズとしてリソースがヒットする根拠となった関連フレーズである。また、検索エンジン20は、リソースのテキスト文に基づく分岐ノードA'から合流ノードB'までの元のパスを辿って、フレーズ‘poly | adp | ribose | polymerase’を特定する（図12B（d））。元のパスの該特定したフレーズが、関連フレーズが生成される基となったテキスト文のフレーズである。

[0094] 以上のようにして、検索エンジン20は、検索根拠となったフレーズを特定した後、該検索根拠となったフレーズを含む検索結果をクライアントに送信する。クライアント30は、これを受けて、例えばウェブブラウザ上に検索結果を表示する。

[0095] 図13A～13Cは、本発明の一実施形態に係る情報検索システムにおけるクライアント上に表示される検索結果画面の例を示す図である。

[0096] 図13Aに示す例では、検索結果画面1300Aは、例えば、検索クエリ入力フィールド1301に入力された検索クエリに基づいて検索エンジン20によって検索されたリソース名1302及びそのリンク1303を含む。

図13Aに示す例では、検索クエリ「“d i p h o s p h a t e r i b o s e” OR “p a r p 阻害剤”」に対してヒットしたリソースとして、「がん治療薬」及びそのURLが表示されている。また、本実施形態では、検索結果画面は、本文表示領域1304及び検索根拠表示領域1305を含む。本文表示領域1304及び検索根拠表示領域1305は、例えば、左右に並べて配置される。本文表示領域1304は、検索クエリに対してヒットしたリソースにおけるテキスト文の少なくとも一部を表示する領域である。表示されるテキスト文は、例えば、検索クエリのフレーズに関連するテキスト文の少なくとも一部であり得る。検索根拠表示領域1305は、検索クエリに対してリソースがヒットした根拠となった関連フレーズを表示する領域である。関連フレーズなどは、ユーザの視認性向上のため、例えば色分け、斜線、太字、斜体、枠付きなどの手法によって強調表示される。図13Aに示す例では、検索クエリのフレーズに対してヒットしたリソースにおけるフレーズが、どの関連フレーズに関連付けられているかが示されている。言い換えれば、リソースにおけるテキスト文のどのフレーズがどのような関連フレーズに置き換わってヒットしたかが示されている。

[0097] また、図13Bでは、本文表示領域1304'に、検索クエリに対してヒットしたリソースにおけるテキスト文の少なくとも一部と、検索クエリに対してリソースがヒットした根拠となった関連フレーズとが表示される検索結果画面1300Bの例が示される。関連フレーズなどは、ユーザの視認性向上のため、例えば、強調表示される。また、本例の検索結果画面1300Bは、検索根拠を表示するか否かをユーザが任意に選択することができるように、チェックボックス1306を含む。チェックボックス1306にチェックが入れている場合、検索結果画面1300Bに、検索根拠が表示される。例えば、ユーザは、検索クエリを入力し、検索結果の提示を受けた後、検索根拠を確認したいと考える場合に、チェックボックス1306にチェックを入れ、それを表示させるようにすることができる。このようにすることで、ユーザは、検索結果画面1300Bを簡略表示させた後、必要な部分の

みの検索根拠を選択的に表示させることができるようになる。

[0098] また、図 13Cでは、本文表示領域 1304 に、検索クエリに対してヒットしたリソースにおけるテキスト文の少なくとも一部が表示され、ユーザがポインタカーソル 1307 を操作して、ポインタカーソル 1307 を関連フレーズ上に重ねると、検索根拠表示領域 1305' がポップアップされる検索結果画面 1300C の例が示される。例えば、どのフレーズに関連フレーズが関連付けられているかをユーザが把握できるように、該フレーズが例えば強調表示される。

[0099] 以上のように、本実施形態の検索結果画面は、検索クエリに対してヒットしたリソースにおけるテキスト文と、該リソースがヒットした根拠となった関連フレーズとを関連付けて表示するため、ユーザは、検索根拠となった関連フレーズを容易に認識できるようになる。

