(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第3627446号
(P3627446)
(45)発行日　平成17年3月9日(2005.3.9)　(24)登録日　平成16年12月17日(2004.12.17)

(51)Int.Cl.[7]　F I
G06F 17/30　G06F 17/30　340C
G06F 17/30　170A

請求項の数 4　(全 12 頁)

(21)出願番号　特願平9-146259
(22)出願日　平成9年6月4日(1997.6.4)
(65)公開番号　特開平10-334115
(43)公開日　平成10年12月18日(1998.12.18)
審査請求日　平成13年7月10日(2001.7.10)

前置審査

(73)特許権者　000005496
富士ゼロックス株式会社
東京都港区赤坂二丁目１７番２２号
(74)代理人　100092152
弁理士　服部　毅巖
(72)発明者　倉持　勉
神奈川県足柄上郡中井町境４３０　グリーンテクなかい　富士ゼロックス株式会社内
(72)発明者　倉橋　政之
神奈川県足柄上郡中井町境４３０　グリーンテクなかい　富士ゼロックス株式会社内

審査官　深津　始

最終頁に続く

(54)【発明の名称】検索式作成装置及び検索式作成方法

(57)【特許請求の範囲】
【請求項１】
文書検索のための検索式を作成する検索式作成装置において、
複数の文書データを保持する文書データ保持手段と、
入力される検索キーワードを保持する検索キーワード保持手段と、
前記検索キーワードの近傍に位置する近傍語のうち、前記検索キーワードの前方から抽出する前方単語数と後方から抽出する後方単語数とが個別に指定された近傍語抽出条件を保持する近傍語抽出条件保持手段と、
前記検索キーワード及び前記近傍語抽出条件に基づいて、前記複数の文書データから、前記検索キーワードから前方に数えて前記前方単語数内に位置する前記近傍語と、前記検索キーワードから後方に数えて前記後方単語数内に位置する前記近傍語とを抽出する近傍語抽出手段と、
前記近傍語から近傍語リストを生成する近傍語リスト生成手段と、
前記近傍語リストを保持する近傍語リスト保持手段と、
表示された前記近傍語リストに含まれた近傍語への選択を受け付けて、選択された近傍語を前記検索キーワードとして追加するキーワード追加手段と、
前記検索キーワードから検索式を作成する検索式作成手段と、
を有することを特徴とする検索式作成装置。
【請求項２】
複数の文書データから文書を検索する文書検索装置において、

前記複数の文書データを保持する文書データ保持手段と、
入力される検索キーワードを保持する検索キーワード保持手段と、
前記検索キーワードの近傍に位置する近傍語のうち、前記検索キーワードの前方から抽出する前方単語数と後方から抽出する後方単語数とが個別に指定された近傍語抽出条件を保持する近傍語抽出条件保持手段と、
前記検索キーワード及び前記近傍語抽出条件に基づいて、前記複数の文書データから、前記検索キーワードから前方に数えて前記前方単語数内に位置する前記近傍語と、前記検索キーワードから後方に数えて前記後方単語数内に位置する前記近傍語とを抽出する近傍語抽出手段と、
前記近傍語から近傍語リストを生成する近傍語リスト生成手段と、
前記近傍語リストを保持する近傍語リスト保持手段と、
表示された前記近傍語リストに含まれた近傍語への選択を受け付けて、選択された近傍語を前記検索キーワードとして追加するキーワード追加手段と、
前記検索キーワードから検索式を作成する検索式作成手段と、
前記検索式にて前記複数の文書データから文書を検索する検索手段と、
を有することを特徴とする文書検索装置。

【請求項３】

文書検索のための検索式を作成する検索式作成方法において、
文書データ保持手段が、複数の文書データを保持するステップと、
検索キーワード保持手段が、入力される検索キーワードを保持するステップと、
近傍語抽出条件保持手段が、前記検索キーワードの近傍に位置する近傍語のうち、前記検索キーワードの前方から抽出する前方単語数と後方から抽出する後方単語数とが個別に指定された近傍語抽出条件を保持するステップと、
近傍語抽出手段が、前記検索キーワード及び前記近傍語抽出条件に基づいて、前記複数の文書データから、前記検索キーワードから前方に数えて前記前方単語数内に位置する前記近傍語と、前記検索キーワードから後方に数えて前記後方単語数内に位置する前記近傍語とを抽出するステップと、
近傍語リスト生成手段が、前記近傍語から近傍語リストを生成するステップと、
近傍語リスト保持手段が、前記近傍語リストを保持するステップと、
キーワード追加手段が、表示された前記近傍語リストに含まれた近傍語への選択を受け付けて、選択された近傍語を前記検索キーワードとして追加するステップと、
検索式作成手段が、前記検索キーワードから検索式を作成するステップと、
を有することを特徴とする検索式作成方法。

