

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3985483号

(P3985483)

(45) 発行日 平成19年10月3日(2007.10.3)

(24) 登録日 平成19年7月20日(2007.7.20)

(51) Int. Cl. F I
G06F 17/30 (2006.01)
 G06F 17/30 330C
 G06F 17/30 170A
 G06F 17/30 180A

請求項の数 14 (全 20 頁)

(21) 出願番号	特願2001-297675 (P2001-297675)	(73) 特許権者	000002369
(22) 出願日	平成13年9月27日(2001.9.27)		セイコーエプソン株式会社
(65) 公開番号	特開2003-108583 (P2003-108583A)		東京都新宿区西新宿2丁目4番1号
(43) 公開日	平成15年4月11日(2003.4.11)	(74) 代理人	110000028
審査請求日	平成15年10月28日(2003.10.28)		特許業務法人明成国際特許事務所
		(72) 発明者	今井 俊
			長野県諏訪市大和三丁目3番5号 セイコーエプソン株式会社内
		審査官	池田 聡史
		(56) 参考文献	特開平10-240536 (JP, A)
			特開平10-232871 (JP, A)
			特開平11-272710 (JP, A)
			特開平10-334101 (JP, A)
			最終頁に続く

(54) 【発明の名称】 言語文を用いた検索装置、検索システム、検索方法、プログラム、および記録媒体

(57) 【特許請求の範囲】

【請求項1】

言語文を用いて検索を行なう装置であって、
 検索用の検索文を入力する検索文入力手段と、
 該入力された検索文を利用して、検索を行なう検索手段と、
 少なくとも前記検索された対象に含まれる文である対象文を解析して、少なくとも一つの述部を含む構文上の最小単位である部分文を抽出すると共に、該抽出された部分文を、文における役割に着目して、少なくとも条件部と結論部とに分類する第1の分類手段と、
 前記検索文を解析して、前記部分文を抽出すると共に、該抽出された部分文を、文における役割に着目して、少なくとも条件部と結論部とに分類する第2の分類手段と、
 前記検索文と前記対象文から抽出された部分文に含まれる自立語が、前記分類された条件部と結論部のいずれに属するかを判定し、前記判定結果に基づいて前記対象文に前記検索文に対する類似度を付与し、前記対象文を前記類似度が大きい順に配列する対象文評価手段と

を備えた検索装置。

【請求項2】

請求項1記載の検索装置であって、

前記第1の分類手段は、

前記対象文を形態素解析して文節を切り出す形態素解析手段と、

該形態素解析により、前記対象文に含まれる部分文の接続関係を示す接続詞、接続助

詞を特定し、該接続詞、接続助詞を用いて、部分文を抽出する部分文抽出手段とを備えた検索装置。

【請求項 3】

請求項 1 記載の検索装置であって、

前記検索手段が検索した対象に含まれる対象文を、前記対象文評価手段が付与した類似度の大きさに応じた態様で表示する対象文表示手段を備える検索装置。

【請求項 4】

請求項 1 記載の検索装置であって、

前記分類手段は、

前記対象文を形態素解析して文節を切り出す形態素解析手段と、

該切り出された文節の係り受け関係を解析して文節間の関係を特定する文節関係特定手段とを備え、

前記対象文評価手段は、前記特定された文節間の前記係り受けの関係に基づいて、前記対象文に前記検索文に対する類似度を付与する手段である

検索装置。

【請求項 5】

前記対象文評価手段は、前記検索文が前記対象文における前記結論部に現われる対象文を類似度の大きな文として配列する請求項 3 記載の検索装置。

【請求項 6】

請求項 1 ないし 5 のいずれか記載の検索装置であって、

前記検索手段は、前記分類された結論部に含まれる単語を用いて検索を行なう手段である

検索装置。

【請求項 7】

前記対象文評価手段は、前記類似度の付与において、検索文に含まれる自立語が対象文に含まれる場合に、前記類似度を大きくする請求項 1 記載の検索装置。

【請求項 8】

前記対象文評価手段は、前記類似度の付与において、検索文に含まれる自立語に基づいてシソーラスを参照して得られた上位概念に相当する語が対象文に含まれる場合に、前記類似度を大きくする請求項 1 記載の検索装置。

【請求項 9】

前記第 2 の分類手段は、前記部分文の抽出に先立って、予め登録した不要文を、前記検索文から取り除く処理を行なう請求項 1 記載の検索装置。

【請求項 10】

ネットワークを介して接続されたクライアントコンピュータとサーバコンピュータとにより実現され、該ネットワークを介して接続された他のコンピュータ上のデータを検索して表示する検索システムであって、

前記サーバコンピュータは、

前記クライアントコンピュータ側で入力された検索用の検索文を利用して、前記ネットワークを介して接続された他のコンピュータ上のデータを検索する検索手段と、

少なくとも前記検索された対象に含まれる文である対象文を解析して、少なくとも一つの述部を含む構文上の最小単位である部分文を抽出すると共に、該抽出された部分文を、文における役割に着目して、少なくとも条件部と結論部とに分類する第 1 の分類手段と、

前記検索文を解析して、前記部分文を抽出すると共に、該抽出された部分文を、文における役割に着目して、少なくとも条件部と結論部とに分類する第 2 の分類手段と、

前記検索文と前記対象文から抽出された部分文に含まれる自立語が、前記分類された条件部と結論部のいずれに属するかを判定し、前記判定結果に基づいて前記対象文に前記検索文に対する類似度を付与し、前記対象文を前記類似度が大きい順に配列する対象文評価手段と、

10

20

30

40

50

該振り分けた対象文を、所定の構造で、前記クライアントコンピュータ上で表示可能な形態で、前記検索文の入力を行なったクライアントコンピュータに送り出す送信手段とを備えた検索システム。

【請求項 1 1】

ネットワークを介して接続されたクライアントコンピュータとサーバコンピュータとにより実現され、該ネットワークを介して接続された他のコンピュータ上のデータを検索して表示する検索システムであって、

前記サーバコンピュータは、

前記クライアントコンピュータ側で入力された検索用の検索文を利用して、前記ネットワークを介して接続された他のコンピュータ上のデータを検索する検索手段を備え、

10

前記クライアントコンピュータは、

前記サーバコンピュータから前記検索の結果として、前記検索された対象に含まれる文である対象文を受け取る受信手段と、

該対象文を解析して、少なくとも一つの述部を含む構文上の最小単位である部分文を抽出すると共に、該抽出された部分文を、文における役割に着目して、少なくとも条件部と結論部とに分類する第 1 の分類手段と、

前記検索文を解析して、前記部分文を抽出すると共に、該抽出された部分文を、文における役割に着目して、少なくとも条件部と結論部とに分類する第 2 の分類手段と、

前記検索文と前記対象文から抽出された部分文に含まれる自立語が、前記分類された条件部と結論部のいずれに属するかを判定し、前記判定結果に基づいて前記対象文に前記検索文に対する類似度を付与し、前記対象文を前記類似度が大きい順に配列する対象文評価手段と、

20

該振り分けた対象文を、所定の構造で表示する表示手段と

を備えた検索システム。

【請求項 1 2】

言語文を用いて、コンピュータが検索を行なう方法であって、

検索用の検索文をキーボードなどの入力手段から入力し、

該入力された検索文を利用して、コンピュータが検索を行ない、

少なくとも前記検索された対象に含まれる文である対象文を解析して、コンピュータが、少なくとも一つの述部を含む構文上の最小単位である部分文を抽出すると共に、該抽出された部分文を、文における役割に着目して、少なくとも条件部と結論部とに分類し、

30

前記検索文を解析して、コンピュータが、前記部分文を抽出すると共に、該抽出された部分文を、文における役割に着目して、少なくとも条件部と結論部とに分類し、

前記検索文と前記対象文から抽出された部分文に含まれる自立語が、前記分類された条件部と結論部のいずれに属するかを判定し、前記判定結果に基づいて前記対象文に前記検索文に対する類似度を付与し、前記対象文を前記類似度が大きい順に配列する

検索方法。

【請求項 1 3】

言語文を用いて検索を行なう機能をコンピュータで実現するプログラムであって、

検索用の検索文を入力する機能と、

40

該入力された検索文を利用して、検索を行なう機能と、

少なくとも前記検索された対象に含まれる文である対象文を解析して、少なくとも一つの述部を含む構文上の最小単位である部分文を抽出すると共に、該抽出された部分文を、文における役割に着目して、少なくとも条件部と結論部とに分類する機能と、