[0100] 次に、図 8 に示した有向グラフの探索及び更新の処理の他の例を説明する。本例では、探索は、有向グラフにおいて未だ探索が行われていないエッジ（トークン）に対してのみ、つまり、既に探索されたトークンに対する再度の探索を回避するように行われる。このために、本例では、順序付きノードにより構成される木構造を用いる。このような木構造は、トライ（trie）木又はプレフィックス木として知られる。すなわち、トライ木では、あるノードの配下の全ノードは、自身に対応する文字列（ここではトークン）に共通するプレフィックスを有する。本例では、トライ木におけるノードとノードとの間のエッジに、辞書 22 に登録されたフレーズを構成する個々の語が割り当てられる。以下では、理解の容易のため、辞書 22 の一部に、図 14 に示すようなフレーズが登録されているものとして説明する。

[0101] 本例の検索エンジン 20 は、辞書 22 に基づいて作成されたトライ木を備える。図 15 は、図 14 に示した辞書 22 に登録されたフレーズに基づいて作成されたトライ木の一例を説明するための図である。トライ木 T は、典型的には、コンピューティングデバイスのプロセッサにより解釈可能なある種のデータ構造ないしはデータ配列であり、例えば、1 つ又は複数のファイル

ないしはテーブルとして構成されるが、これに限られるものではない。また、例えば、トライ木Tは、静的なファイルないしはテーブルとして構成され、その一部又は全部がメモリ上に常駐するのではなく、処理の実行中に、所定のアルゴリズムに従って、動的に構成されても良い。

[0102] 図15に示すように、トライ木Tは、ノードn0をルートとして、ノード間のエッジに、辞書22に登録されているフレーズの個々の単語が順番に割り当てられている。各ノードには例えば一意の番号が割り当てられる。ルートノードn0は、例えば「0」番が割り当てられる。同図では、便宜上、残りの各ノードに1～10の番号が割り当てられているものとする。また、図中、黒色で示されるノードは、フレーズ終端ノードであり、辞書22に登録されたフレーズを識別するためのフラグとして機能する。言い換えれば、ルートノードn0からフレーズ終端ノードまでのパスで示されるフレーズは、辞書22に登録されたフレーズである。例えば、図14に示したように、「栄養」及び「栄養療法」は、それぞれ、辞書22に登録されているため、図15に示すトライ木では、ノードn3及びn9は、それぞれ、フレーズ終端ノードとなる。

[0103] 図16A及び16Bは、本発明の一実施形態に係る検索エンジンによるインデックスの作成処理の他の例を説明するためのフローチャートであり、より具体的には、有向グラフの探索及び更新処理のフローチャートを示している。

[0104] 図16Aに示すように、まず、有向グラフ生成部233は、例えば、初期有向グラフの各ノードに、トライ木Tのルートノードn0を表す番号（ルート番号）「0」を割り当てる（S1601）。なお、代替的に、初期有向グラフの各ノードへのルート番号の割り当ては、初期有向グラフの生成時に行われるようにしても良い（図6のS603参照）。

[0105] 次に、有向グラフ探索部234は、初期有向グラフを探索対象の有向グラフに設定し（S1602）、続いて、有向グラフの先頭ノードを1つ選択し、これを現在の探索ノードとして設定する（S1603）。選択された探索

ノードは、探索済みのノードとして記憶される。このとき、有向グラフ探索部234は、探索ノードの位置及びフレーズを挿入すべき始端位置（以下「フレーズ挿入始端位置」という。）を図示しないキャッシュ領域にキャッシュする（S1604）。これにより、有向グラフ探索部234は、キャッシュされた内容に従って、未探索のノードの位置及び関連フレーズを追加すべき位置を特定することができる。