【請求項４】

複数の文書データから文書を検索する文書検索方法において、
文書データ保持手段が、前記複数の文書データを保持するステップと、
検索キーワード保持手段が、入力される検索キーワードを保持するステップと、
近傍語抽出条件保持手段が、前記検索キーワードの近傍に位置する近傍語のうち、前記検索キーワードの前方から抽出する前方単語数と後方から抽出する後方単語数とが個別に指定された近傍語抽出条件を保持するステップと、
近傍語抽出手段が、前記検索キーワード及び前記近傍語抽出条件に基づいて、前記複数の文書データから、前記検索キーワードから前方に数えて前記前方単語数内に位置する前記近傍語と、前記検索キーワードから後方に数えて前記後方単語数内に位置する前記近傍語とを抽出するステップと、
近傍語リスト生成手段が、前記近傍語から近傍語リストを生成するステップと、
近傍語リスト保持手段が、前記近傍語リストを保持するステップと、
キーワード追加手段が、表示された前記近傍語リストに含まれた近傍語への選択を受け付けて、選択された近傍語を前記検索キーワードとして追加するステップと、
検索式作成手段が、前記検索キーワードから検索式を作成するステップと、
検索手段が、前記検索式にて前記文書データから文書を検索するステップと、

を有することを特徴とする文書検索方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は検索式作成装置及び検索式作成方法に関し、特に文書検索のための検索式を作成する検索式作成装置及び検索式作成方法に関する。

【０００２】

【従来の技術】

多量の文書データから必要に応じて文書データを検索する方法には、シソーラスに基づいて統制されたキーワードや分類コード等の情報を利用する方法のほかに、文書全文を直接参照して検索する全文検索（フルテキストサーチ）方法がある。前者の方法に比べて後者の方法では検索もれが少ない。

【０００３】

しかし、全文検索方法では検索もれが少ない代わりに多量の文書がヒットとなってしまい、この内容を逐一チェックするのには大変な労力を要する。例えば、「有害物質の輸出入の規制」に関する内容の文書を検索する場合に、「”有害物質”、”輸出入”、”規制”という単語を含む」ことを検索条件としてしまうと、これらの単語が散在した文書もヒットとなってしまう場合がある。

【０００４】

そこで、この全文検索方法で情報をより正確に絞り込むために、ユーザが指定した検索語（検索キーワード）の近傍のテキスト情報を利用する方法が提案されている。

【０００５】

例えば、検索者が２つの検索キーワード及びその位置関係を指定すると、その検索条件に合致する検索結果を出力する「文書検索方法および装置」が、特開平４－２９３１６１号公報に示されている。

【０００６】

この方法では、上記で例とした「有害物質の輸出入の規制」に関する内容の文書を抽出する場合、「”有害物質”という単語と”輸出入”という単語とを含み、その間隔は２語以内であり、かつ、”輸出入”という単語と”規制”という単語とを含み、その間隔は２語以内である」ことを検索条件とすることができる。よって、検索キーワードが散在した文書のヒットを防止することができる。

【０００７】

この方法は、検索要求が「有害物質の輸出入の規制」というように明確である場合には有効である。しかし、例えばユーザの検索要求で明らかになっているのが、検索キーワード「規制」のみ、という場合もある。

【０００８】

このように、検索要求が定かでない場合に、ユーザが１つの検索キーワードを指定すると、その検索キーワードが出現する箇所の前後のテキストを抜き出してユーザに提示する「対話式データベース探索方法および装置」が特開平５－１８９４８７号公報に示されている。

【０００９】

この方法では、例えば「規制」という検索キーワードを指定すると「有害物質の輸出入の規制」を始め、「～の規制」、「～に関する規制」といったテキストが列挙され、ユーザはその中から必要な情報を選択的に得ることができる。

【００１０】

【発明が解決しようとする課題】

しかし、上記の方法では提示されるテキストが大量となり、ユーザの選択作業にかかる負担が増大してしまうという問題点があった。なお、この原因の１つは、意味が殆ど同じであっても言い回しの異なるテキストは、すべて区別して提示してしまうことにある。例えば、「排気ガスの規制」と「排気ガスに関する規制」などがこれにあたる。

