

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第3566111号
(P3566111)

(45) 発行日 平成16年9月15日(2004.9.15)

(24) 登録日 平成16年6月18日(2004.6.18)

(51) Int. Cl.⁷

G06F 17/30

F I

G06F 17/30 415
G06F 17/30 170A
G06F 17/30 350A

請求項の数 5 (全 38 頁)

<p>(21) 出願番号 特願平10-340765 (22) 出願日 平成10年11月30日(1998.11.30) (65) 公開番号 特開2000-163442(P2000-163442A) (43) 公開日 平成12年6月16日(2000.6.16) 審査請求日 平成13年2月22日(2001.2.22)</p>	<p>(73) 特許権者 000005821 松下電器産業株式会社 大阪府門真市大字門真1006番地 (74) 代理人 100082692 弁理士 蔵合 正博 (72) 発明者 菅野 祐司 大阪府門真市大字門真1006番地 松下電器産業株式会社内 審査官 辻本 泰隆</p>
--	--

最終頁に続く

(54) 【発明の名称】 記号辞書作成方法及び記号辞書検索方法

(57) 【特許請求の範囲】

【請求項1】

ある定められた文字セットに含まれる文字のN個(一定個数)以下の並びのうち相異なるものを記号として有限個登録した記号データに対して機械検索可能な記号辞書を作成する方法であって、

記号辞書作成の第1ステップでは、記号被覆手段が前記記号データ中の各記号に対して、前記文字セット中の各文字について、当該1文字の文字列だけからなる初期状態のメタ記号として登録した、初期状態のメタ記号辞書を用いた極大単語切り出し法(ある記号Sのさまざまな部分文字列に文字列照合するメタ記号辞書中のメタ記号のうちで、照合文字区間が、他のどの照合文字列にも、完全には含まれない照合だけを取り出す単語切り出し方法)によって照合するメタ記号を検索し、被覆結果(すなわち被覆対象の前記記号中の部分文字列と照合するメタ記号と照合開始文字位置の対を要素とし、前記記号中のどの位置の文字も、被覆結果中のいずれかの前記記号の部分文字列に含まれるような集合)を求め、前記被覆結果をメタ記号集計手段が集計して、前記記号データ中での各メタ記号の総出現度数をメタ記号度数表に蓄積し、メタ記号辞書更新判定手段がメタ記号度数表を参照し、前記メタ記号度数表中の度数が0のメタ記号を削除し、度数が記号内容で定まる所定の定数以上であるものについて、その末尾にメタ記号辞書中の一文字の各メタ記号に対して、その一文字を追加したメタ記号を追加し、

記号辞書作成の第2ステップでは、前記記号被覆手段が前記記号データ中の各記号に対して、現時点のメタ記号辞書を用いた前記極大単語切り出し法によって照合するメタ記号を

10

20

検索して被覆結果を求め、この被覆結果をメタ記号集計手段が集計して、各メタ記号の前記記号データ中での総出現度数をメタ記号度数表に蓄積し、メタ記号辞書更新判定手段が前記メタ記号度数表を参照し、度数が記号データ内容と第2ステップの繰り返し回数で決まる所定の定数以上であるものについて、末尾に、定められた文字セット中の任意の一字を追加するとともに、メタ記号辞書更新判定手段におけるメタ記号の追加と削除がまったくない場合に停止する停止条件を満足するまで、前記第2ステップを繰り返し行い、記号辞書構築の第3ステップでは、前記メタ記号辞書更新判定手段が前記メタ記号度数表を参照し、メタ記号度数表中の度数が所定の定数未満の、2文字以上のメタ記号の削除を行い、

記号辞書構築の第4ステップでは、前記記号被覆手段が第3ステップで算出した前記メタ記号辞書を用いて前記記号データに対して前記極大単語切り出し法によって被覆結果を求め、メタ記号出現情報作成手段が前記被覆結果から各メタ記号が出現する記号の番号と出現文字位置とを記録したメタ記号出現情報を算出し、記号辞書作成の第5ステップでは、記号辞書作成手段が、前記メタ記号辞書と前記メタ記号出現情報とから、メタ記号情報とメタ記号出現情報とを格納した、機械検索可能な記号辞書を作成することを特徴とする記号辞書作成方法。

【請求項2】

請求項1記載の記号辞書作成方法によって作成されたメタ記号情報とメタ記号出現情報とを格納した記号辞書を任意の文字列によって完全一致検索（すなわち質問文字列と同一の記号の検索）する方法であって、記号辞書検索の第1ステップでは、質問文字列被覆手段が、前記記号辞書中のメタ記号情報を検索し、検索条件である質問文字列Qに対して、重複最長一致単語切り出し法によって被覆結果（すなわち被覆対象文字列中の部分文字列と照合するメタ記号m，照合文字開始位置s，照合終了文字位置e（ $1 \leq s < e \leq |Q| + 1$ ）の対（m，s，e）を被覆要素とし、Qのどの文字も1つ以上の被覆要素に含まれるような集合）を求め、被覆結果がない（すなわち被覆対象文字列中の部分文字列と照合するメタ記号m，照合文字開始位置s，照合終了文字位置e（ $1 \leq s < e \leq |Q| + 1$ ）の対（m，s，e）を被覆要素とし、Qのどの文字も1つ以上の被覆要素に含まれるような集合がない）場合には、検索結果なしとして検索処理を終了し、被覆結果がある場合には被覆結果を作業域に格納し、記号辞書検索の第2ステップでは、記号番号集合算定手段が、前記記号辞書中のメタ記号出現情報を検索して、前記被覆結果中の全ての要素に共通して含まれる記号番号（高々1つ）があれば、検索結果として出力して検索処理を終了し、前記被覆結果中の全ての要素に共通して含まれる記号番号がなければ検索結果なしとして検索処理を終了することを特徴とする記号辞書検索方法。

【請求項3】

請求項1記載の記号辞書作成方法によって作成されたメタ記号情報とメタ記号出現情報とを格納した記号辞書を任意の文字列によって前方一致検索（すなわち質問文字列を先頭部分に持つ全ての記号の検索）する方法であって、記号辞書検索の第1ステップでは、質問文字列被覆手段が、前記記号辞書中のメタ記号情報を検索し、検索条件である質問文字列Qに対して、重複最長一致単語切り出し法によって被覆結果（すなわち被覆対象文字列中の部分文字列と照合するメタ記号m，照合文字開始位置s，照合終了文字位置e（ $1 \leq s < e \leq |Q| + 1$ ）の対（m，s，e）を被覆要素とし、Qのどの文字も1つ以上の被覆要素に含まれるような集合）を求め、被覆結果がない（すなわち被覆対象文字列中の部分文字列と照合するメタ記号m，照合文字開始位置s，照合終了文字位置e（ $1 \leq s < e \leq |Q| + 1$ ）の対（m，s，e）を被覆要素とし、Qのどの文字も1つ以上の被覆要素に含まれるような集合がない）場合には、検索結果なしとして検索処理を終了し、被覆結果がある場合には被覆結果を記録し、記号辞書検索の第2ステップでは、右延長メタ記号算定手段が、前記記号辞書のメタ記号情報を検索して、前記被覆結果中で、照合開始文字位置が1である被覆要素のメタ記号Zの延長メタ記号（すなわち、Zを含むメタ記号）のうちで、前記質問文字列の第j最右部分文字列（すなわち、質問文字列の第j文字（ $1 \leq j \leq |Q|$ ）から最終文字までの部分文字列）Rの右延長メタ記号（すなわち文字列Rを先

10

20

30

40

50

頭部分に含むメタ記号) になっているメタ記号 x を全て検索して、要素 $(x, j, |R| + j)$ を前記被覆結果に追加して記録し、記号辞書検索の第3ステップでは、記号番号集合算定手段が、前記質問文字列、またはその任意の右延長文字列を被覆する前記被覆結果中の要素の集合 C を系統的に生成しつつ、前記記号辞書中のメタ記号出現情報を検索して、 C 中の全ての要素に共通して含まれる記号番号集合 SC を求めて検索結果の一部として記録する動作を行って、全ての SC の和集合を最終的な検索結果として出力することを特徴とする記号辞書検索方法。

【請求項4】

請求項1記載の記号辞書作成方法によって作成されたメタ記号情報とメタ記号出現情報とを格納した記号辞書を任意の文字列によって後方一致検索(すなわち質問文字列を末尾部分に持つ全ての記号の検索)する方法であって、記号辞書検索の第1ステップでは、質問文字列被覆手段が、前記記号辞書中のメタ記号情報を検索し、検索条件である質問文字列 Q に対して、重複最長一致単語切り出し法によって被覆結果(すなわち被覆対象文字列中の部分文字列と照合するメタ記号 m , 照合文字開始位置 s , 照合終了文字位置 e ($1 \leq s < e \leq |Q| + 1$) の対 (m, s, e) を被覆要素とし、 Q のどの文字も1つ以上の被覆要素に含まれるような集合)を求め、被覆結果がない(すなわち被覆対象文字列中の部分文字列と照合するメタ記号 m , 照合文字開始位置 s , 照合終了文字位置 e ($1 \leq s < e \leq |Q| + 1$) の対 (m, s, e) を被覆要素とし、 Q のどの文字も1つ以上の被覆要素に含まれるような集合がない)場合には、検索結果なしとして検索処理を終了し、被覆結果がある場合には被覆結果を記録し、記号辞書検索の第2ステップでは、左延長メタ記号算定手段が、前記記号辞書のメタ記号情報を検索して、前記被覆結果のうち、照合終了文字位置が $|Q| + 1$ である被覆要素のメタ記号 Z の延長メタ記号(すなわち、 Z を含むメタ記号)のうちで、前記質問文字列の第 j 最左部分文字列(すなわち、質問文字列の第1文字から第 j 文字 ($1 \leq j \leq |Q|$) までの部分文字列) L の左延長メタ記号(すなわち文字列 L を末尾部分に含むメタ記号)になっているメタ記号 x を全て検索して、要素 $(x, j + 1 - |L|, j + 1)$ を前記被覆結果に追加して記録し、記号辞書検索の第3ステップでは、記号番号集合算定手段が、前記質問文字列、またはその任意の左延長文字列を被覆する前記被覆結果中の要素の集合 C を系統的に生成しつつ、前記記号辞書中のメタ記号出現情報を検索して、 C 中の全ての要素に共通して含まれる記号番号集合 SC を求めて検索結果の一部として記録する動作を行って、全ての SC の和集合を最終的な検索結果として出力することを特徴とする記号辞書検索方法。

【請求項5】

請求項1記載の記号辞書作成方法によって作成されたメタ記号情報とメタ記号出現情報とを格納した記号辞書を任意の文字列によって中間一致検索(すなわち質問文字列を含む全ての記号の検索)する方法であって、記号辞書検索の第1ステップでは、質問文字列被覆手段が、前記記号辞書中のメタ記号情報を検索し、検索条件である質問文字列 Q に対して、重複最長一致単語切り出し法によって被覆結果(すなわち被覆対象文字列中の部分文字列と照合するメタ記号 m , 照合文字開始位置 s , 照合終了文字位置 e ($1 \leq s < e \leq |Q| + 1$) の対 (m, s, e) を被覆要素とし、 Q のどの文字も1つ以上の被覆要素に含まれるような集合)を求め、被覆結果がない(すなわち被覆対象文字列中の部分文字列と照合するメタ記号 m , 照合文字開始位置 s , 照合終了文字位置 e ($1 \leq s < e \leq |Q| + 1$) の対 (m, s, e) を被覆要素とし、 Q のどの文字も1つ以上の被覆要素に含まれるような集合がない)場合には、検索結果なしとして検索処理を終了し、被覆結果がある場合には被覆結果を記録し、記号辞書検索の第2ステップでは、右延長メタ記号算定手段が、前記記号辞書のメタ記号情報を検索して、前記被覆結果中で、照合開始文字位置が1である被覆要素のメタ記号 Z の延長メタ記号(すなわち、 Z を含むメタ記号)のうちで、前記質問文字列の第 j 最右部分文字列(すなわち、質問文字列の第 j 文字 ($1 \leq j \leq |Q|$) から最終文字までの部分文字列) R の右延長メタ記号(すなわち文字列 R を先頭部分に含むメタ記号)になっているメタ記号 x を全て検索して、要素 $(x, j, |R| + j)$ を前記被覆結果に追加して記録し、記号辞書検索の第3ステップでは、左延長メタ記号算定手

段が、前記記号辞書のメタ記号情報を検索して、前記被覆結果のうち、照合終了文字位置が $|Q| + 1$ である被覆要素のメタ記号 Z の延長メタ記号（すなわち、 Z を含むメタ記号）のうちで、前記質問文字列の第 j 最左部分文字列（すなわち、質問文字列の第 1 文字から第 j 文字（ $1 \leq j \leq |Q|$ ）までの部分文字列） L の左延長メタ記号（すなわち文字列 L を末尾部分に含むメタ記号）になっているメタ記号 x を全て検索して、要素 $(x, j + 1 - |L|, j + 1)$ を前記被覆結果に追加して記録し、記号辞書検索の第 4 ステップでは、両延長メタ記号算定手段が前記メタ記号情報を検索して、 Q の両延長メタ記号（すなわち文字列 Q を第 j 文字から第 $j + |Q|$ 文字の部分（ $1 < j$ ）に含むメタ記号） X を全て検索して、要素 $(X, 1 - j, 1 - j + |X|)$ を前記被覆結果に追加して記録し、記号辞書検索の第 5 ステップでは、記号番号集合算定手段が、前記質問文字列、またはその任意の延長文字列を被覆する前記被覆結果中の要素の集合 C を系統的に生成しつつ、前記記号辞書中のメタ記号出現情報を検索して、 C 中の全ての要素に共通して含まれる記号番号集合 SC を求めて検索結果の一部として記録する動作を行って、全ての SC の和集合を最終的な検索結果として出力することを特徴とする記号辞書検索方法。

10

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、電子化されて蓄積された記号情報を、計算機を利用して管理・検索する、データベース装置や文書検索装置等に使用される機械検索可能な記号辞書の作成と検索に関するものである。

20

【0002】

【従来の技術】

近年、ワードプロセッサやパーソナルコンピュータの普及、CD-ROM等の大容量かつ低価格な記憶媒体の普及、イーサネットに代表されるネットワークの発達により、リレーショナルデータベースや全文検索データベース等のデータベースシステムが実用化され、広く利用されるようになってきた。

その中で、人名、地名、組織名、住所、分類コード、部品コード等、数文字～数百文字程度の比較的短い文字列を「記号」として扱い、記号のCSVリスト（カンマ「,」によって連結された記号の列で、例えば取引会社名のフィールドとして「森下電器産業, 森下通信工業, 九州森下電器」などをデータベースの1項目（フィールド）に格納し、このフィールドに対して、記号単位の完全一致照合、前方一致照合、後方一致照合、中間一致照合（部分一致照合、任意一致照合とも言う）によるレコードの検索（前方一致照合の例では「取引会社名フィールドに『森下』で始まる記号を含むレコードを検索せよ」といった条件の検索）を高速に行うことが求められている。

30

【0003】

上記の4種類の照合方法のうち、完全一致照合、および前方一致照合のみ行えば十分である用途には、「アルゴリズム第2巻（R.セジウィック著、野下浩平他訳、近代科学社、1992年刊、ISBN4-7649-0189-7、52頁～72頁）」や「アルゴリズムとデータ構造ハンドブック（G.H.Gonnet著、玄光男他訳、啓学出版、1987年刊、ISBN4-7665-0326-0、111頁～122頁）」等の成書に記述されているTRIE（基数探索木とも呼ぶ）というデータ構造を使用すれば、効率のよい検索処理が実現できる。これに加えて後方一致照合が必要な場合には、記号の文字順序を反転させたデータに対してTRIEを作成し、検索すればよい。

40

【0004】

中間一致照合をも行う必要がある場合には、TRIEでは効率的な検索処理は困難で、従来は、例えば特開平3-42774号広報に開示されているような方法が用いられていた。

【0005】

この方法では、記号辞書作成の際には、記号の文字列を文字単位に分割し、文字毎に、記号の番号と、記号中での当該文字の出現文字位置の対を記録した辞書情報を作成し、記号

50

辞書検索の際には、質問文字列を文字単位に分解し、各文字に対応する辞書情報を検索して、記号番号が同一で、出現文字位置が連続する記号番号の集合を検索結果として出力する。

【0006】

しかしながら、上記の従来の記号辞書作成方法では、記号の種類が数万種類以上に達する場合には、作成する記号辞書ファイルが検索対象の記号データの2倍以上の大きさになり、使用できる記憶装置の容量に制約のある場合に実現が困難である、という課題があった。

【0007】

また、上記の従来の記号辞書検索方法では、文字列が長く、高頻出の文字を多数含む記号を検索する場合には、記号辞書から検索すべきデータ量が多くなり、検索速度が低下するという課題があった。

10

【0008】

この従来の記号辞書検索方法の欠点は、記号辞書の記録を、1文字単位ではなく、2文字以上のN文字連鎖について行うことによってある程度緩和することが可能だが、その場合でも、例えば「19980000123A」のような、先頭に年度を記述し、多数桁の整数値で、ほとんどの記号の先頭部分桁が0の連続であるような記号を多数集めた記号データを検索するような場合には、先頭の10文字以上が一致する記号が数多く存在するため、N文字連鎖のNが2~4程度である場合には、やはり記号辞書から検索すべきデータ量が多くなり、検索速度が低下するという課題があった。

20

【0009】

さらに、文字連鎖数のNを大きくすると、出現するN文字連鎖の種類が急激に増加し、記号辞書の作成が困難になり、また作成した記号辞書の容量も増加するという課題があった。また、この従来の記号辞書検索方法では、文字列が長く、高頻出の文字を多数含む記号を検索する場合には、完全一致検索が最も処理時間がかかるため、完全一致照合が質問の大部分を占めるような用途では、平均検索速度が低下するという課題があった。

【0010】

【発明が解決しようとする課題】

このように、上記の従来の記号辞書作成方法では、作成する記号辞書ファイルが検索対象の記号データの2倍以上の大きさになり、使用できる記憶装置の容量に制約のある場合に実現が困難である、という課題があった。

30

【0011】

また、上記の従来の記号辞書検索方法では、文字列が長く、高頻出の文字、あるいは文字連鎖を含む記号を検索する場合には、記号辞書から検索すべきデータ量が多くなり、検索速度が低下するという課題があった。

【0012】

さらに、文字連鎖数Nを大きくすると、出現するN文字連鎖の種類が急激に増加し、記号辞書の作成が困難になり、また作成した記号辞書の容量も増加するという課題があった。

【0013】

本発明の記号辞書作成方法は、記号データ中の記号を被覆する、記号より短い部分文字列の集合である「メタ記号」と呼ぶ文字列（いわゆる記号）を集めたメタ記号辞書を自動的に作成し、このメタ記号辞書中のメタ記号で記号データ中の各記号を被覆し、各記号がどのように被覆されたか、という情報を、各メタ記号毎に記録したメタ記号出現情報を作成することで、中間一致照合まで含めた高速な検索が可能で、かつ容量の小さな記号辞書の作成を可能にし、本発明の記号辞書検索方法は、質問文字列を、記号辞書に含まれるメタ記号辞書を検索してメタ記号で被覆し、被覆結果に左、右、両延長メタ記号の検索結果を追加し、質問文字列あるいはその左、右、両延長文字列を被覆する被覆結果中の要素集合毎に、共通して含まれる記号番号集合を求めることにより、中間一致照合まで含めた高速な検索が可能で、しかも完全一致照合が質問の大部分を占めるような用途でも、平均検索速度が低下しない記号検索を可能にするものである。

40

50

【 0 0 1 4 】

【課題を解決するための手段】

この目的を達成するため、本発明の第1の様態による記号辞書作成方法は、記号データ中の各記号を、メタ記号辞書を検索して、極大単語切り出し法などの切り出し手法によって被覆結果を求める、記号被覆手段と、被覆結果を集計するメタ記号集計手段と、記号データ中の各メタ記号の総出現度数を蓄積するメタ記号度数表と、メタ記号度数表を参照し、前記メタ記号度数表中の度数が0のメタ記号を削除し、度数が記号内容で定まる所定の定数以上であるものについて、その末尾にメタ記号辞書中の一文字の各メタ記号に対して、その一文字を追加したメタ記号を追加し、また、停止条件を満足するか否かを判定するメタ記号辞書更新判定手段と、被覆結果から各メタ記号が出現する記号の番号と出現文字位置とを記録したメタ記号出現情報を算出するメタ記号出現情報作成手段と、メタ記号辞書とメタ記号出現情報とから、機械検索可能な記号辞書を作成する記号辞書作成手段とを有することにより、文字列が長く、高頻出の文字、あるいは文字連鎖を含む記号を検索する場合でも、中間一致照合まで含めた高速な検索が可能で、かつ容量の小さな記号辞書の作成を可能にするという効果を奏するものである。

10

【 0 0 1 5 】

また、本発明の第2の様態による、完全一致照合の記号辞書検索方法は、記号辞書中のメタ記号辞書を検索し、検索条件である質問文字列に対して、重複最長一致単語切り出し法によって被覆結果を求める質問文字列被覆手段と、前記記号辞書中のメタ記号出現情報を検索し、第1ステップで求めた被覆結果中の各要素に対して、当該メタ記号を当該位置に含む記号番号の集合を求めるメタ記号出現情報検索手段と、被覆結果中の全要素について、対応する記号番号集合の共通部分を求め、求めた共通部分が空でないなら、その要素である記号番号を検索結果として出力して検索処理を終了し、空集合ならば検索結果なしとして検索処理を終了する記号番号算定手段とを有することにより、文字列が長く、高頻出の文字、あるいは文字連鎖を含む記号を検索する場合でも、高速な完全一致照合による記号辞書検索が可能になるという効果を奏するものである。

20

【 0 0 1 6 】

また、本発明の第3の様態による、前方一致照合の記号辞書検索方法は、前記記号辞書中のメタ記号辞書を検索し、検索条件である質問文字列Qに対して、重複最長一致単語切り出し法によって被覆を求め、被覆がない場合には、検索結果なしとして検索処理を終了し、被覆がある場合には被覆結果を記録する質問文字列被覆手段と、前記記号辞書のメタ記号情報を検索して、前記被覆結果中で、照合開始文字位置が1である被覆要素のメタ記号Zの延長メタ記号(すなわち、Zを含むメタ記号)のうちで、前記質問文字列の第j最右部分文字列(すなわち、質問文字列の第j文字(1 ≤ j ≤ |Q|)から最終文字までの部分文字列)Rの右延長メタ記号(すなわち文字列Rを先頭部分に含むメタ記号)になっているメタ記号xを全て検索して、要素(x, j, |R| + j)を前記被覆結果に追加して記録する右延長メタ記号算定手段と、前記質問文字列、またはその任意の右延長文字列を被覆する前記被覆結果中の要素の集合Cを系統的に生成しつつ、前記記号辞書中のメタ記号出現情報を検索して、C中の全ての要素に共通して含まれる記号番号集合SCを求めて検索結果の一部として記録する動作を行って、全てのSCの和集合を最終的な検索結果として出力する記号番号集合算定手段とを有することにより、文字列が長く、高頻出の文字、あるいは文字連鎖を含む記号を検索する場合でも、高速な前方一致照合による記号辞書検索が可能になるという効果を奏するものである。

30

40

【 0 0 1 7 】

また、本発明の第4の様態による、後方一致照合の記号辞書検索方法は、記号辞書中のメタ記号辞書を検索し、検索条件である質問文字列Qに対して、重複最長一致単語切り出し法によって被覆を求め、被覆がない場合には、検索結果なしとして検索処理を終了し、被覆がある場合には被覆結果を記録する質問文字列被覆手段と、前記記号辞書のメタ記号情報を検索して、前記被覆結果のうち、照合終了文字位置が|Q| + 1である被覆要素のメタ記号Zの延長メタ記号(すなわち、Zを含むメタ記号)のうちで、前記質問文字列の第