[0106] 有向グラフ探索部234は、次に、探索ノードに関連付けられたトークン（フレーズ）が、トライ木Tの対応するノードに従って検索（retrieve）されるか否かを判断する（S1605）。より具体的には、有向グラフ探索部234は、探索ノードに接続されたエッジのトークンを特定し、該特定したトークンが、該探索ノードに割り当てられた番号に対応するトライ木Tのノードに接続されたエッジのトークンと一致するか否かを判断する。有向グラフ探索部234は、該特定したトークンがトライ木Tにおいて検索された場合（S1605のYes）、有向グラフにおける該特定したトークンの下流ノードに、トライ木Tにおけるトークンの下流ノードに割り当てられた番号を追加的に割り当てる（図16BのS1606）。

[0107] 例えば、有向グラフにおける探索ノードに接続されたトークンが‘t p n’であったとする。この場合、有向グラフ探索部234は、トライ木Tにおいて、ルートノードn0に接続されたエッジのトークン‘t p n’を見つけ出し、該エッジに接続された下流ノードの番号「2」を、有向グラフにおけるトークン‘t p n’に接続された下流ノードに追加的に割り当てる（図17（b）参照）。

[0108] 次に、有向グラフ探索部234は、トライ木Tの該下流ノードがフレーズ終端ノードであるか否かを判断する（S1607）。有向グラフ探索部234は、該下流ノードが終端であると判断する場合（S1607のYes）、該下流ノードにさらに下流ノードがあるか否かを判断する（S1608）。有向グラフ探索部234が、さらに下流ノードがないと判断する場合（S1608のNo）、これを受けて、有向グラフ生成部233は、辞書22を参

照して、該トークン（フレーズ）の関連フレーズを抽出し（S 1 6 0 9）、該抽出した関連フレーズを有向グラフの対応するノード間に追加的に接続することによって、有向グラフを更新する（S 1 6 1 0）。つまり、有向グラフ生成部 2 3 3 は、関連フレーズを含むパスを、キャッシュされたフレーズ挿入始端位置のノードと有向グラフにおけるトークンの下流ノード（挿入後端位置のノード）とに並列的に接続する。有向グラフ生成部 2 3 3 による有向グラフの更新後、有向グラフ探索部 2 3 4 は、有向グラフにおいて新たに追加された関連フレーズに接続されたノードに、ルートノード n 0 の番号「0」を割り当て（S 1 6 1 1）、S 1 6 0 3（図 1 6 A）の処理に戻る。

[0109] 一方、有向グラフ探索部 2 3 4 は、トライ木 T の該下流ノードがフレーズ終端ノードでないと判断する場合（S 1 6 0 7 の No）、次の探索ノードを設定するため、S 1 6 0 3 の処理に戻る。加えて、有向グラフ探索部 2 3 4 は、トライ木 T の該下流ノードにさらに下流ノードがあると判断する場合もまた（S 1 6 0 8 の Yes）、次の探索ノードを設定するため、S 1 6 0 3 の処理に戻る。

[0110] また、S 1 6 0 5 において、有向グラフ探索部 2 3 4 は、該特定したトークンがトライ木 T において検索できなかった場合（S 1 6 0 5 の No）、有向グラフ内のすべてのノードを探索したか否かを判断する（S 1 6 1 2）。有向グラフ探索部 2 3 4 は、有向グラフ内のすべてのノードを探索していないと判断する場合（S 1 6 1 2 の No）、有向グラフ内の探索を続行するため、S 1 6 0 3 の処理に戻る。一方、有向グラフ探索部 2 3 4 は、有向グラフ内のすべてのノードを探索したと判断する場合（S 1 6 1 2 の Yes）、探索処理を終了する。

[0111] 次に、上記のトライ木 T を用いた、検索エンジン 2 0 による有向グラフの展開処理の一例を説明する。図 1 7 は、本発明の一実施形態に係る検索エンジンにより有向グラフが展開される処理過程の他の例を説明するための図である。なお、同図中、白色で示されるノードは、探索の対象となるノード（未探索ノード）を示し、黒色で示されるノードは、探索済みのノードを示す

