(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第3965798号
(P3965798)
(45)発行日　平成19年8月29日(2007.8.29)　(24)登録日　平成19年6月8日(2007.6.8)

(51)Int.Cl.　F I
G06F 17/30 (2006.01)　G06F 17/30 140
G06F 17/30 419A

請求項の数 9　(全 52 頁)

(21)出願番号　特願平10-253427
(22)出願日　平成10年9月8日(1998.9.8)
(65)公開番号　特開2000-90091(P2000-90091A)
(43)公開日　平成12年3月31日(2000.3.31)
審査請求日　平成16年2月18日(2004.2.18)

(73)特許権者　000005496
富士ゼロックス株式会社
東京都港区赤坂九丁目７番３号
(74)代理人　100092152
弁理士　服部　毅巖
(72)発明者　門馬　敦仁
神奈川県足柄上郡中井町境４３０　グリーンテクなかい　富士ゼロックス株式会社内

審査官　辻本　泰隆

(56)参考文献　特開平０８－０３０６２０（ＪＰ，Ａ）
特開平０４－２８１５６８（ＪＰ，Ａ）
特開平８－１９０５４２（ＪＰ，Ａ）

最終頁に続く

(54)【発明の名称】データ処理装置、文書処理装置、データ処理プログラムを記録したコンピュータ読み取り可能な記録媒体、文書処理プログラムを記録したコンピュータ読み取り可能な記録媒体、データ処理方

(57)【特許請求の範囲】

【請求項１】

有向順序木で表現された情報に対する処理を行うデータ処理装置において、

複数のノードと該ノード間の接続関係とを有向順序木で表現した処理対象情報を格納する情報格納手段と、

有向順序木を構成するノードの内容に関する条件を示す述語と、該述語の条件を満たすノード間の接続関係に関する条件を記述したパターン指定とを含む検索条件が入力され、該検索条件のいずれかのパターン指定において接続関係における下位のノードの置き換えを許容することが示されている場合は、前記情報格納手段に格納された前記処理対象情報の有向順序木に対し、該パターン指定で示される接続関係における上位ノードの直下の処理対象のノードを、該処理対象のノードの直下のノード列に置換する操作を行った結果得られる均質化有向順序木を処理対象として、前記均質化有向順序木内の前記検索条件に適合する有向順序木を抽出するマッチング手段と、

を有することを特徴とするデータ処理装置。

【請求項２】

前記マッチング手段は、前記検索条件のいずれかのパターン指定において接続関係における下位のノードの読み飛ばしを許容することが示されている場合は、前記処理対象情報の有向順序木に対し、該パターン指定で示される接続関係の上位ノードの直下の処理対象のノードを頂点とする部分木の一部を読み飛ばした結果得られる均質化有向順序木を処理対象とすることを特徴とする請求項１記載のデータ処理装置。