50

j 最左部分文字列（すなわち、質問文字列の第 1 文字から第 j 文字（ $1 \leq j \leq |Q|$ ）までの部分文字列）L の左延長メタ記号（すなわち文字列 L を末尾部分に含むメタ記号）になっているメタ記号 x を全て検索して、要素（ $x, j+1 - |L|, j+1$ ）を前記被覆結果に追加して記録する左延長メタ記号算定手段と、前記質問文字列、またはその任意の左延長文字列を被覆する前記被覆結果中の要素の集合 C を系統的に生成しつつ、前記記号辞書中のメタ記号出現情報を検索して、C 中の全ての要素に共通して含まれる記号番号集合 S C を求めて検索結果の一部として記録する動作を行って、全ての S C の和集合を最終的な検索結果として出力する記号番号集合算定手段とを有することにより、文字列が長く、高頻出の文字、あるいは文字連鎖を含む記号を検索する場合でも、高速な後方一致照合による記号辞書検索が可能になるという効果を奏するものである。

10

【0018】

また、本発明の第 5 の様態による、中間一致照合の記号辞書検索方法は、記号辞書中のメタ記号辞書を検索し、検索条件である質問文字列 Q に対して、重複最長一致単語切り出し法によって被覆を求め、被覆がない場合には、検索結果なしとして検索処理を終了し、被覆がある場合には被覆結果を記録する質問文字列被覆手段と、前記記号辞書のメタ記号情報を検索して、前記被覆結果中で、照合開始文字位置が 1 である被覆要素のメタ記号 Z の延長メタ記号（すなわち、Z を含むメタ記号）のうちで、前記質問文字列の第 j 最右部分文字列（すなわち、質問文字列の第 j 文字（ $1 \leq j \leq |Q|$ ）から最終文字までの部分文字列）R の右延長メタ記号（すなわち文字列 R を先頭部分に含むメタ記号）になっているメタ記号 x を全て検索して、要素（ $x, j, |R| + j$ ）を前記被覆結果に追加して記録する右延長メタ記号算定手段と、前記記号辞書のメタ記号情報を検索して、前記被覆結果のうち、照合終了文字位置が $|Q| + 1$ である被覆要素のメタ記号 Z の延長メタ記号（すなわち、Z を含むメタ記号）のうちで、前記質問文字列の第 j 最左部分文字列（すなわち、質問文字列の第 1 文字から第 j 文字（ $1 \leq j \leq |Q|$ ）までの部分文字列）L の左延長メタ記号（すなわち文字列 L を末尾部分に含むメタ記号）になっているメタ記号 x を全て検索して、要素（ $x, j+1 - |L|, j+1$ ）を前記被覆結果に追加して記録する左延長メタ記号算定手段と、前記メタ記号辞書を検索して、Q の両延長メタ記号 x を全て検索して、要素（ $x, 1 - j, 1 - j + |x|$ ）を前記被覆結果に追加して記録する両延長メタ記号算定手段と、前記質問文字列、またはその任意の延長文字列を被覆する前記被覆結果中の要素の集合 C を系統的に生成しつつ、前記記号辞書中のメタ記号出現情報を検索して、C 中の全ての要素に共通して含まれる記号番号集合 S C を求めて検索結果の一部として記録する動作を行って、全ての S C の和集合を最終的な検索結果として出力する記号番号集合算定手段とを有することにより、文字列が長く、高頻出の文字、あるいは文字連鎖を含む記号を検索する場合でも、高速な中間一致照合による記号辞書検索が可能になるという効果を奏するものである。

20

30

【0019】

【発明の実施の形態】

本発明の請求項 1 記載の発明は、記号辞書作成方法として、ある定められた文字セットに含まれる文字の N 個（一定個数）以下の並びのうち相異なるものを記号として有限個登録した記号データに対して機械検索可能な記号辞書を作成する方法であって、記号辞書作成の第 1 ステップでは、記号被覆手段が前記記号データ中の各記号に対して、前記文字セット中の各文字について、当該 1 文字の文字列だけからなる初期状態のメタ記号として登録した、初期状態のメタ記号辞書を用いた極大単語切り出し法（ある記号 S のさまざまな部分文字列に文字列照合するメタ記号辞書中のメタ記号のうちで、照合文字区間が、他のどの照合文字列にも、完全には含まれない照合だけを取り出す単語切り出し方法）によって照合するメタ記号を検索し、被覆結果（すなわち被覆対象の前記記号中の部分文字列と照合するメタ記号と照合開始文字位置の対を要素とし、前記記号中のどの位置の文字も、被覆結果中のいずれかの前記記号の部分文字列に含まれるような集合）を求め、前記被覆結果をメタ記号集計手段が集計して、前記記号データ中での各メタ記号の総出現度数をメタ記号度数表に蓄積し、メタ記号辞書更新判定手段がメタ記号

40

50

度数表を参照し、前記メタ記号度数表中の度数が0のメタ記号を削除し、度数が記号内容で定まる所定の定数以上であるものについて、その末尾にメタ記号辞書中の一文字の各メタ記号に対して、その一文字を追加したメタ記号を追加し、記号辞書作成の第2ステップでは、前記記号被覆手段が前記記号データ中の各記号に対して、現時点のメタ記号辞書を用いた前記極大単語切り出し法によって照合するメタ記号を検索して被覆結果を求め、この被覆結果をメタ記号集計手段が集計して、各メタ記号の前記記号データ中での総出現度をメタ記号度数表に蓄積し、メタ記号辞書更新判定手段が前記メタ記号度数表を参照し、度数が記号データ内容と第2ステップの繰り返し回数で定まる所定の定数以上であるものについて、末尾に、定められた文字セット中の任意の一文字を追加するとともに、メタ記号辞書更新判定手段におけるメタ記号の追加と削除がまったくない場合に停止する停止条件を満足するまで、前記第2ステップを繰り返し行い、記号辞書構築の第3ステップでは、前記メタ記号辞書更新判定手段が前記メタ記号度数表を参照し、メタ記号度数表中の度数が所定の定数未満の、2文字以上のメタ記号の削除を行い、記号辞書構築の第4ステップでは、前記記号被覆手段が第3ステップで算出した前記メタ記号辞書を用いて前記記号データに対して前記極大単語切り出し法によって被覆結果を求め、メタ記号出現情報作成手段が前記被覆結果から各メタ記号が出現する記号の番号と出現文字位置とを記録したメタ記号出現情報を算出し、記号辞書作成の第5ステップでは、記号辞書作成手段が、前記メタ記号辞書と前記メタ記号出現情報とから、メタ記号情報とメタ記号出現情報とを格納した、機械検索可能な記号辞書を作成するようにしたものであり、文字列が長く、高頻出の文字、あるいは文字連鎖を含む記号を検索する場合でも、中間一致照合まで含めた高速な検索が可能で、かつ容量の小さな記号辞書の作成を可能にするという作用を有する。

10

20

【0026】

本発明の請求項2記載の発明は、請求項1記載の記号辞書作成方法によって作成された記号辞書検索方法として、メタ記号情報とメタ記号出現情報とを格納した記号辞書を任意の文字列によって完全一致検索（すなわち質問文字列と同一の記号の検索）する方法であって、記号辞書検索の第1ステップでは、質問文字列被覆手段が、前記記号辞書中のメタ記号情報を検索し、検索条件である質問文字列Qに対して、重複最長一致単語切り出し法によって被覆結果（すなわち被覆対象文字列中の部分文字列と照合するメタ記号m，照合文字開始位置s，照合終了文字位置 $e (1 \leq s < e \leq |Q| + 1)$ の対(m, s, e)を被覆要素とし、Qのどの文字も1つ以上の被覆要素に含まれるような集合)を求め、被覆結果がない（すなわち被覆対象文字列中の部分文字列と照合するメタ記号m，照合文字開始位置s，照合終了文字位置 $e (1 \leq s < e \leq |Q| + 1)$ の対(m, s, e)を被覆要素とし、Qのどの文字も1つ以上の被覆要素に含まれるような集合がない）場合には、検索結果なしとして検索処理を終了し、被覆結果がある場合には被覆結果を作業域に格納し、記号辞書検索の第2ステップでは、記号番号集合算定手段が、前記記号辞書中のメタ記号出現情報を検索して、前記被覆結果中の全ての要素に共通して含まれる記号番号（高々1つ）があれば、検索結果として出力して検索処理を終了し、前記被覆結果中の全ての要素に共通して含まれる記号番号がなければ検索結果なしとして検索処理を終了するにようにしたものであり、文字列が長く、高頻出の文字、あるいは文字連鎖を含む記号を検索する場合でも、高速な完全一致照合による記号辞書検索が可能になるという作用を有する。

30

40

【0027】

本発明の請求項3記載の発明は、請求項1記載の記号辞書作成方法によって作成された記号辞書検索方法として、メタ記号情報とメタ記号出現情報とを格納した記号辞書を任意の文字列によって前方一致検索（すなわち質問文字列を先頭部分に持つ全ての記号の検索）する方法であって、記号辞書検索の第1ステップでは、質問文字列被覆手段が、前記記号辞書中のメタ記号情報を検索し、検索条件である質問文字列Qに対して、重複最長一致単語切り出し法によって被覆結果（すなわち被覆対象文字列中の部分文字列と照合するメタ記号m，照合文字開始位置s，照合終了文字位置 $e (1 \leq s < e \leq |Q| + 1)$ の対(m, s, e)を被覆要素とし、Qのどの文字も1つ以上の被覆要素に含まれるような集合)を求め、被覆結果がない（すなわち被覆対象文字列中の部分文字列と照合するメタ記号m

50

、照合文字開始位置 s 、照合終了文字位置 e ($1 \leq s < e \leq |Q| + 1$) の対 (m, s, e) を被覆要素とし、 Q のどの文字も 1 つ以上の被覆要素に含まれるような集合がない) 場合には、検索結果なしとして検索処理を終了し、被覆結果がある場合には被覆結果を記録し、記号辞書検索の第 2 ステップでは、右延長メタ記号算定手段が、前記記号辞書のメタ記号情報を検索して、前記被覆結果中で、照合開始文字位置が 1 である被覆要素のメタ記号 Z の延長メタ記号 (すなわち、 Z を含むメタ記号) のうちで、前記質問文字列の第 j 最右部分文字列 (すなわち、質問文字列の第 j 文字 ($1 \leq j \leq |Q|$) から最終文字までの部分文字列) R の右延長メタ記号 (すなわち文字列 R を先頭部分に含むメタ記号) になっているメタ記号 x を全て検索して、要素 ($x, j, |R| + j$) を前記被覆結果に追加して記録し、記号辞書検索の第 3 ステップでは、記号番号集合算定手段が、前記質問文字列、またはその任意の右延長文字列を被覆する前記被覆結果中の要素の集合 C を系統的に生成しつつ、前記記号辞書中のメタ記号出現情報を検索して、 C 中の全ての要素に共通して含まれる記号番号集合 SC を求めて検索結果の一部として記録する動作を行って、全ての SC の和集合を最終的な検索結果として出力するようにしたものであり、文字列が長く、高頻出の文字、あるいは文字連鎖を含む記号を検索する場合でも、高速な前方一致照合による記号辞書検索が可能になるという作用を有する。

【0028】

本発明の請求項 4 記載の発明は、請求項 1 記載の記号辞書作成方法によって作成された記号辞書検索方法として、メタ記号情報とメタ記号出現情報とを格納した記号辞書を任意の文字列によって後方一致検索 (すなわち質問文字列を末尾部分に持つ全ての記号の検索) する方法であって、記号辞書検索の第 1 ステップでは、質問文字列被覆手段が、前記記号辞書中のメタ記号情報を検索し、検索条件である質問文字列 Q に対して、重複最長一致単語切り出し法によって被覆結果 (すなわち被覆対象文字列中の部分文字列と照合するメタ記号 m 、照合文字開始位置 s 、照合終了文字位置 e ($1 \leq s < e \leq |Q| + 1$) の対 (m, s, e) を被覆要素とし、 Q のどの文字も 1 つ以上の被覆要素に含まれるような集合) を求め、被覆結果がない (すなわち被覆対象文字列中の部分文字列と照合するメタ記号 m 、照合文字開始位置 s 、照合終了文字位置 e ($1 \leq s < e \leq |Q| + 1$) の対 (m, s, e) を被覆要素とし、 Q のどの文字も 1 つ以上の被覆要素に含まれるような集合がない) 場合には、検索結果なしとして検索処理を終了し、被覆結果がある場合には被覆結果を記録し、記号辞書検索の第 2 ステップでは、左延長メタ記号算定手段が、前記記号辞書のメタ記号情報を検索して、前記被覆結果のうち、照合終了文字位置が $|Q| + 1$ である被覆要素のメタ記号 Z の延長メタ記号 (すなわち、 Z を含むメタ記号) のうちで、前記質問文字列の第 j 最左部分文字列 (すなわち、質問文字列の第 1 文字から第 j 文字 ($1 \leq j \leq |Q|$) までの部分文字列) L の左延長メタ記号 (すなわち文字列 L を末尾部分に含むメタ記号) になっているメタ記号 x を全て検索して、要素 ($x, j + 1 - |L|, j + 1$) を前記被覆結果に追加して記録し、記号辞書検索の第 3 ステップでは、記号番号集合算定手段が、前記質問文字列、またはその任意の左延長文字列を被覆する前記被覆結果中の要素の集合 C を系統的に生成しつつ、前記記号辞書中のメタ記号出現情報を検索して、 C 中の全ての要素に共通して含まれる記号番号集合 SC を求めて検索結果の一部として記録する動作を行って、全ての SC の和集合を最終的な検索結果として出力するようにしたものであり、文字列が長く、高頻出の文字、あるいは文字連鎖を含む記号を検索する場合でも、高速な後方一致照合による記号辞書検索が可能になるという作用を有する。

【0029】

本発明の請求項 5 記載の発明は、請求項 1 記載の記号辞書作成方法によって作成された記号辞書検索方法として、メタ記号情報とメタ記号出現情報とを格納した記号辞書を任意の文字列によって中間一致検索 (すなわち質問文字列を含む全ての記号の検索) する方法であって、記号辞書検索の第 1 ステップでは、質問文字列被覆手段が、前記記号辞書中のメタ記号情報を検索し、検索条件である質問文字列 Q に対して、重複最長一致単語切り出し法によって被覆結果 (すなわち被覆対象文字列中の部分文字列と照合するメタ記号 m 、照合文字開始位置 s 、照合終了文字位置 e ($1 \leq s < e \leq |Q| + 1$) の対 (m, s, e))

を被覆要素とし、Qのどの文字も1つ以上の被覆要素に含まれるような集合)を求め、被覆結果がない(すなわち被覆対象文字列中の部分文字列と照合するメタ記号 m 、照合文字開始位置 s 、照合終了文字位置 e ($1 \leq s < e \leq |Q| + 1$)の対(m, s, e)を被覆要素とし、Qのどの文字も1つ以上の被覆要素に含まれるような集合がない)場合には、検索結果なしとして検索処理を終了し、被覆結果がある場合には被覆結果を記録し、記号辞書検索の第2ステップでは、右延長メタ記号算定手段が、前記記号辞書のメタ記号情報を検索して、前記被覆結果中で、照合開始文字位置が1である被覆要素のメタ記号 Z の延長メタ記号(すなわち、 Z を含むメタ記号)のうちで、前記質問文字列の第 j 最右部分文字列(すなわち、質問文字列の第 j 文字($1 \leq j \leq |Q|$)から最終文字までの部分文字列)Rの右延長メタ記号(すなわち文字列Rを先頭部分に含むメタ記号)になっているメタ記号 x を全て検索して、要素($x, j, |R| + j$)を前記被覆結果に追加して記録し、記号辞書検索の第3ステップでは、左延長メタ記号算定手段が、前記記号辞書のメタ記号情報を検索して、前記被覆結果のうち、照合終了文字位置が $|Q| + 1$ である被覆要素のメタ記号 Z の延長メタ記号(すなわち、 Z を含むメタ記号)のうちで、前記質問文字列の第 j 最左部分文字列(すなわち、質問文字列の第1文字から第 j 文字($1 \leq j \leq |Q|$)までの部分文字列)Lの左延長メタ記号(すなわち文字列Lを末尾部分に含むメタ記号)になっているメタ記号 x を全て検索して、要素($x, j + 1 - |L|, j + 1$)を前記被覆結果に追加して記録し、記号辞書検索の第4ステップでは、両延長メタ記号算定手段が前記メタ記号情報を検索して、Qの両延長メタ記号(すなわち文字列Qを第 j 文字から第 $j + |Q|$ 文字の部分($1 < j$)に含むメタ記号) X を全て検索して、要素($X, 1 - j, 1 - j + |X|$)を前記被覆結果に追加して記録し、記号辞書検索の第5ステップでは、記号番号集合算定手段が、前記質問文字列、またはその任意の延長文字列を被覆する前記被覆結果中の要素の集合Cを系統的に生成しつつ、前記記号辞書中のメタ記号出現情報を検索して、C中の全ての要素に共通して含まれる記号番号集合SCを求めて検索結果の一部として記録する動作を行って、全てのSCの和集合を最終的な検索結果として出力する文字列が長く、高頻出の文字、あるいは文字連鎖を含む記号を検索する場合でも、高速な中間一致照合による記号辞書検索が可能になるという効果を奏するものである。

【0030】

(実施の形態1)

以下、本発明の第1の実施の形態について、図面を参照しながら説明する。図1は本発明の請求項1乃至7にかかる、記号辞書作成装置の一実施の形態の全体構成を表わすブロック図である。図1において、101は辞書の作成対象である記号データ、102はメタ記号辞書、103はメタ記号辞書102を検索して、極大単語切り出し法によって記号データ中の各記号の被覆結果を求める記号被覆手段、104は記号被覆手段103の出力する被覆結果を受け取って、被覆要素として切り出されたメタ記号を集計するメタ記号集計手段、105はメタ記号集計手段104の集計結果を蓄えるメタ記号度数表、106はメタ記号度数表105の集計結果をもとに、メタ記号辞書102へのメタ記号の追加、メタ記号辞書102からのメタ記号の削除、およびメタ記号辞書更新処理の停止条件の判定をおこなうメタ記号辞書更新判定手段、107は記号被覆手段103の出力する被覆結果を受け取って、メタ記号毎に、当該メタ記号が被覆要素として切り出された記号番号と、切り出された文字位置とを記録したメタ記号出現情報を作成するメタ記号出現情報作成手段、108はメタ記号出現情報作成手段107が作成したメタ記号出現情報、109はメタ記号辞書102とメタ記号出現情報108とから、検索可能な記号辞書を作成する記号辞書作成手段、110は記号辞書作成手段109が作成する、検索可能な記号辞書である。また、図6は、記号被覆手段103における記号データ101中の各被覆対象記号から、メタ記号辞書102を用いて、極大単語切り出し法によって被覆結果を求める際の処理の手順を表す流れ図である。

【0031】

以上の構成の記号辞書作成装置において、その動作を、日付、15分刻みの時刻、ローマ字表記の苗字をつなげた記号データから、検索可能な記号辞書を作成する例を用いて説明

する。図15は、記号データの例である。この図では一部の記号を省略しており、実際には合計1000個の相異なる記号が、日付、時刻の順に記号データとして格納されているとする。各記号の最大文字数は1000文字とする。また、第1ステップにおけるメタ記号の追加のしきい値 C_1 は記号数の $1/20$ とする。この例では、記号数が1000なので、 $C_1 = 50$ となる。第2ステップでの k 回目の繰り返しにおけるメタ記号の追加のしきい値 C_k は $k < 10$ では記号数の $1/10$ とし、 $10 \leq k$ では、記号数の $1/5$ とする。この例では、 $k < 10$ では $C_k = 100$ 、 $10 \leq k$ では、 $C_k = 200$ となる。第3ステップでのしきい値 E の値は5と定める。なお、この記号データ中では、「-」、「/」の2種の特殊記号、「0」~「9」の数字、「A」~「Z」までの英字のみを使用するものとし、その他の文字は使用されることはないとする。このような記号を被覆するためのメタ記号辞書として、記号辞書の作成に先立って、予め、図16のような、記号中で出現する可能性のあるすべての1文字からなる内容のメタ記号辞書を用意する。記号被覆手段において、効率よく被覆を求めることができるように、メタ記号辞書は、図30のようなデジタル木のデータ構造(すなわちT R I E)として保持するものとする。ある文字列に最長一致するメタ記号の検索処理は、このようなT R I Eを用いることにより、T R I Eの根(左端の黒丸)から、文字列の1文字目, 2文字目, ...を鍵にして木構造をたどり、最も根から遠い2重丸のノードまでの部分文字列を出力することで、効率よく行うことが可能になる。記号辞書作成の第1ステップでは、記号被覆手段103が、記号データ101に格納された、図15の記号データを、順に読み込んで、メタ記号辞書102に格納された図30の内容のメタ記号辞書を検索して、極大単語切り出し法によって被覆結果を

10

20

求める。「極大単語切り出し法」とは、ある記号 S のさまざまな部分文字列に文字列照合するメタ記号辞書中のメタ記号のうちで、照合文字区間が、他のどの照合文字列にも、完全には含まれない照合だけを取り出す単語切り出しの方法である。例えば、記号「東京都議会」を、6個のメタ記号を含むメタ記号辞書{東, 東京, 東京都, 京都, 都議会, 議会}を検索して極大単語切り出しを行うと、図6に示した流れ図のように被覆処理を行うことにより、

{(東京都, 1, 4), (都議会, 3, 6)}という被覆結果が得られる。

【0032】

ただし、 (X, s, e) は、 X がメタ記号を、 s が照合開始文字位置を、 e が照合終了文字位置(すなわち照合している部分のすぐ右の文字の文字位置)を、それぞれ表し、この3つ組みを「被覆要素」と呼ぶ。「東」や「東京」や「京都」というメタ記号は、「東京都議会」の部分文字列と照合するものの、メタ記号辞書にはあるが、「東京都」という照合部分(1文字目から3文字目まで)に完全に含まれてしまうので、被覆結果には入らない。この「他の照合部分に完全に含まれてしまう照合を取り除く」処理は、図6では603で判定しており、記号 S の i 文字目から $i + j - 1$ 文字目までと照合するメタ記号 M に対して、 $1 \sim i - 1$ 文字目からの、すでに求めた照合結果のうち、照合終了位置(=照合部分右隣の文字位置) e が、 $i + j$ 以上ならば、その照合 (X, s, e) の照合文字区間 $[s, e - 1]$ は文字区間 $[i, i + j - 1]$ を完全に含むことになり、その場合には603の判定が「yes」となって、被覆結果集合 C には、照合した結果は加えない。

30

【0033】

40

さて、図16のメタ記号辞書中のメタ記号はいずれも1文字であるから、第1ステップでの記号被覆手段の被覆処理においては、切り出したメタ記号は必ず1文字であり、図6の603の判定が「yes」になることはなく、例えば、記号

「1998 - JAN - 01 / AM 0200 / KAWAYASU」

は、

{(1, 1, 2), (9, 2, 3), ..., (S, 26, 27), (U, 27, 28)}

のように、1文字ずつ被覆される。記号中に出現する可能性のある文字はすべてメタ記号辞書に含まれているので、必ず空でない被覆結果を求めることができる。各記号の被覆結果は、メタ記号集計手段104に渡され、各メタ記号が、被覆要素として何回切り出されたかが、メタ記号度数表105に記録される。例えば、記号データ中の最初の記号「19