ものとする。

- [0112] 上述したように、入力されたテキスト文に基づいて、初期有向グラフが生成されたとする。同図（a）は、初期有向グラフの一部を示している。初期有向グラフの各ノードには、ルート番号「0」が割り当てられている。有向グラフ探索部234は、有向グラフの探索過程で、白抜き矢印で示されるノードを探索ノードとして設定する。また、有向グラフ探索部234は、上述したように、探索ノード位置及びフレーズ挿入始端位置をキャッシュ領域にキャッシュする。
- [0113] 次に、有向グラフ探索部234は、探索ノードに接続されたエッジのトークン「t p n」が、該探索ノードに割り当てられた番号「0」に対応するトライ木Tのノードに接続されたエッジのトークンと一致するか否かを判断する。本例では、トライ木Tにおいて、ルートノードn0に接続されたエッジの1つにトークン「t p n」があるため、有向グラフ探索部234は、有向グラフの該トークンの下流ノードに、トライ木Tの該トークンの下流ノードに割り当てられた番号「2」を追加的に割り当てる（同図（b））。
- [0114] 次に、有向グラフ探索部234は、トライ木Tにおける該トークンの下流ノードがフレーズ終端ノードであると判断し、これを受けて、有向グラフ生成部233は、辞書22（図14）を参照して、該トークン（フレーズ）の関連フレーズ「total | parenteral | nutrition」を、キャッシュされているフレーズ挿入始端位置に従って有向グラフに追加するとともに、新たに追加されたノードに、ルートノードn0の番号「0」を割り当てる（同図（c））。
- [0115] その後、有向グラフ探索部234は、同様にして、探索の処理を進め、同図（d）に示されるように、黒色で示されるノードについて探索が終了し、白抜き矢印で示されるノードを探索ノードとして設定したとする。ここで、有向グラフ探索部234は、トライ木T内の検索において、ノードn6からのエッジのトークン「nutrition」に従い、下位ノードのノード番号「9」を、有向グラフの下位ノードに割り当てる。なお、有向グラフ探索

部234は、ノードn6がフレーズ終端ノードであることから、辞書22を参照して、‘total | parenteral | nutrition’の関連フレーズ（トークン）‘tpn’のパスを接続することになるところ、関連フレーズ‘tpn’はフレーズ挿入位置にすでに存在するため、有向グラフ探索部234は、このパスの接続処理を行わない。これにより、有向グラフの展開処理において冗長なパスの追加が抑制される。

[0116] 続いて、有向グラフ探索部234は、同様に探索の処理を進め、トライ木のノードn4がフレーズ終端ノードであるため、辞書22を参照して、有向グラフにフレーズ挿入位置にトークン‘栄養’のパスを接続する（同図（e））。

[0117] 上記のようにして、有向グラフ探索部234による有向グラフの探索の結果、有向グラフ生成部233は、最終的に、同図（f）に示すような有向グラフを生成する。

[0118] なお、上記の処理において、有向グラフ生成部233が行う処理の一部を有向グラフ探索部234が行うように構成しても良いし、有向グラフ探索部234が行う処理の一部を有向グラフ生成部233が行うように構成しても良い。

[0119] 以上のように本実施形態によれば、インデックス型検索エンジンにおいて、収集されたリソースを効率的に検索することができるインデックス241が作成される。これにより、ユーザは、インデックス241の作成の基となった収集されたリソースに含まれるフレーズ以外のフレーズによっても、該リソースにアクセスすることができるようになる。したがって、例えば医療分野のような専門分野において、特定のフレーズに基づいて、類義語などの関連フレーズ群を含むリソースを幅広く効率的に検索することができるようになる。

[0120] また、例えばOR検索を含む検索クエリに基づいて本実施形態のインデックス241に対して検索が行われた場合であっても、ユーザに検索の根拠が提示されることで、ユーザは、複数のフレーズの中から、リソースのヒット

に貢献したフレーズを容易に把握することができるようになる。また、検索クエリからユーザが予期しないリソースがヒットされた場合であっても、検索の根拠が提示されるため、ユーザは、検索エンジン 20 が正常に機能していることを認識できるようになる。