【請求項３】
前記マッチング手段は、ノードの属性の指定を含む前記検索条件が入力された場合には、前記処理対象のノードに定義されている属性を、該処理対象のノードから置換されたノード列の属性とみなして、前記均質化有向順序木内の前記検索条件に適合する有向順序木の出力を行うことを特徴とする請求項１記載のデータ処理装置。
【請求項４】
前記マッチング手段は、ノード属性の変換規則を含む前記検索条件が入力された場合には、前記検索条件に適合する有向順序木の各ノードに定義されている属性を対応する変換規則に従って変換した結果得られる有向順序木を出力することを特徴とする請求項１記載のデータ処理装置。
【請求項５】
有向順序木で表現された構造化文書に対する処理を行う文書処理装置において、
複数の文書要素と該文書要素間の接続関係とを有向順序木で表現した処理対象構造化文書を格納する情報格納手段と、
有向順序木を構成する文書要素の内容に関する条件を示す述語と、該述語の条件を満たす文書要素間の接続関係に関する条件を記述したパターン指定とを含む検索条件が入力され、該検索条件のいずれかのパターン指定において接続関係における下位の文書要素の置き換えを許容することが示されている場合は、前記情報格納手段に格納された前記処理対象構造化文書の有向順序木に対し、該パターン指定で示される接続関係における上位文書要素の直下の処理対象の文書要素を、該処理対象の文書要素の直下の文書要素列に置換する操作を行った結果得られる均質化有向順序木を処理対象として、前記均質化有向順序木内の前記検索条件に適合する有向順序木を抽出するマッチング手段と、
前記マッチング手段が抽出した有向順序木を論理構造とする構造化文書に対して既定の処理を実行する文書処理手段と、
を有することを特徴とする文書処理装置。
【請求項６】
有向順序木で表現された情報に対する処理を行うデータ処理プログラムを記録したコンピュータ読み取り可能な記録媒体において、
複数のノードと該ノード間の接続関係とを有向順序木で表現した処理対象情報を格納する情報格納手段、
有向順序木を構成するノードの内容に関する条件を示す述語と、該述語の条件を満たすノード間の接続関係に関する条件を記述したパターン指定とを含む検索条件が入力され、該検索条件のいずれかのパターン指定において接続関係における下位のノードの置き換えを許容することが示されている場合は、前記情報格納手段に格納された前記処理対象情報の有向順序木に対し、該パターン指定で示される接続関係における上位ノードの直下の処理対象のノードを、該処理対象のノードの直下のノード列に置換する操作を行った結果得られる均質化有向順序木を処理対象として、前記均質化有向順序木内の前記検索条件に適合する有向順序木を抽出するマッチング手段、
としてコンピュータを機能させることを特徴とするデータ処理プログラムを記録したコンピュータ読み取り可能な記録媒体。
【請求項７】
有向順序木で表現された構造化文書に対する処理を行う文書処理プログラムを記録したコンピュータ読み取り可能な記録媒体において、
複数の文書要素と該文書要素間の接続関係とを有向順序木で表現した処理対象構造化文書を格納する情報格納手段、
有向順序木を構成する文書要素の内容に関する条件を示す述語と、該述語の条件を満たす文書要素間の接続関係に関する条件を記述したパターン指定とを含む検索条件が入力され、該検索条件のいずれかのパターン指定において接続関係における下位の文書要素の置き換えを許容することが示されている場合は、前記情報格納手段に格納された前記処理対象構造化文書の有向順序木に対し、該パターン指定で示される接続関係における上位文書

要素の直下の処理対象の文書要素を、該処理対象の文書要素の直下の文書要素列に置換する操作を行った結果得られる均質化有向順序木を処理対象として、前記均質化有向順序木内の前記検索条件に適合する有向順序木を抽出するマッチング手段、
前記マッチング手段が抽出した有向順序木を論理構造とする構造化文書に対して既定の処理を実行する文書処理手段、
としてコンピュータを機能させることを特徴とする文書処理プログラムを記録したコンピュータ読み取り可能な記録媒体。
【請求項８】
有向順序木で表現された情報に対する処理を行うためのデータ処理方法において、
情報格納手段が、複数のノードと該ノード間の接続関係とを有向順序木で表現した処理対象情報を格納し、
マッチング手段が、有向順序木を構成するノードの内容に関する条件を示す述語と、該述語の条件を満たすノード間の接続関係に関する条件を記述したパターン指定とを含む検索条件が入力され、該検索条件のいずれかのパターン指定において接続関係における下位のノードの置き換えを許容することが示されている場合は、前記情報格納手段に格納された前記処理対象情報の有向順序木に対し、該パターン指定で示される接続関係における上位ノードの直下の処理対象のノードを、該処理対象のノードの直下のノード列に置換する操作を行った結果得られる均質化有向順序木を処理対象として、前記均質化有向順序木内の前記検索条件に適合する有向順序木を抽出する、
ことを特徴とするデータ処理方法。
【請求項９】
有向順序木で表現された情報に対する処理を行うための文書処理方法において、
情報格納手段が、複数の文書要素と該文書要素間の接続関係とを有向順序木で表現した処理対象構造化文書を格納し、
マッチング手段が、有向順序木を構成する文書要素の内容に関する条件を示す述語と、該述語の条件を満たす文書要素間の接続関係に関する条件を記述した検索条件が入力され、該検索条件のいずれかのパターン指定において接続関係における下位の文書要素の置き換えを許容することが示されている場合は、前記情報格納手段に格納された前記処理対象構造化文書の有向順序木に対し、該パターン指定で示される接続関係における上位文書要素の直下の処理対象の文書要素を、該処理対象の文書要素の直下の文書要素列に置換する操作を行った結果得られる均質化有向順序木を処理対象として、前記均質化有向順序木内の前記検索条件に適合する有向順序木を抽出し、
文書処理手段が、前記マッチング手段が抽出した有向順序木を論理構造とする構造化文書に対して既定の処理を実行する、
ことを特徴とする文書処理方法。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】
本発明は有向順序木で表現された情報に対する処理を行うデータ処理装置および構造化文書を対象とする文書処理装置に関し、特に大量の情報に対して処理を行うデータ処理装置および大量の構造化文書から必要な部分構造を検索して処理する文書処理装置に関する。
【０００２】
【従来の技術】
オフィスなどで利用される文書の再利用に関する利便性を向上させるために、文書を電子的に管理することが広く行われている。さらに、章、節、段落、図表などの文書要素レベルでの再利用を可能とするために、文書要素と要素間の論理的な関係（論理構造）を伴う文書（構造化文書）による文書管理が先進的なユーザによって行われている。なお、以下単に「文書」といった場合には、すべて構造化文書を指すものとする。
【０００３】
図６８は、構造化文書の第１の例を示す図である。図６８に示す文書９１の矩形は文書要

