(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2002－132807

（Ｐ２００２－１３２８０７Ａ）

(43)公開日　平成14年５月10日(2002.5.10)

(51)Int.Cl.[7]	識別記号	ＦＩ			ﾃｰﾏｺｰﾄﾞ (参考)
G06F 17/30	320	G06F 17/30	320	C	5B075
	140		140		5B082
	230		230	Z	
	419		419	A	
12/00	520	12/00	520	P	

審査請求　有　請求項の数４　ＯＬ　（全８頁）

(21)出願番号　特願2000－326409(Ｐ2000－326409)

(22)出願日　平成12年10月26日(2000.10.26)

(71)出願人　301022471
独立行政法人通信総合研究所
東京都小金井市貫井北町４－２－１

(72)発明者　村田　真樹
兵庫県神戸市西区岩岡町岩岡588－２　郵政省通信総合研究所　関西先端研究センター内

(72)発明者　内山　将夫
兵庫県神戸市西区岩岡町岩岡588－２　郵政省通信総合研究所　関西先端研究センター内

(74)代理人　100087848
弁理士　小笠原　吉義

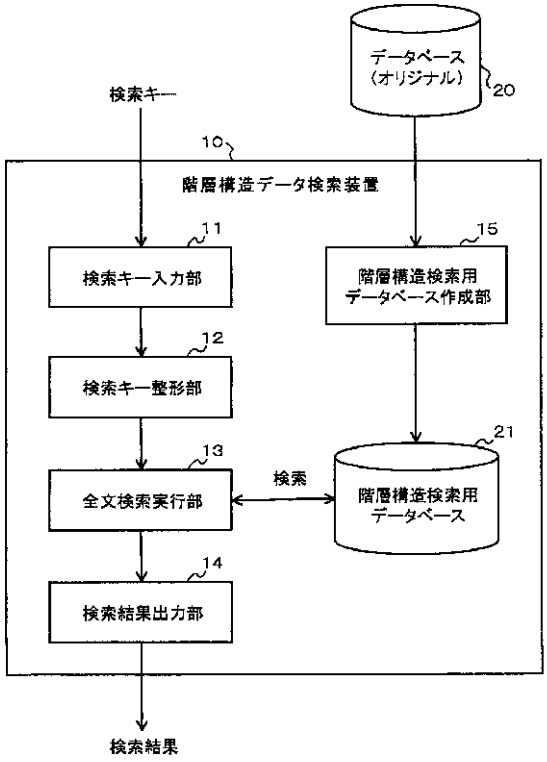
最終頁に続く

(54)【発明の名称】階層構造データ検索システム，階層構造データ検索処理方法およびそのプログラム記録媒体

(57)【要約】

【課題】　階層構造データを持つデータベースの検索において，データの階層構造の特徴を利用することによって，抽象化されたデータを効率よく検索する。

【解決手段】　検索対象のデータベースを，抽象度の低い下位階層のデータが抽象度の高い上位階層のデータによって挟まれる形でデータを保持する構造にする。検索キー整形部１２は，入力検索キーが複数の連続したデータを指定するキーであって，それぞれ下位階層から上位階層までのいずれかの階層のデータを指定するものである場合に，複数の入力検索キーの間に含まれる階層のデータを補うことにより検索キーを整形する。全文検索実行部１３は，その検索キーを用いてデータベースを全文検索する。



1

【特許請求の範囲】

【請求項１】　入力した検索キーにより階層構造データを持つデータベースを検索するシステムにおいて，前記データベースは，抽象度の低い下位階層のデータが抽象度の高い上位階層のデータによって挟まれる形でデータを保持するように構成され，入力検索キーが複数のデータを指定するキーであって，それぞれ前記下位階層から上位階層までのいずれかの階層を含むデータを指定するものである場合に，前記入力検索キーにおける複数のデータの間に含まれる階層のデータを補うことにより検索キーを整形する検索キー整形手段と，整形した検索キーを用いて前記データベースを全文検索する全文検索実行手段とを備えることを特徴とする階層構造データ検索システム。

【請求項２】　階層構造データを持つオリジナルのデータベースを入力し，抽象度の低い下位階層のデータが抽象度の高い上位階層のデータによって挟まれる形でデータを保持するようにデータベースを作り替える階層構造検索用データベース作成手段を備えることを特徴とする階層構造データ検索システム。