前記検索文を解析して、前記部分文を抽出すると共に、該抽出された部分文を、文における役割に着目して、少なくとも条件部と結論部とに分類する機能と、

前記検索文と前記対象文から抽出された部分文に含まれる自立語が、前記分類された条件部と結論部のいずれに属するかを判定し、前記判定結果に基づいて前記対象文に前記検索文に対する類似度を付与し、前記対象文を前記類似度が大きい順に配列する機能と

をコンピュータ上で実現するプログラム。

50

【請求項 1 4】

請求項 1 3 記載のプログラムを記録した記録媒体。

【発明の詳細な説明】**【0001】****【発明の属する技術分野】**

本発明は、言語文を用いて検索を行なう技術に関し、詳しくは検索された情報と検索文との相関を評価する技術に関する。

【0002】**【従来の技術】**

データベースなどの検索は、通常、検索用の単語を入力し、この単語が含まれるデータ、あるいはその単語が含まれないデータといった条件で行なわれる。一つの単語で検索した結果、あまりに多くのデータがヒットした場合には、さらに単語を追加して絞り込み検索を行ったり、いくつかの単語による各々の検索結果を対象として、アンド条件やオア条件などを指定して、検索対象を絞り込むといったことも行なわれる。こうした単語を用いて所望の結果を得るには、ある程度の訓練を必要とした。

10

【0003】

そこで、従来から、よりよい検索手法を求めて、様々な提案がなされている。例えば、検索すべき単語が指定されると、その単語と意味を同じくする単語をシソーラスを用いて特定し、その単語についても検索を行なうことで、高精度の検索を行なおうとする技術が提案されている。シソーラスを用いれば、検索しようとする単語の上位概念で検索することもできる。このほか、自然言語を用いて検索を行なおうとするもの（例えば、特開平 1 - 1 8 0 0 4 6 号公報に開示された「自然言語理解方法および情報検索装置」、特開 2 0 0 1 - 1 4 1 6 5 号公報に開示された応答装置など）が提案されている。これらは、あらかじめ検索の対象（プラントの監視や保守など）に即して、検索のシナリオを作り、これに沿って検索を進めるといった技術である。こうした検索技術では、単に単語を用いたものと比べると、検索しようとするものは、自然な言語文で検索作業を進めることができる。

20

【0004】**【発明が解決しようとする課題】**

しかしながら、かかる検索技術では、次の点で問題があり、特に大量のデータ、例えばインターネットなどのネットワークに接続されたサイトの情報を検索するといったシステムでは、未だ十分な検索技術が提案されているとは言えなかった。まず、単語とシソーラスを用いた検索では、結局単語による検索であることに変わりはないので、多数のデータが検索により見いだされた場合など、絞り込みを行なわねばならず、検索に熟練を要する点は従前と同じであった。このため、シソーラスを用いて、検索精度を高めることは困難であった。

30

【0005】

また、自然言語文を用いて検索を行なうものは、検索対象の特徴などを生かしたシナリオを事前に作成した上で検索を行っており、自然言語を用いた検索が、事前に想定したパターンをはずれると、対応できないという問題があった。このため、例えばインターネット上のサイト検索などのように、事前のシナリオが想定できない対象に対しては、自然言語文を用いた検索ができなかった。

40

【0006】

本発明の装置は、こうした問題を解決し、自然言語文を用いて、高精度の検索を実現することを目的とする。

【0007】**【課題を解決するための手段およびその作用・効果】**

上記課題の少なくとも一部を解決する本発明の装置は、
言語文を用いて検索を行なう装置であって、
検索用の検索文を入力する検索文入力手段と、
該入力された検索文を利用して、検索を行なう検索手段と、

50

少なくとも前記検索された対象に含まれる文である対象文を解析して、少なくとも一つの述部を含む構文上の最小単位である部分文を抽出すると共に、該抽出された部分文を、文における役割に着目して、少なくとも条件部と結論部とに分類する第1の分類手段と、

前記検索文を解析して、前記部分文を抽出すると共に、該抽出された部分文を、文における役割に着目して、少なくとも条件部と結論部とに分類する第2の分類手段と、

前記検索文と前記対象文から抽出された部分文に含まれる自立語が、前記分類された条件部と結論部のいずれに属するかを判定し、前記判定結果に基づいて前記対象文に前記検索文に対する類似度を付与し、前記対象文を前記類似度が大きい順に配列する対象文評価手段と

を備えたことを要旨としている。

10

【0008】

また、この装置に対応した方法の発明は、

言語文を用いて、コンピュータが検索を行なう方法であって、

検索用の検索文をキーボードなどの入力手段から入力し、

該入力された検索文を利用して、コンピュータが検索を行ない、

少なくとも前記検索された対象に含まれる文である対象文を解析して、コンピュータが、少なくとも一つの述部を含む構文上の最小単位である部分文を抽出すると共に、該抽出された部分文を、文における役割に着目して、少なくとも条件部と結論部とに分類し、

前記検索文を解析して、コンピュータが、前記部分文を抽出すると共に、該抽出された部分文を、文における役割に着目して、少なくとも条件部と結論部とに分類し、

20

前記検索文と前記対象文から抽出された部分文に含まれる自立語が、前記分類された条件部と結論部のいずれに属するかを判定し、前記判定結果に基づいて前記対象文に前記検索文に対する類似度を付与し、前記対象文を前記類似度が大きい順に配列すること

を要旨としている。

【0009】

かかる装置および方法によれば、検索をしようとする者により入力された検索用の検索文を利用して、検索を行なう。このとき、検索された対象に含まれる文である対象文が取得され、この対象文を解析して、少なくとも一つの述部を含む構文上の最小単位である部分文を抽出すると共に、該抽出された部分文を、文における役割に着目して、少なくとも条件部と結論部とに分類する。同様に、検索文を解析して、部分文を抽出すると共に、抽出された部分文を、文における役割に着目して、少なくとも条件部と結論部とに分類し、検索文と対象文から抽出された部分文に含まれる自立語が、分類された条件部と結論部のいずれに属するかを判定することにより、対象文に検索文に対する類似度を付与するから、この類似度に基づいて対象文を配列することができる。

30

【0010】

【発明の他の態様】

また、こうした検索装置の発明は、その実現形態として、サーバ上で実現したり、サーバコンピュータとクライアントコンピュータとが協働するシステムとして実現することもできる。また、コンピュータに上で動作するプログラムにより上記の検索方法を実現することができるので、本発明をプログラムとして、あるいはそのプログラムが記録された記録媒体（例えばフレキシブルディスク、CD-ROM、DVD-ROM、磁気テープなど）として把握することもできる。プログラムは、記録媒体に記録して扱うこともできるが、ネットワーク上のサーバなどにおき、これをネットワークを介してダウンロードして、クライアント側のコンピュータで実行するという扱いにすることもできる。

40

【0011】

【発明の実施の形態】

以下、本発明の実施の形態について説明する。図1は、本発明の実施の形態の一つとしての検索システム100を示すブロック図である。図示する各ブロックは、実際には、サーバコンピュータ200とクライアントコンピュータ300とから構成されている。両コンピュータは、ネットワーク110を介して接続されている。サーバコンピュータ200は

50

、検索用エンジンを搭載しており、クライアントコンピュータ300は、サーバコンピュータ200に対して検索の要求を出力し、検索結果をサーバ200から受け取って表示する。サーバコンピュータ200やクライアントコンピュータ300の具体的なハードウェア構成の詳細などは後述する実施例に譲り、ここではブロックレベルで構成とその作用を説明する。

【0012】

図1に示すように、クライアントコンピュータ300は、検索用の検索文を自然言語の一つである日本語で受け付ける検索文入力部310、この検索文を解析する検索文解析部320、解析された検索文から検索用の単語列を取り出してこれをサーバコンピュータ200に出力する検索語出力部330、更にサーバコンピュータ200からの検索結果を受け取り画面に表示する結果表示部340を有する。他方、サーバコンピュータ200は、クライアントコンピュータ300からの検索語を受け取る検索語受信部210、受け取った検索語を用いて検索を行なう検索エンジン220、検索結果を文単位で取り出し形態素解析などを行なって解析する対象文解析部230、解析した対象文と検索語との対比を行なう対比実行部240、対比結果に従って対象文を配列する配列部250、配列済みの対象文を順次クライアントコンピュータ300に送信する検索結果出力部260などを備える。