50

「98 - JAN - 01 / AM0200 / KAWAYASU」の処理の後では、メタ記号度数表の内容は、「-」が2回、「/」が2回、「0」が4回、「1」が2回、「2」が1回、「8」が1回、「9」が2回、「A」が5回、「J」が1回、「K」が1回、「M」が1回、「N」が1回、「S」が1回、「U」が1回、「W」が1回、「Y」が1回となる。その他の文字の度数は0回である。記号データ101中の1000個の記号を処理した後のメタ記号度数表の内容は図17のようになる。この例では、どの記号も「yyyy - mmm - dd / XXhhmm / name」の形をしているので、「-」と「/」は2000度数(1記号に2回*1000記号分)となる。また、メタ記号「H」、「I」、「Q」、「X」、「Z」は図15の記号データ中には1度も表れないことがわかる。この時点で、メタ記号表度数表中の、度数が0である上記の5種のメタ記号が、メタ記号辞書から削除される。図18は、削除後の各メタ記号についての度数表である。しきい値C1が50であり、図18中の全てのメタ記号が文字数が1と、最大文字数1000未満なので、メタ記号辞書更新判定手段106は、図18中の全てのメタ記号について、図18中の各メタ記号を末尾に追加した「0」、「- /」、「- 0」などの2文字のメタ記号を追加し、図19のようなメタ記号を含むメタ記号辞書に更新する。メタ記号辞書の実際の構造は、図31のようなデジタル木のデータ構造(すなわちTRIE)として構築し、保持される。以上で記号辞書作成の第1ステップを終了する。

【0034】

記号辞書作成の第2ステップでは、上述した第1ステップでの各記号の被覆と切り出したメタ記号の度数の集計が、図19のメタ記号辞書を用いて再度行われる。例えば、記号「1998 - JAN - 01 / AM0200 / KAWAYASU」

は、

{ (19, 1, 3), (99, 2, 4), ..., (AS, 25, 27), (SU, 26, 28) }

のように、2文字ずつ被覆される。記号中に出現する可能性のある文字はすべてメタ記号辞書に含まれているので、必ず空でない被覆結果を求めることができる。なお、第2ステップでの被覆処理では、文字数の異なるメタ記号がメタ記号辞書102に混在するため、第1ステップと異なり、図6の流れ図において、ステップ603の判定が「yes」になる場合があり、最長一致で切り出した全てのメタ記号が被覆結果に入るわけではない(上記の例では、末尾の(U, 27, 28)は、(SU, 26, 28)があるので、被覆結果には入らない)。記号データ101中の1000個の記号を処理した後のメタ記号度数表のうち、度数が0でないメタ記号の部分は図20のようになる。これを図18と比べると、例えば図18で2000度数のメタ記号「-」は、図20では「- 0」~「- S」の「-」で始まる12種の2文字のメタ記号と、「8 -」、「B -」、「C -」、「G -」、「L -」、「N -」、「P -」、「R -」、「T -」、「V -」、「Y -」の「-」で終わる11種のメタ記号の、合計23種のメタ記号に度数が分散していることがわかる。「-」で始まる12種の2文字のメタ記号の度数の合計値と、「-」で終わる11種の2文字のメタ記号の度数の合計値は、共にちょうど2000で、記号中の文字「-」は、この文字「-」を共有する「-」で始まるメタ記号と、「-」で終わるメタ記号の2つのメタ記号に含まれることが見て取れる。しきい値C2が100であり、図20中の全てのメタ記号が文字数が2と、最大文字数1000未満なので、メタ記号辞書更新判定手段106は、図20中の「- 0」、「- 1」、「- 2」、「- A」等、度数が100以上の、62種のメタ記号について、図18中の各メタ記号(すなわち、1文字のメタ記号)を末尾に追加した「0」、「- 0 /」、「- 0 0」などの3文字のメタ記号を追加し、メタ辞書を更新する。また同時に、メタ記号辞書更新判定手段106は、上記のように、メタ記号の追加が1つ以上発生しているので、第2ステップを終了せず、引き続き、更新したメタ記号辞書を使って同様な被覆・集計・更新処理を行うという判定を下す。更新したメタ記号辞書102を使って同様な被覆・集計処理を行った後のメタ記号度数表のうち、度数が1以上のメタ記号の部分は図21のようになる。例えば、記号「1998 - JAN - 01 / AM0200 / KAWAYASU」

は、

{ (1 9 9 , 1 , 4) , (9 9 8 , 2 , 5) , (9 8 - , 3 , 6) , (8 - J , 4 , 7) , (- J A , 5 , 8) , (A N - , 7 , 1 0) , (N - 0 , 8 , 1 1) , ... , (Y A S , 2 4 , 2 7) , (A S U , 2 5 , 2 8) }

のように被覆され、文字「J」からの最長一致メタ記号「JA」は被覆結果には含まれないことがわかる。また、図21では、2文字のメタ記号と3文字のメタ記号が混在しており、記号を被覆する際に、出現度数の大きい部分は3文字のメタ記号で、出現度数の比較的少ない部分は2文字のメタ記号を使って被覆が行われていることが確認できる。しきい値C3が100であり、図21中の全てのメタ記号が文字数が3文字以下と、最大文字数1000未満なので、メタ記号辞書更新判定手段106は、図21中の「-01」, 「-02」, 「-03」, 「-04」等、度数が100以上の、42種のメタ記号について、図18中の各メタ記号(すなわち、1文字のメタ記号)を末尾に追加した「01」, 「-01/」, 「-010」などのメタ記号を追加し、メタ辞書を更新する。また同時に、メタ記号辞書更新判定手段106は、上記のように、メタ記号の追加が1つ以上発生しているので、第2ステップを終了せず、引き続き、更新したメタ記号辞書を使って同様な被覆・集計・更新処理を行うという判定を下す。更新したメタ記号辞書102を使って同様な被覆・集計処理を行った後のメタ記号度数表のうち、度数が1以上のメタ記号の部分は図22のようになる。しきい値C4が100であり、図22中の全てのメタ記号が文字数が4文字以下と、最大文字数1000未満なので、メタ記号辞書更新判定手段106は、図22中の「-NOV」, 「/AM0」, 「/AM1」, 「/KAW」等、度数が100以上の、31種のメタ記号について、図18中の各メタ記号(すなわち、1文字のメタ記号)を末尾に追加した「NOV」, 「-NOV/」, 「-NOV0」などのメタ記号を追加し、メタ辞書を更新する。

【0035】

また同時に、メタ記号辞書更新判定手段106は、上記のように、メタ記号の追加が1つ以上発生しているので、第2ステップを終了せず、引き続き、更新したメタ記号辞書を使って同様な被覆・集計・更新処理を行うという判定を下す。更新したメタ記号辞書102を使って同様な被覆・集計処理を行った後のメタ記号度数表のうち、度数が1以上のメタ記号の部分は図23のようになる。図23を図22と比較すると、度数が1以上のメタ記号の種類は、メタ記号を追加したにも関わらず、2種類減少しており、文字数の多いメタ記号による極大単語切り出しによって、より文字数の少ないメタ記号が切り出し結果から「締め出されていく」ことが確認できる。しきい値C5が100であり、図23中の全てのメタ記号が文字数が5文字以下と、最大文字数1000未満なので、メタ記号辞書更新判定手段106は、図23中の「-NOV-」, 「/KAWA」, 「/SUDA」, 「/SUKA」等、度数が100以上の、20種のメタ記号について、図18中の各メタ記号(すなわち、1文字のメタ記号)を末尾に追加した「NOV」, 「-NOV-」, 「-NOV-0」などのメタ記号を追加し、メタ記号辞書を更新する。

【0036】

また同時に、メタ記号辞書更新判定手段106は、上記のように、メタ記号の追加が1つ以上発生しているので、第2ステップを終了せず、引き続き、更新したメタ記号辞書を使って同様な被覆・集計・更新処理を行うという判定を下す。更新したメタ記号辞書102を使って同様な被覆・集計処理を行った後のメタ記号度数表のうち、度数が1以上のメタ記号の部分は図24のようになる。図24を図22と比較すると、度数が1以上のメタ記号の種類はさらに減少しており、文字数の多いメタ記号による極大単語切り出しによって、より文字数の少ないメタ記号が切り出し結果から「締め出されていく」ことが確認できる。しきい値C6が100であり、図24中の全てのメタ記号が文字数が6文字以下と、最大文字数1000未満なので、メタ記号辞書更新判定手段106は、図24中の度数が100以上の、16種のメタ記号について、図18中の各メタ記号(すなわち、1文字のメタ記号)を末尾に追加したメタ記号を追加し、メタ記号辞書を更新する。

【0037】

10

20

30

40

50

また同時に、メタ記号辞書更新判定手段106は、上記のように、メタ記号の追加が1つ以上発生しているので、第2ステップを終了せず、引き続き、更新したメタ記号辞書を使って同様な被覆・集計・更新処理を行うという判定を下す。更新したメタ記号辞書102を使って同様な被覆・集計処理を行った後のメタ記号度数表のうち、度数が1以上のメタ記号の部分は図25のようになる。しきい値C7が100であり、図25中の全てのメタ記号が文字数が7文字以下と、最大文字数1000未満なので、メタ記号辞書更新判定手段106は、図25中の度数が100以上の、9種のメタ記号について、図18中の各メタ記号(すなわち、1文字のメタ記号)を末尾に追加したメタ記号を追加しようとするが、「/SUDA」、「/SUKAWA」、「/SUWA」、「98-NOV」、「WADA」の5種のメタ記号については、すでにメタ記号辞書中に末尾に1文字を追加したメタ記号がすべて含まれているので、残りの4種のメタ記号について追加処理を行い、メタ記号辞書を更新する。また同時に、メタ記号辞書更新判定手段106は、上記のように、メタ記号の追加が1つ以上発生しているので、第2ステップを終了せず、引き続き、更新したメタ記号辞書を使って同様な被覆・集計・更新処理を行うという判定を下す。なお、図25と図24とを比較すると、メタ記号「SUKAWA」の度数が187から81と減少している。これは図25のメタ記号「/SUKAWA」の存在で、「1998.../SUKAWA」という記号の「SUKAWA」が被覆結果から除かれ、「1998...YASUKAWA」という記号の「SUKAWA」のみの度数になったことによることが確認できる。更新したメタ記号辞書102を使って同様な被覆・集計処理を行った後のメタ記号度数表のうち、度数が1以上のメタ記号の部分は図26のようになる。しきい値C8が100であり、図26中の全てのメタ記号が文字数が8文字以下と、最大文字数1000未満なので、メタ記号辞書更新判定手段106は、図26中の度数が100以上の、6種のメタ記号について、図18中の各メタ記号(すなわち、1文字のメタ記号)を末尾に追加したメタ記号を追加しようとするが、「/SUDA」、「/SUKAWA」、「/SUWA」、「998-NOV」、「WADA」の5種のメタ記号については、すでにメタ記号辞書中に末尾に1文字を追加したメタ記号がすべて含まれているので、残りのメタ記号「1998-NOV」について追加処理を行い、メタ記号辞書を更新する。

【0038】

また同時に、メタ記号辞書更新判定手段106は、上記のように、メタ記号の追加が1つ以上発生しているので、第2ステップを終了せず、引き続き、更新したメタ記号辞書を使って同様な被覆・集計・更新処理を行うという判定を下す。更新したメタ記号辞書102を使って同様な被覆・集計処理を行った後のメタ記号度数表のうち、度数が1以上のメタ記号の部分は図27のようになる。しきい値C9が100であり、図27中の全てのメタ記号が文字数が9文字以下と、最大文字数1000未満なので、メタ記号辞書更新判定手段106は、図27中の度数が100以上の、5種のメタ記号について、図18中の各メタ記号(すなわち、1文字のメタ記号)を末尾に追加したメタ記号を追加しようとするが、「/SUDA」、「/SUKAWA」、「/SUWA」、「WADA」の4種のメタ記号については、すでにメタ記号辞書中に末尾に1文字を追加したメタ記号がすべて含まれているので、残りのメタ記号「1998-NOV-」について追加処理を行い、メタ記号辞書を更新する。また同時に、メタ記号辞書更新判定手段106は、上記のように、メタ記号の追加が1つ以上発生しているので、第2ステップを終了せず、引き続き、更新したメタ記号辞書を使って同様な被覆・集計・更新処理を行うという判定を下す。更新したメタ記号辞書102を使って同様な被覆・集計処理を行った後のメタ記号度数表のうち、度数が1以上のメタ記号の部分は図28のようになる。しきい値C10が100であり、図28中の全てのメタ記号が文字数が10文字以下と、最大文字数1000未満なので、メタ記号辞書更新判定手段106は、図28中の度数が100以上の、4種のメタ記号について、図18中の各メタ記号(すなわち、1文字のメタ記号)を末尾に追加したメタ記号を追加しようとするが、「/SUDA」、「/SUKAWA」、「/SUWA」、「WADA」の4種のメタ記号は、すでにメタ記号辞書中に末尾に1文字を追加したメタ記号がすべて含まれているので、追加処理は行われず、メタ記号辞書は更新されない。メタ記号

10

20

30

40

50

辞書更新判定手段106は、上記のように、メタ記号の追加が発生していないので、第2ステップを終了する。

【0039】

記号辞書作成の第3ステップでは、メタ記号辞書参照更新手段106が、メタ記号度数表105を参照し、しきい値E(すなわち5)未満の度数を持つ、2文字以上のメタ記号を、メタ記号辞書102から削除する。図28を見ると、度数が1以上のメタ記号のうち、度数が5未満のものはなく、この場合には、度数が0である、2文字以上のメタ記号がすべて削除され、結果的にメタ記号辞書102の内容は、図28のメタ記号に、図18のメタ記号を加えたものになる。メタ記号辞書の実際の構造は、図32のようなデジタル木のデータ構造(すなわちT R I E)として構築し、保持される。以上で記号辞書作成の第3ステップを終了する。

10

【0040】

記号辞書作成の第4ステップでは、記号被覆手段103が第3ステップで算出した図32のメタ記号辞書102を用いて記号データ101中の各記号データを被覆して被覆結果を求め、メタ記号出現情報作成手段108が、被覆結果から各メタ記号が出現する記号の番号と、出現文字位置とを記録したメタ記号出現情報108を作成する。この場合には、図29のようなメタ記号出現情報が作成される。ただし、図29では、解釈を容易にするため、記号の番号の代わりに記号の文字列を用いている。この作成処理は、本質的にはいわゆる転置処理(inversion)であり、情報検索システムで通常用いられる手法によって、効率よく行うことができる。照合文字位置は、照合部分の左側の文字数(図29の「左文字数」と、照合部分の右側の文字数(図29の「右文字数」)によって表されている。図29のような内容が、メタ記号ごとにまとめた表として記録され、メタ記号と、照合文字位置を鍵にして検索することで、指定したメタ記号を指定した文字位置に含む記号の番号の列(集合)を効率よく求めることができる。以上で記号辞書作成の第4ステップを終了する。

20

【0041】

記号辞書作成の第5ステップでは、記号辞書作成手段109が、メタ記号辞書102と、メタ記号出現情報108とから、機械検索可能な記号辞書110を作成する。その際、メタ記号出現情報108は、図29のような内容の表をそのまま記号辞書110内に格納するが、メタ記号辞書102については、図32のようなT R I E構造の情報の他に、図33のようなメタ記号の延長情報を付加したメタ記号延長表を、メタ記号情報として、記号辞書110に格納する。図33のメタ記号延長表は、メタ記号辞書中の各メタ記号M毎に、Mを文字列として含んだメタ記号辞書中のメタ記号と、左右の延長部分の文字数の3つ組を記録した表で、例えばメタ記号「-」の延長情報は、

30

{(-, 0, 0), (-01, 0, 2), ..., (-29, 0, 2), (-3, 0, 1),
...,
(R-0, 1, 1), (R-3, 1, 1)}

のようになる。このメタ記号の延長表の作成も、先のメタ記号出現情報108の作成処理と同様にして行うことができる。以上で記号辞書作成の第4ステップが終了し、記号辞書110が作成され、記号辞書作成が完了する。

40

以上の説明のように、本発明の第1の実施の形態の記号辞書作成方法によれば、記号データ中で高頻度で出現する部分文字列については、より多くの文字数を持つメタ記号を持ったメタ記号辞書を作成し、このメタ記号辞書を使って記号の被覆情報を記録するので、より少ない情報量で記号辞書が作成でき、記号辞書検索の際にも、高頻度で出現する部分文字列を含んだ記号の検索が、従来の記号辞書の検索に比べて高速に行えるようになる。さらに、このメタ記号辞書の作成は、しきい値を設定すれば機械的に実行可能なので、人手をかけずに、記号データの文字列分布の偏りに適応した、最適な記号辞書を作成することが可能になる。

【0042】

(実施の形態2)

50

以下、本発明の第2の形態について、図面を参照しながら説明する。図2は、本発明の請求項8にかかる、記号辞書検索装置の全体構成を表わすブロック図である。図2において、201はメタ記号情報とメタ記号出現情報とを格納した記号辞書、202は検索条件である文字列を入力する検索条件入力手段、203は、検索条件入力手段202から入力された検索条件である質問文字列を、記号辞書201を使い、重複最長一致単語切り出し法によって被覆し、被覆結果を求める質問文字列被覆手段、204は質問文字列被覆手段203の求めた被覆結果、205は被覆結果204と、記号辞書201のメタ記号出現情報とから、質問文字列と完全一致する、すなわち、質問文字列をまったく同一な記号の番号を算定する記号番号算定手段、206は記号番号算定手段205の算定した記号の番号などを出力する検索結果出力手段である。

10

【0043】

以上の構成の記号辞書検索装置において、その動作を、実施の形態1で例にあげた記号辞書と、簡単な検索条件を例に、図面を参照して説明する。図7は質問文字列被覆手段203における被覆結果を求める処理の手順を記述した流れ図、図8は記号番号算定手段205における記号番号の算定処理の手順を記述した流れ図、図34は、検索条件として、質問文字列「1998 - NOV - 01 / PM1030 / KAWAYASU」と完全一致する記号の番号を求めよ、という条件を与えた場合の記号辞書検索の過程での主要な中間データを記述した概念図である。

【0044】

まず、記号辞書検索の第1ステップでは、質問文字列被覆手段203が、記号辞書201中のメタ記号情報を検索し、質問文字列「1998 - NOV - 01 / PM1030 / KAWAYASU」に対して、重複最長一致単語切り出し法によって被覆を求め、図34の*STEP1のCの被覆結果を得る。「重複最長一致単語切り出し法」とは、被覆対象文字列の左側から最長一致するメタ記号を求めていくが、メタ記号同士の部分的重複を許し、かつ、あるメタ記号Aの照合文字区間が、1つ以上のほかのメタ記号群B, ..., Xの照合文字区間の和からなる区間に完全に含まれる場合には、そのメタ記号Aは被覆要素として記録しない、という被覆方法である。具体的には、図7のステップ702で、直前の切り出し結果より末尾側で、「間があかない」ような照合文字区間をもつメタ記号のうちで、最も末尾側までカバーするメタ記号の集合Hをまず求め、Hの中から、照合開始位置がもっとも先頭側のもの、すなわち最も文字数の多いメタ記号を被覆要素として用い、この被覆要素の照合文字区間を基準にして、次の末尾側の被覆要素をさらに決定する、という、先頭から末尾への一連の切り出し処理で、極大単語切り出し法で得られる被覆結果の部分集合が得られる被覆方法である。この質問文字列の場合には被覆結果204が空でないので記号番号算定手段205での処理が開始されるが、被覆結果なしの場合には、ただちに処理を停止し、検索結果なしとする。以降の処理は、図8に沿って進められる。まずステップ801で、被覆結果中で、照合開始文字位置sが1である要素(C中で高々1つ)を探す。この例では(1998 - NOV - 0, 1, 11)が見つかる。続いて記号辞書中のメタ記号出現情報の中で、M = 「1998 - NOV - 0」の出現記号情報で(X, 0, n - e + 1)の形のもの(nは質問文字列の文字数で、この例では14)を残らず探し、その記号Xの記号番号の集合をAとして記録する。なお、図34では、読みやすさを考慮して、記号番号でなく、記号文字列を用いて集合A, Bを記述している。この例では1998 - NOV - 01 / AM0830 / NODA等の記号の記号番号が求まる。Aが求まった後、Cから今選択した要素(1998 - NOV - 0, 1, 11)を削除する。Cは、図34の*STEP2のCのようになり、図8のステップ802の「Cが空集合か?」という条件判定は「no」となって、ステップ803に進む。ステップ803では、この例の場合には、Cの先頭の要素(1/P, 11, 14)を選択し、Bとして、「1998 - JAN - 01 / PM065 / NODA」などを含む、図34の*STEP3のBのような記号(に対応する記号の番号)を求める。その後、AとBとの共通部分を求めて、Aに格納する。すなわちAの内容を、Bにあるものだけに絞り込む。この例の場合には、Aの内容は、4つの記号(の番号)に絞り込まれる。続いて、図8のステップ804で判定

20

30

40

50

処理が行われ、Aが空ではないので、ステップ805に進み、Cからステップ803で選択した要素(1/P, 11, 14)を削除し、ステップ802に戻る。以降、*STEP4~*STEP6まで、同様にして次々とCから要素を選択して、メタ記号出現情報からBを求め、中間結果Aを絞り込んでいく。その間、AもCも空にはならないので、処理が途中で終了することはない。

【0045】

最後に、*STEP7の処理が終了した、図8のステップ805の終了時点で、Cが空となって、ステップ802の判定が「yes」になって、記号番号算定手段205における処理が終了し、Aの要素である、記号「1998-NOV-01/PM1030/KAWAYASU」の番号が、検索結果表示手段206に出力され、記号検索処理が終了する。

10

【0046】

以上の説明のように、本発明の第2の実施の形態の記号辞書検索方法によれば、記号データ中で高頻度で出現する部分文字列については、より多くの文字数を持つメタ記号を持ったメタ記号情報が作成されており、このメタ記号情報を使って、質問文字列から一旦被覆結果を構成し、この被覆結果と、メタ記号出現情報とを使って検索処理を行うので、高頻度で出現する部分文字列を含んだ記号の検索時でも、従来の記号辞書の検索に比べて高速に行えるようになる。

【0047】

(実施の形態3)

以下、本発明の第3の形態について、図面を参照しながら説明する。図3は、本発明の請求項9にかかる、記号辞書検索装置の全体構成を表わすブロック図である。図3において、301はメタ記号情報とメタ記号出現情報とを格納した記号辞書、302は検索条件である文字列を入力する検索条件入力手段、303は、検索条件入力手段302から入力された検索条件である質問文字列を、記号辞書301を使い、極大単語切り出し法によって被覆し、被覆結果を求める質問文字列被覆手段、304は質問文字列被覆手段303の求めた被覆結果、305は被覆結果304と、記号辞書301のメタ記号出現情報とから、質問文字列と前方一致する、すなわち、質問文字列を先頭部分に含む記号の番号の集合を算定する記号番号集合算定手段、307は、記号辞書301のメタ記号情報を検索して、被覆結果304のうちで、最も照合開始文字位置の大きい被覆要素のメタ記号Zの延長メタ記号(すなわち、Zを含むメタ記号)のうちで、質問文字列の最右部分文字列Rの右延長メタ記号(すなわちRを先頭部分に含むメタ記号)になっているメタ記号の番号と照合位置との組を、残らず求めて被覆結果304に追加格納する右延長メタ記号算定手段、306は記号番号集合算定手段305の算定した記号の番号などを出力する検索結果出力手段である。なお、図3の構成要素301乃至304は、第2の実施の形態のブロック図である図2の構成要素201乃至204と対応する。

20

30

【0048】

以上の構成の記号辞書検索装置において、その動作を、実施の形態1で例にあげた記号辞書と、簡単な検索条件を例に、図面を参照して説明する。図7は質問文字列被覆手段303における被覆結果を求める処理の手順を記述した流れ図、図9、図10は記号番号集合算定手段305における記号番号集合の算定処理の手順を記述した流れ図、図35、図36は、検索条件として、質問文字列「1998-NOV-01/PM」と前方一致する記号の番号の集合を求めよ、という条件を与えた場合の記号辞書検索の過程での主要な中間データを記述した概念図である。

40

【0049】

まず、記号辞書検索の第1ステップでは、質問文字列被覆手段303が、記号辞書301中のメタ記号情報を検索し、質問文字列「1998-NOV-01/PM」に対して、重複最長一致単語切り出し法によって被覆を求め、図35の*STEP1のCの被覆結果を得る。被覆処理の手順は、実施の形態2で示した記号の被覆処理の手順と同一である。この質問文字列の場合には被覆結果304が空でないで右延長メタ記号算定手段307での処理が開始されるが、被覆結果なしの場合には、ただちに処理を停止し、検索結果なし