【請求項３】　抽象度の低い下位階層のデータが抽象度の高い上位階層のデータによって挟まれる形でデータを保持するように構成された階層構造データを持つデータベースを検索する検索処理方法であって，入力検索キーが複数のデータを指定するキーであって，それぞれ前記下位階層から上位階層までのいずれかの階層を含むデータを指定するものである場合に，前記入力検索キーにおける複数のデータの間に含まれる階層のデータを補うことにより検索キーを整形する過程と，整形した検索キーを用いて前記データベースを全文検索する過程とを有することを特徴とする階層構造データ検索処理方法。

【請求項４】　抽象度の低い下位階層のデータが抽象度の高い上位階層のデータによって挟まれる形でデータを保持するように構成された階層構造データを持つデータベースを，コンピュータによって検索するためのプログラムを記録した記録媒体であって，入力検索キーが複数のデータを指定するキーであって，それぞれ前記下位階層から上位階層までのいずれかの階層を含むデータを指定するものである場合に，前記入力検索キーにおける複数のデータの間に含まれる階層のデータを補うことにより検索キーを整形する処理と，整形した検索キーを用いて前記データベースを全文検索する処理とを，コンピュータに実行させるためのプログラムを記録したことを特徴とする階層構造データ検索用プログラム記録媒体。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は情報検索に関する。詳しくは，言語コーパスなどの階層構造をなすデータベースを高速に検索することを可能にした階層構造データ検索システムに関する。

2

【０００２】

【従来の技術】データベースの検索で最も多く用いられている全文検索では，データベース中の全文字情報中に，検索キーとして指定された文字列が現れるかどうかを調べ，現れた場合にはその位置またはその文字列が含まれる部分の文字情報を出力する。このような全文検索処理を高速に実行するための検索エンジンについては，従来から多くの研究，開発が行われてきている。しかし，データが階層構造をなすデータベースの検索を，階層構造の特徴を利用して高速に実行するシステムの研究はあまり行われていない。

【０００３】例えば言語コーパスのような階層構造をなすデータの検索において，抽象化データである「組織名」に何らかの「助詞」が付加された文字情報を検索したいというような場合，従来の単なる全文検索処理では，二つの検索キー「組織名」，「助詞」についてＡＮＤ検索を行わなければならなかった。例えば検索キー「組織名」を用いて検索した結果について，さらに「組織名」の後に「助詞」が続くかどうかを調べ，「組織名」＋「助詞」に適合するかどうかをチェックするという２段階の検索を行う。

【０００４】

【発明が解決しようとする課題】図７に示す具体例に従って，従来技術の問題点を説明する。以下では，階層構造データを持つデータベースとして言語コーパスの例を用いて説明する。言語コーパスとは，自然言語分析用の電子化された言語資料のデータベースである。

【０００５】言語コーパスには，新聞記事や論文などの多数の文章が格納されており，それらの文章が形態素に分解され，例えば図７（Ａ）に示すように，各形態素の文データに対して，読み，品詞細分類，品詞のような階層構造をなす付属情報が付けられて蓄積されている。図７（Ｂ）は，言語コーパスのデータ例であって，「郵政省の技術。」という文データを形態素に分解して格納した例を示している。ここで，「／」，「：」は区切り記号を表しており，もちろんこのデータの区切りは，他の文字・記号その他で置き換えることができる。

【０００６】図７（Ｂ）に示すデータベースについて，図７（Ｃ）に示すような二つの検索キー「組織名」＋「助詞」を与えて全文検索するとする。この場合，図７（Ｄ）に示すように，まず「組織名」を検索キーとしてデータベースを検索し，その検索結果についてさらに「助詞」の語が後続するかどうかをチェックし，後続しない場合にはさらに「組織名」を検索キーとする検索を続け，その検索結果に「助詞」の語が後続するものを「組織名」＋「助詞」の検索結果として出力する。

【０００７】このように，従来の全文検索では，「組織名」＋「助詞」のような検索キーの場合には，「組織名」と「助詞」の２段階の検索を行っており，検索処理に時間がかかるといった問題があった。

【０００８】本発明は上記問題点の解決を図り，データの階層構造の特徴を利用することによって，抽象化されたデータを効率よく検索する技術を提供することを目的とする。特に，入力検索キーが複数の連続した抽象化データを指定するキーであるような場合に，全文検索を１回実行するだけで，求める検索結果を得ることができるようにし，検索処理の高速化を図ることを目的とする。