10

【0013】

クライアントコンピュータ300の検索文入力部310は、利用者がキーボードなどを用いて入力した日本語を受け付ける。インターネットなどのネットワークに接続されたサイトの検索を行なう場合には、この検索文入力部310は、通常のブラウザにより表示された検索語の入力ボックスに、IME（日本語入力メソッド）などを用いて日本語文を入力する処理に相当する。検索文入力部310を介して、例えば、「電源を入れると壊れた」などの自然な言語文が入力される。なお、本実施の形態では、インターネットを介して接続した故障診断サイトで、コンピュータの故障についての診断（原因や対処）を受ける場合を想定している。

20

【0014】

利用者は、通常、自分のコンピュータの故障の状態については、言葉で表現できるが、その原因を特定して検索語を設定したり、単語を複数入力して、徐々に検索範囲を絞り込んだりすることは、困難なことが多い。そこで、この実施の形態では、利用者が通常している自然言語（この例では日本語）を用いて、自分で表現できる文の形で問い合わせを行なっている。こうした日本語により検索文が入力されると、検索文解析部320がこれを解析する。解析の内容については、後述する実施例で詳しく説明するが、通常は、まず検索文を形態素解析し、日本語として自然な文を構成する文節に切り分ける処理を行なう。文節に切り分けた後、検索に用いるべき単語を解析する。例えば、検索文が、上記の「電源を入れると壊れた」であれば、これを形態素解析して、「電源を」（名詞+助詞）、「入れると」（動詞連用形+助詞）、「壊れた」（動詞終止形+過去を示す助動詞）、といった文節に分け、更にここから検索語として、「電源」、「入れる」、「壊れる」などを抽出するのである。検索文解析部320では、こうした単語の抽出に加えて、シソーラスを参照して、同義語や類義語（例えば、「電源」に対する「パワーサプライ」や、「壊れる」に対する「破損する」など）も、検索語として抽出するものとしても良い。また、形態素解析に加えて、文節間の係り受けや、更に文を構成する句や節などの部分文の構成をも解析するものとしても良い。

30

40

【0015】

こうして抽出された検索語は、検索語出力部330により、ネットワークを介して出力され、パケットに付されたIPアドレスなどを用いて、故障診断プログラムが動作しているサーバコンピュータ200に届けられる。サーバコンピュータ200は、こうしてネットワークに出力された検索語を、検索語受信部210により受け取り、これを検索エンジン220に受け渡す。検索エンジン220は、受け取った検索語（通常複数個）を用いて、故障診断用の知識データベース225をアクセスし、検索語とを含む文を見いだすとこれ

50

を取り出す。上記の例であれば、「電源」「入れる」「壊れる」などの語のすべてを含む文、あるいは少なくとも一つ以上を含む文を、検索された対象文として取り出す。例えば、知識データベース225に収録されたデータベース内に、「電源を入れると、OSが起動する前に、『NoDisk』が表示されて、止まってしまいます。」や「コンピュータの使用中にハングアップして、コンピュータの電源を切ることもできません。」などの文があれば、検索エンジン220はこれらの文を、該当する検索対象文として検索することになる。

【0016】

こうして得られた検索対象文を、対象文解析部230が解析する。この解析は、先に説明した検索文解析部320による解析とほぼ同一であり、形態素解析を基礎として、係り受けによる句構造の解析や、更に、少なくとも一つの述部を含む構文上の最小単位である部分文などが分析される。更に、対象文解析部230では、こうして取り出された部分文を文における役割に着目して分類する。役割としては、文における条件部か結論部かといった区分でも良いし、更にこれを細かく分けて例えば、条件部であれば、「条件」「理由」「逆接」「並列」などに分類しても良い。こうして検索対象文を分類した後、この分類に従って、検索対象文と検索語との対比を、対比実行部240が行なう。すなわち、検索語が、検索対象文の結論部に現われているか、条件部に現われているか、などを考慮しつつ、両者の対比するのである。なお、この例では、サーバコンピュータ200は、複数の検索語を検索語受信部210により受信して検索エンジン220を動かすものとしたので、対比は、検索語と検索対象文とを単純に対比しているが、検索文を解析した結果も併せて受け取り、検索文の解析結果も用いて、対比を行なうものとしても良い。例えば、検索文における各語の構成（係り受けや節の役割など）を考慮して、検索対象文との対比を行なっても良い。なお、ここで言う「部分文」は、少なくとも一つの述部を含む構文上の最小単位であって、文法上は、主節や従属節、あるいは条件節などと言われる単位に、ほぼ相当する概念である。

【0017】

こうして対比を行なった後、サーバコンピュータ200は、配列部250により検索対象文を振り分けて配列し、これを検索結果出力部260からクライアントコンピュータ300側に返送する。配列としては、検索語に対して、より重要な相関を有すると考えられる検索対象文を上位に並べるのが好適である。単純に、相関の高い情報上位に配列するだけでなく、階層化して出力することも好適である。検索結果の出力は、検索語を送信してきたクライアントコンピュータ300のIPアドレスを、パケットに付して、検索対象文をインターネットなどのネットワークに出力することにより行なわれる。こうしてネットワークに流されたデータは、IPアドレスを手がかりに、検索語を出力したクライアントコンピュータ300に確実に返送される。

【0018】

この検索結果を受け取ったクライアントコンピュータ300は、検索対象文が配列された検索結果を、ブラウザなどを用いて一覧表示する。通常、こうした故障診断では、検索結果には、URLなどが付随しており、利用者は、検索結果を読んで、更に詳しく内容を知りたい場合には、このURLをクリックすることで、直ちにサーバコンピュータ200内の必要な情報にジャンプして、詳しい故障診断の情報（故障の原因や対処方法など）を知ることができる。しかも、利用者からすると、いくつかの検索語を特定したり、これを順次入力して検索結果を絞り込んだりする必要がなく、自分か把握した範囲で、自然な日本語で状況を説明する文を入力すれば足りる、という大きな利点がある。更に、検索の結果も、自然な日本語文で、しかもより関連性が高いと考えられる文が上位に表示されるので、短時間に必要な情報にたどり着けるという利点が得られる。

【0019】

上述した実施の形態において、検索対象文を形態素解析して文節を切り出したとき、この対象文に含まれる部分文の接続関係を示す接続詞、接続助詞を特定し、接続詞、接続助詞を用いて、部分文を抽出するものとしても良い。日本語では、接続詞や接続助詞が用いられ

10

20

30

40

50

る箇所までで部分文が構成されることが多く、しかも接続助詞などに着目すれば、その前が条件を示す部分文であるか、理由を示す部分文であるか、などを容易に認識することができる。

【0020】

更に、上記の実施の形態において、検索文を解析する場合には、少なくとも一つの述部を含む構文上の最小単位である部分文を抽出すると共に、抽出された部分文から、結論部に相当する部分を特定し、検索自体を、特定された結論部に含まれる単語を用いて行なうものとしても良い。例えば「スイッチを入れたら、電源が壊れた」というような検索文の場合、「電源が壊れた」という結論部の方が故障診断にとっては有用なことが多いので、結論部の単語「電源」「壊れる」を用いて検索を行なうのである。もとより、アプリケーションによっては、条件部に相当する部分を特定して、条件部に含まれる単語を用いて検索するものとしても良い。例えば、中毒診断用のシステムでは、「乾電池を飲んだので、腹が痛い」といった検索文が入力された場合、条件部の方が有用と見なして「乾電池」「飲む」などの単語を抽出し、これを用いて検索を行えばよい。

10

【0021】

上記の実施の形態では、検索の対象は、データベースでとしてが、検索の対象は、ネットワーク上に置かれたサイトに含まれる情報であっても差し支えない。いわゆるネットワーク上の検索エンジンに適用すれば、多数に上る関連サイトを、より関連の高いものを優先して表示することも容易である。