50

とする。次に、右延長メタ記号算定手段307が、記号辞書301のメタ記号情報を検索して、被覆結果304のうち、最も照合開始位置の大きい被覆要素のメタ記号Zに対して、その延長メタ記号(すなわちZを含む文字列のメタ記号)を求め、求めた延長メタ記号のうちで、質問文字列の第j最右部分文字列R(すなわち質問文字列の第j文字から最終文字までの部分文字列)の右延長メタ記号(すなわち文字列Rを先頭部分に含むメタ記号)になっているメタ記号Xだけを選別し、

(X, j, |R| + j)

を被覆結果304に追加する。この例の場合には、Z = 「M」となり、その延長メタ記号は、「/AM01」, ... , 「/AM12」, 「/PM01」, 「/PM02」, ... , 「/PM12」, 「1998 - MAR」, 「1998 - MAY」

10

の26種が求まる。このうちで、質問文字列「1998 - NOV - 01 / PM」の最右部分列の右延長メタ記号にもなっている、

「/PM01」, 「/PM02」, ... , 「/PM12」

の12種のメタ記号が、右延長メタ記号算定手段307によって、被覆結果304に追加される。図35の*STEP2が、この様子を示している。こうして、右延長メタ記号まで含めた被覆が求まったならば、記号番号集合算定手段305が、記号番号集合を求め、以降の処理は、図9, 図10に沿って進められる。まずステップ901で、被覆結果中で、照合開始文字位置sが1である要素からなる集合Dを求める。この例ではD = { (1998 - NOV - 0, 1, 11) }となる。また、最終結果の集合SCを空に初期化する。

20

Dが空でないので、ステップ903に進み、Dからただ1つの要素(1998 - NOV - 0, 1, 11)を選択し、記号辞書301中のメタ記号出現情報の中で、M = 「1998 - NOV - 0」の出現記号情報で(X, 0, *)の形のものを残らず探し、その記号Xの記号番号の集合をAとして記録する。ただし、「*」は、任意の値(don't care)を表す。なお、図35, 図36では、読みやすさを考慮して、記号番号でなく、記号文字列を用いてA, C, D等の集合を記述している。この例では*STEP4で、1998 - NOV - 01 / AM0830 / NODA等の記号の記号番号が求まる。Aが求まった後、Dから今選択した要素(1998 - NOV - 0, 1, 11)を削除する。ステップ904の「Aが空集合か?」という条件判定は「no」となって、ステップ905に進む。ステップ905では、qが質問文字列の文字数nより大きいかが判定し、大きければ最終結果の集合SCに、この時点での集合Aの要素を加え、大きくなければ、図10の手続き select_cover1(A, p, q) を呼び出す。この場合には、n = 14, q = 11なので q < nであり、ステップ907に移って、図10の手続き select_cover1(A, p, q) を呼び出して、処理を進める。図10のステップ908では、手続きの引数である文字位置p, qに対して、照合開始文字位置sがpより大きく、q以下であるような被覆結果C中の要素からなる集合Dpを求める。

30

【0050】

この例の場合には、p = 1, q = 11であり、Cから1 < s < 11である要素を求めると、図35の*STEP5のように、D1 = { (1/P, 11, 14) }となる。D1が空ではないので、ステップ909は「no」となり、ステップ910に進む。D1から、最初の要素(1/P, 11, 14)を選び、メタ記号「1/P」の出現メタ記号情報から、(X, 10, *)の形の要素をもれなく探して、Aとの共通部分をとって、中間結果Aを絞り込んで、結果をA1に格納する。ただし、「*」は、任意の値を表す。この例の場合には、図35の*STEP6のように、Aが3要素にまで絞り込まれる。さらに、D1から、今選択したD1の要素(1/P, 11, 14)を削除する。A1が空でないので、ステップ911の判定は「no」となり、ステップ912に進む。n = 14に対して、u = 12であり、ステップ912の判定は「no」となる。

40

【0051】

ステップ914では、A1, t = 11, u = 14を引数として、図10の手続き select_cover1を再帰的に呼び出し、処理を続け、中間結果であるApを図35, 図36のように絞り込んでいく。図36の*STEP18では、u = 17と、n = 14以

50

上なので、A 1 1 が最終結果 S C の一部として記録され、別の結果を探するため、再び検索処理が続行される。このようにして、被覆結果 3 0 4 から、系統的に被覆要素の組み合わせを生成しつつ、記号辞書 3 0 1 中のメタ記号出現情報を検索して、生成した被覆要素の組に共通して含まれる記号の番号の集合を求め、最終結果の集合 S C に記録していく。すべての被覆要素の組み合わせを処理した * S T E P 2 0 の時点で処理を終了し、その時の S C が、検索結果となる。

【 0 0 5 2 】

以上の説明のように、本発明の第 3 の実施の形態の記号辞書検索方法によれば、記号データ中で高頻度で出現する部分文字列については、より多くの文字数を持つメタ記号を持ったメタ記号情報が作成されており、このメタ記号情報を使って、質問文字列から一旦被覆結果を構成し、この被覆結果と、右延長メタ記号算定手段が追加した要素を加えた被覆結果から、メタ記号出現情報とを使って検索処理を行うので、高頻度で出現する部分文字列を含んだ記号の前方一致検索時でも、従来の記号辞書の検索に比べて高速に行えるようになる。

【 0 0 5 3 】

(実施の形態 4)

以下、本発明の第 4 の形態について、図面を参照しながら説明する。図 4 は、本発明の請求項 1 0 にかかる、記号辞書検索装置の全体構成を表わすブロック図である。図 4 において、4 0 1 はメタ記号情報とメタ記号出現情報とを格納した記号辞書、4 0 2 は検索条件である文字列を入力する検索条件入力手段、4 0 3 は、検索条件入力手段 4 0 2 から入力された検索条件である質問文字列を、記号辞書 4 0 1 を使い、重複最長一致単語切り出し法によって被覆し、被覆結果を求める質問文字列被覆手段、4 0 4 は質問文字列被覆手段 4 0 3 の求めた被覆結果、4 0 8 は、記号辞書 4 0 1 のメタ記号情報を検索して、被覆結果 4 0 4 のうちで照合開始文字位置が 1 である被覆要素のメタ記号 Z の延長メタ記号 (すなわち、Z を含むメタ記号) のうちで、質問文字列の最左部分文字列 L の左延長メタ記号 (すなわち L を末尾部分に含むメタ記号) になっているメタ記号の番号と照合位置との組を、残らず求めて被覆結果 4 0 4 に追加格納する左延長メタ記号算定手段、4 0 5 は被覆結果 4 0 4 と、記号辞書 4 0 1 のメタ記号出現情報とから、質問文字列と後方一致する、すなわち、質問文字列を末尾部分に含む記号の番号の集合を算定する記号番号集合算定手段、4 0 6 は記号番号集合算定手段 4 0 5 の算定した記号の番号などを出力する検索結果出力手段である。なお、図 4 の構成要素 4 0 1 乃至 4 0 4 は、第 3 の実施の形態のブロック図である図 3 の構成要素 3 0 1 乃至 3 0 4 と対応する。以上の構成の記号辞書検索装置において、その動作を、実施の形態 1 で例にあげた記号辞書と、簡単な検索条件を例に、図面を参照して説明する。

【 0 0 5 4 】

図 7 は質問文字列被覆手段 4 0 3 における被覆結果を求める処理の手順を記述した流れ図、図 1 1 , 図 1 2 は記号番号集合算定手段 4 0 5 における記号番号集合の算定処理の手順を記述した流れ図、図 3 7 は、検索条件として、質問文字列「K A W A」と後方一致する記号の番号の集合を求めよ、という条件を与えた場合の記号辞書検索の過程での主要な中間データを記述した概念図である。まず、記号辞書検索の第 1 ステップでは、質問文字列被覆手段 4 0 3 が、記号辞書 4 0 1 中のメタ記号情報を検索し、質問文字列「1 K A W A」に対して、重複最長一致単語切り出し法によって被覆を求め、図 3 7 の * S T E P 1 の C の被覆結果を得る。被覆処理の手順は、実施の形態 3 で示した記号の被覆処理の手順と同一である。この質問文字列の場合には被覆結果 4 0 4 が空でないので左延長メタ記号算定手段 4 0 8 での処理が開始されるが、被覆結果なしの場合には、ただちに処理を停止し、検索結果なしとする。次に、左延長メタ記号算定手段 4 0 8 が、記号辞書 4 0 1 のメタ記号情報を検索して、被覆結果 4 0 4 のうち、照合開始位置が 1 である被覆要素のメタ記号 Z に対して、その延長メタ記号 (すなわち Z を含む文字列のメタ記号) を求める。求めた延長メタ記号のうちで、質問文字列の第 j 最左部分文字列 L (すなわち質問文字列の第 1 文字から第 j 文字までの部分文字列) の左延長メタ記号 (すなわち文字列 L を末尾部分

10

20

30

40

50

に含むメタ記号) になっているメタ記号 X だけを選別し、

$(X, j + 1 - |L|, j + 1)$

を被覆結果 4 0 4 に追加する。この例の場合には、 $Z = \text{「KAWA」}$ となり、その延長メタ記号は、

「 / S U K A W A 」 , 「 0 / K A W A D 」 , 「 0 / K A W A N 」 , 「 0 / K A W A Y 」 ,
「 5 / K A W A D 」 , 「 5 / K A W A N 」 , 「 5 / K A W A Y 」 ,
「 K A W A 」 , 「 S U K A W A 」

の 9 種が求まる。このうちで、質問文字列「 K A W A 」の最左部分列の左延長メタ記号にもなっている、

「 / S U K A W A 」 , 「 / K A W A 」

の 2 種のメタ記号が、左延長メタ記号算定手段 4 0 8 によって、被覆結果 4 0 4 に追加される。図 3 7 の * S T E P 2 が、この様子を示している。こうして、左延長メタ記号まで含めた被覆が求まったならば、記号番号集合算定手段 4 0 5 が、記号番号集合を求める。以降の処理は、図 1 1 , 図 1 2 に沿って進められる。まずステップ 1 0 0 1 で、被覆結果中で、照合終了文字位置 e が n である要素からなる集合 D を求める。この例では

$D = \{ (KAWA, 1, 5), (/SUKAWA, -2, 5), (SUKAWA, -1, 5) \}$

となる。また、最終結果の集合 S C を空に初期化する。D が空でないので、ステップ 1 0 0 3 に進み、D から要素 (K A W A , 1 , 5) を選択し、記号辞書 4 0 1 中のメタ記号出現情報の中で、 $M = \text{「KAWA」}$ の出現記号情報で $(X, *, 0)$ の形のものを残らず探し、その記号 X の記号番号の集合を A として記録する。なお、図 3 7 では、読みやすさを考慮して、記号番号でなく、記号文字列を用いて A , C , D 等を記述している。この例では * S T E P 4 で、 1 9 9 8 - J A N - 1 7 / P M 0 9 3 0 / N O K A W A 等の記号の記号番号が求まる。

【 0 0 5 5 】

A が求まった後、D から今選択した要素 (K A W A , 1 , 5) を削除する。ステップ 1 0 0 4 の「 A が空集合か? 」という条件判定は「 n o 」となって、ステップ 1 0 0 5 に進む。ステップ 1 0 0 5 では、選択した被覆要素の照合開始位置 t が 1 以下であるか否かを調べ、1 以下ならば、最終結果の集合 S C に、この時点での集合 A の要素を加え、2 以上ならば、図 1 2 の手続き `select_cover2(A, p, q)` を呼び出す。この場合には、 $t = 1$ なので、最終結果の集合 S C に、この時点での集合 A の要素を加え、別の結果を探すため、再び検索処理が実行される。このようにして、被覆結果 4 0 4 から、系統的に被覆要素の組み合わせを生成しつつ、記号辞書 4 0 1 中のメタ記号出現情報を検索して、生成した被覆要素の組に共通して含まれる記号の番号の集合を求め、最終結果の集合 S C に記録していく。すべての被覆要素の組み合わせを処理した * S T E P 6 の時点で処理を終了し、その時の S C が、検索結果となる。以上の説明のように、本発明の第 4 の実施の形態の記号辞書検索方法によれば、記号データ中で高頻度で出現する部分文字列については、より多くの文字数を持つメタ記号を持ったメタ記号情報が作成されており、このメタ記号情報を使って、質問文字列から一旦被覆結果を構成し、この被覆結果と、左延長メタ記号算定手段が追加した要素を加えた被覆結果から、メタ記号出現情報とを使って検索処理を行うので、高頻度で出現する部分文字列を含んだ記号の後方一致検索時でも、従来の記号辞書の検索に比べて高速に行えるようになる。

【 0 0 5 6 】

(実施の形態 5)

以下、本発明の第 5 の形態について、図面を参照しながら説明する。図 5 は、本発明の請求項 1 1 にかかる、記号辞書検索装置の全体構成を表わすブロック図である。図 5 において、5 0 1 はメタ記号情報とメタ記号出現情報とを格納した記号辞書、5 0 2 は検索条件である文字列を入力する検索条件入力手段、5 0 3 は、検索条件入力手段 5 0 2 から入力された検索条件である質問文字列を、記号辞書 5 0 1 を使い、重複最長一致単語切り出し法によって被覆し、被覆結果を求める質問文字列被覆手段、5 0 4 は質問文字列被覆手段

10

20

30

40

50

503の求めた被覆結果、507は、記号辞書501のメタ記号情報を検索して、被覆結果504のうちで、最も照合開始文字位置の大きい被覆要素のメタ記号Zの延長メタ記号(すなわち、Zを含むメタ記号)のうちで、質問文字列の最右部分文字列Rの右延長メタ記号(すなわちRを先頭部分に含むメタ記号)になっているメタ記号の番号と照合位置との組を、残らず求めて被覆結果504に追加格納する右延長メタ記号算定手段、508は、記号辞書501のメタ記号情報を検索して、被覆結果504のうちで照合開始文字位置が1である被覆要素のメタ記号Zの延長メタ記号(すなわち、Zを含むメタ記号)のうちで、質問文字列の最左部分文字列Lの左延長メタ記号(すなわちLを末尾部分に含むメタ記号)になっているメタ記号の番号と照合位置との組を、残らず求めて被覆結果504に追加格納する左延長メタ記号算定手段、509は記号辞書501のメタ記号情報を検索して、質問文字列の両延長メタ記号(すなわち質問文字列Qを第j文字から第j+|Q|文字の部分(1<j)に含むメタ記号)Xを全て検索して、要素(X, 1-j, 1-j+|X|)を被覆結果504に追加格納する両延長メタ記号算定手段、505は被覆結果504と、記号辞書501のメタ記号出現情報とから、質問文字列と中間一致する、すなわち、質問文字列を含む記号の番号の集合を算定する記号番号集合算定手段、506は記号番号集合算定手段505の算定した記号の番号などを出力する検索結果出力手段である。なお、図5の構成要素501乃至504および506は、第2の実施の形態のブロック図である図2の構成要素201乃至204および206と対応し、図5の構成要素507は第3の実施の形態のブロック図である図3の構成要素307と対応し、図5の構成要素508は第4の実施の形態のブロック図である図4の構成要素408と対応する。以上の構成の記号辞書検索装置において、その動作を、実施の形態1で例にあげた記号辞書と、簡単な検索条件を例に、図面を参照して説明する。

【0057】

図7は質問文字列被覆手段503における被覆結果を求める処理の手順を記述した流れ図、図13、図14は記号番号集合算定手段505における記号番号集合の算定処理の手順を記述した流れ図、図38、図39は、検索条件として、質問文字列「KAWADA」と中間一致する記号の番号の集合を求めよ、という条件を与えた場合の記号辞書検索の過程での主要な中間データを記述した概念図である。まず、記号辞書検索の第1ステップでは、質問文字列被覆手段503が、記号辞書501中のメタ記号情報を検索し、質問文字列「KAWADA」に対して、重複最長一致単語切り出し法によって被覆を求め、図38の*STEP1のCの被覆結果を得る。被覆処理の手順は、実施の形態3で示した記号の被覆処理の手順と同一である。この質問文字列の場合には被覆結果504が空でないので右延長メタ記号算定手段507での処理が開始されるが、被覆結果なしの場合には、ただちに処理を停止し、検索結果なしとする。次に、右延長メタ記号算定手段507が、記号辞書501のメタ記号情報を検索して、被覆結果504のうち、最も照合開始位置の大きい被覆要素のメタ記号Zに対して、その延長メタ記号(すなわちZを含む文字列のメタ記号)を求める。求めた延長メタ記号のうちで、質問文字列の第j最右部分文字列R(すなわち質問文字列の第j文字から最終文字までの部分文字列)の右延長メタ記号(すなわち文字列Rを先頭部分に含むメタ記号)になっているメタ記号Xだけを選別し、

(X, j, |R|+j)

を被覆結果504に追加する。この例の場合には、Z=「WADA」となり、その延長メタ記号として「WADA」ただ1種が求まる。これは、質問文字列「KAWADA」の最右部分列の右延長メタ記号にもなっているので、右延長メタ記号算定手段507は、被覆結果504に追加するが、すでに被覆結果504に同一の被覆要素が含まれているため被覆結果504は変化しない。図38の*STEP2が、この様子を示している。次に、左延長メタ記号算定手段508が、記号辞書501のメタ記号情報を検索して、被覆結果504のうち、照合開始位置が1である被覆要素のメタ記号Zに対して、その延長メタ記号(すなわちZを含む文字列のメタ記号)を求める。求めた延長メタ記号のうちで、質問文字列の第j最左部分文字列L(すなわち質問文字列の第1文字から第j文字までの部分文字列)の左延長メタ記号(すなわち文字列Lを末尾部分に含むメタ記号)になっているメ

タ記号 X だけを選別し、

$(X, j + 1 - |L|, j + 1)$

を被覆結果 504 に追加する。この例の場合には、 $Z = \text{「KAWA」}$ となり、その延長メタ記号は、

「 / S U K A W A 」 , 「 0 / K A W A D 」 , 「 0 / K A W A N 」 , 「 0 / K A W A Y 」 , 「 5 / K A W A D 」 ,

「 5 / K A W A N 」 , 「 5 / K A W A Y 」 , 「 K A W A 」 , 「 S U K A W A 」

の 9 種が求まる。このうちで、質問文字列「 K A W A D A 」の最左部分列の左延長メタ記号にもなっている、

「 / S U K A W A 」 , 「 0 / K A W A D 」 , 「 5 / K A W A D 」 , 「 K A W A 」 , 「 S U K A W A 」 10

の 5 種のメタ記号が、左延長メタ記号算定手段 508 によって、被覆結果 504 に追加される。図 38 の * S T E P 3 が、この様子を示している。次に、両延長メタ記号算定手段 509 が、記号辞書 501 のメタ記号情報を検索して、質問文字列「 K A W A D A 」の両延長メタ記号 (すなわち文字列「 K A W A D A 」を第 j 文字から第 j + 6 文字の部分 ($1 < j$) に含むメタ記号) X を全て検索して、要素 $(X, 1 - j, 1 - j + |X|)$ を被覆結果 504 に追加する。この例の場合には、「 K A W A D A 」を含むメタ記号は、記号辞書 501 のメタ記号情報中にはなく、したがって被覆結果 504 には何も追加されない。

図 42 の * S T E P 4 が、この様子を示している。こうして、右延長メタ記号、左延長メタ記号、両延長記号まで含めた被覆が求まったならば、記号番号集合算定手段 505 が、記号番号集合を求める。以降の処理は、図 13, 図 14 に沿って進められる。まずステップ 1101 で、被覆結果中で、照合開始文字位置 s が 1 以下である要素からなる集合 D を求める。この例では

$D = \{ (KAWA, 1, 5), (/SUKAWA, -1, 5), (0/KAWAD, -1, 6), (5/KAWAD, -1, 6), (SUKAWA, -1, 5) \}$

となる。また、最終結果の集合 SC を空に初期化する。

【 0058 】

図 38 の * S T E P 5 がこの様子を示している。D が空でないので、ステップ 1103 に進み、D から要素 $(KAWA, 1,)$ を選択し、記号辞書 501 中のメタ記号出現情報の中で、 $M = \text{「KAWA」}$ の各出現記号情報 (X, L, R) に対して、 (X, L) の形の要素の集合を求めて A として記録する。なお、図 38, 図 39 では、読みやすさを考慮して、記号番号でなく、記号文字列を用いて A, C, D 等を記述している。この例では * S T E P 6 で、 $(1998 - JAN - 17 / PM0930 / NOKAWA, 21)$ 等の要素が求まる。A が求まった後、D から今選択した要素 $(KAWA, 1, 5)$ を削除する。

ステップ 1104 の「A が空集合か?」という条件判定は「no」となって、ステップ 1105 に進む。ステップ 1105 では、選択した被覆要素の照合終了位置 q が n より大きいかなかを調べ、n より大きければ、最終結果の集合 SC に、この時点での集合 A の要素を加え、n 以下ならば、図 14 の手続き `select_cover3(A, p, q)` を呼び出す。この場合には、 $q = 5$ なので、手続き `select_cover3(A, 1, 5)` を呼び出す。図 14 のステップ 1108 では、照合位置 (s, e) が $1 < s < e$ である要素を C から全て選び、D1 とする。この例では、 $D1 = \{ (WADA, 3, 7) \}$ となる。D1 が空でないので、1109 の判定は「no」となり、ステップ 1110 に進む。ステップ 1110 では、D1 から唯一の要素 $(WADA, 3, 7)$ を選択し、 $M = \text{「WADA」}$ の各出現記号情報 (X, L, R) に対して、 $(X, L - 2)$ を B に記録し、選択した要素を D1 から削除する。さらに、 $A \cap B$ を計算するが、共通部分がなく、A1 は空となって、ステップ 1111 の判定結果が「yes」となり、ステップ 1109 に戻る。ところが、D1 は空なので、ステップ 1109 での判定結果は「yes」となり、`select_cover3` から戻る。図 38 の * S T E P 8 がこの様子を示している。図 13 のステップ 1102 では、D が空でないため、再びステップ 1103 に進む。ステップ 1103 では、D から要素 $(/SUKAWA, 2, 5)$ 30

を呼び出す。この場合には、 $q = 5$ なので、手続き `select_cover3(A, 1, 5)` を呼び出す。図 14 のステップ 1108 では、照合位置 (s, e) が $1 < s < e$ である要素を C から全て選び、D1 とする。この例では、 $D1 = \{ (WADA, 3, 7) \}$ となる。D1 が空でないので、1109 の判定は「no」となり、ステップ 1110 に進む。ステップ 1110 では、D1 から唯一の要素 $(WADA, 3, 7)$ を選択し、 $M = \text{「WADA」}$ の各出現記号情報 (X, L, R) に対して、 $(X, L - 2)$ を B に記録し、選択した要素を D1 から削除する。さらに、 $A \cap B$ を計算するが、共通部分がなく、A1 は空となって、ステップ 1111 の判定結果が「yes」となり、ステップ 1109 に戻る。ところが、D1 は空なので、ステップ 1109 での判定結果は「yes」となり、`select_cover3` から戻る。図 38 の * S T E P 8 がこの様子を示している。図 13 のステップ 1102 では、D が空でないため、再びステップ 1103 に進む。ステップ 1103 では、D から要素 $(/SUKAWA, 2, 5)$ 40

を呼び出す。この場合には、 $q = 5$ なので、手続き `select_cover3(A, 1, 5)` を呼び出す。図 14 のステップ 1108 では、照合位置 (s, e) が $1 < s < e$ である要素を C から全て選び、D1 とする。この例では、 $D1 = \{ (WADA, 3, 7) \}$ となる。D1 が空でないので、1109 の判定は「no」となり、ステップ 1110 に進む。ステップ 1110 では、D1 から唯一の要素 $(WADA, 3, 7)$ を選択し、 $M = \text{「WADA」}$ の各出現記号情報 (X, L, R) に対して、 $(X, L - 2)$ を B に記録し、選択した要素を D1 から削除する。さらに、 $A \cap B$ を計算するが、共通部分がなく、A1 は空となって、ステップ 1111 の判定結果が「yes」となり、ステップ 1109 に戻る。ところが、D1 は空なので、ステップ 1109 での判定結果は「yes」となり、`select_cover3` から戻る。図 38 の * S T E P 8 がこの様子を示している。図 13 のステップ 1102 では、D が空でないため、再びステップ 1103 に進む。ステップ 1103 では、D から要素 $(/SUKAWA, 2, 5)$ 50