[0121] 上記各実施形態は、本発明を説明するための例示であり、本発明をこれらの実施形態にのみ限定する趣旨ではない。本発明は、その要旨を逸脱しない限り、さまざまな形態で実施することができる。

[0122] 例えば、本明細書に開示される方法においては、その結果に矛盾が生じない限り、ステップ、動作又は機能を並行して又は異なる順に実施しても良い。説明されたステップ、動作及び機能は、単なる例として提供されており、ステップ、動作及び機能のうちのいくつかは、発明の要旨を逸脱しない範囲で、省略でき、また、互いに結合させることで一つのものとしてもよく、また、他のステップ、動作又は機能を追加しても良い。

[0123] また、本明細書では、さまざまな実施形態が開示されているが、一の実施形態における特定のフィーチャ（技術的事項）を、適宜改良しながら、他の実施形態に追加し、又は該他の実施形態における特定のフィーチャと置換することができる、そのような形態も本発明の要旨に含まれる。

産業上の利用可能性

[0124] 本発明は、情報検索技術の分野に広く利用することができる。

符号の説明

- [0125] 1…情報検索システム
- 10…コンピュータネットワーク
 - 20…検索エンジン
 - 21…クローラ
 - 22…辞書
 - 23…インデクサ
 - 231…入力インターフェース部
 - 232…テキスト処理部

- 2 3 3 …有向グラフ生成部
- 2 3 4 …有向グラフ探索部
- 2 3 5 …インデックス作成部
- 2 4 …データベース
- 2 4 1 …インデックス
- 2 5 …クエリサーバ
- 3 0 …クライアント

請求の範囲

- [請求項1] テキスト文を含むリソースに基づいて、データベースを検索するためのインデックスを作成する装置であって、
- 特定の概念ごとに、相互に関連付けられた複数のフレーズを含む辞書と、
- 前記リソースの入力を受け付ける入力インターフェース部と、
- 前記受け付けたリソースのテキスト文から複数の語をトークンとして切り出すための処理を行うテキスト処理部と、
- 前記テキスト文に従って、前記切り出した複数のトークンどうしの接続関係を表現した有向グラフを生成する有向グラフ生成部と、
- 探索対象フレーズに基づいて前記有向グラフ内を探索し、前記探索対象フレーズが前記辞書中に発見される場合に、前記有向グラフ内における前記探索対象フレーズが顕出した箇所を特定する有向グラフ探索部と、
- 前記有向グラフ内における複数のトークンに基づいて、前記リソースとの関連付けを示すインデックスを作成するインデックス作成部と、を備え、
- 前記有向グラフ探索部は、
- 前記有向グラフ内における少なくとも1つのトークンを前記探索対象フレーズに設定し、前記設定した探索対象フレーズに基づいて前記辞書を参照し、
- 前記有向グラフ生成部は、
- 前記辞書に基づいて、前記探索対象フレーズに関連付けられた少なくとも1つのフレーズに基づく新たなトークンを前記有向グラフに追加することによって、前記有向グラフを更新する、装置。
- [請求項2] 前記有向グラフ生成部は、前記複数のトークンのうちの隣接するトークンどうしをノードによって接続した前記有向グラフを生成する、