【0022】

また、上記の実施の形態では、検索システムは、サーバコンピュータ200とクライアントコンピュータ300とから構成したが、利用者が使用するコンピュータにデータベースや検索エンジンが置かれたいわゆるスタンドアロンの使用形態でも差し支えない。また、上記の実施の形態では、クライアントコンピュータ300は、検索文を検索文解析部320により解析し、検索語として、サーバコンピュータ200側に渡しているが、検索文をそのままクライアントコンピュータ300側に出し、サーバコンピュータ200側で検索文の解析処理から行なうものとしても良い。この場合、検索文の解析を行なう能力はサーバコンピュータ200側のプログラムやデータベースにより決定されるので、サーバコンピュータ200側のプログラムを入れ替えるだけで、解析能力をアップグレードできるという利点が見られる。また、クライアントコンピュータ300毎に解析能力が異なるということもない。もとより、検索文解析部320は、ブラウザにプラグインにより追加されるように構成し、これをサーバコンピュータ200側からクライアントコンピュータ300側に送信するものとしても良い。こうすれば、クライアントコンピュータ300によらず、ほぼ同等の解析能力を用意することができる。しかも、多数のクライアントコンピュータ300からアクセスされるサーバコンピュータ200側の負担を減らすことができる。

20

30

【0023】

上述した実施の態様では、検索エンジン220により得られた検索対象文の解析をサーバコンピュータ200側の対象文解析部230で行なっているが、解析をサーバコンピュータ200側では行わず、検索対象文をそのままクライアントコンピュータ300に出し、クライアントコンピュータ300側で解析と対比を行なうものとしても良い。クライアントコンピュータ300は、検索文が入力されたマシンであり、利用者に最も近い側に位置するので、利用者の要求にそって検索対象文を解析し、検索文との相関を判定して、所望の順序で表示することができる。例えば、音声認識を用いて検索文を入力するような構成を採用すれば、音声入力時の抑揚や強調された単語などの情報を、検索対象文と検索文との相関の判断において、考慮すると言ったことも可能である。また、検索対象文の解析をクライアントコンピュータ300側で行なうものとするれば、複数のサーバコンピュータ200上で検索エンジンを動かし、複数のサーバコンピュータ200からの検索結果を受け取って、これをまとめて解析して、相関の程度により順序付けして表示するといったことも可能となる。

40

50

【 0 0 2 4 】

上記の実施の形態は、サーバコンピュータ200とクライアントコンピュータ300からなる検索システムとして説明したが、これらの検索などの機能をコンピュータ上で実現するプログラムを、CD-ROMなどの記録媒体上に記録した形態で、本発明を実施することも可能である。この場合、上述したように、サーバ側のプログラムとクライアント側のプログラムに分けて、それぞれ記録媒体上に記録して実施しても良いし、一つのプログラムあるいはプログラム群として記録しても良い。更には、サーバ側に必要なプログラムをおき、このサーバ側のプログラムと協働して動作するプログラムを、サーバ側にダウンロード可能に用意し、検索を行なおうとするクライアント側から読み出して、実行する形態で実施することも可能である。

10

【 0 0 2 5 】

【実施例】

以上説明した実施の形態を更に具体的に説明するために、その実施例について説明する。

(1) 実施例の構成：

はじめに、実施例のハードウェア構成について、図2の概略構成図を用いて説明する。図2に示した実施例では、インターネットのようなネットワーク10に接続されたサーバコンピュータ20にプログラムをインストールし、このプログラムを実行することで、クライアントコンピュータ30からの検索要求に応じて検索を実行する検索システムが具現化されている。検索システム50におけるサーバコンピュータ20（以下、これを検索用サーバと呼ぶ）は、それ自身スタンドアロンの検索装置として使用可能であるが、以下で説明するように、サーバとして他のクライアントコンピュータ（以下、これを単にクライアントと呼ぶ）30から利用することが可能である。すなわち、ネットワーク10に接続された多数のクライアント30の利用者が、ネットワーク10を介して検索用サーバ20にアクセスすることで、自然言語を用いた検索とその結果の提供を受けることができる。入力部分については、検索用サーバ20とクライアント30はほぼ同じなので、ハードウェア構成については、検索用サーバ20を例として説明を行なう。

20

【 0 0 2 6 】

検索用サーバ20の内部構成を図2に基づいて説明する。検索用サーバ20は、モデムやルータ18を介してネットワーク10とのデータのやり取りを制御するネットワークインタフェース（NT-I/F）21、処理を行なうCPU22、処理プログラムや固定的なデータを記憶するROM23、ワークエリアとしてのRAM24、時間を管理するタイマ25、モニタ29への表示を司る表示回路26、テキストデータをデータベースとして蓄積しているハードディスク（HD）27、キーボード11、マウス12、マイク13とのインタフェースを司る入力インタフェース（I/F）28等を備える。なお、ハードディスク27は、固定式のものとして記載したが、着脱式のものでも良いし、着脱式の記憶装置（例えばCD-ROM、CD-R、CD-RW、DVD-ROM、DVD-RAM、フレキシブルディスクなど）を併用することも可能である。また、この実施例では、検索用サーバ20の処理プログラムは、ROM23内に記憶されているものとしたが、ハードディスク27に記憶しておき、起動時にRAM24上に展開して実行するものとしても良い。あるいは、上述した着脱式の記録媒体から読み込むものとしても良い。更には、ネットワーク10を介して、他のサーバから読み込んで実行するものとしても良い。同様に、以下に説明するように、ハードディスク27に必要なデータの総てが記憶されている必要はなく、ネットワーク10により接続される他のサーバに膨大なデータを分散して記憶、更新、管理する構成としても良い。

30

40

【 0 0 2 7 】

ハードディスク27には、形態素解析辞書IDC、文判定ルールSDI、シソーラスTSRおよび検索対象データベースDBが記憶されている。この形態素解析辞書IDCは、いわゆる仮名漢字変換辞書とほぼ同一の内容を記憶した辞書であり、仮名漢字変換辞書とは、見出しと読みが逆になっている。このため、キーボード11やネットワーク10を介して入力されたかな文字列を解析して、仮名漢字文字列に変換する仮名漢字変換辞書をその

50

まま用い、読みと見出しの関係だけインデックスの形でもった形態とすることもできる。形態素解析辞書 I D C の一例を、図 3 に示した。この例では、読みと表記と文法情報のみを示されているが、実際の形態素解析辞書 I D C には、単語やその他の単語に関する読み、表記、文法情報の他に、同一意味の口語、同意語、類義語、省略語、更には係り受けの情報などが関連付けられて記憶されている。この解析辞書 I D C は、検索用サーバ 2 0 では、クライアント 3 0 から受け取った検索文を形態素解析する際に用いられる。この解析辞書 I D C を用いることで、検索用サーバ 2 0 は、受け取った検索文を精度良く解析することができる。例えば、解析するかな文字列が口語体の自然言語であっても、その口語体を、正確に解析することが可能である。

【 0 0 2 8 】

文判定ルール S D I は、係り受けや部分文の関係を規定するルールを記憶した辞書である。係り受けについては、形態素解析でも用いられるが、ここでは、形態素解析により得られた文節同士の間を特定するのに用いられている。更に、部分文の関係を規定するルールとは、大きくは、条件部か結論部かを特定するルールであり、条件部については、更に、条件、理由、逆接、並列などを区別するルールが格納されている。また、結論部については、結論に影響を与えない不要部を取り除くためのルールなども記憶されている。

【 0 0 2 9 】

単語シソーラス T S R は、意味的な関係のある単語（例えば類義語、反意語など）を、その概念関係に従って整理した辞書である。概念関係としては、上位、下位、並列といった関係の他に、様々な関係が設けられ、多数の単語がこうした概念関係で整理されている。例えば「入れる」「切る」「回す」「ひねる」といった動詞について、「人間の動作」といった観点から、類義語として整理されている。

【 0 0 3 0 】

検索対象データベース D B は、利用者が検索しようとする対象そのものであり、この実施例では、故障解析・診断用のデータベースである。なお、こうしたデータベース D B は、本実施例では、ハードディスク 2 7 内に記録されているが、インターネット上に存在する多数のサイトなどを検索対象データベース D B として扱うことも勿論可能である。こうした場合には、巡回型検索エンジンにより、インターネット上のサイトのデータを検索し、インデックスの形で、検索用サーバ 2 0 内のハードディスク 2 7 にデータを蓄えておいても良いし、その都度、検索を行なっても良い。

【 0 0 3 1 】

(2) 検索システムの動作 - 解析処理 :