【０００９】

【課題を解決するための手段】本発明は，上記課題を解決するため，階層構造データを持つオリジナルのデータベースを，抽象度の低い下位階層のデータが抽象度の高い上位階層のデータによって挟まれる形でデータを保持するように作り替える。すなわち，階層構造検索用データベースのデータ構造を，抽象度が最も低いものを中心に記述し，その周りに抽象度の低い順に記述していく構造にする。

【００１０】そして，入力検索キーが複数の連続したデータを指定するキーであって，それぞれ下位階層から上位階層までのいずれかの階層を含むデータを指定するものである場合に，複数の入力検索キーの間に含まれる階層のデータを補うことにより検索キーを整形し，一つの検索キーにまとめる。その整形した一つの検索キーを用いて階層構造検索用データベースを全文検索すれば，１回の全文検索で求める検索結果が得られることになる。

【００１１】以上の各処理手段をコンピュータによって実現するためのプログラムは，コンピュータが読み取り可能な可搬媒体メモリ，半導体メモリ，ハードディスクなどの適当な記録媒体に格納することができる。

【００１２】

【発明の実施の形態】以下，本発明の実施の形態を，図面を用いて説明する。

【００１３】図１に，本発明の構成例を示す。階層構造データ検索装置１０は，ＣＰＵやメモリなどからなるコンピュータであって，検索キーを入力する検索キー入力部１１と，入力した検索キーが複数の連続したデータを指定するキーであって，それぞれ下位階層から上位階層までのいずれかの階層を含むデータを指定するものである場合に，その複数の入力検索キーの間に含まれる階層のデータを補うことにより検索キーを整形する検索キー整形部１２と，検索キー整形部１２で整形した検索キーを用いて階層構造検索用データベース２１を全文検索する全文検索実行部１３と，全文検索実行部１３の検索結果を出力する検索結果出力部１４とを備える。

【００１４】また，階層構造データ検索装置１０は，階層構造データを持つオリジナルのデータベース２０を入力し，これから階層構造検索用データベース２１を作成する階層構造検索用データベース作成部１５を備える。階層構造検索用データベース２１は，抽象度の低い下位階層のデータが抽象度の高い上位階層のデータによって挟まれる形でデータを保持するように構成される。なお，階層構造検索用データベース作成部１５は，検索処理を実行する階層構造データ検索装置１０とは別の装置に設けてもよい。

【００１５】図１に示す階層構造データ検索装置１０による検索処理の概要を，図２に示す処理フローチャートを用いて説明する。あらかじめ階層構造検索用データベース作成部１５により，検索対象となるデータが格納されたデータベース２０のデータ構造を変形し，階層構造検索用データベース２１を作成しておく。

【００１６】検索を行う利用者またはアプリケーションプログラムからの検索要求を待ち，検索要求があったならば（ステップＳ１），ステップＳ３へ進む。また，検索終了の指示があったかどうかを判断し（ステップＳ２），検索終了の指示があったならば，処理を終了する。

【００１７】ステップＳ３では，検索キーを入力として受け取る。次に，入力した検索キーが整形可能であるかどうかを判断する（ステップＳ４）。ここでは，入力した検索キーが一つの最上位から最下位までの階層のデータ群（カテゴリ）を単位として連続したデータ（例えば形態素連続のデータ）を指定するキーであって，それぞれが下位階層から上位階層までのいずれかの階層を含むデータを指定するものである場合に，整形可能と判断する。厳密には，複数のデータからなる検索キーの両端以外のデータが最下位階層のデータであり，両端のデータがいずれかの階層のデータである場合に，整形可能と判断する。

【００１８】検索キーが整形可能でなければ，本発明による検索を行わないで，従来技術による通常の検索を行い（ステップＳ５），ステップＳ８へ進む。検索キーが整形可能であれば，検索キー整形部１２により入力した検索キーを整形する（ステップＳ６）。その整形された検索キーを用いて全文検索を行い（ステップＳ７），全文検索の結果を出力する（ステップＳ８）。

【００１９】なお，上記ステップＳ４における整形可能かどうかの判定では，検索キーを複数に分割して，分割されたそれぞれのキーまたは分割された一部のキーに対して本発明を適用できる場合に，その部分について整形可能として扱い，本発明による検索と従来技術による検索とを併用するような実施も可能である。

【００２０】例えば，上位，中位，下位の３階層データからなるデータベース検索において，入力検索キーが「上位階層データ」＋「中位階層データ１」＋「中位階層データ２」であるような場合に，中央のデータ（両端以外のデータ）が「下位階層データ」ではないため，検索キーの全体に対して本発明を適用することができない。しかし，この検索キーを，①「上位階層データ」＋「中位階層データ１」と，②「中位階層データ２」とに分割すれば，前者の①のキーについては，検索キーの整形が可能であり，①のキー部分について，本発明を適用