【００１１】
本発明は以上のような点に鑑みてなされたものであり、文書の検索要求を絞り込むための情報提示を効率良く行い、検索式作成にかかるユーザの負担を軽減した検索式作成装置及び検索式作成方法を提供することを目的とする。
【００１２】
【課題を解決するための手段】
本発明では上記課題を解決するために、文書検索のための検索式を作成する検索式作成装置において、複数の文書データを保持する文書データ保持手段と、入力される検索キーワードを保持する検索キーワード保持手段と、前記検索キーワードの近傍に位置する近傍語のうち、前記検索キーワードの前方から抽出する前方単語数と後方から抽出する後方単語数とが個別に指定された近傍語抽出条件を保持する近傍語抽出条件保持手段と、前記検索キーワード及び前記近傍語抽出条件に基づいて、前記複数の文書データから、前記検索キーワードから前方に数えて前記前方単語数内に位置する前記近傍語と、前記検索キーワードから後方に数えて前記後方単語数内に位置する前記近傍語とを抽出する近傍語抽出手段と、前記近傍語から近傍語リストを生成する近傍語リスト生成手段と、前記近傍語リストを保持する近傍語リスト保持手段と、表示された前記近傍語リストに含まれた近傍語への選択を受け付けて、選択された近傍語を検索キーワードとして追加するキーワード追加手段と、前記検索キーワードから検索式を作成する検索式作成手段と、を有することを特徴とする検索式作成装置が提供される。
【００１３】
このような検索式作成装置にて文書検索のための検索式を作成する場合、文書データ保持手段が複数の文書データを保持する。また、検索キーワード保持手段が入力される検索キーワードを保持する。近傍語抽出条件保持手段は入力される近傍語抽出条件を保持する。近傍語抽出手段は検索キーワード及び近傍語抽出条件に基づいて、複数の文書データから近傍語を抽出する。近傍語リスト生成手段は近傍語から近傍語リストを生成する。近傍語リスト保持手段は、生成された近傍語リストを保持する。キーワード追加手段は表示された前記近傍語リストに含まれた近傍語への選択を受け付けて、選択された近傍語を検索キーワードとして追加する。そして検索式作成手段は検索キーワードから検索式を作成する。
【００１４】
このようにして文書検索のための検索式を作成すると、検索要求を絞り込むための情報、すなわち検索キーワード及び近傍語抽出条件に基づいて抽出される近傍語の提示が効率良く行われ、この近傍語を検索式を作成するための検索キーワードとして追加できるので、検索式作成にかかるユーザの負担を軽減することが可能となる。
【００１５】
【発明の実施の形態】
以下、本発明の実施の形態を図面を参照して説明する。
図１は本発明の検索式作成装置の原理構成を示したブロック図である。
【００１６】
本発明の検索式作成装置は、文書データ保持手段１と、検索キーワード保持手段２と、近傍語抽出条件保持手段３と、近傍語抽出手段４と、近傍語リスト生成手段５と、近傍語リスト保持手段６と、キーワード追加手段７と、検索式作成手段８と、から構成される。
【００１７】
ここで、文書データ保持手段１は、複数の文書データを保持する。また、検索キーワード保持手段２は、入力される検索キーワードを保持する。近傍語抽出条件保持手段３は、入力される近傍語抽出条件を保持する。近傍語抽出手段４は、検索キーワード及び近傍語抽出条件に基づいて、文書データから近傍語を抽出する。近傍語リスト生成手段５は、近傍語リストを生成する。近傍語リスト保持手段６は、生成された近傍語リストを保持する。キーワード追加手段７は、近傍語リストから選択された近傍語を検索キーワードとして追加する。そして検索式作成手段８は、検索キーワードから検索式を作成する。

【００１８】
このような検索式作成装置にて文書検索のための検索式を作成する場合、ユーザが入力する検索キーワードは、検索キーワード保持手段２に保持される。すなわち、検索キーワード保持手段２は、入力される検索キーワードが「ａａａａ」ならば、これを保持する。また同様に、ユーザが入力する近傍語抽出条件は、近傍語抽出条件保持手段３に保持される。
【００１９】
そして近傍語抽出手段４は、ユーザの指示を受けると、検索キーワード及び近傍語抽出条件に基づいて、文書データ保持手段１に保持された文書データから、近傍語を抽出する。例えば検索キーワード「ａａａａ」を含む文書データとして「ａａａａ、ｂｂｂｂ、・・・」及び「ａａａａ、ｃｃｃｃ、・・・」が保持されている場合、近傍語抽出条件に合致することを確認してから、「ｂｂｂｂ」や「ｃｃｃｃ」を近傍語として抽出する。
【００２０】
この際、抽出元となる文書データのＩＤ等を抽出された近傍語に関連付けておくが、複数の文書データから同じ近傍語が抽出される場合もあり、近傍語と抽出元文書データとの関係は１：１とは限らない。なお、本明細書中における近傍語とは、検索キーワードの近傍（範囲は近傍語抽出条件にて定められる）に存在し、単独で意味をなす自立語を指す。
【００２１】
近傍語リスト生成手段５は、抽出された近傍語にて近傍語リストを生成する。すなわち、抽出された近傍語が「ｂｂｂｂ」と「ｃｃｃｃ」であれば、これらをリストアップする。
近傍語リスト保持手段６は、生成された近傍語リストを保持する。表示される近傍語リストを見たユーザが、リスト中の任意の近傍語を選択すると、キーワード追加手段７はこの選択を受け付け、選択された近傍語を、検索キーワード保持手段２に検索キーワードとして追加する。ここでユーザが近傍語「ｂｂｂｂ」を選択すれば、この単語が新たな検索キーワードとして追加される。
【００２２】
そして検索式作成手段８は、ユーザの指示を受けると、検索キーワード保持手段２に保持された検索キーワードから検索式を作成する。すなわち、この時点で検索キーワードとして保持されているのが「ａａａａ」と「ｂｂｂｂ」であれば、これらの単語から検索式を作成する。
【００２３】
このようにして検索式を作成すると、検索要求を絞り込むための情報、すなわち近傍語の提示が効率良く行われ、この近傍語を検索式を作成するための検索キーワードとして追加できるので、検索式作成にかかるユーザの負担を軽減することが可能となる。
【００２４】
ここで、本発明の検索式作成装置を文書検索装置と組み合わせて実施する場合の構成及び動作について説明する。
図２は、本発明の検索式作成装置を適用した文書検索装置のハードウェア構成図である。
【００２５】
本発明の検索式作成装置を適用した文書検索装置１０は文書情報記憶部２０を持ち、入力装置３０を介してユーザからの入力を受け付け、出力装置４０を介してユーザに情報を出力する。ここで、入力装置３０としてはマウス等のポインティングデバイスやキーボード等が適用可能である。また、出力装置４０としてはモニタやプリンタ等が適用可能である。
【００２６】
文書検索装置１０は、テキスト検索処理部１１と、検索結果記憶部１２と、近傍語抽出処理部１３と、近傍語記憶部１４と、同一文内近傍語検出処理部１５と、品詞分類処理部１６と、から構成されている。
【００２７】
ここで、テキスト検索処理部１１はＣＰＵ等で構成され、入力装置３０を介して入力され