検索用サーバ 2 0 とこれに接続されたクライアント 3 0 からなる検索システム 5 0 の動作について説明する。この実施例では、クライアント 3 0 では、インターネット上のサイトの情報をブラウズするブラウザが動作しており、利用者が、検索用サーバ 2 0 から送られたデータに基づいてこのブラウザに表示した検索用のボックスに、検索しようとする内容を自然な日本語で入力すると、これを解析することなく、そのままネットワーク 1 0 を介して、検索用サーバ 2 0 に送信している。実施の形態では、検索文の解析はクライアント 3 0 側で行なったが、この実施例では、検索文の解析から、すべて検索用サーバ 2 0 側で行なっている。クライアント 3 0 側は、検索文の入力およびその出力と、検索結果の表示のみを担当している。

【 0 0 3 2 】

そこで、クライアント 3 0 側の動作についての説明は簡略にとどめ、図 4 の説明図を用いて、検索用サーバ 2 0 側の動作について詳しく説明する。検索用サーバ 2 0 は、ネットワーク 1 0 を介したクライアント 3 0 側からの要求を受け取ると、図 4 に示した処理を開始する。検索用サーバ 2 0 が実行する処理は大きくは解析処理と照合処理である。解析処理は、形態素解析処理（ステップ S 1 0 0 ）、係り受け解析（ステップ S 1 1 0 ）、部分文の判定（ステップ S 1 2 0 ）から構成されている。他方、照合処理は、単語照合（ステップ S 1 3 0 ）、係り受け照合（ステップ S 1 4 0 ）、および部分文の照合（ステップ S 1 5 0 ）から構成されている。

10

20

30

40

50

【0033】

図4に示した処理は、クライアント30から検索文を受け取ったときに開始され、まず形態素解析処理が行なわれる(ステップS100)。形態素解析処理は、上述したように、形態素解析辞書IDCを参照して行なわれる処理であり、クライアント30から受け取った検索文から単語と文節を取り出す処理である。形態素解析処理(ステップS100)の詳細を図5のフローチャートに示した。

【0034】

形態素解析処理が開始されると、クライアント30から受け取った検索文が解析の対象として特定され、この文の先頭からM文字目(M=1, 2, ...)からL文字分(L=1, 2, ...)を取り出して解析辞書IDCを引く処理を行なう(ステップS102) 10。Mは、着目している文字列の先頭位置を、Lは、取り出す文字数を、それぞれ示していることになる。解析辞書の参照の手法は、まずM=1、即ち先頭位置から、L=1、即ち1文字分の文字を取り出し、辞書を参照して該当語を取り出す処理から開始する。Lを順次インクリメントしながら辞書IDCを参照し、該当する見出し語がなくなれば、着目する文字列の先頭位置Mをインクリメントし、再度文字数Lを1に戻して、辞書の検索を行なう。こうして着目する文字の位置か、解析しようとする文の文字数を越えたところで、辞書の参照をうち切る。

【0035】

例えば、クライアント30から「電源を入れたら壊れた」という検索文が入力された場合を想定すると、解析辞書IDCを参照すると、「電源を」「電」「源」「源を」「を」「入れたら」「入れた」「ら」「入れ」「たら」「た」「入」「れたら」「壊れた」「壊れ」「た」「壊」「れた」「れ」といった語を切り出すことができる。ここで、「た」などの仮名一音も、語として切り出しているのは、過去形の助動詞「た」などが、文中に現れる可能性があるからである。 20

【0036】

解析辞書IDCには、これらの語がその文法情報と共に記憶されている。そこで、切り出した語を次に文法情報に従って並べて、破綻しない配列を見い出す処理を行なう。かかる解析は、例えば複数文節最長一致法や最小コスト法といった手法が知られており、所定の語の組み合わせのうちどれが最も日本語としてもっとももらしいかを検定するのである。本実施例では、最小コスト法を採用しているため、こうして得られた多数の文字列を対象として、次にコスト計算を行なう(ステップS104) 30。コスト計算とは、文字列の配列に対して、日本語らしい配列ほど点数が低くなるように予め用意された文字列のコストを計算する処理である。その規則は大まかに言えば、自立語はコスト2、これに付属語が付属する場合はコスト0、といったものである。例えば、「電源を」を例にとると、「電源」+「を」ではあれば、自立語+付属語(助詞)の結びつきとなって、コスト2、「電」+「源」+「を」であれば、自立語+自立語+付属語(助詞)となってコストは4となるのである。最小コスト法のルールは、現実の日本語にあわせてチューニングされており、「まったく」+「ない」などの共起関係にある単語が文中に生じる場合は、コスト「-1」など、様々な規則が用意されている。

【0037】

こうして、逆引き辞書の参照により得られた全ての単語について、上記のコストを計算し、そのうちで最小のコストになる文を特定する処理を行なう(ステップS106) 40。上記の例では、「電」(自立語・名詞)+「源」(自立語・名詞)+「を」(付属語・助詞)よりも、「電源」(自立語・名詞)+「を」(付属語・助詞)の方が、日本語として確からしいと判断するのである。もとより、この計算は、少なくとも文を単位として行なわれ、文全体で、コストが最小になるような単語の配列を選択する。従って、例えば共起関係によるコストの低減などがあれば、異なる組み合わせが選択される場合も存在する。

【0038】

こうして最小コスト法により最小コスト文が特定されると、結局検索文を構成する文節の組み合わせが、その文法情報と共に得られたことになるので、次に、得られた文節を、図6 50

に示す配列に格納する処理を行なう（ステップS108）。図6は、検索文を解析する際に用いられる配列の一例を示す説明図である。検索文は、全体としては、単語情報（図6）、文節情報（図7）、部分文情報（図8）という形態で解析され、記憶される。このうち、図6は、単語情報の内容（配列）を示しており、この配列は、単語、単語の読み、品詞から構成されている。以下、単語の配列は、 $T[t]$ ($t = 0, 1, \dots$)として参照するものとする。

【0039】

こうして形態素解析を完了すると、次に係り受け解析（ステップS110）を行なう（図4参照）。係り受け解析とは、文を構成する各文節の関係を特定する処理である。係り受け解析は、文節情報を特定するための処理である。係り受け解析を行なうことにより、文節間の関係を知ることができる。即ち、ある文節がどの文節に係っているかを知ることができる。例えば、名詞+「を」（助詞）は後方の最も近い述部にかかる、というルールから、「電源を」「切ると」という関係が特定される。こうした係り受け解析により得られた文節情報は、配列Bに格納される。この配列B[b] ($b = 0, 1, \dots$)の一例を図7に示した。この文節情報は、単語を示すインデックスである配列B[b]、この配列B[b]に所属している単語の番号t、係り先文節の番号b、係りもと文節の番号bから構成されている。図7の表中、「-」は該当する文節が存在しないことを示している。配列に所属している単語の番号tが与えられれば、図6に示した配列T[t]を参照して、実際の単語を取得することができる。

【0040】

係り受け解析（ステップS110）が完了すると、次に部分文の判定処理を行なう（ステップS120）。この処理は、係り受け解析により解析した文節同士の関係を利用して、1以上の文節からなる部分文同士の関係を特定するものである。ここで部分文とは、少なくとも一つの述部を含み構文上の最小単位である節とほぼ等しい概念である。部分文同士の関係は、図8に示したように、配列S[s] ($s = 0, 1, \dots$)として与えられ、配列S[s]には、所属する文節の番号b、結論部からの距離、条件部の意味が対応づけられる。これら、単語の配列T[t]、文節の配列B[b]、部分文の配列S[s]の関係を図9に示した。図示するように、これらは、上位-下位の構成となっており、一つの部分文から、これに含まれる文節、単語などを自由に参照することができる。

【0041】

部分文の切出しの処理を図10に示す。この処理は、文判定ルールSDIを参照することにより行なわれる。文判定ルールSDIの一例を図11に示す。図11は、図10のフローチャートにおける判定単語列Rmを示したものである。また、各見だしは、条件部の意味を示している。図11の表中における「*」は、いわゆるワイルドカードを示しており、どんな単語でも当てはまることを示している。また、「*:*:動詞」は品詞が動詞の単語であれば、読みや見出しは問わず当てはまることを示している。例えば、図11中、符号INで示した文型は、条件部の意味は「条件」であり、「(*:*:動詞、*:*:活用語尾、と:*:接続助詞)」という文型を指定しているから、動詞の後に活用語尾がついた上で、接続助詞「と」が接続される総ての部分文を示していることになる。動詞「入れ」+活用形「る」+接続助詞「と」は、この文型に一致することになる。

