

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第3552842号
(P3552842)

(45) 発行日 平成16年8月11日(2004.8.11)

(24) 登録日 平成16年5月14日(2004.5.14)

(51) Int. Cl.⁷

F I

G06F 17/22

G06F 17/22 508A

G06F 17/30

G06F 17/30 170J

請求項の数 4 (全 9 頁)

(21) 出願番号	特願平8-71479	(73) 特許権者	000006747 株式会社リコー
(22) 出願日	平成8年2月29日(1996.2.29)		東京都大田区中馬込1丁目3番6号
(65) 公開番号	特開平9-237272	(74) 代理人	100079843 弁理士 高野 明近
(43) 公開日	平成9年9月9日(1997.9.9)	(72) 発明者	山形 寛子 東京都大田区中馬込1丁目3番6号 株式 会社リコー内
審査請求日	平成13年7月11日(2001.7.11)	審査官	成瀬 博之
		(56) 参考文献	特開平08-044728 (JP, A) 特開平08-016571 (JP, A) 特開平04-160558 (JP, A) 特開平05-342402 (JP, A) 最終頁に続く

(54) 【発明の名称】 単漢字検索装置

(57) 【特許請求の範囲】

【請求項1】

単漢字の構成部分を用いて表現した検索式を入力する入力部と、単漢字に対して、該単漢字の漢字表記、表記コード、該単漢字の構成部分を表す漢字構成情報とを関連付けて格納する辞書部と、前記検索式を前記辞書部の漢字構成情報と照合して、一致した単漢字に対する漢字表記または表記コードを検索結果とする検索部と、前記検索結果を出力する出力部を有する単漢字検索装置において、前記辞書部および前記検索式における漢字構成情報は、単漢字の各構成部分について、該構成部分と該構成部分の相対的位置情報の組み合わせ、或いは、該構成部分をさらに分解した分解構成部分の漢字構成情報で置き換えて、該構成部分の漢字構成情報を入れ子状に表現し、前記検索部は、検索式と前記辞書部の漢字構成情報とを照合し、一致した場合には、一致した単漢字に対する漢字表記または表記コードを検索結果とし、一致しなかった場合には、検索式中に入れ子状になっている各構成部分の漢字構成情報について前記辞書部を検索し、得られた結果を該入れ子の構成部分の漢字構成情報と置き換えた置換済み検索式で再検索を行ない、前記置換済み検索式と前記辞書部の漢字構成情報とが一致するまで検索を繰り返すことを特徴とする単漢字検索装置。

10

【請求項2】

請求項1において、ある漢字Yと共通する構成部分が存在する漢字Xがある場合、該共通する構成部分を漢字の構成部分として指定できるようにしたことを特徴とする単漢字検索装置。

20

【請求項 3】

請求項 2 において、前記共通する構成部分の指定は、該共通部分が存在する前記漢字 X の上下または左右または中外のいずれの位置にあるかという相対的位置情報と該漢字 X との組み合わせで指定することを特徴とする単漢字検索装置。

【請求項 4】

請求項 1 ないし 3 のいずれかにおいて、前記辞書部に読み情報を有し、前記検索部は、検索結果として得たい漢字の読みも出力することを特徴とする単漢字検索装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、単漢字検索装置に係り、例えば、ワープロ、フロントエンド、文書作成支援装置、漢字辞書検索等に適用して好適なものである。

【0002】

【従来の技術】

WS (ワークステーション) 上などにおける日本語文書作成において不便に感じている事柄の一つに、所望の漢字が得られないということがある。

仮名漢字変換用辞書には単漢字も登録されているが、数は限られている。仮名漢字変換で所望の漢字が得られない場合、JIS (区点) コードなどで検索し漢字を得ることになるが、JIS 漢字は第一、第二を合わせると 6000 文字以上になるため、コードをひく手間がこれまた大変である。

通常、読みや画数などをキーにコード本 (書籍) をひきコードを得て仮名漢字変換の JIS (区点) コード入力モードで変換するか、部首検索モードがある場合はそれを利用して

【0003】

漢字検索に関する従来技術には下記のようなものがある。

・部首検索モード及びその類

部首コードは 214 もあるため (JIS X0208)、部首一覧表を提示されてもユーザにとってこの中から探す作業は負担である。また部首によっては該当漢字を 200 ~ 300 字くらい有するものもあり効率がよいとはいえない。また、漢字によっては複数の部首に属するものもあるし、どの部首に属するかを直観的に当てること自体も難しかったり