る検索キーワードや検索式に基づいて、文書情報記憶部２０から文書情報を検索する。検索にあたっては全文検索等の公知技術が適用される。

【００２８】

検索結果記憶部１２はメモリ等で構成され、テキスト検索処理部１１の検索結果を記憶する。すなわち、検索キーワードや検索式と合致した文章の文番号及びその文章を含む文書のＩＤ等を関連付けて記憶する。また、ここに記憶される検索結果は、出力装置４０を介して出力される。

【００２９】

近傍語抽出処理部１３はＣＰＵ等で構成され、入力装置３０を介して入力される近傍語抽出条件に基づいて、検索結果記憶部１２に記憶されている検索結果の文書データから近傍語を抽出する。近傍語を抽出するには、検索キーワードや検索式と合致した文章を形態素解析し、近傍語抽出条件に合った単語を検出すればよい。

【００３０】

なお、近傍語抽出条件は「［ｍ］検索キーワード［ｎ］」といった形式で入力される。これは、検索キーワードの前方に位置し、単独で意味を持つ自立語ｍ個と、検索キーワードの後方に位置し、単独で意味を持つ自立語ｎ個とを、近傍語として抽出することを示す。また、「検索キーワード１＊［ｍ］検索キーワード２［ｎ］」（＊は論理積）と入力された場合には、検索キーワード１と検索キーワード２との両方を含む文書を対象として、「［ｍ］検索キーワード２［ｎ］」を満たす近傍語を抽出すればよい。

【００３１】

また、この際行われる形態素解析は自然言語処理の基本技術として広く知られる技術で、「自然言語処理の基礎技術」（野村浩郷著、社団法人電子情報通信学会、１９８８）等に詳しい。

【００３２】

近傍語記憶部１４はメモリ等で構成され、近傍語抽出処理部１３にて抽出される近傍語を記憶する。なお、近傍語は、その近傍語を抽出した元の文章の文番号及び文書のＩＤ等と関連付けて記憶する。また、ここに記憶された近傍語は品詞分類処理部１６に入力される。

【００３３】

同一文内近傍語検出処理部１５は、出力装置４０を介して出力された近傍語のリストからユーザが任意の近傍語を選択して検索キーワードを追加した場合に、追加された検索キーワードを含む文を近傍語記憶部１４で検索し、この文に含まれる近傍語を検出する。検出した近傍語は品詞分類処理部１６に入力される。

【００３４】

品詞分類処理部１６は、入力された近傍語を品詞、すなわち、名詞、動詞・・・等に分類して近傍語リストを生成する。生成した近傍語リストは出力装置４０を介して出力する。

【００３５】

次に、この文書検索装置１０にて文書検索を行う手順を説明する。

図３は、本発明の検索式作成装置を適用した文書検索装置にて文書を検索する手順を示したフローチャートである。以下、ステップ番号に沿って説明する。

［Ｓ１］テキスト検索処理部１１は、入力装置３０を介して入力される検索キーワードを受け付ける。

［Ｓ２］テキスト検索処理部１１は、受け付けた検索キーワードにて文書情報記憶部２０を検索し、検索キーワードを含んだ文章の文番号及びその文章を含む文書データのＩＤ等を検索結果として検索結果記憶部１２に記憶させる。