【0042】

図10に示した部分文の解析処理ルーチンについて説明する。このルーチンが開始されると、まず検索文から不要文を削除する処理を行なう（ステップS200）。不要文とは、「どうしたらよいですか」と言った検索しようとする内容そのものとは関係がない部分である。これらの部分は、予め不要文のリストの形で記憶しておき、該当する文を削除するものとすればよい。例えば「電源を入れると壊れたのですが、どうしたらよいですか」といった検索部が与えられている場合には、形態素解析および係り受けの解析により、こうした不要文に相当する部分特定することができるので、これを削除するのである。削除した文節は、単語の配列（図6参照）や文節の配列（図7参照）や部分文の配列（図8参照）などから削除される。

10

20

30

40

50

【0043】

次に、部分文の解析を開始するものとして、解析処理を行なう検索文を構成する全単語数 n を設定し、着目する従属節の数を示す変数を初期化 ($j = 0$) する処理を行なう (ステップ S 2 1 0)。次のステップ S 2 2 0 では、図 1 1 に示した条件部を示す文例の数を示す変数 m を初期化し ($m = 0$)、以下、変数 m が図 1 1 に示した文例の総数になるまで、以下の処理を繰り返す。図 1 1 に示した文例は、一つの文例が () に括られている部分であり、先頭から順に $m = 1, 2, \dots$ として指定することができる。そこで、まず m 番目の文例を、判定単語列 R_m として取得し、併せて k に判定単語列 R_m のを構成する単語数を設定する処理を行なう (ステップ S 2 3 0)。例えば、上述した (* : * : 動詞、* : * : 活用語尾、と : * : 接続助詞) という文例では、構成単語数 k は、値 3 となる。

10

【0044】

次に、検索文にその最後尾から着目し、その $n - k + 1$ 番目から n までの単語列 $W(n - k + 1, n)$ を取得する処理を行なう (ステップ S 2 4 0)。対比する文例が、単語数 k なので、検索文からも単語数 k 個の単語からなる単語列を取り出すのである。単語列の取出は、単語を示す配列 $T[t]$ を用いて容易に取り出すことができる。例えば「電源を入れると壊れた」という文が検索文として入力された場合には、後方から 3 個の単語として、「と」+「壊れ」+「た」が取得されることになる。こうして比較用の単語列が取得されると、次に、両者を照合する処理を行ない (ステップ S 2 5 0)、両者が一致しなければ、次の文例を取得するために変数 m を値 1 だけインクリメントして (ステップ S 2 6 0)、図 1 1 に示した文例が尽きるまで (ステップ S 2 7 0)、ステップ S 2 3 0 に戻って処理を

20

【0045】

そこで、次に着目する単語列を末尾から一つ手前に移動するために変数 n を値 1 だけデクリメントし (ステップ S 2 8 0)、この変数 n が値 0 より小さくなるまで (ステップ S 2 9 0)、ステップ S 2 2 0 に戻って、変数 m を初期化する処理から上記の各処理を繰り返す。この処理を繰り返す結果、やがて末尾から 3 番目の単語「と」からの k 個の単語を取得するようになると、何番目かの文例 R_m 「(* : * : 動詞、* : * : 活用語尾、と : * : 接続助詞)」が、検索文の単語列 $W(n - k + 1, n)$ である「入れ」+「る」+「と」と一致する (ステップ S 2 5 0)。このとき、処理はステップ S 3 0 0 以下に分岐し、従属節が一つ見つかったとして、従属節を示す変数 j を値 1 だけインクリメントし (ステップ S 3 0 0)、この従属節に関する情報を設定する処理を行なう (ステップ S 3 1 0)。従属節に関する情報の設定については、次の段落で説明する。この処理の後、着目している単語の位置を、 $k - 1$ 個分だけ進め (ステップ S 3 2 0)、更に上述した変数 n のデクリメント (ステップ S 2 8 0) から、上記の処理を繰り返す。従属節が一つ見つかった後も更に処理を継続するのは、自然言語では、従属節が複数許されているからである。例えば、「突然パソコンが終了したので、電源を入れると壊れた」という検索文が入力された場合を想定すると、「突然パソコンが終了したので」と「電源を入れると」の二つが条件を示す従属節として設定されることになる。

30

40

【0046】

従属節に関する情報の設定は、図 8 に示した結論部からの距離と条件部の意味の二つである。結論部自身は、距離 0 であり、ここから文頭に向けて、結論部 (この文例では、「壊れた」) に近い従属節から、距離 1、2、 \dots となる。また一致した判定単語列 R_m に付与されていた分類に従い、「条件」「理由」「逆接」「並列」などの区別が、従属節に関する情報として、配列 $S[s]$ に対応づけて記憶される。

【0047】

(3) 検索システムの処理 - 照合処理 :

以上の処理により、図 4 に示した解析処理が完了する。次に照合処理が行なわれる。照合処理は、入力された検索文と、これに基づいてデータベース DB から検索した検索対象文

50

との照合を行なう処理であり、まず単語の照合処理を行なう(ステップS130)。ここでは、基本的には検索文に含まれていた単語を用いてデータベースDBを引く処理が行なわれるが、検索語についてはシソーラスTSRを参照し、類義語や同意語などを広く検索する。例えば、「電源」「入れる」という単語のみならず、「パワースイッチ」や「パワーサプライ」などの類義語や、「切る」に対して人間の身体動作として同じカテゴリに分類されている「切る」や「回す」なども検索の対象とされる。こうした検索処理により、データベースDBから多数の検索対象文が広汎に得られるから、自然言語により検索文が入力されても、検索漏れを生じることが少ない。

【0048】

単語の照合処理は、更に次のように行なわれる。検索対象となった文に、

10

1 検索文に含まれてる自立語が存在する場合には、類似点として値1を与え、2 シソーラスTSRにより上位概念が一致する単語が存在する場合には、値0.9を与える。

例えば、「電源を入れると壊れた」という検索文に対して、「PCの電源を切ると」という文がデータベースDB内に存在した場合には、単語「電源」については類似点として値1が与え「入れる」と「切る」については、共通の上位概念「身体動作」を持つので、類似点として値0.9を与えるのである。従って、この両文の類似点は、 $1 + 0.9 = 1.9$ となる。なお、こうした類似点の付与は、更に文末表現などに応じて細かく調整するものとしてもよい。例えば「壊れるようだ」とか「壊れるらしい」といった文が見い出された場合には、文末の関係表現に着目して、伝聞や推量であれば、値0.1ないし0.3を

20

【0049】

次に係り受けの照合の処理を行なう(ステップS140)。この処理は、ある単語に着目したとき、その単語の係り先の単語も一致する場合には、その単語についての類似点を増加するのである。例えば、「電源を入れる」と「電源を切る」という二つの文の場合、「電源を」という文節を構成する単語「電源」は、「入れる」と「切る」の両方に係り受けの関係を持っており、しかも、「入れる」と「切る」は身体動作という点で同一のカテゴリに属する。こうした場合には、「電源」についての類似点として与えられた値1を50%増加し、値1.5とする。なお、増加の仕方は、こうした50%アップなどに限られる

30

ものではなく、所定の値(例えば0.5)を付与するといった手法でも差し支えない。また、係り受けの係り先の単語が完全一致の場合には、更に高い値を与えるようにすることも望ましい。この結果、先の単語の照合とあわせると、「電源を入れる」と「電源を切る」との類似点は、 $1.5 + 0.9 = 2.4$ となる。

【0050】

係り受けの照合を行なった後、次に部分文の照合を行なう(ステップS150)。部分文の照合は、着目している部分文が、結論部に相当するか条件部に相当するかにより、類似点の増加を異ならせることにより、行なっている。この関係を図12に示した。「電源を入れる」と「電源を切る」とを例文として用いるものとして、