請求項 1 に記載の装置。

- [請求項3] 前記有向グラフ生成部は、前記有向グラフ内における前記顕出した箇所の前後のノードに、前記辞書中に発見された前記探索対象フレーズに関連付けられた前記少なくとも 1 つのフレーズを接続する、請求項 2 に記載の装置。
- [請求項4] 前記有向グラフ探索部は、前記有向グラフ生成部によって更新された有向グラフを探索する、請求項 1 に記載の装置。
- [請求項5] 前記有向グラフ探索部は、前記更新された有向グラフにおいて既に探索し終えたトークンに対する再探索が回避されるように、前記更新された有向グラフ内を探索する、請求項 4 に記載の装置。
- [請求項6] 前記有向グラフは、前記複数のトークンのうちの隣接するトークンどうしを接続する複数のノードを含み、
前記有向グラフ探索部は、前記探索をし終えたトークンに関連付けられたノードを探索済みノードとして記憶し、前記探索済みノード以外のノードに接続されたトークンに対して再探索する、
請求項 5 に記載の装置。
- [請求項7] 前記装置は、前記辞書に含まれる前記複数のフレーズに基づいて構成される木構造を示すデータ配列をさらに備え、前記木構造は、前記複数のフレーズのそれぞれに対応する順序付きノードを含み、
前記有向グラフ探索部は、前記木構造を参照しながら、前記有向グラフ内を探索する、
請求項 1 に記載の装置。
- [請求項8] 前記有向グラフ探索部は、前記探索対象フレーズが前記木構造における少なくとも 1 つの前記順序付きノードに対応する少なくとも 1 つのフレーズと一致する場合に、前記有向グラフ内に前記探索対象フレーズが顕出したと判断する、請求項 7 に記載の装置。
- [請求項9] 前記有向グラフ生成部は、前記一致する探索対象フレーズに関連付けられた、前記辞書における少なくとも 1 つのフレーズに基づく新た

なトークンを前記有向グラフに追加することによって、前記有向グラフを更新する、請求項8に記載の装置。

[請求項10] 前記インデックス作成部は、更新された前記有向グラフ内における複数のトークンに基づいてN-gramを抽出し、前記抽出したN-gramに基づいて、前記インデックスを作成する、請求項1に記載の装置。

[請求項11] 前記インデックス作成部は、前記N-gramとして、単語ベースのN-gramを抽出する、請求項10に記載の装置。

[請求項12] 請求項1に記載された装置により作成されたインデックスと、
前記インデックスに関連付けられた前記リソースに関するリソース情報ファイルと、

検索クエリに基づいて前記インデックスを検索し、該検索の結果に基づいて前記リソース情報ファイルから少なくとも1つのリソースを特定し、前記特定した少なくとも1つのリソースを示す検索結果を出力するクエリサーバと、を備える検索装置であって、

前記クエリサーバは、前記特定した少なくとも1つのリソースから得られる、前記少なくとも1つのリソースにおけるフレーズに関連する所定の関連フレーズを含む有向グラフに基づいて、前記少なくとも1つのリソースが特定された根拠となった根拠フレーズ及び前記フレーズに関連付けられた前記少なくとも1つのリソースにおける元フレーズを特定し、

前記検索結果は、前記根拠フレーズ及び前記元フレーズを含む、検索装置。

[請求項13] コンピューティングデバイスにより実行される、テキスト文を含むリソースに基づいてデータベースを検索するためのインデックスを作成するための方法であって、

前記コンピューティングデバイスのプロセッサの制御の下、前記コンピューティングデバイスが、

特定の概念ごとに、相互に関連付けられた複数のフレーズを含む辞書を、前記プロセッサによってアクセスされるメモリに展開することと、

入力インターフェース部を介して、前記リソースの入力を受け付けることと、

前記受け付けたリソースのテキスト文から複数の語をトークンとして切り出すためのテキスト処理を行うことと、

前記テキスト文に従って、前記切り出した複数のトークンどうしの接続関係を表現した有向グラフを生成し、前記プロセッサによってアクセスされるメモリに展開することと、

探索対象フレーズに基づいて前記辞書を参照しながら前記有向グラフ内を探索することと、

前記有向グラフ内における複数のトークンに基づいて、前記リソースとの関連付けを示すインデックスを作成して前記データベースに登録することと、を実行することを含み、

前記有向グラフ内を探索することは、

前記有向グラフ内における少なくとも1つのトークンを前記探索対象フレーズに設定し、前記設定した探索対象フレーズに基づいて前記辞書を参照し、前記探索対象フレーズが前記辞書中に発見された場合に、前記有向グラフ内における前記探索対象フレーズが顕出した箇所を特定することを含み、