5

することができる。したがって，このような場合には，①のキー部分についてステップＳ６，Ｓ７による検索を実行し，その検索結果の中で②のキーに合致するものを最終的な検索結果とする。

【００２１】〔第１の実施の形態〕まず，階層構造検索用データベース２１の作成処理について説明する。図３は，階層構造検索用データベース２１の作成例を示す図である。

【００２２】検索対象のデータベースが，図３（Ａ）に示すような，文データ，読み，品詞細分類，品詞からなる階層構造データを持つ言語コーパスであるとする。図３（Ａ）のデータベースのデータ構造において，上から３行目の品詞細分類および４行目の品詞は，上から１行目の文データをいくぶん抽象化したものである。すなわち，「組織名」は「郵政省」より抽象度の高い上位階層データ，「名詞」は「組織名」より抽象度の高い上位階層データである。

【００２３】図３（Ｂ）は，図３（Ａ）に示す階層構造データを持つデータベース（オリジナル）２０の記述例を示す図である。オリジナルのデータベース２０では，文データごとに，「文データ，読み，品詞細分類，品詞」のように，抽象度の低い順，すなわち下位階層から上位階層の順にデータが記述されている。なお，それぞれのデータの階層レベルが明らかであれば，データベース中のデータの並びの順は，必ずしも階層レベルの順でなくてもよい。

【００２４】階層構造検索用データベース作成部１５では，図３（Ｂ）のオリジナルのデータベース２０の内容を取り込み，まず，一番階層の低い（一番抽象化されていない）文データを中心に記述し，その文データの前後に抽象度の低いデータから順番に既に記述したデータを挟むような形で記述したデータ構造の階層構造検索用データベース２１を作成する。このとき，あらかじめ検索キーにならないとわかっている階層データがあれば，必要に応じてそれを削除してもよい。この例では「読み」についての検索は不要であるとして，階層構造検索用データベース２１では削除している。

【００２５】例えば，図３（Ｂ）に示すデータベース２０における「郵政省：ゆうせいしょう：組織名：名詞」のデータについては，一番抽象度の低い「文データ」の「郵政省」を中心に記述し，「読み」の「ゆうせいしょう」については削除し，次に抽象度の低い「品詞細分類」である「組織名」で「郵政省」の前後を挟むように記述し，さらに抽象度の高い「品詞」である「名詞」で「組織名：郵政省：組織名」の前後を挟むように記述して，「名詞：組織名：郵政省：組織名：名詞」と変形する。

【００２６】この処理を文データごとに繰り返し実行し，その結果を連結することにより，階層構造検索用データベース２１を作成する。図３（Ｃ）に，図３（Ｂ）

6

のデータベース２０から作成した階層構造検索用データベース２１の記述例を示す。

【００２７】次に，入力した検索キーの整形処理について説明する。検索したいデータは，図３（Ａ）に示すような階層構造をなすとし，図３（Ｃ）に示す階層構造検索用データベース２１を使用するものとする。

【００２８】図４に，検索キーの整形の例を示す。ユーザ（アプリケーションプログラムを含む）が「組織名：名詞」＋「の：接続助詞：助詞」＋「技術：普通名詞：名詞」を検索キーとして入力したとする。検索キー整形部１２は，図４（Ａ）に示す３個のデータを含む検索キーを整形して，図４（Ｂ）に示す「組織名：名詞／助詞：接続助詞：の：接続助詞：助詞／名詞：普通名詞：技術：普通名詞：名詞」という一つの検索キーを作成する。

【００２９】整形方法は，以下のとおりである。ユーザが指定した検索キーのデータが抽象化していないもの（例：郵政省，技術），すなわち最下位階層のデータである場合には，階層構造検索用データベース２１の作成と同様に，一番抽象度の低いデータを中心に記述し，それの前後に抽象度の低いものから順にデータを補って記述していく。

【００３０】ユーザが指定した検索キーのデータが抽象化されたもので，複数の検索キーの並びの一番左にある場合には，それを記述し，それより抽象化されたものを順にその右側に記述する。すなわち，検索キーの先頭（一番左側）のものについては，その記述の後のみ上位階層データの記述を順に補充していく。なお，検索キーの並びの一番左にある抽象化されたデータの中に，その上位階層データの記述も含まれる場合には，上位階層データがそれより下位の階層データよりも右側にくるように必要に応じてデータの並べ替えを行う。