下記の出願公開公報に記載された発明は、部首を直接入力しないで、ある漢字 X を入力すると同じ部首を持つものが出力されるため、表をひく手間が省けるというメリットがあるが、結果として、該当するものが沢山でてしまうという問題がある。

【0004】

特開平 2 - 255934 号公報 (部首別漢字入力装置) :

部首一覧に限らず、システム側から提示された内容から所望の情報を選択するというやり方は多かれ少なかれユーザにとって負担になってしまうという問題がある。

・画数検索

画数は本出願人の手持ちデータ (JIS 漢字 1, 2) では、画数 4 ~ 21 まで、ずっと該当漢字が 3 桁で存在する。絞り込み能力が弱いためあまり有効な検索手段とはいえない。また、画数を得るために、一旦エディティング作業を中断する必要がある。

【0005】

画数検索で、入力画数の近傍画数を自動検索するという内容で公開公報に記載された発明が、下記になされているが、本出願人の手持のデータでは、画数 11 を指定してその近傍 + - 1 即ち、画数 10 ~ 12 とすると 1500 字以上も該当してしまい絞り込み能力にかけられると思われる。

【0006】

特開昭 61 - 180362 号公報 (漢字処理装置) :

・画数及び部首検索

10

20

30

40

50

本出願人の手持ちデータ（JIS漢字1, 2）では、画数と部首を組あわせれば、かなりの絞り込みが期待できるが、部首表を検索する、画数を調べるという作業が、一旦エディティング作業を中断させ負担になってしまうという問題は残る。

【0007】

・その他

その他の検索方法としては、漢字の構成部分（部首ではなく）を取り上げて、それをキーとして検索するものである。

例えば、特願平3 - 140887号に記載の発明は、ユーザの知識を使用して所定の方式（検索式）で既知の漢字で所望漢字の構成を指定することで所望漢字を得るというものであり、読み、画数は一切使用せず、また部首表の提示、キーリストの提示、パターンテーブルの提示などもせず、ユーザの知識を所定の検索式の形態で表現してもらうだけで、所望の漢字を検索する漢字検索装置を提示するという点で、本発明に最も類似している。

10

【0008】

【発明が解決しようとする課題】

特願平3 - 140887号（漢字入力装置）の漢字入力装置の実現手段は、下記の問題点を含んでいる。

A. 情報量が膨大である漢字構成辞書を有する。

特願平3 - 140887号の漢字構成辞書は、差分辞書で記された構成部分ごとに漢字集合を作ったものである。例えば、“諮”について考える。

差分辞書において下記のように記されるとすると、

20

差分辞書における“諮”の記述

諮，言，次，口

そして更に、漢字構成辞書において下記のように記されることになる。

漢字構成辞書の記述

言 記，語，諮...

次 姿，諮...

口 吹，咽，告，諮...

諮 諮...

即ち、“諮”は差分辞書において、構成情報として3エントリ有し、さらに、漢字構成辞書において、4回出現する。

30

すべての漢字について同様のことをすると本来の漢字数の何倍もの情報量になってしまう。また、それに伴って、計算量もふえ効率をさげてしまう。

【0009】

B. 相対的位置情報を用いていないためヒット精度が落ちる。

特願平3 - 140887号では、相対的位置情報を用いていないため木へんのものを探す場合でも検索式に“...木+...”と記述することになる。同検索手段では、“木”がどの位置にあるかということは考慮されないのので、漢字構成部分に“木”を含むもの全部を探してくる。例えば、構成部分に“木”を含む漢字集合例：案，杏，椅，楽，禁，襟，困，梱，床，湘，槻，嘛，噤，婪，嫻，孀，悠...となり、これは数百を下らない件数になり部首入力などで、木へんのものを得た方が効率がいいことになってしまう。

40

【0010】

C. 遑って検索するという手段がないため、特願平3 - 140887号の差分生成辞書においては、情報量が膨大、繁雑になってしまう。

特願平3 - 140887号の記述方式では、例えば、吾、梧、聒を例にとれば、

吾：五，口

梧：木，五，口，吾

聒：耳，千，口，舌

となり、想定されるものを全部記述しなくてはならない。そのため、記述量も多く、位置情報や漢字構成部ごとのまとまりに関する情報もないため繁雑になってしまうおそれがある。