前記有向グラフを生成することは、

前記辞書中に発見された前記探索対象フレーズに関連付けられた少なくとも1つのフレーズに基づく新たなトークンを前記有向グラフに追加することによって、前記有向グラフを更新することを含む、方法。

[請求項14]

前記有向グラフを生成することは、前記複数のトークンのうちの隣接するトークンどうしをノードによって接続した前記有向グラフを生

成することを含む、請求項 13 に記載の方法。

[請求項15] 前記有向グラフを生成することは、前記有向グラフ内における前記頭出した箇所の前後のノードに、前記辞書中に発見された前記探索対象フレーズに関連付けられた前記少なくとも 1 つのフレーズを接続する、請求項 14 に記載の方法。

[請求項16] 前記有向グラフを探索することは、前記更新された有向グラフを新たな探索対象の有向グラフとして探索する、請求項 13 に記載の方法。

[請求項17] 前記インデックスを作成することは、更新された前記有向グラフ内における複数の語に基づいて $N-g\ r\ a\ m$ を抽出し、前記抽出した $N-g\ r\ a\ m$ に基づいて、前記インデックスを作成することを含む、請求項 13 に記載の方法。

[請求項18] 前記インデックスを作成することは、前記 $N-g\ r\ a\ m$ として、単語ベースの $N-g\ r\ a\ m$ を抽出する、請求項 17 に記載の方法。

[請求項19] 検索エンジンを含む検索装置により実行される情報検索方法であって、

前記検索装置のプロセッサの制御の下、前記検索エンジンが、

クライアントコンピュータから受信した検索クエリに基づいて、前記検索装置のデータベース上のインデックスを検索し、該検索の結果に基づいてリソース情報ファイルから少なくとも 1 つのリソースを特定することと、

前記特定した少なくとも 1 つのリソースから得られる、前記少なくとも 1 つのリソースにおけるフレーズに関連する所定の関連フレーズを含む有向グラフを前記プロセッサによってアクセスされるメモリに展開し、前記メモリに展開された前記有向グラフに基づいて、前記少なくとも 1 つのリソースが特定された根拠となった根拠フレーズ及び前記フレーズに関連付けられた前記所定のリソースにおける元フレーズを特定することと、

前記少なくとも 1 つのリソース並びに前記根拠フレーズ及び前記元フレーズを示す情報を検索結果として出力し、前記クライアントコンピュータに送信することと、を実行することを、
を含む情報検索方法。

[請求項20]

テキスト文を含むリソースに基づいて、データベースを検索するためのインデックスを作成する方法をコンピューティングデバイスに実現させるためのコンピュータ読み取り可能プログラムを非一時的に記録した記録媒体であって、

前記方法は、

特定の概念ごとに、相互に関連付けられた複数のフレーズを含む辞書を、前記装置のプロセッサによってアクセスされるメモリに展開することと、

入力インターフェース部を介して、前記リソースの入力を受け付けることと、

前記受け付けたリソースのテキスト文から複数の語をトークンとして切り出すためのテキスト処理を行うことと、

前記テキスト文に従って、前記切り出した複数のトークンどうしの接続関係を表現した有向グラフを生成することと、

探索対象フレーズに基づいて前記辞書を参照しながら前記有向グラフ内を探索することと、

前記有向グラフ内における複数のトークンに基づいて、前記リソースとの関連付けを示すインデックスを作成することと、を含み、

前記有向グラフ内を探索することは、

前記有向グラフ内における少なくとも 1 つのトークンを前記探索対象フレーズに設定し、前記設定した探索対象フレーズに基づいて前記辞書を参照し、前記探索対象フレーズが前記辞書中に発見された場合に、前記有向グラフ内における前記探索対象フレーズが顕出した箇所を特定することを含み、

前記有向グラフを生成することは、

前記辞書中に発見された前記探索対象フレーズに関連付けられた少なくとも 1 つのフレーズに基づく新たなトークンを前記有向グラフに追加することによって、前記有向グラフを更新することを含む、記録媒体。