を呼び出す。この場合には、 $q = 5$ なので、手続き `select_cover3(A, 1, 5)` を呼び出す。図 14 のステップ 1108 では、照合位置 (s, e) が $1 < s < e$ である要素を C から全て選び、D1 とする。この例では、 $D1 = \{ (WADA, 3, 7) \}$ となる。D1 が空でないので、1109 の判定は「no」となり、ステップ 1110 に進む。ステップ 1110 では、D1 から唯一の要素 $(WADA, 3, 7)$ を選択し、 $M = \text{「WADA」}$ の各出現記号情報 (X, L, R) に対して、 $(X, L - 2)$ を B に記録し、選択した要素を D1 から削除する。さらに、 $A \cap B$ を計算するが、共通部分がなく、A1 は空となって、ステップ 1111 の判定結果が「yes」となり、ステップ 1109 に戻る。ところが、D1 は空なので、ステップ 1109 での判定結果は「yes」となり、`select_cover3` から戻る。図 38 の * S T E P 8 がこの様子を示している。図 13 のステップ 1102 では、D が空でないため、再びステップ 1103 に進む。ステップ 1103 では、D から要素 $(/SUKAWA, 2, 5)$ 50

を選択し、*STEP 9のようなAを算出し、ステップ1107まで進み、再び図14の手続き `select_cover3(A, 1, 5)` を呼び出す。図14のステップ1108では、照合位置(s, e)が $1 < s < 5 < e$ である要素をCから全て選び、D1とする。この例では、 $D1 = \{(WADA, 3, 7)\}$ となる。図38の*STEP 10がこの様子を示している。D1が空でないので、1109の判定は「no」となり、ステップ1110に進む。ステップ1110では、D1から唯一の要素(WADA, 3, 7)を選択し、M = 「WADA」の各出現記号情報(X, L, R)に対して、(X, L - 2)をBに記録し、選択した要素をD1から削除する。さらに、AとBを計算するが、共通部分がなく、A1は空となって、ステップ1111の判定結果が「yes」となり、ステップ1109に戻る。ところが、D1は空なので、ステップ1109での判定結果は「yes」となり、`select_cover3`から戻る。図38の*STEP 11がこの様子を示している。図13のステップ1102では、Dが空でないため、再びステップ1103に進む。ステップ1103では、Dから要素(0 / KAWAD, - 1, 6)を選択し、*STEP 12のようなAを算出し、ステップ1107まで進み、再び図14の手続き `select_cover3(A, , 6)`を呼び出す。図14のステップ1108では、照合位置(s, e)が $1 < s < 6 < e$ である要素をCから全て選び、D1とする。この例では、 $D1 = \{(WADA, 3, 7)\}$ となる。図39の*STEP 13がこの様子を示している。D1が空でないので、1109の判定は「no」となり、ステップ1110に進む。

【0059】

ステップ1110では、D1から唯一の要素(WADA, 3,)を選択し、M = 「WADA」の各出現記号情報(X, L, R)に対して、(X, L - 2)をBに記録し、選択した要素をD1から削除する。さらに、AとBを計算する。この共通部分A1が空でなく、かつu = 7はn = 6より大きいので、ステップ1113に進み、最終結果の一部として、A1をSCに加える。ステップ1109に戻ると、D1が空なので、`select_cover3`から戻る。図39の*STEP 14がこの様子を示している。図13のステップ1102では、Dが空でないため、再びステップ1103に進む。ステップ1103では、Dから要素(5 / KAWAD, - 1, 6)を選択し、*STEP 15のようなAを算出し、ステップ1107まで進み、再び図14の手続き `select_cover3(A, 1, 6)`を呼び出す。図14のステップ1108では、照合位置(s, e)が $1 < s < 6 < e$ である要素をCから全て選び、D1とする。この例では、 $D1 = \{(WADA, 3, 7)\}$ となる。図39の*STEP 16がこの様子を示している。D1が空でないので、1109の判定は「no」となり、ステップ1110に進む。ステップ1110では、D1から唯一の要素(WADA, 3, 7)を選択し、M = 「WADA」の各出現記号情報(X, L, R)に対して、(X, L - 2)をBに記録し、選択した要素をD1から削除する。さらに、AとBを計算する。この共通部分A1が空でなく、かつu = 7はn = 6より大きいので、ステップ1113に進み、最終結果の一部として、A1をSCに加える。ステップ1109に戻ると、D1が空なので、`select_cover3`から戻る。図39の*STEP 17がこの様子を示している。図13のステップ1102では、Dが空でないため、再びステップ1103に進む。ステップ1103では、Dから要素(SUKAWA, - 1, 5)を選択し、*STEP 18のようなAを算出し、ステップ1107まで進み、再び図14の手続き `select_cover3(A, 1, 5)`を呼び出す。図14のステップ1108では、照合位置(s, e)が $1 < s < 5 < e$ である要素をCから全て選び、D1とする。この例では、 $D1 = \{(WADA, 3, 7)\}$ となる。図38の*STEP 19がこの様子を示している。D1が空でないので、1109の判定は「no」となり、ステップ1110に進む。ステップ1110では、D1から唯一の要素(WADA, 3, 7)を選択し、M = 「WADA」の各出現記号情報(X, L, R)に対して、(X, L - 2)をBに記録し、選択した要素をD1から削除する。さらに、AとBを計算するが、共通部分がなく、A1は空となって、ステップ1111の判定結果が「yes」となり、ステップ1109

10

20

30

40

50

に戻る。ところが、D1は空なので、ステップ1109での判定結果は「yes」となり、select_cover3から戻る。図38の*STEP20がこの様子を示している。図13のステップ1102では、Dが空であるので、記号番号集合の算定処理を終了する。この時点における、SCが、中間一致照合する記号(の番号)と、照合部分の左側(記号の先頭側)の文字数の組の集合をもれなく保持しているので、各組の第1要素を取り出せば、中間一致検索結果が得られる。このようにして、被覆結果504から、系統的に被覆要素の組み合わせを生成しつつ、記号辞書501中のメタ記号出現情報を検索して、生成した被覆要素の組に共通して含まれる記号の番号の集合を求めることができる。

【0060】

以上の説明のように、本発明の第5の実施の形態の記号辞書検索方法によれば、記号データ中で高頻度で出現する部分文字列については、より多くの文字数を持つメタ記号を持ったメタ記号情報が作成されており、このメタ記号情報を使って、質問文字列から一旦被覆結果を構成し、この被覆結果と、右延長メタ記号算定手段、左延長メタ記号算定手段、両延長メタ記号算定手段の3つの手段が追加した要素を加えた被覆結果から、メタ記号出現情報とを使って検索処理を行うので、高頻度で出現する部分文字列を含んだ記号の中間一致検索時でも、従来の記号辞書の検索に比べて高速に行えるようになる。

【0061】

以上、本発明の5種類の実施の形態における説明のように、本発明の記号辞書作成方法および記号辞書検索方法によれば、

(1) 記号データ中の記号を被覆するより短い「メタ記号」と呼ぶ記号を集めたメタ記号辞書を自動的に作成し、このメタ記号辞書中のメタ記号で記号データ中の各記号を被覆し、各記号がどのように被覆されたか、という情報を、各メタ記号毎に記録したメタ記号出現情報を作成する

(2) 質問文字列を、記号辞書に含まれるメタ記号辞書を検索してメタ記号で被覆し、被覆結果に左、右、両延長メタ記号の検索結果を追加し、質問文字列あるいはその左、右、両延長文字列を被覆する被覆結果中の要素集合毎に、共通して含まれる記号番号集合を求めることにより、従来の記号辞書作成方法および記号辞書検索方法における、

(1) 作成する記号辞書ファイルが検索対象の記号データの2倍以上の大きさになり、使用できる記憶装置の容量に制約のある場合に実現が困難である。

(2) 文字列が長く、高頻出の文字、あるいは文字連鎖を含む記号を検索する場合には、記号辞書から検索すべきデータ量が多くなり、検索速度が低下する。

(3) 文字連鎖を利用する方法において、文字連鎖数Nを大きくすると、出現するN文字連鎖の種類が急激に増加し、記号辞書の作成が困難になり、また作成した記号辞書の容量も増加する

という様々な課題を解決できる。

【0062】

このように、従来の記号辞書作成、検索技術では困難であった、中間一致照合まで含めた高速な検索が可能で、かつ容量の小さな記号辞書の作成を可能になり、しかも完全一致照合が質問の大部分を占めるような用途でも、平均検索速度が低下しない記号検索が可能になるので、実用上、非常に大きな効果が得られる。

【0063】

なお、上記の5種の実施の形態において、文字セットとして、英数字と特殊記号対象としたが、漢字やギリシャ文字を追加した文字セットを対象としても、同様に実施できる。また、第1の実施の形態において、記号辞書の作成に先立って図16のような1文字のみのメタ記号辞書を用意したが、図16の内容に加えて「1998-」、「AM」など予め出現が予測できる2文字以上のメタを含む、メタ記号辞書を用意してもよく、その場合も上記の説明と同様な手順で記号辞書が作成できる。また、メタ記号情報の格納形式として、TRIE構造と表構造を用いたが、有限状態機械やPATRICIA木、ハッシュ表など、他のデータ構造を用いた場合にも、上述の説明と同様の手順で実施できる。メタ記号出現情報の格納形式についても、表ではなく、TRIEやハッシュ表など他のデータ構造を

10

20

30

40

50

用いた場合にも、やはり上述の説明と同様の手順で実施できる。

【0064】

記号辞書検索においても、説明の便宜のため被覆結果などを、集合を用いて表現したが、リンクリストやヒープ、木構造、ハッシュ表など、他のデータ構造を用いた場合にも、やはり上述の説明と同様の手順で実施できる。

【0065】

【発明の効果】

以上のように本発明の記号辞書作成方法は記号データを完全一致照合、前方一致照合、後方一致照合、中間一致照合によって機械検索するための記号辞書の作成に際して、記号データ中の記号を被覆するより短い「メタ記号」と呼ぶ記号を集めたメタ記号辞書を自動的に作成し、このメタ記号辞書中のメタ記号で記号データ中の各記号を被覆し、各記号がどのように被覆されたか、という情報を、各メタ記号毎に記録したメタ記号出現情報を作成することにより、中間一致照合まで含めた高速な検索が可能で、かつ容量の小さな記号辞書の作成が可能になるという、優れた効果を奏するものである。

10

【0066】

また、本発明の記号辞書検索方法は、記号辞書の完全一致照合、前方一致照合、後方一致照合、中間一致照合による機械検索に際して、質問文字列を、記号辞書に含まれるメタ記号辞書を検索してメタ記号で被覆し、被覆結果に左、右、両延長メタ記号の検索結果を追加し、質問文字列あるいはその左、右、両延長文字列を被覆する被覆結果中の要素集合毎に、共通して含まれる記号番号集合を求めることにより、容量の小さな記号辞書を用いて、中間一致照合まで含めた高速な検索が可能で、しかも完全一致照合が質問の大部分を占めるような用途でも、平均検索速度が低下しない記号検索が可能になるという、優れた効果を奏するものである。

20

【0067】

本発明の効果は、検索対象となる記号データが数万種類以上の記号を含んでおり、各記号の文字数が多く、数多くの記号に共通に含まれる部分文字列が存在するような、分布に偏りのある大規模な記号データから記号辞書を作成し、検索する場合に、極めて顕著に現れる。例えば、各記号が100桁の数字で、その上位90桁分がどの記号でも等しく、下位10桁が各記号ですべて異なるような記号を100万個含んだ記号データから、記号辞書を作成する実験を行うと、従来のn-gram方式の記号辞書では、少なくとも1億個の記号番号・出現文字位置情報と、文字接続のための情報が必要で、そのサイズは400メガバイト以上となるが、本発明の記号辞書作成方法で作成した記号辞書では、約5万種類のメタ記号の情報と、約400万個のメタ記号出現情報を記録すればよく、そのサイズは40メガバイト以下であり、従来の1/10以下の容量になる。さらに、完全一致照合による検索を行う場合には、従来のn-gram方式の記号辞書では、文字接続の接続数nを41以上にとらない限り、上位40桁分の中間結果が常に100万記号分となって、検索速度が大幅に低下するが、本発明の記号辞書の検索では、記号データの分布の偏りに適応した、40+桁のメタ記号を含む、メタ記号辞書が自動的に作成され、質問文字列に関連するメタ記号の出現情報だけを参照して記号番号を求めるため、極めて高速に検索が行える。このように、従来は困難であった、偏りのある記号データに対しても、中間一致まで含めて高速な検索処理が可能となり、実用上、非常に大きな効果が得られる。

30

40

【図面の簡単な説明】

【図1】第1の実施の形態における記号辞書作成装置の全体構成を表わすブロック図

【図2】第2の実施の形態における記号辞書検索装置の全体構成を表わすブロック図

【図3】第3の実施の形態における記号辞書検索装置の全体構成を表わすブロック図

【図4】第4の実施の形態における記号辞書検索装置の全体構成を表わすブロック図

【図5】第5の実施の形態における記号辞書検索装置の全体構成を表わすブロック図

【図6】第1の実施の形態における記号被覆手段での被覆処理の手順を記述した流れ図

【図7】第2～第5の実施の形態における質問文字列被覆手段での被覆処理の手順を記述した流れ図

50

【図 8】第 2 の実施の形態における記号番号算定手段での記号番号の算定処理の手順を記述した流れ図

【図 9】第 3 の実施の形態における記号番号集合算定手段での記号番号集合の算定処理の手順を記述した流れ図

【図 10】第 3 の実施の形態における記号番号集合算定手段での記号番号集合の算定処理の手順を記述した流れ図

【図 11】第 4 の実施の形態における記号番号集合算定手段での記号番号集合の算定処理の手順を記述した流れ図

【図 12】第 4 の実施の形態における記号番号集合算定手段での記号番号集合の算定処理の手順を記述した流れ図

10

【図 13】第 5 の実施の形態における記号番号集合算定手段での記号番号集合の算定処理の手順を記述した流れ図

【図 14】第 5 の実施の形態における記号番号集合算定手段での記号番号集合の算定処理の手順を記述した流れ図

【図 15】第 1 の実施の形態における記号データの例

【図 16】第 1 の実施の形態における初期段階のメタ記号辞書の登録内容の例

【図 17】第 1 の実施の形態におけるメタ記号度数表の内容の例

【図 18】第 1 の実施の形態におけるメタ記号度数表の内容の例

【図 19】第 1 の実施の形態におけるメタ記号辞書の例

【図 20】第 1 の実施の形態におけるメタ記号辞書の一部の例

20

【図 21】第 1 の実施の形態におけるメタ記号度数表の内容の例

【図 22】第 1 の実施の形態におけるメタ記号度数表の内容の例

【図 23】第 1 の実施の形態におけるメタ記号度数表の内容の例

【図 24】第 1 の実施の形態におけるメタ記号度数表の内容の例

【図 25】第 1 の実施の形態におけるメタ記号度数表の内容の例

【図 26】第 1 の実施の形態におけるメタ記号度数表の内容の例

【図 27】第 1 の実施の形態におけるメタ記号度数表の内容の例

【図 28】第 1 の実施の形態におけるメタ記号度数表の内容の例

【図 29】第 1 の実施の形態におけるメタ記号出現情報の内容の例

【図 30】第 1 の実施の形態におけるメタ記号辞書のデータ構造の例を表す概念図

30

【図 31】第 1 の実施の形態におけるメタ記号辞書のデータ構造の例を表す概念図

【図 32】第 1 の実施の形態におけるメタ記号辞書のデータ構造の例を表す概念図

【図 33】第 1 の実施の形態におけるメタ記号の延長情報の内容の例

【図 34】第 2 の実施の形態における記号辞書検索の過程での主要な中間データを記述した概念図

【図 35】第 3 の実施の形態における記号辞書検索の過程での主要な中間データを記述した概念図

【図 36】第 3 の実施の形態における記号辞書検索の過程での主要な中間データを記述した概念図

【図 37】第 4 の実施の形態における記号辞書検索の過程での主要な中間データを記述した概念図

40

【図 38】第 5 の実施の形態における記号辞書検索の過程での主要な中間データを記述した概念図

【図 39】第 5 の実施の形態における記号辞書検索の過程での主要な中間データを記述した概念図

【符号の説明】

101 記号データ

102 メタ記号辞書

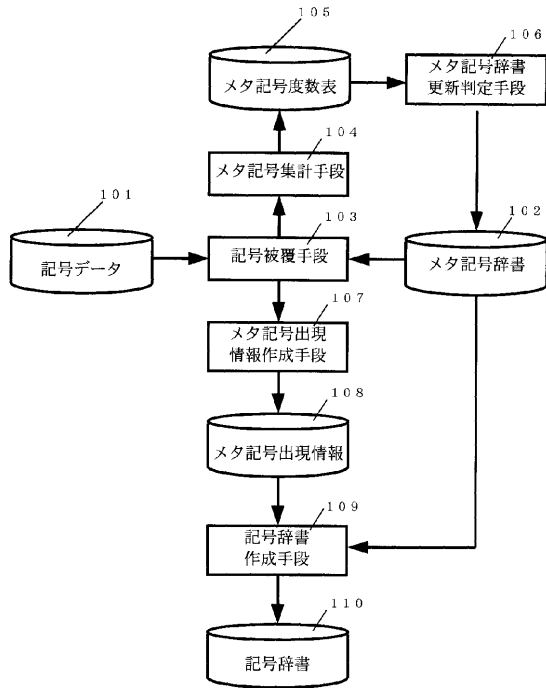
103 記号被覆手段

104 メタ記号集計手段

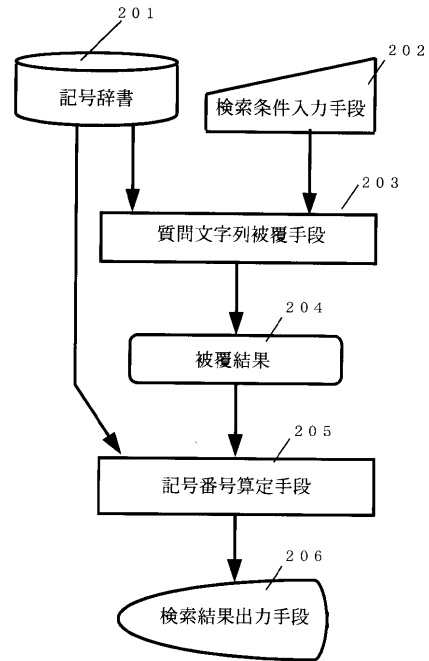
50

1 0 5	メ夕記号度数表	
1 0 6	メ夕記号辞書更新判定手段	
1 0 7	メ夕記号出現情報作成手段	
1 0 8	メ夕記号出現情報	
1 0 9	記号辞書作成手段	
1 1 0	記号辞書	
2 0 1	記号辞書	
2 0 2	検索条件入力手段	
2 0 3	質問文字列被覆手段	
2 0 4	被覆結果	10
2 0 5	記号番号算定手段	
2 0 6	検索結果出力手段	
3 0 1	記号辞書	
3 0 2	検索条件入力手段	
3 0 3	質問文字列被覆手段	
3 0 4	被覆結果	
3 0 5	記号番号集合算定手段	
3 0 6	検索結果出力手段	
3 0 7	右延長メ夕記号算定手段	
4 0 1	記号辞書	20
4 0 2	検索条件入力手段	
4 0 3	質問文字列被覆手段	
4 0 4	被覆結果	
4 0 5	記号番号集合算定手段	
4 0 6	検索結果出力手段	
4 0 8	左延長メ夕記号算定手段	
5 0 1	記号辞書	
5 0 2	検索条件入力手段	
5 0 3	質問文字列被覆手段	
5 0 4	被覆結果	30
5 0 5	記号番号集合算定手段	
5 0 6	検索結果出力手段	
5 0 7	右延長メ夕記号算定手段	
5 0 8	左延長メ夕記号算定手段	
5 0 9	両延長メ夕記号算	