【００３１】一方，ユーザが指定した検索キーのデータが抽象化されたもので，複数の検索キーの並びの一番右にある場合には，それを記述し，それより抽象化されたものを順にその左側に記述する。すなわち，検索キーの後尾（一番右側）のものについては，その記述の前方にのみ上位階層データの記述を順に補充していく。なお，検索キーの並びの一番右にある抽象化されたデータの中に，その上位階層データの記述も含まれる場合には，上位階層データがそれより下位の階層データよりも左側にくるように必要に応じてデータの並べ替えを行う。

【００３２】例えば図４（Ａ）に示す検索キーの整形では，検索キーの先頭（左側）の「組織名：名詞」については，すでに「組織名」の最上位階層の「名詞」が記述されているので，補充しない。また，検索キーの後尾（右側）の「技術：普通名詞：名詞」については，「技術」の前に１つ上位の階層の「普通名詞」を補充し，さらにその上位階層の「名詞」も補充する。検索キーの内部の「の：接続助詞：助詞」については，「の」の前に

１つ上位階層の「接続助詞」を補充し，さらにその前に１つ上位階層の「助詞」を補充する。これらを区切り記号「／」で連結する。

【００３３】なお，検索キー中で抽象化されるデータ，すなわち上位階層データは，検索キーの先頭（一番左側）または後尾（一番右側）のものに限られる。検索キー「組織名：名詞」＋「助詞」＋「名詞」のように，検索キーの内部に位置するものとして抽象化した上位階層データを用いることはできない。このような検索を行いたい場合には，ＡＮＤ検索等を用いて「組織名：名詞」＋「助詞」を検索した後に，それが「組織名：名詞」＋「助詞」＋「名詞」を満たしているかどうかを調べて検索する。この場合でも，３段階の検索を行う従来方式に比べて高速に検索することが可能である。

【００３４】また，図４（Ｃ）に示すように，ユーザが「組織名：名詞」＋「助詞」で検索したいとする。この場合には，検索キー整形部１２では，入力した検索キーを整形し，図４（Ｄ）に示すように「組織名：名詞／助詞」とする。

【００３５】入力した検索キーの「組織名：名詞」の部分については，検索キーの先頭に位置し，かつ，「名詞」が最上位階層であるので記述を補充しない。また，「助詞」の部分については，検索キーの後尾に位置し，かつ「助詞」が最上位階層であるので，同様に記述を補充しない。

【００３６】〔第２の実施の形態〕ユーザが，検索キーを「組織名：名詞」＋「助詞」のように指定せずに，「組織名」＋「助詞」のように指定して検索したい場合もあると考えられる。

【００３７】そのため，第２の実施の形態においては，図５（Ａ）に示すような階層構造のデータを持つ図５（Ｂ）のオリジナルのデータベース２０から，それぞれのデータに階層レベルを付加した階層構造検索用データベース２１を作成しておく。

【００３８】すなわち，階層構造検索用データベース作成部１５では，第１の実施の形態の場合と同様の処理を行い，オリジナルのデータベース２０を入力して，下位階層のデータを中心に記述し，その記述の前後に階層の低いものから順に，階層レベルと上位階層データとを補充していく。これによって，図５（Ｃ）に示すような階層構造検索用データベース２１を作成し，これを本実施の形態における検索に用いる。図５（Ｃ）の例では，階層レベルを下位階層から順番に「１：」「２：」「３：」で表している。

【００３９】検索キー入力部１１では，ユーザから検索キーを入力するときに，階層レベルを検索キーに含めて指定してもらう。これは例えば「組織名」という検索キーが指定された場合に，この語が抽象化されたデータであることをユーザが意図しているか，通常の単語としての抽象化されていないデータであることをユーザが意図しているかを判別できないからである。階層レベルをユーザに指定させることにより，ユーザが「２：組織名」と入力した場合には前者の抽象化されたデータ，「１：組織名」と入力した場合には後者の抽象化されていないデータであると判別することができる。

【００４０】図６（Ａ）に示すように，入力した検索キーが「２：組織名」＋「３：助詞」であるとする。検索キー整形部１２では，データベースから「２：組織名」より抽象度の高いものを調べて，これより抽象度の高いものとして「３：名詞」を得る。また，「３：助詞」より抽象度の高いものを調べて，これより抽象度の高いものがないことを知る。その結果，検索内容が，「２：組織名：３：名詞」＋「３：助詞」であることがわかる。