1 この両文が、検索文および対象文の結論部に存在していれば、類似点を100%増加するものとし、

40

2 一方が結論部に、他方が条件部に存在していれば、類似点を50%増加するものとし、

3 両文が、共に条件部に存在していれば、更に、両者の結論部からの距離を判定し、距離jが一致していれば、類似点を20%増加するものとし、

4 両文が共に条件部に存在しており、かつ結論部からの距離jが異なっていれば、類似点を10%増加するものとする、

のである。

【0051】

この結果、「電源を入れる」と「電源を切る」が共に結論部にあれば、類似点は、2.4

50

$\times 2 = 4.8$ となり、一方が結論部に他方が条件部にあれば、 $2.4 \times 1.5 = 3.6$ となり、共に条件部にあって結論部からの距離が等しければ、 $2.4 \times 1.2 = 2.88$ となり、結論部からの距離が異なれば、 2.64 となる。

【0052】

もう少し複雑な例文を例に挙げて、類似点の計算したものを以下に説明する。検索文としてクライアント30から入力した文が、「コンピュータの使用中にハングアップして、コンピュータの電源を切ることもできません」であり、データベースDBから、次の二つの文(A)(B)が、検索により取り出されたとする。

(A) PCの電源をいれると、オペレーティングシステムが起動する前に、「No System Disk」が表示されて起動が止まってしまいます。

10

(B) コンピュータの電源が切れません。

この二つの文について、単語の照合を行なうと、例文(A)については、「電源」が完全に一致、「コンピュータ」と「PC」、「切る」と「入れる」が、シソーラスTSRを参照して類似となる。従って、単語における類似点は、 $0.9 + 1 + 0.9 = 2.8$ となる。他方、例文(B)については、「コンピュータ」「電源」「切る(否定)」が完全一致するので、類似点は3となる。

【0053】

次に、係り受けによる照合を行なうと、例文(A)については、「PC」と「電源」の係り先が同一カテゴリと判断できるので、両者の類似点を50%増加して、 $0.9 \times 1.5 + 1 \times 1.5 + 0.9 = 3.75$ となる。他方、例文(B)については、「コンピュータ」と「電源」の係り先が同一と判断されるので、同様に50%増加して、 $1 \times 1.5 + 1 \times 1.5 + 1 = 4$ となる。

20

【0054】

更に、部分文の一致について照合すると、例文(A)については、「コンピュータを、使用中に ハングアップして」は条件部にあり、「PCの 電源を 切る(否定)」は結論部にあることから、類似点の総和3.75を50%増加して、最終的な類似点は、 5.63 となる。従って、この例文(A)と検索文との類似度は、類似点 $5.63 + 1 = 6.63$ として与えられる。他方、例文(B)については、単語が類似した部分文が共に結論部にあることから、類似点の総和4を100%増加して、 $4 \times 2 = 8$ となり、検索文(類似点1)との類似度は、 $8 + 1 = 9$ となる。

30

【0055】

この結果、例文(B)の方が例文(A)より、検索文により高い相関を示すと判断して、検索用サーバ20は、例文(B)を例文(A)により上位に配列して、クライアント30に出力する。検索用サーバ20からのデータを受けて、クライアント30上で動作しているブラウザは、図13に例示するように、例文(B)を例文(A)より上位に表示することになる。従って、検索を行なおうとしたものと、より相関の高い検索結果から順に参照することができ、所望の情報を一層容易に入力することができる。なお、上記の実施例では、検索の結果、類似度を判断して、より相関の高いと考えられる情報を上位に表示しているが、この場合に、類似点をあわせて表示したり、結論部で一致したか、条件部で一致したか等の情報を加えて表示するものとしてもよい。こうすれば、利用者は、検索結果を単に上位から順に眺めるだけでなく、どのような条件で一致した情報かを判断することができ、好適である。

40

【0056】

また、上記実施例では、検索文の解析も検索用サーバ20で行なったが、検索文の解析をクライアント30側で行なうものとすることもできる。あるいは、検索用サーバ20は、クライアント30側から受け取った単語による検索だけを行ない、検索語に部分一致が見いだされたデータをすべてクライアント30側に渡し、クライアント30側で、図4に示した解析処理および照合処理のすべてを行なうものとしても良い。解析処理と照合処理を、検索用サーバ20側とクライアント30側にわけても良い。あるいは、クライアント30と検索用サーバ20との間に専用のサーバを設けて、ここで、解析処理や照合処理を行

50

なっても良い。

【 0 0 5 7 】

上記実施例では、シソーラス T S R を設けて、検索文に含まれる単語の類義語などを含めて広く検索を行ない、検索語の偏りなどによる検索漏れを防止しているが、検索を実行する前に、検索文を標準化することで検索漏れを防止しても良い。こうした標準化の処理としては、半角/全角文字の統一などの文字の標準化、送りがなや長音記号の有無などの表記の標準化、同一の意味の他の自立語への統一など自立語の標準化など、種々のレベルを考慮することができる。検索前にこうした標準化を行なっておけば、シソーラス T S R の参照を行なわないか、行なうとしても限定的なものにとどめることができる。

【 0 0 5 8 】

以上、本発明の実施の形態について説明したが、本発明はこうした実施の形態に何等限定されるものではなく、本発明の要旨を逸脱しない範囲内において、更に種々なる形態で実施し得ることは勿論である。例えば、本実施例の検索システムは、クライアント - サーバシステムとして実現したが、スタンドアロンのコンピュータで実現しても差し支えない。また、検索対象としては、ネットワーク上のサイトなどを対象とすることも可能である。更に、上記実施の形態や実施例では、検索結果を評価して対象文の振り分けを行なった後、これをクライアント側に出力しているが、検索結果の評価と対象文の振り分けまでとどめても差し支えない。評価され振り分けられた対象文を単に出力するだけでなく、評価され振り分けられた対象文を推論エンジンの推論対象として利用するなど、多様な応用が可能である。マイク 1 3 を用いて検索文を音声認識により入力する構成や、検索結果を音声により報知する構成も可能である。