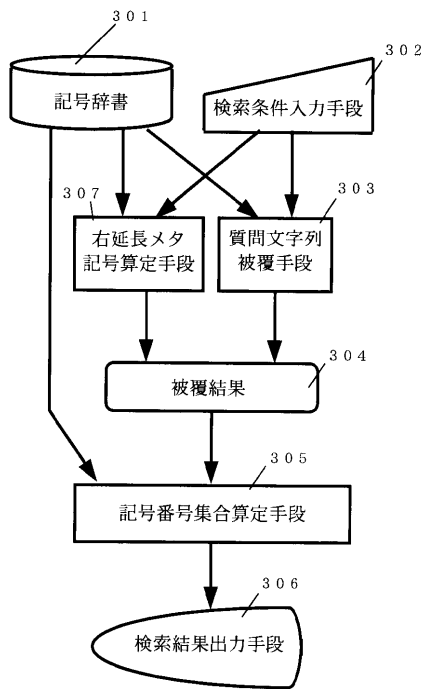
【 図 1 】



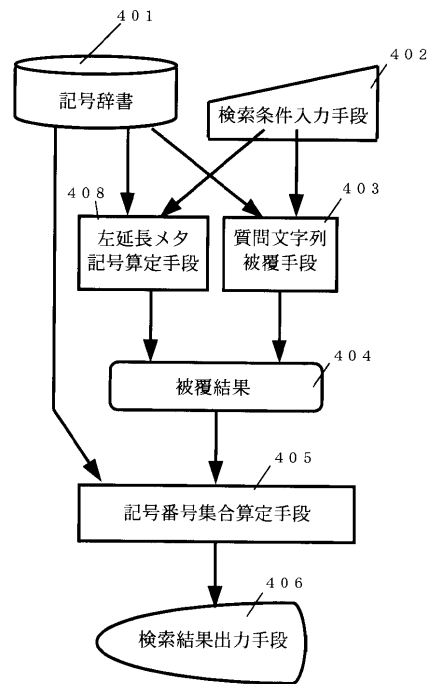
【 図 2 】



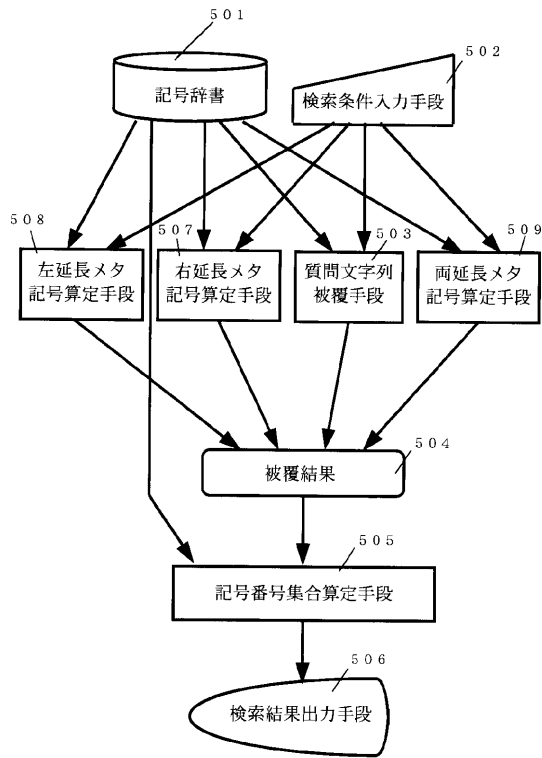
【 図 3 】



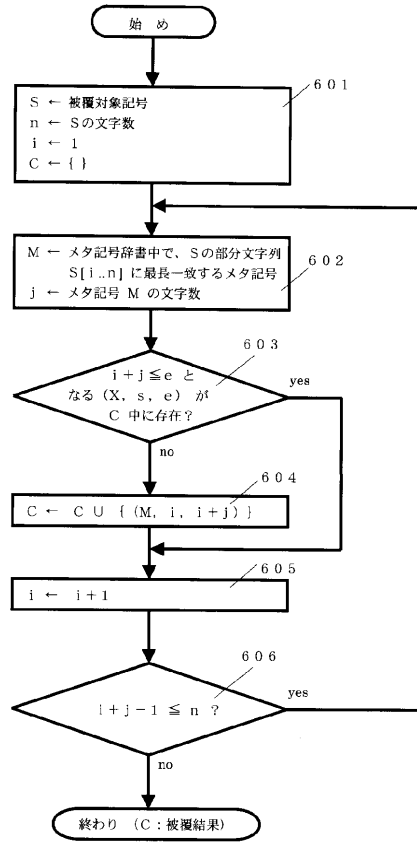
【 図 4 】



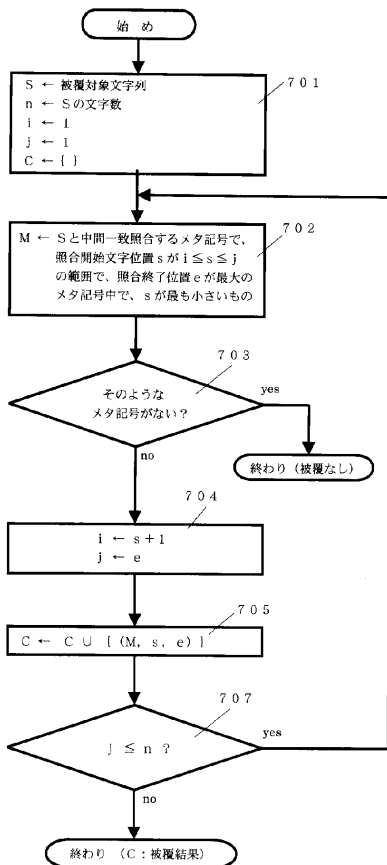
【 図 5 】



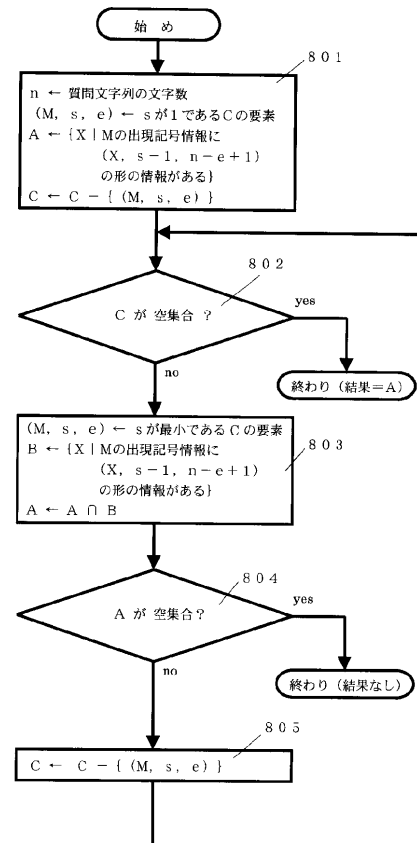
【 図 6 】



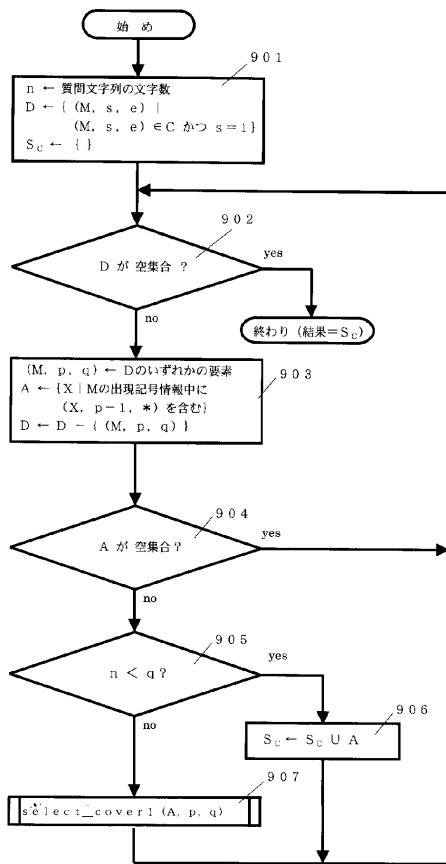
【 図 7 】



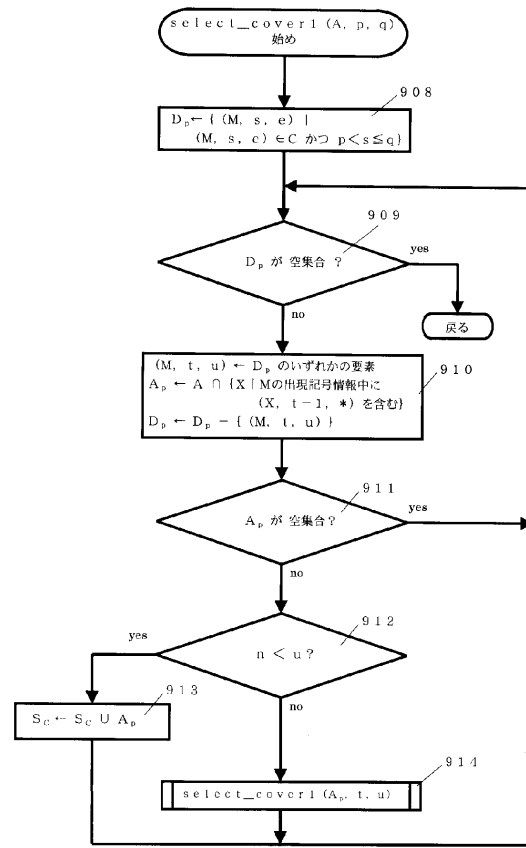
【 図 8 】



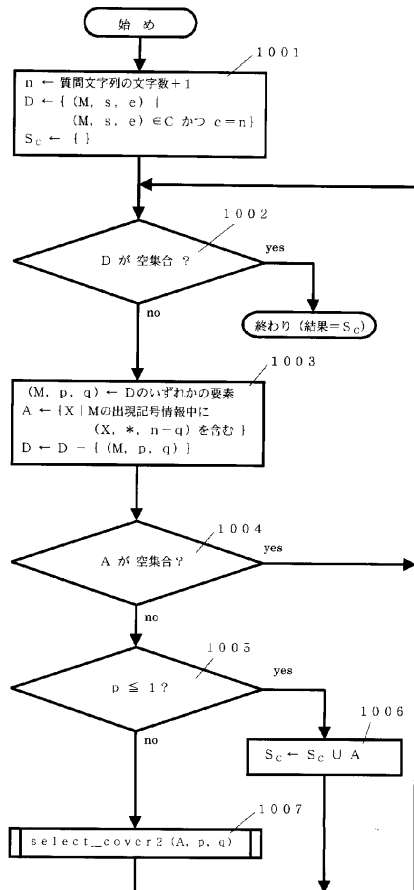
【図9】



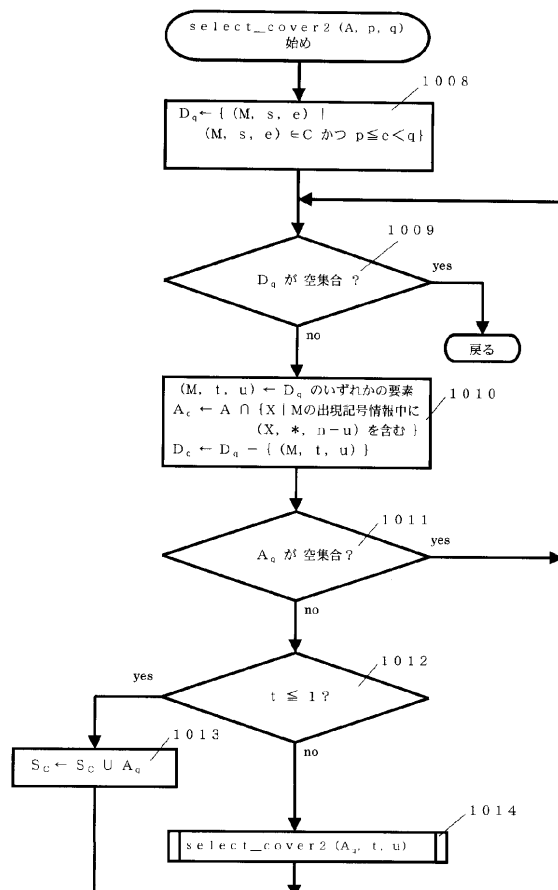
【図10】



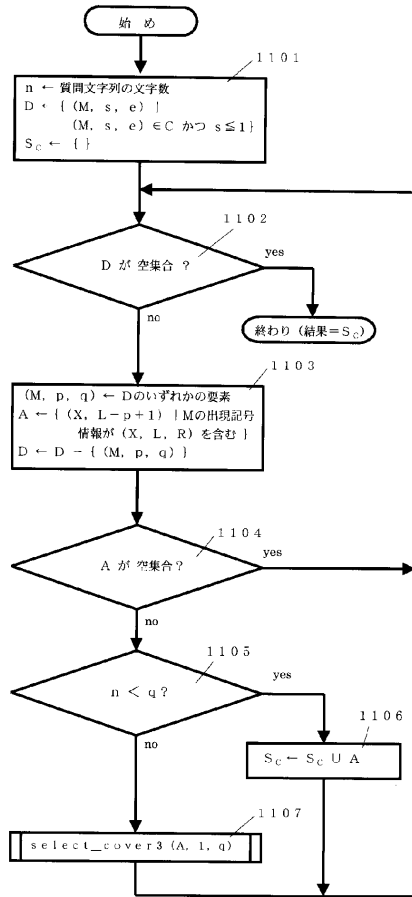
【図11】



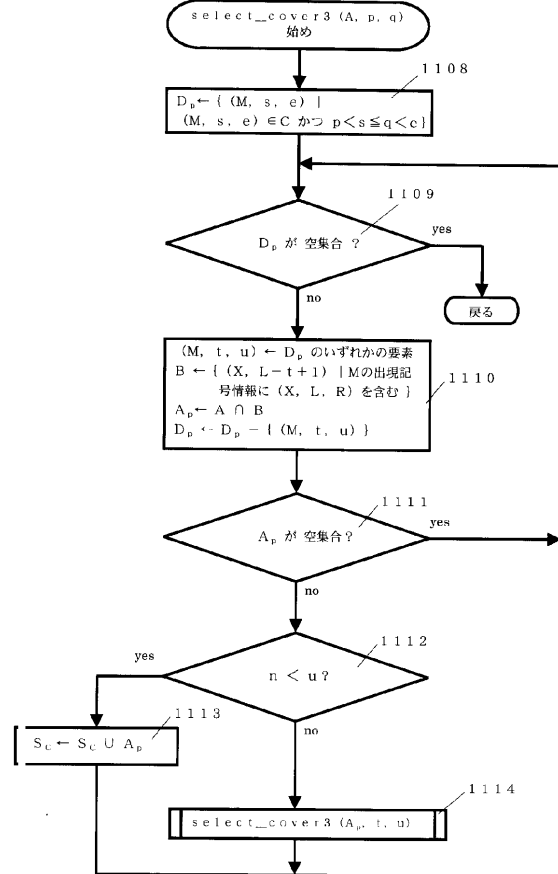
【図12】



【 図 1 3 】



【 図 1 4 】



【 図 1 5 】

1998-JAN-01/AM0200/KAWAYASU
 1998-JAN-01/AM0815/WADA
 1998-JAN-01/PM0615/NODA
 1998-JAN-01/PM1045/KAWADA
 1998-JAN-01/PM1130/NOKAWA
 1998-JAN-01/PM1215/KAWAYASU
 1998-JAN-03/AM0615/WADA
 1998-JAN-03/AM0745/YASUDA
 1998-JAN-04/AM0615/SUKAWA
 1998-JAN-04/AM1600/SUWA
 1998-JAN-05/AM1000/SUKAWA
 1998-JAN-05/PM0700/YASUDA
 1998-JAN-06/AM0430/YASUKAWA
 1998-JAN-06/AM0645/KAWAYASU
 1998-JAN-07/AM1130/WADA
 1998-JAN-08/AM1111/NODA
 1998-JAN-08/PM0315/SUWA
 1998-JAN-09/PM0730/SUWA
 1998-JAN-09/PM0845/NODA
 1998-JAN-10/AM1100/KAWAYASU
 1998-JAN-10/PM0245/YASUDA
 1998-JAN-11/PM0945/WADA
 1998-JAN-11/PM1145/NODA
 1998-JAN-12/AM0715/WADA
 1998-JAN-12/PM0730/SUDA
 1998-JAN-13/AM0345/SUKAWA
 1998-JAN-13/PM0145/SUWA
 1998-JAN-13/PM0930/SUKAWA
 1998-JAN-14/AM0645/YASUKAWA
 .
 .
 1998-JAN-29/PM1015/SUKAWA
 1998-JAN-29/PM1045/SUWA
 1998-JAN-30/AM0915/KAWAYASU
 1998-JAN-30/AM1130/YASUDA
 1998-JAN-30/PM0945/KAWADA
 1998-JAN-30/PM1045/KAWADA
 1998-JAN-31/AM0330/SUWA
 1998-JAN-31/AM0630/SUKAWA
 1998-JAN-31/PM0445/KAWADA
 1998-JAN-31/PM0800/SUDA
 1998-FEB-01/PM0300/SUKAWA
 1998-FEB-01/PM1230/YASUKAWA
 1998-FEB-01/PM1245/KAWAYASU
 1998-FEB-02/AM0645/KAWADA
 1998-FEB-02/PM1215/SUDA
 1998-FEB-03/AM1130/WADA
 1998-FEB-03/AM1145/YASUKAWA
 1998-FEB-04/AM0645/SUKAWA
 1998-FEB-04/AM1215/KAWADA

1998-FEB-04/PM0115/NOKAWA
 1998-FEB-05/AM0315/SUDA
 1998-FEB-05/AM0930/KAWADA
 1998-FEB-05/PM0130/NOKAWA
 1998-FEB-05/PM1045/SUWA
 1998-FEB-06/AM0715/NOKAWA
 1998-FEB-06/AM0815/KAWANO
 1998-FEB-06/AM0830/SUDA
 1998-FEB-06/AM1130/NOKAWA
 1998-FEB-07/AM0500/KAWANO
 1998-FEB-07/AM0715/SUWA
 1998-FEB-07/AM0745/YASUDA
 1998-FEB-07/AM0845/SUWA
 1998-FEB-07/PM0345/KAWADA
 1998-FEB-07/PM1200/SUKAWA
 1998-FEB-08/AM0445/SUWA
 1998-FEB-08/AM1015/KAWADA
 1998-FEB-08/AM1100/NODA
 1998-FEB-09/AM0300/KAWAYASU
 .
 .
 1998-DEC-25/AM1100/YASUDA
 1998-DEC-25/AM1230/SUDA
 1998-DEC-25/PM0915/KAWAYASU
 1998-DEC-26/AM1045/NOKAWA
 1998-DEC-26/AM1000/KAWANO
 1998-DEC-27/AM1000/WADA
 1998-DEC-27/AM1100/KAWADA
 1998-DEC-27/AM1145/SUDA
 1998-DEC-27/AM1200/SUKAWA
 1998-DEC-27/PM0215/SUWA
 1998-DEC-27/PM0415/NOKAWA
 1998-DEC-27/PM0800/KAWANO
 1998-DEC-27/PM0900/YASUKAWA
 1998-DEC-27/PM0945/YASUDA
 1998-DEC-28/AM1245/SUDA
 1998-DEC-28/PM0630/YASUDA
 1998-DEC-28/PM0730/NOKAWA
 1998-DEC-29/AM0315/KAWADA
 1998-DEC-29/AM0330/SUWA
 1998-DEC-29/AM0845/KAWANO
 1998-DEC-29/PM1030/NODA
 1998-DEC-30/AM0845/KAWAYASU
 1998-DEC-30/AM1030/KAWAYASU
 1998-DEC-30/AM1100/SUKAWA
 1998-DEC-31/AM1115/YASUDA
 1998-DEC-31/AM1200/NOKAWA
 1998-DEC-31/PM0400/SUKAWA

【 図 1 6 】

-
/
0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z

【 図 17 】

メタ記号	度数
-	2000
/	2000
0	1996
1	2130
2	584
3	491
4	452
5	677
6	168
7	191
8	1167
9	2164
A	2917
B	88
C	181
D	532
E	266
F	88
G	99
H	0
I	0
J	224
K	559
L	78
M	1145
N	505
O	443
P	636
Q	0
R	134
S	640
T	84
U	800
V	107
W	754
X	0
Y	334
Z	0

【 図 18 】

メタ記号	度数
-	2000
/	2000
0	1996
1	2130
2	584
3	491
4	432
5	677
6	168
7	191
8	1167
9	2164
A	2917
B	88
C	181
D	532
E	266
F	88
G	99
J	224
K	559
L	78
M	1145
N	505
O	443
P	636
R	134
S	640
T	84
U	800
V	107
W	754
Y	334

【 図 19 】

-
--
-/
-0
-1
-2
-3
-4
-5
-6
-7
-8
-9
-A
-B
-C
-D
-E
-F
-G
-J
-K
-L
-M
-N
-O
-P
-R
-S
-T
-U
-V
-W
-Y
/
/-
//
/0
/1
/2
/3
/4
/5
/6
/7
/8
/9
/A

/B
/C
/D
/E
/F
/G
/J
/K
/L
/M
/N
/O
/P
/R
/S
/T
/U
/V
/W
/Y
0
0-
0/
00
01
02
03
04
05
06
07
08
09
0A
0B
0C
0D
0E
0F
0G
0J
0K
0L
0M
0N
0O
0P
0R

0S
0T
0U
0Y
0W
0Y
1
1/
10
11
.
.
Y
Y-
Y0
Y1
Y2
Y3
Y4
Y5
Y6
Y7
Y8
Y9
YA
YB
YC
YD
YE
YF
YG
YJ
YK
YL
YM
YN
YO
YP
YR
YS
Y-
YU
YV
YW
YY

【 図 20 】

メタ記号	度数
-0	294
-1	309
-2	328
-3	69
-A	174
-D	97
-F	88
-J	224
-M	145
-N	107
-O	84
-S	81
/A	520
/K	273
/N	169
/P	480
/S	311
/W	95
/Y	152
0/	608
00	281
01	157
02	108
03	132
04	144
05	124
06	113
07	121
08	106
09	102
1/	145
10	169
11	156
12	114
13	80
14	79
15	265
16	27
17	30
18	35
19	1030
2/	101
20	61
21	83
22	34
23	71

メタ記号	度数
24	70
25	38
26	28
27	40
28	26
29	32
3/	93
30	309
31	53
33	16
34	20
4/	96
40	21
41	18
43	24
44	23
45	259
5/	550
50	24
51	21
53	23
54	19
6/	82
60	23
61	23
63	24
64	16
7/	105
70	18
71	24
73	18
74	26
8-	1000
8/	88
80	21
81	22
83	18
84	18
9/	92
90	25
91	13
93	15
94	17
98	1000
99	1000
AD	189

メタ記号	度数
AM	520
AN	165
AP	75
AR	59
AS	248
AU	99
AW	550
AY	182
B-	88
C-	97
CT	84
DA	435
DE	97
EB	88
EC	97
EP	81
FE	83
G-	93
JA	82
JU	142
KA	559
L-	78
MO	749
MI	251
MA	145
N-	145
NO	339
OC	84
OD	70
OK	99
OV	107
P-	81
PW	480
PK	75
R-	134
SE	81
SU	559
T-	84
UD	175
UG	99
UX	187
UL	78
UN	64
UV	160
V-	157
WA	754

メタ記号	度数
Y-	86
YA	248

【図 2 1】

メタ記号	度数	メタ記号	度数	メタ記号	度数	メタ記号	度数	メタ記号	度数
-01	41	000	19	10/	37	AM0	388	R-1	41
-02	32	01/	41	100	08	AM1	132	R-2	40
-03	32	010	24	101	19	AN-	82	R-3	11
-04	33	011	24	103	20	ANO	83	SE	81
-05	37	018	26	104	25	ASU	248	STD	176
-06	27	014	23	11/	27	ANA	569	SUK	187
-07	35	015	19	110	24	AY-	85	SLW	100
-08	27	02/	32	111	16	AYA	96	T-	84
-09	30	020	17	113	23	B-	88	IDA	178
-10	37	021	16	114	25	C-	97	UG	99
-11	27	023	21	115	40	CT	84	UKA	187
-12	35	024	22	12/	35	DA	70	UWA	100
-13	31	02/	32	120	17	DE	97	V-0	32
-14	31	030	47	121	26	EE	88	V-1	28
-15	26	031	17	123	20	EC	57	V-2	41
-16	27	033	18	124	16	EP	81	V-3	6
-17	30	034	23	15/	26	FE	88	WAD	189
-18	35	04/	33	19/	30	G-	99	WAN	83
-19	30	040	21	199	1006	JUL	78	WAY	96
-20	27	041	18	2/	57	JUN	64	YAS	248
-21	41	043	24	2/	44	KAW	559		
-22	34	044	23	3/	61	L-	78		
-23	30	045	25	30/	282	MO1	97		
-24	32	05/	37	300	27	MO2	76		
-25	38	050	24	31	36	MO3	80		
-26	28	051	21	4/	63	MO4	86		
-27	40	053	23	45/	250	MO5	87		
-28	28	054	19	5/	54	MO6	86		
-29	32	05/	27	5/	130	MO7	86		
-3	40	060	23	5/	88	MO8	79		
-AP	75	061	23	5/P	47	MO9	72		
-AD	99	063	24	5/S	156	MO0	83		
-JA	82	064	16	5/W	48	MI1	89		
-JU	142	07/	35	5/Y	67	MI2	79		
-MA	145	070	18	6/	55	MAE	98		
-NO	107	071	24	7/	50	MAY	86		
/AM	520	073	18	7/P	55	N-0	38		
/KA	273	074	23	8-A	174	N-1	40		
/NO	168	08/	27	8-D	97	N-2	55		
/PM	480	080	21	8-F	88	N-3	12		
/SL	311	081	22	8-J	224	NOD	70		
/YA	152	083	18	8-M	145	NOK	99		
0/A	47	084	18	8-N	107	NOV	107		
0/K	143	08/	30	8-O	84	OC	84		
0/N	81	090	26	8-S	81	OV-	107		
0/P	50	091	13	8/	61	P-	81		
0/S	155	093	16	9/	32	PM0	361		
0/Y	85	1/	17	98	1000	PR	119		
0/	262	1/P	77	ADA	189	R-0	42		