［Ｓ３］近傍語抽出処理部１３は、入力装置３０を介して入力される近傍語抽出条件を受け付ける。

［Ｓ４］近傍語抽出処理部１３は、入力された検索キーワードもしくは検索式を含んだ文章を１つ選択し、この文章を形態素解析する。なお、この形態素解析の際、名詞が連続して出現した場合には、その部分を単独の複合語として扱うこととする。よって、例えば「

形態素」という単語と「解析」という単語とが連続して出現した場合には、「形態素解析」という単語が出現したものとして扱う。
［Ｓ５］近傍語抽出処理部１３は、検索キーワード及び近傍語抽出条件に基づいて、形態素解析した文章から近傍語を抽出し、その文章の文番号及びその文章を含む文書のＩＤ等を、近傍語と関連付けて近傍語記憶部１４に記憶させる。
［Ｓ６］近傍語抽出処理部１３は、検索結果記憶部１２にステップＳ４～Ｓ５の処理の済んでいない文書データがあるか否か判断する。未処理の文書データがあればステップＳ４へ再度進み、未処理の文章を選択する。また、未処理の文書データがない場合はステップＳ７へ進む。
［Ｓ７］近傍語記憶部１４は、記憶した近傍語をすべて品詞分類処理部１６に入力する。品詞分類処理部１６は、この近傍語を品詞分類して近傍語リストを生成する。
［Ｓ８］品詞分類処理部１６は、生成した近傍語リストを出力装置４０を介して出力する。
［Ｓ９］同一文内近傍語検出処理部１５は、入力装置３０を介してユーザから、追加する検索キーワードが選択されたか否か判断する。ここで選択されるのは、ステップＳ８で表示した近傍語リストに含まれた近傍語の１つである。追加する検索キーワードが選択された場合はステップＳ１０に進む。また、追加する検索キーワードが選択されない場合はステップＳ１１に進む。
［Ｓ１０］検索キーワードが複数存在するので、検索式を作成する。この詳細については後に説明する。検索式が作成した後、ステップＳ４に再度進む。
［Ｓ１１］ここまでの処理で作成された検索式があれば、その検索式はテキスト検索処理部１１に入力される。テキスト検索処理部１１は、その検索式に基づいて文書情報記憶部２０を検索する。検索結果は検索結果記憶部１２に記憶される。なお、検索式が作成されていない場合には、この処理を省いてもよい。
［Ｓ１２］検索結果記憶部１２は、記憶した検索結果を出力装置４０を介して出力する。
【００３６】
ここで、ステップＳ１０で行う検索式の作成について説明する。出力装置４０としてモニタ等の表示装置が適用されている場合、この時点でこの表示装置には近傍語のリストが表示されている。
【００３７】
図４は、近傍語リストの表示画面例を示す図である。
表示画面５０には、近傍語リスト５１と、開始ボタン５６と、終了ボタン５７とが表示されている。
【００３８】
ユーザは、検索キーワードを追加して同一文内の近傍語を検出したい場合に、マウス等のポインティングデバイスで開始ボタン５６を選択する。また、検索キーワードの追加の必要がない場合に、終了ボタン５７を選択する。
【００３９】
近傍語リスト５１は、近傍語抽出条件表示欄５２と、近傍語＜名詞＞表示欄５３と、近傍語＜動詞＞表示欄５４と、近傍語＜その他＞表示欄５５と、から構成されている。また、近傍語＜名詞＞表示欄５３及び近傍語＜動詞＞表示欄５４には、それぞれスクロールバーが付随している。
【００４０】
近傍語抽出条件表示欄５２には、入力された近傍語抽出条件が表示される。また、近傍語＜名詞＞表示欄５３には、抽出された近傍語のうち、名詞が表示される。近傍語＜動詞＞表示欄５４には、抽出された近傍語のうち、動詞が表示される。そして、近傍語＜その他＞表示欄５４には、抽出された近傍語のうち、名詞でも動詞でもないものが表示される。
【００４１】
なお、ここでは検索キーワードを「規制」、近傍語抽出条件を「［２］規制」として文書検索を行った場合の例を示している。これは、近傍語抽出条件「［２］規制［０］」と等

価である。

【００４２】

検索キーワード「規制」を図２に示した入力装置３０を介して入力した場合、テキスト検索処理部１１は、文書情報記憶部２０に記憶されている文書データに対し、この検索キーワード「規制」に基づいた文書検索を行う。この検索結果は検索結果記憶部１２に記憶される。