50

【 0 0 1 1 】

D. また、ある漢字 X の一部分を指定することができないため実際には検索できないものがでてしまう。所望漢字が、仮に X Z という構成をしていて、ユーザが X W という漢字を知っていた場合に、平 3 - 1 4 0 8 8 7 号のアルゴリズムでは X W - W + Z という指定の仕方をとる。しかし、漢字の構成要素がすべて登録されていることはまずない。J I S 漢字には偏や旁に相当する字が含まれているがそれもほんの一部である。従って、差を示すべく入力する W や Z に相当するものが入力できない場合があるという問題点がある。

例えば、J I S 漢字には部首である“辶 / 邇 / 透”の下部共通部分(しんにゅうの一種)や“病 / 癌 / 痕”の上部共通部分(やまいだれ)すらも対応するコードがないため、J I S 内漢字環境で対応しようと思うとそもそも検索式を作ることができないということになる。

10

【 0 0 1 2 】

本発明は、以上のような問題点に鑑みてなされたもので、ユーザの知識を使用して所定の方式で既知の漢字並びに構成を指定することで所望の漢字を得るというものであり、読み、画数は一切使用せず、また部首表の提示、キーリストの提示、パターンテーブルの提示などもせず、ユーザの知識を所定の検索式の形態で表現してもらっただけで、所望の漢字を検索する漢字検索装置で、且つ特願平 3 - 1 4 0 8 8 7 号よりも辞書容量、計算量共に小さく効率的で、且つ柔軟で頑強な漢字検索装置の提供を目的とする。

【 0 0 1 3 】

【課題を解決するための手段】

20

請求項 1 の発明は、単漢字の構成部分を用いて表現した検索式を入力する入力部と、単漢字に対して、該単漢字の漢字表記、表記コード、該単漢字の構成部分を表す漢字構成情報とを関連付けて格納する辞書部と、前記検索式を前記辞書部の漢字構成情報と照合して、一致した単漢字に対する漢字表記または表記コードを検索結果とする検索部と、前記検索結果を出力する出力部を有する単漢字検索装置において、前記辞書部および前記検索式における漢字構成情報は、単漢字の各構成部分について、該構成部分と該構成部分の相対的位置情報の組み合わせ、或いは、該構成部分をさらに分解した分解構成部分の漢字構成情報で置き換えて、該構成部分の漢字構成情報を入れ子状に表現し、前記検索部は、検索式と前記辞書部の漢字構成情報とを照合し、一致した場合には、一致した単漢字に対する漢字表記または表記コードを検索結果とし、一致しなかった場合には、検索式中に入れ子状
になっている各構成部分の漢字構成情報について前記辞書部を検索し、得られた結果を該入れ子の構成部分の漢字構成情報と置き換えた置換済み検索式を生成して再検索を行ない、前記置換済み検索式と前記辞書部の漢字構成情報とが一致するまで検索を繰り返すようにしたものである。

30

【 0 0 1 5 】

請求項 2 の発明は、請求項 1 において、ある漢字 Y と共通する構成部分が存在する漢字 X がある場合、該共通する構成部分を漢字の構成部分として指定できるようにしたものである。

【 0 0 1 6 】

請求項 3 の発明は、請求項 2 において、前記共通する構成部分の指定は、該共通部分が存在する前記漢字 X の上下または左右または中外のいずれの位置にあるかという相対的位置情報と該漢字 X との組み合わせで指定するようにしたものである。

40

【 0 0 1 7 】

請求項 4 の発明は、請求項 1 ないし 3 のいずれかにおいて、前記辞書部に読み情報を有し、前記検索部は、検索結果として得たい漢字の読みも出力するようにしたものである

【 0 0 1 8 】

【発明の実施の形態】

図 1 は、本発明の全体の構成図であり、図中、1 は入力部で、入力には音声認識入力、キーボード入力などが考えられる。2 は検索部、3 は辞書部で、辞書には単漢字とその構成情報が最低限搭載される。また、4 は出力部で、出力装置には、ディスプレイを想定して

50

いるが、音声合成なども可能である。

【0019】

本発明では、“ある漢字を入力したいのだが、読みがわからない、また、知っている読みでは仮名漢字変換できない。”という場合を想定している。

これは、例えば、名刺などに記載されている人名・地名をインプットしたい時などにおこることが想定される。例えば、“林 惇猷(あつみち)”という名前をインプットしたい時など読みではまず変換されない。