【図 2 2】

メタ記号	度数	メタ記号	度数	メタ記号	度数	メタ記号	度数	メタ記号	度数
-01	41	00/K	84	091	13	5/P	34	EP	81
-02	32	00/N	31	093	16	5/SU	166	FE	88
-03	32	00/S	81	094	17	6/	55	G-	99
-04	33	00/W	24	1/	68	7/A	50	KAWA	187
-05	37	00/Y	42	1/P	77	7/P	55	L-	78
-06	27	00/	19	10/	37	8-AP	75	N-0	39
-07	35	01/	41	100	68	8-AL	99	N-1	40
-08	27	010	24	101	19	8-JA	82	N-2	55
-09	30	011	24	103	20	8-JU	142	N-3	12
-10	37	013	26	104	25	8-MA	145	NOV-	107
-11	27	014	23	11/	27	8-NO	107	OC	84
-12	35	015	19	110	25	8/	61	OV-0	32
-13	31	02/	32	111	16	9/	32	OV-1	28
-14	31	020	17	113	23	98-A	174	OV-2	41
-15	26	021	16	114	25	98-B	97	OV-3	6
-16	27	023	21	115	40	98-F	98	P-	81
-17	30	024	22	12/	35	98-I	224	PM01	40
-18	35	03/	32	120	17	98-M	145	PM02	38
-19	30	030	47	121	23	98-N	107	PM03	40
-20	27	031	17	123	20	98-O	84	PM04	42
-21	41	033	16	124	16	98-S	81	PM05	45
-22	34	034	20	15/	13	98-	1000	PM06	40
-23	30	04/	33	15/K	59	AM01	57	PM07	46
-24	32	040	21	15/N	49	AM02	38	PM08	30
-25	38	041	18	15/P	13	AM03	40	PM09	40
-26	28	043	24	15/S	32	AM04	44	PM10	43
-27	40	044	23	15/W	27	AM05	42	PM11	37
-28	26	045	25	15/Y	32	AM06	46	PM12	39
-29	32	05/	37	19/	30	AM07	40	PR	75
-3	40	050	24	199/	1000	AM08	49	R-0	42
-JUL	78	051	21	2/A	57	AM09	32	R-1	41
-JUN	64	053	23	2/P	44	AM10	40	R-2	40
-MAR	59	054	19	2/	61	AM11	52	R-3	11
-MAY	86	06/	27	30/A	21	AM12	40	SE	81
-NOV	107	060	23	30/K	59	AN-	82	SUDA	176
/AM0	388	061	23	30/N	50	ANO	83	SUKA	187
/AM1	132	063	24	30/P	12	ASUD	71	SUWA	100
/KAW	273	064	16	30/S	74	ASUX	81	T-	84
/NOD	70	07/	35	30/W	23	ARAD	94	UG	99
/NOK	99	070	18	30/Y	43	AWAN	83	UKAW	187
/PM0	361	071	24	30/	27	AWAY	96	WADA	189
/PM1	119	073	18	31	36	AY-	86	YASU	248
/SD	105	074	26	4/	63	AYA	96		
/SUK	106	08/	27	45/K	71	B-	88		
/SUD	100	080	21	45/N	39	C-	97		
/YAS	152	081	22	45/S	74	CT	84		
0/A	47	083	18	45/W	31	DA	70		
0/KA	143	084	18	45/Y	45	DE	97		
0/P	50	08/	30	5/A	41	ER	85		
0/SU	155	090	25	5/KA	150	EL	97		

【図 2 3】

メタ記号	度数	メタ記号	度数	メタ記号	度数	メタ記号	度数	メタ記号	度数
-01	41	/PM05	45	054	19	3/	61	B-	88
-02	32	/PM06	40	06/	27	30/A	21	C-	97
-03	32	/PM07	46	080	23	30/K	59	CT	84
-04	33	/PM08	30	061	23	30/N	50	DA	70
-05	37	/PM09	40	063	24	30/P	12	DE	97
-06	27	/PM10	43	064	16	30/S	74	ER	88
-07	35	/PM11	37	07/	25	30/W	23	EC	97
-08	27	/SUDA	35	070	18	30/Y	43	EP	81
-09	30	/SUKA	105	071	24	30/	27	FE	88
-10	37	/SUKA	105	073	18	31	35	G-	99
-11	27	/SUWA	100	074	26	4/	63	KAWA	99
-12	35	/YASU	152	080	27	45/K	71	KAWAD	94
-13	31	0/A	26	080	21	45/N	39	KAWAN	83
-14	31	0/KAW	143	081	22	45/S	74	KAWAY	96
-15	26	0/P	38	083	18	45/W	21	L-	78
-16	27	0/SD	45	084	18	45/Y	45	N-0	39
-17	30	0/SUK	71	08/	30	5/A	41	N-1	40
-18	35	0/SUW	38	090	26	5/KAW	130	N-2	55
-19	30	00/K	84	091	13	5/P	34	N-3	12
-20	27	00/N	31	093	16	5/SUD	59	NOV-0	32
-21	41	00/S	81	094	17	5/SUK	35	NOV-1	28
-22	34	00/W	24	1/	68	5/SUW	62	NOV-2	41
-23	30	00/Y	42	1/P	77	6/	55	NOV-3	6
-24	32	00/	19	10/	37	7/A	30	OC	84
-25	38	01/	41	100	65	7/P	55	P-	81
-26	28	010	24	101	15	8-JUL	78	PK	75
-27	40	011	24	103	20	8-JUN	94	R-0	42
-28	26	013	25	104	25	8-MAR	59	R-1	41
-29	32	014	23	11/	27	8-MAY	86	R-2	40
-3	40	015	19	110	25	8-NOV	107	R-3	11
-NOV-	107	02/	32	111	16	8/	61	SE	81
/AM01	57	020	17	113	23	9/	32	SUDA	71
/AM02	38	021	16	114	25	98-AP	75	SUKAW	187
/AM03	40	023	21	115	40	98-AU	99	T-	84
/AM04	44	024	22	12/	35	98-JA	82	UG	99
/AM05	42	03/	32	120	17	98-JU	142	UKAWA	187
/AM06	46	030	47	121	26	98-MA	145	WADA	189
/AM07	40	031	17	123	20	98-NO	107	YASU	96
/AM08	49	033	18	124	16	98-SA	174	YASUD	71
/AM09	22	034	20	15/A	13	98-D	97	YASUK	81
/AM10	40	04/	33	15/K	59	98-F	88		
/AM11	52	040	21	15/N	49	98-J	224		
/AM12	40	041	18	15/P	13	98-M	145		
/KAWA	273	043	24	15/S	82	98-N	107		
/NOD	70	044	23	15/W	27	98-O	84		
/NOK	99	045	25	15/Y	22	98-S	81		
/PM01	40	05/	37	19/	30	AN-	82		
/PM02	38	050	24	199-	1000	ASU	53		
/PM03	40	051	21	2/A	57	AY	86		
/PM04	42	053	23	2/P	44	ATA	96		

【図 2 4】

メタ記号	度数	メタ記号	度数	メタ記号	度数	メタ記号	度数	メタ記号	度数
-01	41	/NOK	99	044	23	15/W	27	98-MA	145
-02	32	/PM01	40	045	25	15/Y	22	98-NO	107
-03	32	/PM02	38	05/	37	19/	30	AN-	82
-04	33	/PM03	40	050	24	1998-A	174	ANO	83
-05	37	/PM04							

【 図 2 5 】

メタ記号	度数	メタ記号	度数	メタ記号	度数	メタ記号	度数	メタ記号	度数
-01	41	/PM07	46	053	23	1998-D	97	ANO	83
-02	32	/PM08	30	054	19	1998-F	88	AY-	86
-03	32	/PM09	40	067	27	1998-JA	82	AYA	96
-04	33	/PM10	43	066	23	1998-JU	142	B-	88
-05	37	/PM11	37	061	23	1998-MA	145	C-	97
-06	27	/PM12	39	063	24	1998-NO	107	CT	84
-07	35	/SIDA	105	064	16	1998-O	84	DA	70
-08	27	/SUKAWA	106	077	35	1998-S	81	DE	97
-09	30	/SUWA	100	070	18	2/A	57	EB	88
-10	37	/YASUD	71	071	24	2/P	44	EC	97
-11	27	/YASUK	81	073	18	3/	61	EP	81
-12	35	0/A	26	074	26	30/A	21	FE	88
-13	31	0/KAWAD	47	087	27	30/K	59	G-	99
-14	31	0/KAWAN	47	080	21	30/N	50	KAWA	99
-15	26	0/KAWAY	49	081	22	30/P	12	L	78
-16	27	0/P	38	083	18	30/S	74	N-0	39
-17	30	0/SUD	46	084	18	30/W	23	N-1	40
-18	35	0/SUK	71	087	30	30/Y	43	N-2	55
-19	30	0/SUN	38	089	26	300	27	N-3	12
-20	27	00/K	84	091	13	31	36	OC	84
-21	41	00/N	31	093	16	4/	63	P-	81
-22	34	00/S	81	094	17	45/K	71	PR	75
-23	30	00/W	24	1/A	68	45/N	39	R-0	42
-24	32	00/Y	42	1/P	77	45/S	74	R-1	41
-25	38	000	19	107	37	45/W	21	R-2	40
-26	28	01/	41	100	68	45/Y	45	R-3	11
-27	40	010	24	101	19	5/A	41	SE	81
-28	26	011	24	103	20	5/KAWAD	47	SIDA	71
-29	32	013	26	104	25	5/KAWAN	36	SUKAWA	81
-3	40	014	23	117	27	5/KAWAY	47	T-	84
/AM01	57	015	19	110	25	5/P	34	UG	99
/AM02	38	027	32	111	16	5/SUD	59	WADA	189
/AM03	40	020	17	113	23	5/SUK	35	YASU	96
/AM04	44	021	16	114	25	5/SUN	62		
/AM05	42	023	21	115	40	6/	55		
/AM06	46	024	22	127	35	7/A	50		
/AM07	40	037	32	120	17	7/P	55		
/AM08	49	030	47	121	26	8-NOV-0	32		
/AM09	32	031	17	123	20	8-NOV-1	28		
/AM10	40	033	16	124	16	8-NOV-2	41		
/AM11	52	034	20	15/A	13	8-NOV-3	6		
/AM12	40	047	33	15/K	59	8/	61		
/AND	70	040	21	15/N	49	9/	32		
/NOK	99	041	18	15/P	13	98-NOV	107		
/PM01	40	043	24	15/S	82	98-JUL	78		
/PM02	38	044	23	15/W	27	98-NOV	64		
/PM03	40	045	25	15/Y	22	98-MAR	95		
/PM04	42	057	37	19/	30	98-MAY	86		
/PM05	45	050	24	1998-AP	75	98-NOV	107		
/PM06	40	051	21	1998-AU	99	AN-	82		

【 図 2 6 】

メタ記号	度数	メタ記号	度数	メタ記号	度数	メタ記号	度数	メタ記号	度数
-01	41	/PM07	46	053	23	1998-D	97	C-	97
-02	32	/PM08	30	054	19	1998-F	88	CT	84
-03	32	/PM09	40	067	27	1998-JUL	78	DA	70
-04	33	/PM10	43	066	23	1998-JUN	64	DE	97
-05	37	/PM11	37	061	23	1998-MAR	59	EB	88
-06	27	/PM12	39	063	24	1998-MAY	86	EC	97
-07	35	/SIDA	105	064	16	1998-NOV	107	EP	81
-08	27	/SUKAWA	106	077	35	1998-O	84	FE	88
-09	30	/SUWA	100	070	18	1998-S	81	G-	99
-10	37	/YASUD	71	071	24	2/A	57	KAWA	99
-11	27	/YASUK	81	073	18	2/P	44	L-	78
-12	35	0/A	26	074	26	3/	61	N-0	39
-13	31	0/KAWAD	47	087	27	2/A	57	N-1	40
-14	31	0/KAWAN	47	080	21	2/P	44	N-2	55
-15	26	0/KAWAY	49	081	22	3/	61	N-3	12
-16	27	0/P	38	083	18	30/A	21	OC	84
-17	30	0/SUD	46	084	18	30/K	59	P-	81
-18	35	0/SUK	71	087	30	30/W	23	PR	75
-19	30	0/SUN	38	089	26	30/Y	43	R-0	42
-20	27	00/K	84	091	13	300	27	R-1	41
-21	41	00/N	31	093	16	31	36	R-2	40
-22	34	00/S	81	094	17	4/	63	R-3	11
-23	30	00/W	24	1/A	68	45/K	71	SE	81
-24	32	00/Y	42	1/P	77	45/N	39	SIDA	71
-25	38	000	19	107	37	45/S	74	SUKAWA	81
-26	28	01/	41	100	68	45/W	21	T-	84
-27	40	010	24	101	19	45/Y	45	UG	99
-28	26	011	24	103	20	5/A	41	WADA	189
-29	32	013	26	104	25	5/KAWAD	47	YASU	96
3	40	014	23	117	27	5/KAWAN	36		
/AM01	57	015	19	110	25	5/KAWAY	47		
/AM02	38	027	32	111	16	5/P	34		
/AM03	40	020	17	113	23	5/SUD	59		
/AM04	44	021	16	114	25	5/SUN	62		
/AM05	42	023	21	115	40	6/	55		
/AM06	46	024	22	127	35	7/A	50		
/AM07	40	037	32	120	17	7/P	55		
/AM08	49	030	47	121	26	7/P	55		
/AM09	32	031	17	123	20	8/	61		
/AM10	40	033	16	124	16	9/	32		
/AM11	52	034	20	15/A	13	98-NOV-0	32		
/AM12	40	047	33	15/K	59	98-NOV-1	28		
/AND	70	040	21	15/N	49	98-NOV-2	41		
/NOK	99	041	18	15/P	13	98-NOV-3	6		
/PM01	40	043	24	15/S	82	98-NOV	107		
/PM02	38	044	23	15/W	27	AN-	82		
/PM03	40	045	25	15/Y	22	ANO	83		
/PM04	42	057	37	19/	30	AY-	86		
/PM05	45	050	24	1998-AP	75	AYA	96		
/PM06	40	051	21	1998-AU	99	B-	88		

【 図 2 7 】

メタ記号	度数	メタ記号	度数	メタ記号	度数	メタ記号	度数	メタ記号	度数
-01	41	/PM07	46	053	23	1998-D	97	CT	84
-02	32	/PM08	30	054	19	1998-F	88	DA	70
-03	32	/PM09	40	067	27	1998-JUL	78	DE	97
-04	33	/PM10	43	066	23	1998-JUN	64	EB	88
-05	37	/PM11	37	061	23	1998-MAR	59	EC	97
-06	27	/PM12	39	063	24	1998-MAY	86	EP	81
-07	35	/SIDA	105	064	16	1998-NOV	107	FE	88
-08	27	/SUKAWA	106	077	35	1998-O	84	G-	99
-09	30	/SUWA	100	070	18	1998-S	81	KAWA	99
-10	37	/YASUD	71	071	24	2/A	57	L-	78
-11	27	/YASUK	81	073	18	2/P	44	N-0	39
-12	35	0/A	26	074	26	3/	61	N-1	40
-13	31	0/KAWAD	47	087	27	30/A	21	N-2	55
-14	31	0/KAWAN	47	080	21	30/N	50	N-3	12
-15	26	0/KAWAY	49	081	22	30/Y	43	OC	84
-16	27	0/P	38	083	18	30/W	23	P-	81
-17	30	0/SUD	46	084	18	30/X	59	PR	75
-18	35	0/SUK	71	087	30	30/Y	43	R-0	42
-19	30	0/SUN	38	089	26	30/Z	23	R-1	41
-20	27	00/K	84	091	13	300	27	R-2	40
-21	41	00/N	31	093	16	31	36	R-3	11
-22	34	00/S	81	094	17	4/	63	SE	81
-23	30	00/W	24	1/A	68	45/K	71	SIDA	71
-24	32	00/Y	42	1/P	77	45/N	39	SUKAWA	81
-25	38	000	19	107	37	45/S	74	T-	84
-26	28	01/	41	100	68	45/W	21	UG	99
-27	40	010	24	101	19	45/Y	45	WADA	189
-28	26	011	24	103	20	5/A	41	YASU	96
-29	32	013	26	104	25	5/KAWAD	47		
-3	40	014	23	117	27	5/KAWAN	36		
/AM01	57	015	19	110	25	5/KAWAY	47		
/AM02	38	027	32	111	16	5/P	34		
/AM03	40	020	17	113	23	5/SUD	59		
/AM04	44	021	16	114	25	5/SUN	62		
/AM05	42	023	21	115	40	6/	55		
/AM06	46	024	22	127	35	7/A	50		
/AM07	40	037	32	120	17	7/P	55		
/AM08	49	030	47	121	26	7/P	55		
/AM09	32	031	17	123	20	8/	61		
/AM10	40	033	16	124	16	9/	32		
/AM11	52	034	20	15/A	13	98-NOV-0	32		
/AM12	40	047	33	15/K	59	98-NOV-1	28		
/AND	70	040	21	15/N	49	98-NOV-2	41		
/NOK	99	041	18	15/P	13	98-NOV-3	6		
/PM01	40	043	24	15/S	82	AN-	82		
/PM02	38	044	23	15/W	27	ANO	83		
/PM03	40	045	25	15/Y	22	AY-	86		
/PM04	42	057	37	19/	30	AYA	96		
/PM05	45	050	24	1998-AP	75	B-	88		
/PM06	40	051	21	1998-AU	99	C-	97		

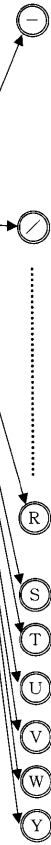
【 図 2 8 】

メタ記号	度数	メタ記号	度数	メタ記号	度数	メタ記号	度数	メタ記号	度数
-01	41	/PM07	46	053	23	1998-D	97	DA	70
-02	32	/PM08	30	054	19	1998-F	88	DE	97
-03	32	/PM							

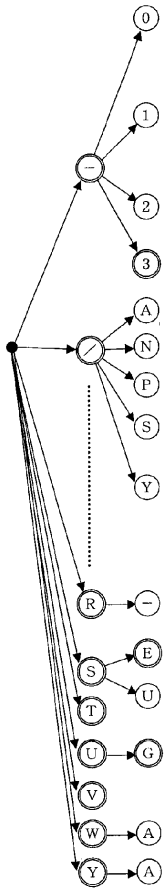
【 図 29 】

夕記号 (記号, 左文字数, 右文字数)	メ夕記号 (記号, 左文字数, 右文字数)
01 (1998-JAN-01/AM0200/KAWAYASU, 8, 15) (1998-JAN-01/AM0815/WADA, 8, 12) (1998-JAN-01/PM0615/NODA, 8, 12) (1998-JAN-01/PM1045/KAWADA, 8, 14) (1998-JAN-01/PM1130/KAWADA, 8, 14) (1998-JAN-01/PM1215/KAWAYASU, 8, 16) (1998-FEB-01/PM0300/SUKAWA, 8, 14) (1998-FEB-01/PM1230/YASUKAWA, 8, 16) (1998-FEB-01/PM1245/KAWAYASU, 8, 16) . . (1998-DEC-01/AM0100/YASUDA, 8, 14) (1998-DEC-01/AM0115/NOKAWA, 8, 14) (1998-DEC-01/PM0400/SUDA, 8, 12) (1998-DEC-01/PM1000/YASUKAWA, 8, 16)	100 (1998-JAN-04/AM1000/SUWA, 14, 6) (1998-JAN-06/AM0100/SUKAWA, 15, 7) (1998-JAN-10/AM1100/KAWAYASU, 15, 9) . . (1998-DEC-25/AM1100/YASUDA, 15, 7) (1998-DEC-26/AM1000/KAWANO, 14, 8) (1998-DEC-27/AM0100/WADA, 15, 5) (1998-DEC-27/AM1100/KAWADA, 15, 7) (1998-DEC-30/AM1100/SUKAWA, 15, 7) . . .
02 (1998-FEB-02/AM0645/KAWADA, 8, 14) (1998-FEB-02/PM1215/SUDA, 8, 12) . . (1998-DEC-02/PM1030/SUKAWA, 8, 14) (1998-DEC-02/PM1045/YASUKAWA, 8, 16)	DA (1998-JAN-01/PM0615/NODA, 21, 0) (1998-JAN-08/AM1115/NODA, 21, 0) . . (1998-DEC-25/AM0800/NODA, 21, 0) (1998-DEC-29/PM1030/NODA, 21, 0)
29 (1998-JAN-29/AM0800/WADA, 8, 12) (1998-JAN-29/AM0845/KAWADA, 8, 14) . . (1998-DEC-29/PM1030/NODA, 8, 12)	WADA (1998-JAN-01/AM0815/WADA, 16, 0) (1998-JAN-01/PM1045/KAWADA, 21, 0) (1998-JAN-01/PM1130/KAWADA, 21, 0) (1998-JAN-03/AM0615/WADA, 19, 0) . . (1998-DEC-25/PM0530/KAWADA, 21, 0) (1998-DEC-27/AM0100/WADA, 19, 0) (1998-DEC-27/AM1100/KAWADA, 21, 0) (1998-DEC-29/AM0315/KAWADA, 21, 0)
3 (1998-JAN-30/AM0915/KAWAYASU, 8, 17) (1998-JAN-30/AM1130/YASUDA, 8, 15) (1998-JAN-30/PM0945/KAWADA, 8, 15) . . (1998-DEC-30/AM1100/SUKAWA, 8, 15) (1998-DEC-31/AM1115/YASUDA, 8, 15) (1998-DEC-31/AM1200/NOKAWA, 8, 15) (1998-DEC-31/PM0400/SUKAWA, 8, 15) . . .	YASU (1998-JAN-01/AM0200/KAWAYASU, 23, 0) (1998-JAN-01/PM1215/KAWAYASU, 25, 0) (1998-JAN-06/AM0645/KAWAYASU, 23, 0) . . (1998-DEC-25/PM0915/KAWAYASU, 23, 0) (1998-DEC-30/AM0845/KAWAYASU, 23, 0) (1998-DEC-30/AM1020/KAWAYASU, 23, 0)

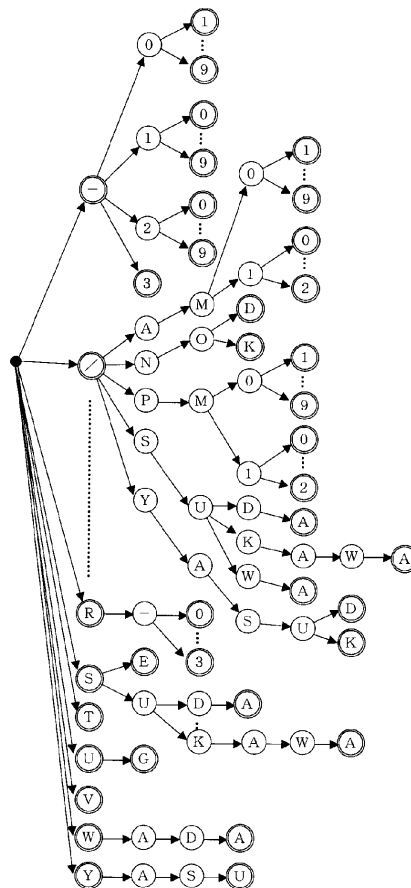
【 図 30 】



【 図 31 】



【 図 32 】



【 3 3 】

メタ記号 (延長メタ記号, 左文字数, 右文字数)	メタ記号 (延長メタ記号, 左文字数, 右文字数)
- (-, 0, 0), (-01, 0, 2), ..., (-29, 0, 2), (-3, 0, 1), (1998-AP, 4, 2), (1998-AL, 4, 2), (1998-B, 4, 1), ..., (1998-NOV-3, 4, 5), (1998-NOV-3, 8, 1), (1998-0, 4, 1), (1998- S, 4, 1), (AN, 2, 0), (AP, 2, 0), (C, 1, 0), (N, 0, 1, 1), ..., (R-3, 1, 1), (R-, 1, 0), (R-0, 1, 1), ..., (R-3, 1, 1)	SUDA (SUDA, 0, 0), (SUDA, 1, 0) SUKAWA (SUKAWA, 0, 0), (SUKAWA, 1, 0) T (T, 0, 0), (T-, 0, 1) T- (T-, 0, 0) U (U, 0, 0), (SUDA, 2, 2), (SUKAWA, 2, 4), (SUDA, 2, 2), (YASUD, 4, 1), (YASUK, 4, 1), (SUD, 3, 1), (SUK, 5, 1), (SUN, 3, 1), (1998 AL, 6, 0), (1998-JU, 6, 0), (SUD, 3, 1), (SUK, 3, 1), (SUN, 3, 1), (1998-JUL, 5, 1), (1998-JUN, 5, 1), (SUDA, 1, 2), (SUKAWA, 1, 4), (CIG, 0, 1), (YASU, 3, 0)
-01 (-01, 0, 0)	UG (UG, 0, 0)
-29 (-29, 0, 0)	Y (8-NOV-0, 4, 2), (8-NOV-1, 4, 2), (8-NOV-2, 4, 2), (8-NOV-3, 4, 2), (8-NOV-6, 0)
-3 (-3, 0, 0), (1998-NOV-3, 8, 0), (N-3, 2, 0), (R-3, 2, 0)	W (W, 0, 0), (SUKAWA, 5, 1), (SUDA, 3, 1), (KAWAD, 4, 2), (KAWA, 4, 2), (KAWAY, 4, 2), (SUN, 3, 0), (OD, 3, 0), (15, 3, 0), (30, 3, 0), (45, 3, 0), (S, KAWAD, 4, 2), (S, KAWA, 4, 2), (S, KAWAY, 4, 2), (S, SUN, 3, 0), (KAWA, 3, 1), (SUKAWA, 4, 1), (WADA, 0, 2)
/ (/, 0, 0), (AM01, 0, 4), ..., (AM12, 0, 4), (X0D, 0, 3), (X0K, 0, 3), (PM01, 0, 4), ..., (PM12, 0, 4), (SUDA, 0, 4), ..., (R/, 1, 0)	WADA (WADA, 0, 0) Y (Y, 0, 0), (YASUD, 1, 4), (YASUK, 1, 4), (KAWAY, 6, 0), (OD, 3, 0), (15, 3, 0), (30, 3, 0), (45, 3, 0), (S, KAWAY, 5, 0), (1998-MAY, 6, 0), (AY-, 1, 1), (AYA, 1, 1), (YASU, 0, 2)
/AM01 (AM01, 0, 0)	YASU (YASU, 0, 0), (YASUD, 1, 1), (YASUK, 1, 1)
/AM12 (AM12, 0, 0)	
/X0D (X0D, 0, 0)	
/X0K (X0K, 0, 0)	
/PM01 (PM01, 0, 0)	
/PM12 (PM12, 0, 0)	
/SUDA (SUDA, 0, 0)	
/SUKAWA (SUKAWA, 0, 0)	
/SUWA (SUWA, 0, 0)	