【００４３】

近傍語抽出条件「［２］規制」を入力装置３０を介して入力された近傍語抽出処理部１３は、検索結果記憶部１２に記憶された検索結果の文書データを形態素解析する。そして、検索キーワード「規制」の出現の度に、この検索キーワード「規制」より前方に出現し、単独で意味を持つ単語を２つ抽出し、近傍語とする。抽出した近傍語は近傍語記憶部１４に記憶され、品詞分類処理部１６にて品詞分類されて近傍語リスト５１が生成される。

【００４４】

このように、抽出される近傍語を品詞別に分類することで、ユーザは検索キーワードを追加するにあたって、適当な単語を探しやすくなる。

なお、ここでは検索キーワードの前方に出現する単語のみを近傍語とする場合を説明したが、近傍語抽出条件によって検索キーワードの後方に出現する単語を近傍語として抽出する場合もある。その場合には、前方に出現した近傍語と後方に出現した近傍語とを区別して表示させてもよい。

【００４５】

また、ここでは近傍語を品詞別に分けて表示したが、これを分けずに表示することも可能である。

次に、ユーザが検索キーワードを追加する場合を説明する。

【００４６】

「譲渡、取引、輸出入に関する規制」に関心を持った場合、ユーザはマウス等のポインティングデバイスで近傍語「譲渡」、「取引」、「輸出入」を指定してから、開始ボタン５６を選択する。

【００４７】

図５は、追加する検索キーワードを指定し、同一文内の近傍語を検出した様子を示す図である。

表示画面５０ａには、近傍語リスト５１ａが表示される。

【００４８】

近傍語記憶部１４では先に近傍語抽出条件「［２］規制」に基づいて近傍語を抽出した際に、各近傍語と、その近傍語を抽出した元文書のＩＤ等とを対応付けて記憶しているので、検索キーワードを追加されると同一文内近傍語検出処理部１５は、近傍語記憶部１４から同一文内に存在する近傍語を検出する。

【００４９】

従って検索キーワードとして「譲渡」、「取引」、「輸出入」が追加指定されると、「規制」という単語を含み、なおかつ「譲渡」もしくは「取引」もしくは「輸出入」という単語を含む文書から、「規制」という単語の前方に出現し、単独で意味を持つ単語２つが近傍語として検出され、近傍語リスト５１ａに表示される。近傍語リスト５１ａに提示されている近傍語の数は、近傍語リスト５１に提示されている近傍語の数と比較すると少なくなっている。

【００５０】

ここでユーザがさらに「野生生物、希少野生動物の譲渡、取引、輸出入に関する規則」に関心を持った場合、ユーザは近傍語「野生生物」、「希少野生動物」を指定してから、開始ボタン５６を選択する。