【００４１】検索キー整形部１２は，この結果から第１の実施の形態で説明した整形方法と同じ整形方法を用いることにより，図６（Ｂ）に示すように「２：組織名」の後に「３：名詞」を付加し，これと「３：助詞」とを結合して，「２：組織名：３：名詞／３：助詞」という検索キーを作成する。これを用いて図５（Ｃ）に示す階層構造検索用データベース２１を全文検索すれば，求める検索結果が得られることになる。

【００４２】ところで，「２：組織名」より抽象度の高いものとして，「３：名詞」だけでなく「３：記号」などが設定されているなど，上位階層が複数存在する場合も考えられる。このような場合には，

「２：組織名：３：名詞／３：助詞」

「２：組織名：３：記号／３：助詞」

の２つの検索キーを用いてＯＲ検索を行って，検索結果を出力する。

【００４３】また，例えばユーザが「１：技術」とだけ指定して検索したい場合もあると考えられる。この場合には，「１：技術」より抽象度の高い上位階層のデータの記述を補充することなく，「１：技術：」とだけ整形して，検索を行うようにしてもよい。

【００４４】以上のようにデータの階層構造を利用して整形した検索キーを用いて検索することにより，例えば従来の方法では，「組織名：名詞」＋「助詞」を検索する場合に「組織名：名詞」と「助詞」の２段階の検索が必要であったのに対し，本発明を用いた方法によれば，「組織名：名詞」と「助詞」とを連結した一つの検索キーで検索することが可能になり，検索の処理時間を大幅に短縮することができる。

【００４５】特に，言語分析のための言語コーパスのようなデータベースの検索では，種々の検索パターンによる検索が必要で検索回数が非常に多いだけでなく，検索結果の個数も数千に及ぶことが多いため，本発明を用いることによる処理時間の短縮の効果は大きい。もちろん，本発明は言語コーパスに限らず，階層構造データを持つデータベースの検索に同様に適用することができる。

【００４６】

【発明の効果】以上説明したように，本発明は，階層構造データを持つデータベースの抽象化されたデータを含むデータの検索を，効率よく高速に実行することができるという効果がある。