【 3 4 】

質問文字列=1998-NOV-01/PM1030/KAWAYASU (完全一致照合)

*STEP 1
C={ (1998-NOV-0, 1, 11), (1/P, 11, 14), (PM10, 12, 17), (30/K, 17, 21),
(O/KAWAY, 18, 25), (YASU, 24, 28) }

*STEP 2
(M, s, e)=(1998-NOV-0, 1, 11)
A={ 1998-NOV-01/AM0830/WADA, 1998-NOV-01/AM1115/SUDA, 1998-NOV-01/PM1030/KAWAYASU,
1998-NOV-01/PM1130/SUWA, 1998-NOV-01/PM1145/KAWADA, 1998-NOV-02/AM0300/KAWADA,
1998/NOV-09/AM0945/SUDA, 1998-NOV-09/PM0515/KAWAYASU }
C={ (1/P, 11, 14), (PM10, 12, 17), (30/K, 17, 21), (O/KAWAY, 18, 25), (YASU, 24, 28) }

*STEP 3
(M, s, e)=(1/P, 11, 14)
B={ 1998-JAN-01/PM0615/WADA, 1998-JAN-01/PM1045/KAWADA, 1998-JAN-11/PM0945/WADA, ...,
1998-NOV-01/PM1030/KAWAYASU, 1998-NOV-01/PM1130/SUWA, 1998-NOV-01/PM1145/KAWADA,
1998-NOV-11/PM0230/WADA, 1998-NOV-21/PM0230/SUDA, ..., 1998-DEC-31/PM0400/STAKFA }
A={ 1998-NOV-01/PM1030/KAWAYASU, 1998-NOV-01/PM1130/SUWA, 1998-NOV-01/PM1145/KAWADA }
C={ (PM10, 12, 17), (30/K, 17, 21), (O/KAWAY, 18, 25), (YASU, 24, 28) }

*STEP 4
(M, s, e)=(PM10, 12, 17)
B={ 1998-JAN-01/PM1045/KAWADA, 1998-JAN-29/PM1015/SUKAWA, 1998-JAN-29/PM1045/SUWA, ...,
1998-NOV-01/PM1030/KAWAYASU, 1998-NOV-06/PM1045/YASURAWA, 1998-NOV-17/PM1045/SUWA,
1998-NOV-20/PM1045/SUWA, 1998-NOV-23/PM1045/WADA, ..., 1998-DEC-21/PM1045/SUDA }
A={ 1998-NOV-01/PM1030/KAWAYASU }
C={ (30/K, 17, 21), (O/KAWAY, 18, 25), (YASU, 24, 28) }

*STEP 5
(M, s, e)=(30/K, 17, 21)
B={ 1998-JAN-01/PM1130/KAWADA, 1998-JAN-17/PM0330/KAWADA, 1998-JAN-22/AM0530/KAWAYASU, ...,
1998-NOV-01/PM1030/KAWAYASU, 1998-NOV-03/AM0300/KAWADA, 1998-NOV-04/PM0630/KAWAYASU,
..., 1998-DEC-23/PM0530/KAWADA, 1998-DEC-30/AM1030/KAWAYASU }
A={ 1998-NOV-01/PM1030/KAWAYASU }
C={ (O/KAWAY, 18, 25), (YASU, 24, 28) }

*STEP 6
(M, s, e)=(O/KAWAY, 18, 25)
B={ 1998-JAN-01/AM0200/KAWAYASU, 1998-JAN-10/AM1100/KAWAYASU, ...,
1998-NOV-01/PM1030/KAWAYASU, 1998-NOV-03/AM0200/KAWAYASU, ...,
1998-DEC-17/AM0230/KAWAYASU, 1998-DEC-30/AM1030/KAWAYASU }
A={ 1998-NOV-01/PM1030/KAWAYASU }
C={ (YASU, 24, 28) }

*STEP 7
(M, s, e)=(YASU, 24, 28)
B={ 1998-JAN-01/AM0200/KAWAYASU, 1998-JAN-01/PM1215/KAWAYASU, ...,
1998-NOV-01/PM1030/KAWAYASU, 1998-NOV-03/AM0200/KAWAYASU,
1998-NOV-03/AM0415/KAWAYASU, 1998-NOV-04/PM0630/KAWAYASU, ...,
1998-DEC-30/AM0845/KAWAYASU, 1998-DEC-30/AM1030/KAWAYASU }
A={ 1998-NOV-01/PM1030/KAWAYASU }
C={ }

【 3 5 】

質問文字列 = 1998-NOV-01/PM (前方一致照合)

*STEP 1
C={ (1998-NOV-0, 1, 11), (1/P, 11, 14), (M, 14, 15) }

*STEP 2
C={ (1998-NOV-0, 1, 11), (1/P, 11, 14), (M, 14, 15), (PM01, 12, 17), (PM02, 12, 17),
(PM03, 12, 17), (PM04, 12, 17), (PM05, 12, 17), (PM06, 12, 17), (PM07, 12, 17),
(PM08, 12, 17), (PM09, 12, 17), (PM10, 12, 17), (PM11, 12, 17), (PM12, 12, 17) }

*STEP 3
n=14
D={ (1998-NOV-0, 1, 11) }
S_C={ }

*STEP 4
(M, p, q)=(1998-NOV-0, 1, 11)
A={ 1998-NOV-01/AM0330/WADA, 1998-NOV-01/AM1115/SUDA, 1998-NOV-01/PM1030/KAWAYASU,
1998-NOV-01/PM1130/SUWA, 1998-NOV-01/PM1145/KAWADA, 1998-NOV-02/AM0300/KAWADA,
1998/NOV-09/AM0945/SUDA, 1998-NOV-09/PM0515/KAWAYASU }
D={ }

*STEP 5
D₁={ (1/P, 11, 14) }

*STEP 6
(M, t, u)=(1/P, 11, 14)
A₁={ 1998-NOV-01/PM1030/KAWAYASU, 1998-NOV-01/PM1130/SUWA, 1998-NOV-01/PM1145/KAWADA }
D₁={ }

*STEP 7
D₁₁={ (M, 14, 15), (PM01, 12, 17), (PM02, 12, 17), (PM03, 12, 17), (PM04, 12, 17),
(PM05, 12, 17), (PM06, 12, 17), (PM07, 12, 17), (PM08, 12, 17), (PM09, 12, 17),
(PM10, 12, 17), (PM11, 12, 17), (PM12, 12, 17) }

*STEP 8
(M, t, u)=(M, 14, 15)
A₁₁={ }
D₁₁={ (PM01, 12, 17), (PM02, 12, 17), (PM03, 12, 17), (PM04, 12, 17), (PM05, 12, 17), (PM06, 12, 17),
(PM07, 12, 17), (PM08, 12, 17), (PM09, 12, 17), (PM10, 12, 17), (PM11, 12, 17), (PM12, 12, 17) }

*STEP 9
(M, t, u)=(PM01, 12, 17)
A₁₁₁={ }
D₁₁₁={ (PM02, 12, 17), (PM03, 12, 17), (PM04, 12, 17), (PM05, 12, 17), (PM06, 12, 17), (PM07, 12, 17),
(PM08, 12, 17), (PM09, 12, 17), (PM10, 12, 17), (PM11, 12, 17), (PM12, 12, 17) }

*STEP 10
(M, t, u)=(PM02, 12, 17)
A₁₁₁₁={ }
D₁₁₁₁={ (PM03, 12, 17), (PM04, 12, 17), (PM05, 12, 17), (PM06, 12, 17), (PM07, 12, 17),
(PM08, 12, 17), (PM09, 12, 17), (PM11, 12, 17), (PM12, 12, 17) }

【 3 6 】

*STEP 11
(M, t, u)=(PM03, 12, 17)
A₁₁₁₁₁={ }
D₁₁₁₁₁={ (PM04, 12, 17), (PM05, 12, 17), (PM06, 12, 17), (PM07, 12, 17), (PM08, 12, 17),
(PM09, 12, 17), (PM10, 12, 17), (PM11, 12, 17), (PM12, 12, 17) }

*STEP 12
(M, t, u)=(PM04, 12, 17)
A₁₁₁₁₁₁={ }
D₁₁₁₁₁₁={ (PM05, 12, 17), (PM06, 12, 17), (PM07, 12, 17), (PM08, 12, 17), (PM09, 12, 17),
(PM10, 12, 17), (PM11, 12, 17), (PM12, 12, 17) }

*STEP 13
(M, t, u)=(PM05, 12, 17)
A₁₁₁₁₁₁₁={ }
D₁₁₁₁₁₁₁={ (PM06, 12, 17), (PM07, 12, 17), (PM08, 12, 17), (PM09, 12, 17), (PM10, 12, 17),
(PM11, 12, 17), (PM12, 12, 17) }

*STEP 14
(M, t, u)=(PM06, 12, 17)
A₁₁₁₁₁₁₁₁={ }
D₁₁₁₁₁₁₁₁={ (PM07, 12, 17), (PM08, 12, 17), (PM09, 12, 17), (PM10, 12, 17), (PM11, 12, 17), (PM12, 12, 17) }

*STEP 15
(M, t, u)=(PM07, 12, 17)
A₁₁₁₁₁₁₁₁₁={ }
D₁₁₁₁₁₁₁₁₁={ (PM08, 12, 17), (PM09, 12, 17), (PM10, 12, 17), (PM11, 12, 17), (PM12, 12, 17) }

*STEP 16
(M, t, u)=(PM08, 12, 17)
A₁₁₁₁₁₁₁₁₁₁={ }
D₁₁₁₁₁₁₁₁₁₁={ (PM09, 12, 17), (PM10, 12, 17), (PM11, 12, 17), (PM12, 12, 17) }

*STEP 17
(M, t, u)=(PM09, 12, 17)
A₁₁₁₁₁₁₁₁₁₁₁={ }
D₁₁₁₁₁₁₁₁₁₁₁={ (PM10, 12, 17), (PM11, 12, 17), (PM12, 12, 17) }

*STEP 18
(M, t, u)=(PM10, 12, 17)
A₁₁₁₁₁₁₁₁₁₁₁₁={ 1998-NOV-01/PM1030/KAWAYASU }
D₁₁₁₁₁₁₁₁₁₁₁₁={ (PM11, 12, 17), (PM12, 12, 17) }
S_C={ } ∩ A₁₁₁₁₁₁₁₁₁₁₁₁={ 1998-NOV-01/PM1030/KAWAYASU }

*STEP 19
(M, t, u)=(PM11, 12, 17)
A₁₁₁₁₁₁₁₁₁₁₁₁₁={ 1998-NOV-01/PM1030/KAWAYASU }
D₁₁₁₁₁₁₁₁₁₁₁₁₁={ (PM12, 12, 17) }
S_C={ 1998-NOV-01/PM1030/KAWAYASU, 1998-NOV-01/PM1130/SUWA, 1998-NOV-01/PM1145/KAWADA }

*STEP 20
(M, t, u)=(PM12, 12, 17)
A₁₁₁₁₁₁₁₁₁₁₁₁₁₁={ }
D₁₁₁₁₁₁₁₁₁₁₁₁₁₁={ }

【 3 7 】

質問文字列 = KAWA (後方一致照合)

*STEP 1
C={ (KAWA, 1, 5) }

*STEP 2
C={ (KAWA, 1, 5), (/SUKAWA, -2, 5), (SUKAWA, -1, 5) }

*STEP 3
n=5
D={ (KAWA, 1, 5), (/SUKAWA, -2, 5), (SUKAWA, -1, 5) }
S_c={ }

*STEP 4
(M, p, q)=(KAWA, 1, 5)
A={ 1998-JAN-17/PM0930/NOKAWA, 1998-JAN-18/PM0300/NOKAWA, 1998-JAN-19/AM1200/NOKAWA, 1998-JAN-20/AM0930/NOKAWA, 1998-DEC-26/AM0145/NOKAWA, 1998-DEC-27/PM0415/NOKAWA, 1998-DEC-28/PM0730/NOKAWA, 1998-DEC-31/AM1200/NOKAWA }
D={ (/SUKAWA, -2, 5), (SUKAWA, -1, 5) }
S_c={ 1998-JAN-17/PM0930/NOKAWA, 1998-JAN-18/PM0300/NOKAWA, 1998-JAN-19/AM1200/NOKAWA, 1998-JAN-20/AM0930/NOKAWA, 1998-DEC-26/AM0145/NOKAWA, 1998-DEC-27/PM0415/NOKAWA, 1998-DEC-28/PM0730/NOKAWA, 1998-DEC-31/AM1200/NOKAWA }

*STEP 5
(M, p, q)=(/SUKAWA, -2, 5)
A={ 1998-JAN-04/AM0615/SUKAWA, 1998-JAN-05/AM0100/SUKAWA, 1998-JAN-13/AM0345/SUKAWA, 1998-JAN-14/PM0930/SUKAWA, 1998-DEC-21/PM0530/SUKAWA, 1998-DEC-27/AM1200/SUKAWA, 1998-DEC-30/AM1100/SUKAWA, 1998-DEC-31/PM0400/SUKAWA }
D={ (SUKAWA, -1, 5) }
S_c={ 1998-JAN-17/PM0930/NOKAWA, 1998-JAN-18/PM0300/NOKAWA, 1998-JAN-19/AM1200/NOKAWA, 1998-JAN-20/AM0930/NOKAWA, 1998-DEC-26/AM0145/NOKAWA, 1998-DEC-27/PM0415/NOKAWA, 1998-DEC-28/PM0730/NOKAWA, 1998-DEC-31/AM1200/NOKAWA, 1998-JAN-04/AM0615/SUKAWA, 1998-JAN-05/AM0100/SUKAWA, 1998-JAN-13/AM0345/SUKAWA, 1998-JAN-14/PM0930/SUKAWA, 1998-DEC-21/PM0530/SUKAWA, 1998-DEC-27/AM1200/SUKAWA, 1998-DEC-30/AM1100/SUKAWA, 1998-DEC-31/PM0400/SUKAWA }

*STEP 6
(M, p, q)=(SUKAWA, -1, 5)
D={ }

A={ 1998-JAN-05/AM0100/YASUKAWA, 1998-JAN-14/0645/YASUKAWA, 1998-JAN-18/AM0130/YASUKAWA, 1998-DEC-27/PM0900/YASUKAWA, 1998-DEC-29/PM1130/YASUKAWA }
S_c={ 1998-JAN-17/PM0930/NOKAWA, 1998-JAN-18/PM0300/NOKAWA, 1998-JAN-19/AM1200/NOKAWA, 1998-JAN-20/AM0930/NOKAWA, 1998-DEC-26/AM0145/NOKAWA, 1998-DEC-27/PM0415/NOKAWA, 1998-DEC-28/PM0730/NOKAWA, 1998-DEC-31/AM1200/NOKAWA, 1998-JAN-04/AM0615/SUKAWA, 1998-JAN-05/AM0100/SUKAWA, 1998-JAN-13/AM0345/SUKAWA, 1998-JAN-14/PM0930/SUKAWA, 1998-DEC-21/PM0530/SUKAWA, 1998-DEC-27/AM1200/SUKAWA, 1998-DEC-30/AM1100/SUKAWA, 1998-DEC-31/PM0400/SUKAWA, 1998-JAN-05/AM0100/YASUKAWA, 1998-JAN-14/0645/YASUKAWA, 1998-JAN-18/AM0130/YASUKAWA, 1998-DEC-27/PM0900/YASUKAWA, 1998-DEC-29/PM1130/YASUKAWA }

【 3 8 】

質問文字列 = KAWADA (中間一致照合)

*STEP 1
C={ (KAWA, 1, 5), (WADA, 3, 7) }

*STEP 2
C={ (KAWA, 1, 5), (WADA, 3, 7) }

*STEP 3
C={ (KAWA, 1, 5), (WADA, 3, 7), (/SUKAWA, -2, 5), (/KAWAD, -1, 6), (/KAWAD, -1, 6), (SUKAWA, -1, 5) }

*STEP 4
C={ (KAWA, 1, 5), (WADA, 3, 7), (/SUKAWA, -2, 5), (/KAWAD, -1, 6), (/KAWAD, -1, 6), (SUKAWA, -1, 5) }

*STEP 5
n=6
D={ (KAWA, 1, 5), (/SUKAWA, -2, 5), (/KAWAD, -1, 6), (/KAWAD, -1, 6), (SUKAWA, -1, 5) }
S_c={ }

*STEP 6
(M, p, q)=(KAWA, 1, 5)
A={ (1998-JAN-17/PM0930/NOKAWA, 21), (1998-JAN-18/PM0300/NOKAWA, 21), (1998-JAN-19/AM1200/NOKAWA, 21), (1998-JAN-20/AM0930/NOKAWA, 21), (1998-DEC-26/AM0145/NOKAWA, 21), (1998-DEC-27/PM0415/NOKAWA, 21), (1998-DEC-28/PM0730/NOKAWA, 21), (1998-DEC-31/AM1200/NOKAWA, 21) }
D={ (/SUKAWA, -1, 5), (/KAWAD, -1, 6), (/KAWAD, -1, 6), (SUKAWA, -1, 5) }

*STEP 7
D₁={ (WADA, 3, 7) }

*STEP 8
(M, t, u)=(WADA, 3, 7)
B={ (1998-JAN-01/AM0815/WADA, 17), (1998-JAN-01/PM1045/KAWADA, 19), (1998-JAN-01/PM1130/KAWADA, 19), (1998-JAN-03/AM0615/WADA, 17), (1998-DEC-23/PM0530/KAWADA, 19), (1998-DEC-27/AM0100/WADA, 17), (1998-DEC-27/AM1100/KAWADA, 19), (1998-DEC-29/AM0315/KAWADA, 19) }
A₁={ }
D₁={ }

*STEP 9
(M, p, q)=(/SUKAWA, -2, 5)
A={ (1998-JAN-04/AM0615/SUKAWA, 21), (1998-JAN-05/AM0100/SUKAWA, 21), (1998-DEC-30/AM1100/SUKAWA, 21), (1998-DEC-31/PM0400/SUKAWA, 21) }
D={ (/KAWAD, -1, 6), (/KAWAD, -1, 6), (SUKAWA, -1, 5) }

*STEP 10
D₁={ (WADA, 3, 7) }

*STEP 11
(M, t, u)=(WADA, 3, 7)
B={ (1998-JAN-01/AM0815/WADA, 17), (1998-DEC-29/AM0315/KAWADA, 19) }
A₁={ }
D₁={ }

【 3 9 】

*STEP 12
(M, p, q)=(/KAWAD, -1, 6)
A={ (1998-JAN-01/PM1130/KAWADA, 19), (1998-JAN-17/PM0330/KAWADA, 19), (1998-DEC-23/PM0930/KAWADA, 19), (1998-DEC-27/AM1100/KAWADA, 19) }
D={ (/KAWAD, -1, 6), (SUKAWA, -1, 5) }

*STEP 13
D₁={ (WADA, 3, 7) }

*STEP 14
(M, t, u)=(WADA, 3, 7)
B={ (1998-JAN-01/AM0815/WADA, 17), (1998-DEC-29/AM0315/KAWADA, 19) }
A₁={ (1998-JAN-01/PM1130/KAWADA, 19), (1998-JAN-17/PM0330/KAWADA, 19), (1998-DEC-23/PM0930/KAWADA, 19), (1998-DEC-27/AM1100/KAWADA, 19) }
D₁={ }
S_c={ (1998-JAN-01/PM1130/KAWADA, 19), (1998-JAN-17/PM0330/KAWADA, 19), (1998-DEC-23/PM0930/KAWADA, 19), (1998-DEC-27/AM1100/KAWADA, 19) }

*STEP 15
(M, p, q)=(/KAWAD, -1, 6)
A={ (1998-JAN-01/PM1045/KAWADA, 19), (1998-JAN-21/AM0315/KAWADA, 19), (1998-DEC-22/PM0945/KAWADA, 19), (1998-DEC-29/AM0315/KAWADA, 19) }
D={ (SUKAWA, -1, 5) }

*STEP 16
D₁={ (WADA, 3, 7) }

*STEP 17
(M, t, u)=(WADA, 3, 7)
B={ (1998-JAN-01/AM0815/WADA, 17), (1998-DEC-29/AM0315/KAWADA, 19) }
A₁={ (1998-JAN-01/PM1045/KAWADA, 19), (1998-JAN-21/AM0315/KAWADA, 19), (1998-DEC-22/PM0945/KAWADA, 19), (1998-DEC-29/AM0315/KAWADA, 19) }
D₁={ }
S_c={ (1998-JAN-01/PM1130/KAWADA, 19), (1998-JAN-17/PM0330/KAWADA, 19), (1998-DEC-23/PM0930/KAWADA, 19), (1998-DEC-27/AM1100/KAWADA, 19), (1998-JAN-01/PM1045/KAWADA, 19), (1998-JAN-21/AM0315/KAWADA, 19), (1998-DEC-22/PM0945/KAWADA, 19), (1998-DEC-29/AM0315/KAWADA, 19) }

*STEP 18
(M, p, q)=(SUKAWA, -1, 5)
A={ (1998-JAN-04/AM0615/SUKAWA, 21), (1998-JAN-05/AM0100/SUKAWA, 21), (1998-DEC-30/AM1100/SUKAWA, 21), (1998-DEC-31/PM0400/SUKAWA, 21) }
D={ }

*STEP 19
D₁={ (WADA, 3, 7) }

*STEP 20
(M, t, u)=(WADA, 3, 7)
B={ (1998-JAN-01/AM0815/WADA, 17), (1998-DEC-29/AM0315/KAWADA, 19) }
A₁={ }
D₁={ }

フロントページの続き

- (56)参考文献 特開平11-073429(JP,A)
特開平10-334118(JP,A)
特開平08-249354(JP,A)
野口直彦、外2名、極大単語索引と複合語辞書を用いた高精度な全文検索、情報処理学会第56回(平成10年前期)全国大会講演論文集(3)、1998年 3月19日、p.118-119
野口直彦、外5名、極大単語索引を用いた知的検索ソフトウェアMEISTER、情報処理学会第55回(平成9年後期)全国大会講演論文集(3)、1997年 9月26日、p.89-102
倉知一晃、外3名、日本語文書に対する新しい索引検索方式、情報処理学会第50回(平成7年前期)全国大会講演論文集(4)、1995年 3月17日、p.41-44

- (58)調査した分野(Int.Cl.⁷, DB名)
G06F17/30