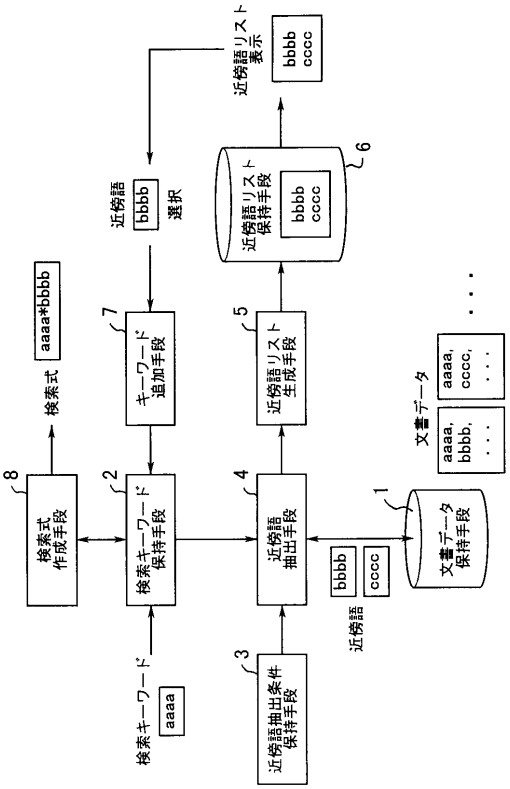
【００５１】

図６は、追加する検索キーワードを指定し、同一文内の近傍語を検出した様子を示す図である。

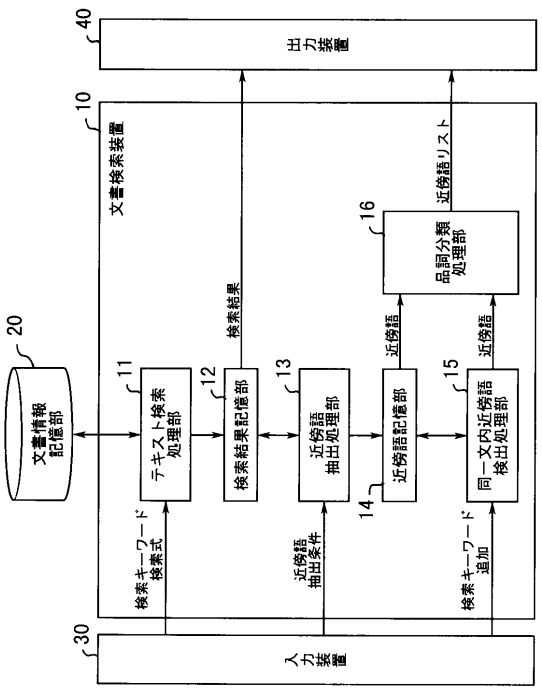
表示画面５０ｂには、近傍語リスト５１ｂが表示される。
【００５２】
検索キーワードとして「野生生物」、「希少野生動物」が追加指定されると、「規制」という単語を含み、「譲渡」もしくは「取引」もしくは「輸出入」という単語を含み、なおかつ「野生生物」もしくは「希少野生動物」という単語を含む文書から、「規制」という単語の前方に出現し、単独で意味を持つ単語２つが近傍語として検出され、近傍語リスト５１ｂに表示される。近傍語リスト５１ｂに提示されている近傍語の数は、近傍語リスト５１ａに提示されている近傍語の数と比較するとさらに少なくなっている。
【００５３】
従って、この時点で、文書情報記憶部２０には検索式「規制＊（譲渡＋取引＋輸出入）＊（野生生物＋希少野生動物）」（＊は論理積を、＋は論理和を示す）を満たす文書が記憶されていることが判る。ユーザは、このように近傍語の数が十分に減ったと判断した時点で、表示されている近傍語抽出条件から検索式を作成すればよい。
【００５４】
この検索式は、近傍語抽出条件表示欄５２、５２ａ、５２ｂ・・・に表示されている文字列の［ｍ］及び［ｎ］を削除し、「：」を「＊」に、「、」を「＋」に変換すれば作成できるので、ＣＰＵ等に処理させてもよいし、ユーザが入力してもよい。
【００５５】
このように、本発明の検索式作成装置を適用した文書検索装置１０では、検索キーワードに対し近傍語が抽出、表示され、ユーザが関心のある近傍語を次々に絞り込んで指定することができるので、繁雑な操作を必要とせずに適切な検索式を作成することが可能である。
【００５６】
また、上記で説明したように、近傍語を段階的に選択していくことができるので、入力する検索キーワードは、広い概念を持つものから順次、狭い概念を持つものに絞り込んでいくことができるので、ユーザ自身が検索意図を掴めていないような場合にも、必要な文書の検索が可能となる。
【００５７】
なお、図４～図６で示した例では、図５から図６へ移った時点で、「輸出入」という近傍語が消えている。このことから、「輸出入」という単語が、「野生生物」や「希少野生動物」という単語と、近傍語として両立している文書が存在しないことが判る。よって、この場合の検索式は「規制＊（譲渡＋取引）＊（野生生物＋希少野生動物）」でもよい。
【００５８】
【発明の効果】
以上説明したように本発明の検索式作成装置は、検索要求を絞り込むための情報、すなわち検索キーワード及び近傍語抽出条件に基づいて抽出される近傍語の提示を効率良く行い、この近傍語を検索式を作成するための検索キーワードとして追加できる構成としたので、検索式作成にかかるユーザの負担を軽減することが可能となる。
【図面の簡単な説明】
【図１】本発明の検索式作成装置の原理構成を示したブロック図である。
【図２】本発明の検索式作成装置を適用した文書検索装置のハードウェア構成図である。
【図３】本発明の検索式作成装置を適用した文書検索装置にて文書を検索する手順を示したフローチャートである。
【図４】近傍語リストの表示画面例を示す図である。
【図５】追加する検索キーワードを指定し、同一文内の近傍語を検出した様子を示す図（その１）である。
【図６】追加する検索キーワードを指定し、同一文内の近傍語を検出した様子を示す図（その２）である。
【符号の説明】
１　文書データ保持手段

2　検索キーワード保持手段
3　近傍語抽出条件保持手段
4　近傍語抽出手段
5　近傍語リスト生成手段
6　近傍語リスト保持手段
7　キーワード追加手段
8　検索式作成手段