【0020】

今、記号“+”を使用し、“A+B”と表現したら、AとBから構成されるという意味を表し、また、記号“,”を使用し、“A,B”と表現したら構成部分として、AまたはBが存在する、という意味を表すとする。また、“上,下,左,右,中,外”といった相対的位置情報をアルファベット“T,D,L,R,I,O”でそれぞれ表し、これらの記号を使用して辞書の構成情報にもこの位置情報を付与することで実現できる。

10

【0021】

例えば、辞書に“榿”に対して、“L木+R蜜”という構成情報を付与していたとすると、ユーザが入れた検索式が「L木+R蜜」であれば、検索式に含まれる構成情報と辞書に搭載されている構成情報を照合した結果“榿”という所望の漢字を得ることができる。なお、木へんを有するもののみを得たければユーザは「L木」をいれればよい。また、“木,蜜”という式を与えたら“木”または“蜜”を構成要素として含むものを検索する。検索では、検索式に含まれる構成情報が辞書の構成情報に含まれるものを検索させる。従

20

って、複数の候補がでることもある。検索を、検索式と辞書の構成情報が同内容の場合のみ成功させることにしてもかまわない。その場合は、ワイルドカードを設定するなどの必要がある。例えば、“*”を0個以上の要素を表すワイルドカードと想定すると、木へんを有するもののみを得たければユーザは「L木+*」を指示することになる。

【0022】

他にも、「O門+I馬」で「闖」を得ることができる。

《辞書例》

漢字：JIS：構成情報

.....

榑：5c2d：L木+R府

控：5b7d：L木+R空

檸：5d26：L木+R寧

檳：5d27：L木+R賓

櫛：5c69：L木+R密

棕：5c23：L木+R宗

榕：5c57：L木+R容

櫛：5c6a：L木+R密

榿：5c43：L木+R室

.....

検索式：「L木+R密」がマッチ。

30

【0023】

《検索方法》

検索方法は簡単である。検索式の“+”は“and”で解釈し、“,”は“or”で解釈して検索式中の条件と辞書中の構成情報とを照合すればよい。

また、検索式中の構成情報にアルファベットが含まれていない場合は、どの位置にきてもいいものと解釈して検索する。

【0024】

例えば、“榿”を入力するのに「左部分は“木”で右部分は上が“白”で下が“七”」と表現したとすると丸括弧(,)を使用して下記のように表現することができる。

ex. 「左部分は“木”で右部分は上が“白”で下が“七”」

- 「L木+

40

50

R (T 白 + D 七) 」

また、例えば、“梧”を検索する場合は、下記のどちらの検索式でも検索可能である。

検索式 1 : 「 L 木 + R 吾 」

検索式 2 : 「 L 木 + R (T 五 + D 口) 」

【 0 0 2 5 】

辞書構成について説明する。

“椈”のように、右構成部分が、一文字として独立しない場合は、辞書の構成情報も、検索式と同様「R (T 白 + D 七) 」のように記す。

“梧”のように、右構成部分が、一文字として独立可能な場合は、辞書の構成情報は、独立した文字の方で表現する(下記辞書例参照)。無論、“R 吾”と共に“T 五 + D 口”の情報

10

を“梧”に与えていてもよいが、メモリをとられたり、辞書情報や検索が繁雑になる恐れがあるのでここでは、“R 吾”のみを記述する方式で以下説明する。

【 0 0 2 6 】

《辞書例》

漢字：J I S：構成情報

.....

吾：3 8 6 3：T 五 + D 口 ←検索式 2 はまずこれにマッチし、“R 吾”
という情報を得てから“梧”にマッチする。

20

.....

.....

梧：3 8 6 8：L 木 + R 吾 ←上記検索式 1, 2 共にマッチ。

.....

櫛：5 c 6 a：L 木 + R 密

櫛：5 c 4 3：L 木 + R 室

椈：5 b 7 4：L 木 + R (T 白 + D 七) ←検索式「L 木 + R (T 白 + D 七) 」
がマッチ。

30

.....