【図面の簡単な説明】

【図１】本発明の構成例を示す図である。

【図２】本発明による検索処理の概要を示す処理フローチャートである。

【図３】第１の実施の形態における階層構造検索用データベースの作成例を示す図である。

【図４】第１の実施の形態における検索キーの整形の例を示す図である。

【図５】第２の実施の形態における階層構造検索用データベースの作成例を示す図である。

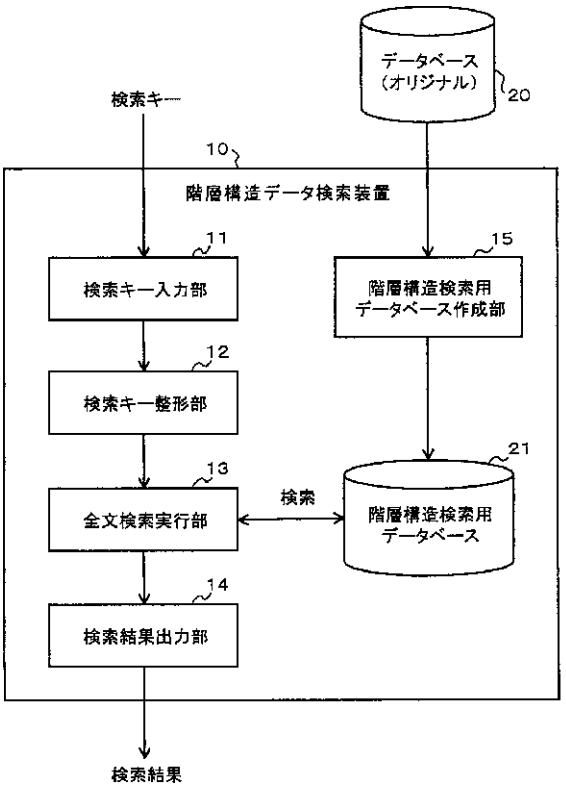
【図６】第２の実施の形態における検索キーの整形の例を示す図である。

【図７】従来技術の問題点を説明するための図である。

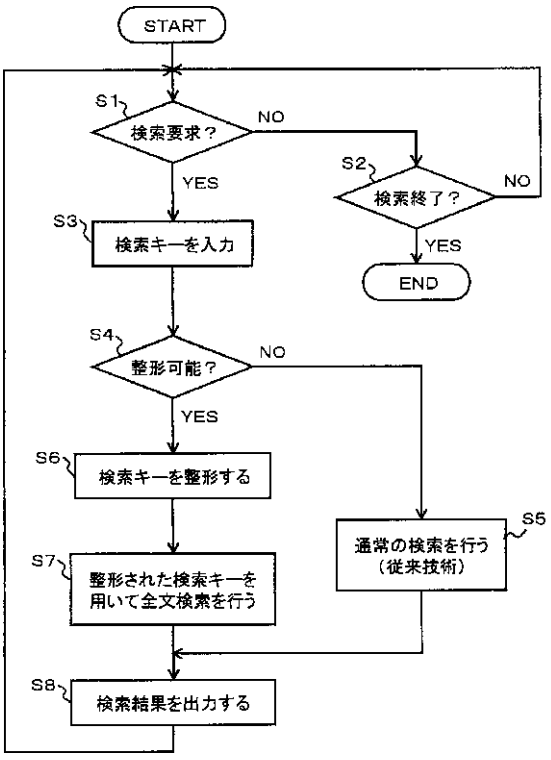
【符号の簡単な説明】

１０　階層構造データ検索装置
１１　検索キー入力部
１２　検索キー整形部
１３　全文検索実行部
１４　検索結果出力部
１５　階層構造検索用データベース作成部
２０　データベース（オリジナル）
２１　階層構造検索用データベース