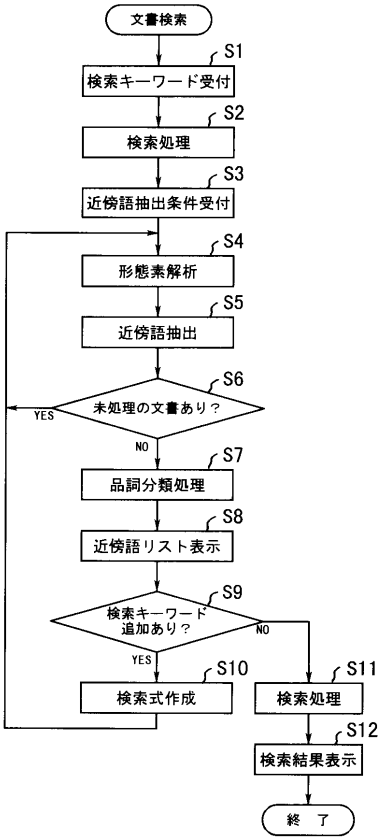
【図１】



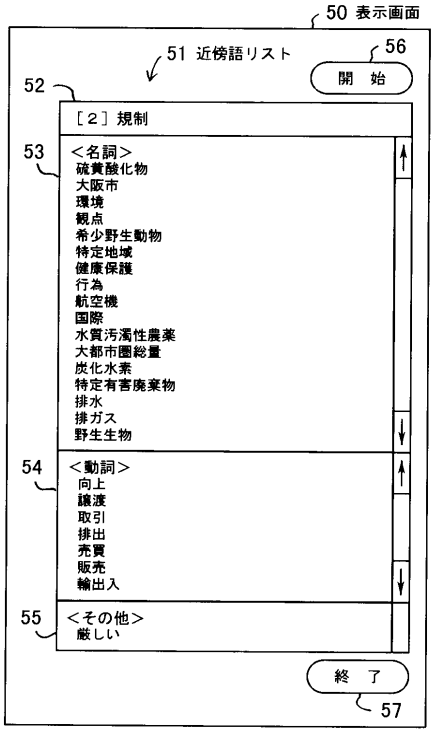
【図２】



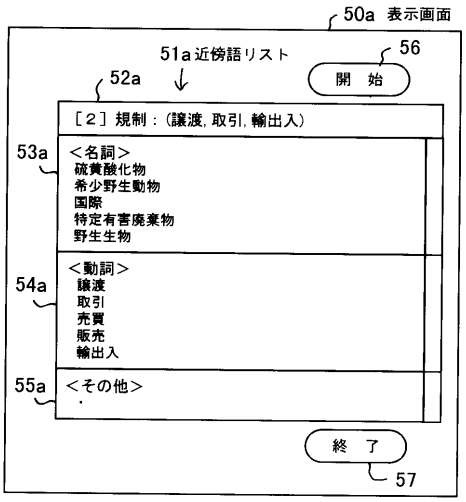
【図３】


文書検索
S1
検索キーワード受付
S2
検索処理
S3
近傍語抽出条件受付
S4
形態素解析
S5
近傍語抽出
S6
未処理の文書あり？
YES
NO
S7
品詞分類処理
S8
近傍語リスト表示
S9
検索キーワード追加あり？
NO
YES
S10
検索式作成
S11
検索処理
S12
検索結果表示
終了

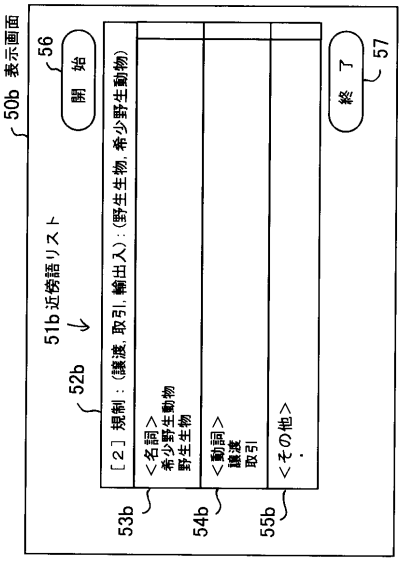
【図４】


50 表示画面
51 近傍語リスト
56
開始
52
［２］規制
53
＜名詞＞
硫黄酸化物
大阪市
環境
観点
希少野生動物
特定地域
健康保護
行為
航空機
国際
水質汚濁性農薬
大都市圏総量
炭化水素
特定有害廃棄物
排水
排ガス
野生生物
54
＜動詞＞
向上
譲渡
取引
排出
売買
販売
輸出入
55
＜その他＞
厳しい
終了
57

【図５】


50a 表示画面
51a近傍語リスト
56
開始
52a
［２］規制：(譲渡, 取引, 輸出入)
53a
＜名詞＞
硫黄酸化物
希少野生動物
国際
特定有害廃棄物
野生生物
54a
＜動詞＞
譲渡
取引
売買
販売
輸出入
55a
＜その他＞
・
終了
57

【図６】


50b 表示画面
56
開始
51b近傍語リスト
52b
［２］規制：(譲渡, 取引, 輸出入)：(野生生物, 希少野生動物)
＜名詞＞
希少野生動物
野生生物
＜動詞＞
譲渡
取引
＜その他＞
・
終了
57
53b
54b
55b

フロントページの続き

(56)参考文献 特開平０５－１８９４８７（ＪＰ，Ａ）
特開平０９－１０１９６９（ＪＰ，Ａ）
特開平０８－２１２２２８（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.[7]，ＤＢ名)
G06F 17/30 -G06F 17/30 419