【 0 0 2 7 】

《検索方法》

検索式 1 の場合は、問題ない。

検索式 2 は、まず最初は、「L 木 + R (T 五 + D 口) 」という内容で辞書をスキャンする。しかし、該当するものが得られないので、丸括弧の内容「T 五 + D 口」で検索する。これで、“吾”を得ることができるので、検索式の丸括弧部分を“吾”に置き換え、今度は

40

「L 木 + R 吾」で再び辞書をスキャンし“梧”を得る。

他にも、「L 波 1 + R (O 門 + I 王) 」で「潤」を得ることができる。

【 0 0 2 8 】

例えば、“猷”を求めるのに、上記の方式で表現すると「L 酋 + R 犬」または「L (T ソ + D 酉) + R 犬」という検索式になるが、もっと簡単に「“櫛”の字の右側に、犬」という表現がしたいことがある。

この場合、例えば、位置情報の“上, 下, 左, 右, 中, 外”を“t, d, l, r, i, o”で表現し、当該漢字の右側に記すことにしたとする。そうすると、「L 櫛 r + R 犬」のように表現することができる。

これは、得たい漢字の左構成部分には、“櫛”の字の右部分がきて、得たい漢字の右構成

50

部分には、“犬”がくるということを指示している。

検索式「L 槽 r + R 犬」 “ 猷 ”

【0029】

辞書構成について説明する。検索式で使用している右側の添字による指示方法は、辞書記述においては極力使用しないようにする。

《辞書例》

漢字：J I S：構成情報

．．．．

槽：4 6 6 a：L 木 + R 酋

．．．．

猷：4 d 3 2：L 酋 + R 犬

．．．．

【0030】

《検索方法》

まず、右側の添字の解釈を行なう。

“槽”の辞書記述は“槽：4 6 6 a：L 木 + R 酋”となっているので、この右構成要素“酋”を得る。そして、今度は「L 酋 + R 犬」で検索を行ない、結果として“猷”を得る。他にも、「O 区 o + I 王」で「匡」を、「L 情 l + R 閱 i」で「悦」を得ることができる。

【0031】

上記で使用している漢字辞書に読み情報を付与し、検索結果として読み情報も得られるようにすればよい。

【0032】

【発明の効果】

本発明によれば、ユーザが通常人に口頭で伝えている表現を数個の記号で表現するだけで、所望の漢字が得られるので、部首表の類をひいたり、画数をカウントしたりしなくてよく、また柔軟な表現が可能であるため、ユーザの負担が非常に少なく済むので、エディティング作業の効率をアップすることができる。

【0033】

また、得たい漢字の形が明確にわかっているときには、絞り込みの精度が高いので効率よく検索ができる。また、あいまいな場合でも検索でき、且つ、得たい漢字の形が明確にわかっているときには、絞り込みの精度が高い検索ができる。また、より柔軟な検索が可能であるので、人が通常行なっている直観的な表現に近い検索が行なえる。また、漢字だけでなく、その読みも得ることができる。

【0034】

更に、本発明と前記特願平3 - 1 4 0 8 8 7号の発明に対比してみると、

A：本発明では、特願平3 - 1 4 0 8 8 7号の差分生成辞書に相当するもののみで済む。

B：本発明では、相対的位置情報を使用するので、木へんのものだということが予めわかっているならば、“L 木”（左側に“木”がくるという意味）という指定ができ、“案，杏，楽，禁，襟，困...”などは最初から候補にあがらないため、検索効率、ヒット精度がよく、候補が絞られるため、後に続く計算量も少なくて済む。

C：本発明の記述 特願平3 - 1 4 0 8 8 7号の記述方式

吾：T 五 + D 口

吾：五，口

梧：L 木 + R 吾

梧：木，五，口，吾

聒：L 耳 + R 舌

聒：耳，千，口，舌

となり、本発明では、検索式1：「L 木 + R 吾」，検索式2：「L 木 + R (T 五 + D 口)」のどちらでも検索可能であるが、辞書では“梧”に「L 木 + R 吾」の方の構成情報のみを登録しておくだけで済む。

本発明では、入力された検索式でみつからなかった場合、丸括弧の内容「T 五 + D 口」で検索し、“吾”を得てから再度検索することができるので、検索式2を入力されても結局

10

20

30

40

50

、検索式 1 を得ることができるからである。

D：本発明では、ある漢字 X の一部分を指定することができるため、例えば、“讠 d ”で、“讠 / 邇 / 透”の下部共通部分を、“病 t ”で“病 / 癌 / 痕”の上部共通部分を指定することができる。

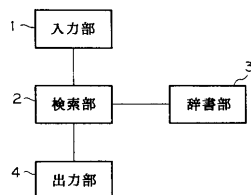
【図面の簡単な説明】

【図 1】本発明の実施例を説明するための全体構成図である。

【符号の説明】

1 ... 入力部、2 ... 検索部、3 ... 辞書部、4 ... 出力部。

【図 1】



フロントページの続き

(58)調査した分野(Int.Cl.⁷, DB名)

G06F 17/21-17/24

G06F 17/30 170