【 図面の簡単な説明 】

【 図 1 】 本発明の実施の形態としての検索システム 1 0 0 の概略構成を示すブロック図である。

【 図 2 】 本発明の一実施例としての検索システム 5 0 の構成を示す概略構成図である。

【 図 3 】 形態素解析辞書 I D C の一部を例示する説明図である。

【 図 4 】 検索用サーバ 2 0 が実行する検索処理の概要を示す説明図である。

【 図 5 】 形態素解析処理ルーチンを示すフローチャートである。

【 図 6 】 形態素解析により得られる単語の配列 T [t] の一例を示す説明図である。

【 図 7 】 係り受け解析により得られる文節の配列 B [b] の一例を示す説明図である。

【 図 8 】 部分文の解析により得られる部分文の配列 S [s] の一例を示す説明図である。

【 図 9 】 単語、文節、部分文の構成礼を示す説明図である。

【 図 1 0 】 部分文の解析ルーチンを示すフローチャートである。

【 図 1 1 】 部分文の解析に用いられる判定単語列 R m を例示する説明図である。

【 図 1 2 】 部分文の照合時における類似点の増加の条件とその割合を示す説明図である。

【 図 1 3 】 検索結果の表示例を示す説明図である。

【 符号の説明 】

1 0 ... ネットワーク

1 1 ... キーボード

1 2 ... マウス

1 3 ... マイク

1 8 ... ルータ

2 0 ... 検索用サーバ

2 2 ... C P U

2 3 ... R O M

2 4 ... R A M

2 5 ... タイマ

2 6 ... 表示回路

2 7 ... ハードディスク

2 9 ... モニタ

10

20

30

40

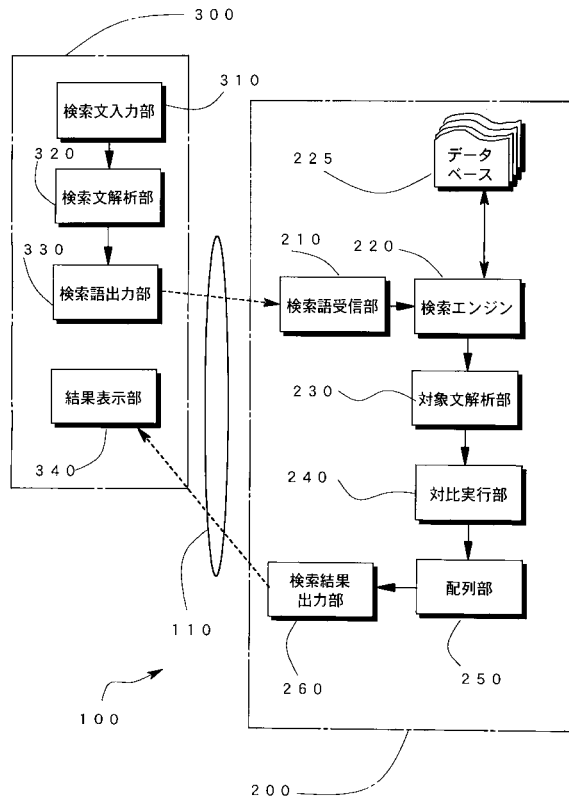
50

- 3 0 ...クライアント
- 3 0 ...検索用サーバ
- 5 0 ...検索システム
- 1 0 0 ...検索システム
- 1 1 0 ...ネットワーク
- 2 0 0 ...サーバコンピュータ
- 2 1 0 ...検索語受信部
- 2 2 0 ...検索エンジン
- 2 2 5 ...知識データベース
- 2 3 0 ...対象文解析部
- 2 4 0 ...対比実行部
- 2 5 0 ...配列部
- 2 6 0 ...検索結果出力部
- 3 0 0 ...クライアントコンピュータ
- 3 1 0 ...検索文入力部
- 3 2 0 ...検索文解析部
- 3 3 0 ...検索語出力部
- 3 4 0 ...結果表示部
- D B ...検索対象データベース
- I D C ...形態素解析辞書
- R m ...判定単語列
- S D I ...文判定ルール
- T S R ...単語シソーラス

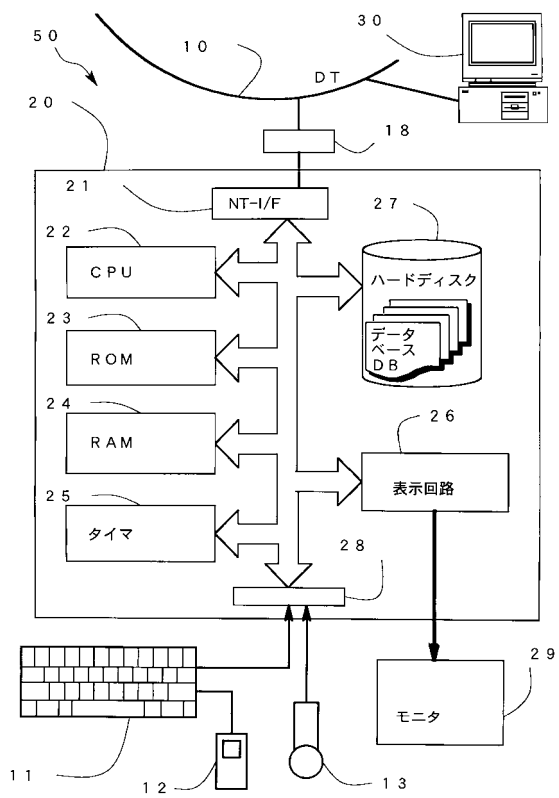
10

20

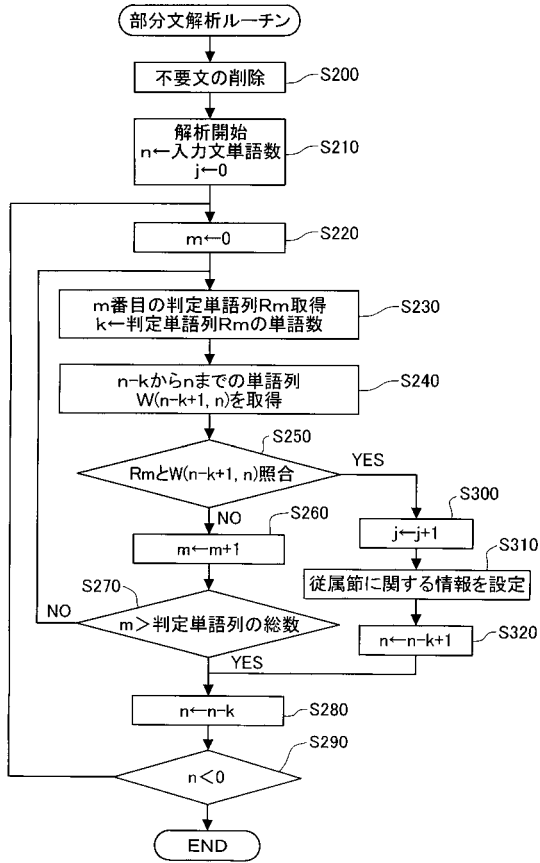
【図1】



【図2】



【 図 1 0 】



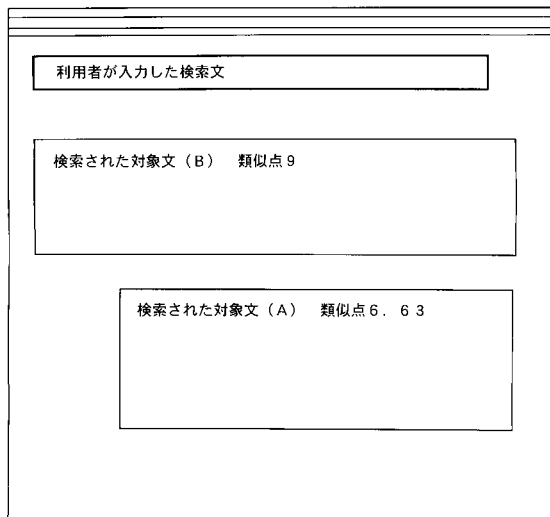
【 図 1 1 】

	Rm	IN
条件	(**動詞, **活用語尾, ば*接続助詞), (**動詞, ば*接続助詞), (**動詞, **活用語尾, と*接続助詞), (**動詞, と*接続助詞), (**動詞, **活用語尾, て*接続助詞), (**動詞, て*接続助詞), (**動詞, **活用語尾, ば*接続助詞), (**動詞, ば*接続助詞), (**動詞, たら*助動詞たら終止形), (**動詞, とぎ*形式名詞, に*格助詞), (**動詞, 前に*形式名詞, に*格助詞), etc.	(**動詞, ば*接続助詞), (**動詞, ば*接続助詞), (**動詞, と*接続助詞), (**動詞, と*接続助詞), (**動詞, て*接続助詞), (**動詞, て*接続助詞), (**動詞, ば*接続助詞), (**動詞, ば*接続助詞), (**動詞, たら*助動詞たら終止形), (**動詞, とぎ*形式名詞, に*格助詞), (**動詞, 前に*形式名詞, に*格助詞), etc.
理由	(**動詞, **活用語尾, ので*接続助詞), (**動詞, ので*接続助詞), (**動詞, **活用語尾, ため*形容名詞), (**動詞, ため*接続助詞), (**動詞, **活用語尾, ため*形容名詞, に*格助詞), (**動詞, ため*接続助詞, に*格助詞), etc.	(**動詞, **活用語尾, ので*接続助詞), (**動詞, ので*接続助詞), (**動詞, **活用語尾, ため*形容名詞), (**動詞, ため*接続助詞), (**動詞, **活用語尾, ため*形容名詞, に*格助詞), (**動詞, ため*接続助詞, に*格助詞), etc.
逆接	(**動詞, **活用語尾, が*接続助詞), (**動詞, が*接続助詞), (**動詞, **活用語尾, ても*接続助詞), (**動詞, ても*接続助詞), etc.	(**動詞, **活用語尾, が*接続助詞), (**動詞, が*接続助詞), (**動詞, **活用語尾, ても*接続助詞), (**動詞, ても*接続助詞), etc.
並列	(**動詞, **活用語尾, とともに*接続助詞と相当), (**動詞, とともに*接続助詞と相当), (**動詞, **活用語尾, ほか*形式名詞, に*格助詞), (**動詞, ほか*接続助詞, に*格助詞), (**動詞, **活用語尾, たり*接続助詞), (**動詞, たり*接続助詞), etc.	(**動詞, **活用語尾, とともに*接続助詞と相当), (**動詞, とともに*接続助詞と相当), (**動詞, **活用語尾, ほか*形式名詞, に*格助詞), (**動詞, ほか*接続助詞, に*格助詞), (**動詞, **活用語尾, たり*接続助詞), (**動詞, たり*接続助詞), etc.

【 図 1 2 】

	条件部	結論部
条件部	20% (距離一致), 10% (距離不一致)	50%
結論部	50%	100%

【 図 1 3 】



フロントページの続き

(58)調査した分野(Int.Cl. , D B名)

G06F 17/30