【図１】

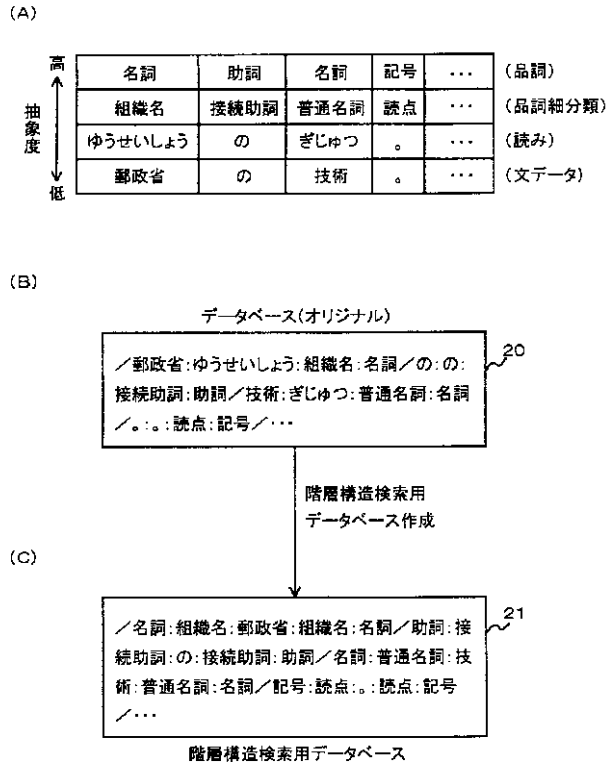


【図２】



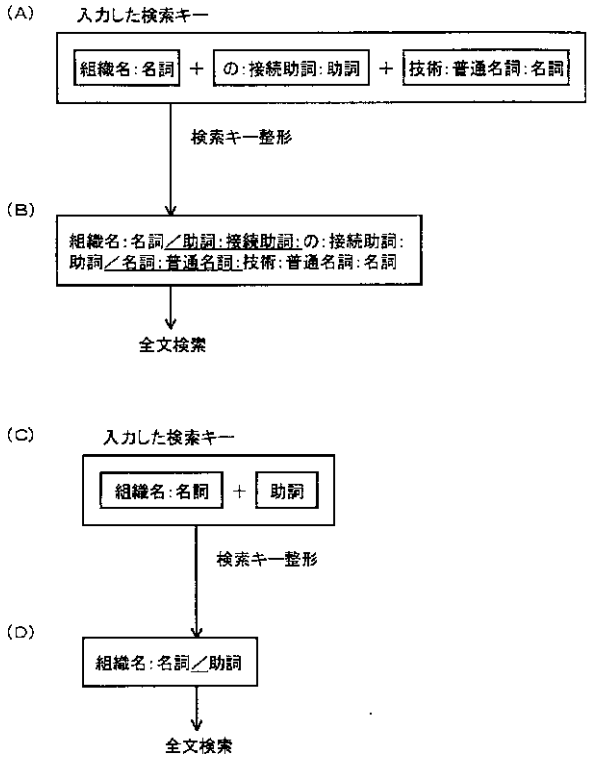
【図３】

階層構造検索用データベースの作成例



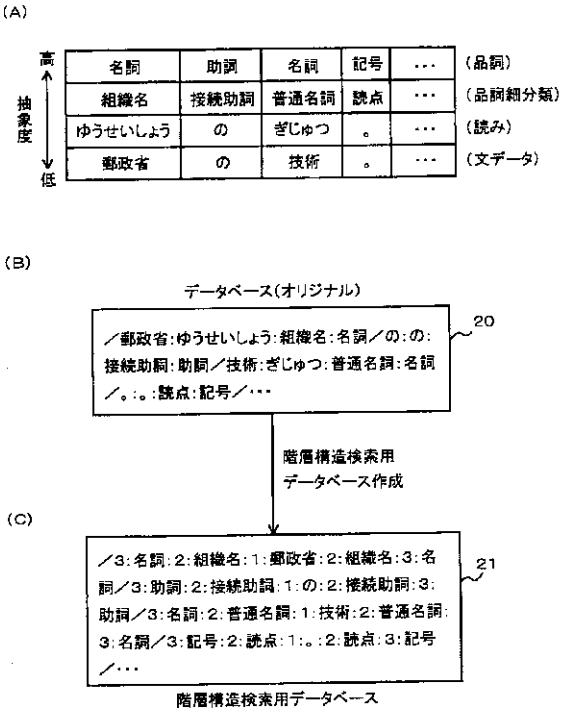
【図４】

検索キーの整形の例



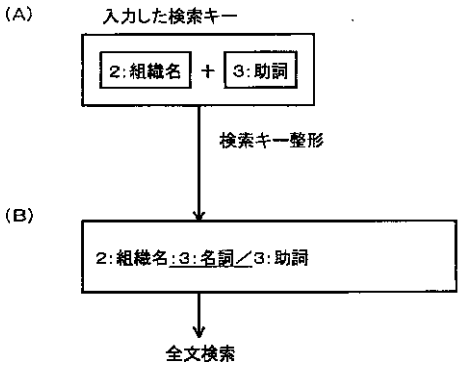
【図５】

階層構造検索用データベースの作成例

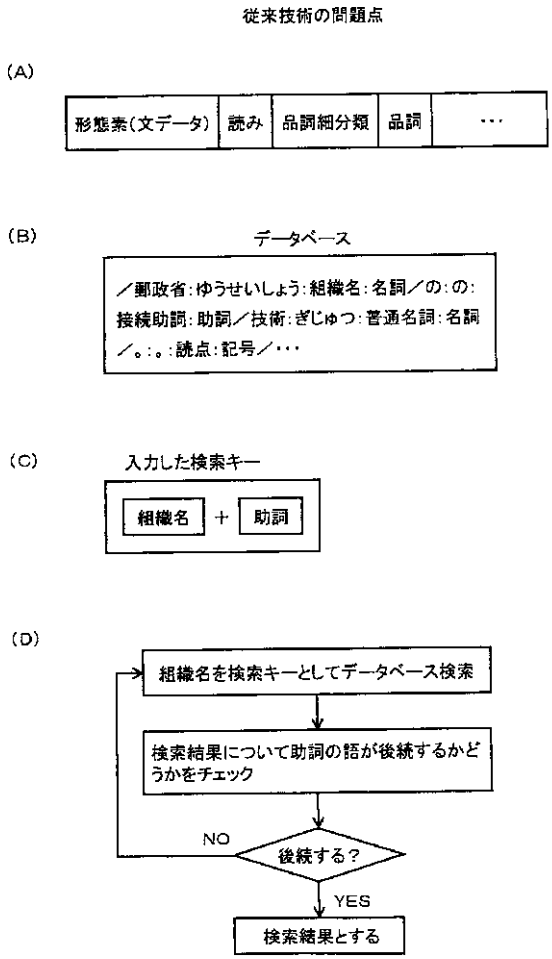


【図６】

検索キーの整形の例



【図７】



フロントページの続き

(72)発明者　井佐原　均
兵庫県神戸市西区岩岡町岩岡588－２　郵政省通信研合研究所　関西先端研究センター内

Ｆターム(参考)　5B075 ND35 PP22
5B082 EA08 GC04