素を表わし、矩形中の文字列は文書要素の名前を表わす。各文書要素には、「001 」～「012 」の識別子が付与されている。この図に示すように、構造化文書の論理構造は有向順序木として表現される。また、文書要素に接続された楕円は、文書要素に対応する内容を表わす。

【０００４】

なお、図６８では、文書要素には名前のみが対応付けられているが、名前以外にもさまざまな属性を対応付けることが広く行われている。構造化文書フォーマットの標準としては、SGML(Standard Generalized Markup Language)とXML(eXtensible Markup Language) が広く知られている。SGMLは１９８６年にISO(International Standard Organization)の標準として規定され、現在に至るまで主に出版分野における電子文書フォーマットとして利用されてきた。いっぽうXML は、インターネット上の構造化文書フォーマットの標準としてSGMLの一部の機能を継承しつつインターネット文書フォーマットの実質的な標準であるHTML(Hyper Text Markup Language)による運用の知見を取り入れて、１９９７年の２月にW3C(World Wide Web Consortium)の勧告として制定された。W3C は、インターネット関連の標準を制定するための国際的な非営利団体である。

【０００５】

XML 文書は、図６８の文書のように有向順序木で表現される論理構造を持つ。HTML文書も有向順序木で表現される論理構造を備えているが、HTMLは文書の表示処理を主眼において設計されているため、表示処理では不要な要素名や要素間の関係を利用できないように、文書構造に対してあらかじめ制限が加えられている。いっぽう、XML 文書では、論理構造が有向順序木で表現されていればよく、要素名や要素間の関係に関する制限を文書作成者自身が定義できる。とはいえ、文書作成者が好き勝手に要素名を決めてしまっては、文書を利用するシステムが文書要素を誤った用途に用いることにもなりかねない。このため、W3C では、XMＬの拡張として、XML 文書の要素名の意味を文書処理システムが一意に特定するための機構の標準化活動を進めている。この標準化が達成されれば、インターネット上の構造化文書を文書要素単位で再利用するためのインフラが整うことになる。以上のことから、インターネット上に存在する大量かつ多様な構造をもつ構造化文書から、文書利用者が必要とする文書要素を検索する必要性が増大することは明らかである。

【０００６】

ここで、構造化文書を対象とした検索処理について説明する。例えば、検索対象となる構造化文書として、図６８に示した文書９１以外に以下のような文書を想定する。

【０００７】

図６９は、構造化文書の第２の例を示す図である。図６９の文書９２は、図６８の文書９１と類似した内容を、別の構造で表したものである。この文書９２は、根となる識別子「001 」の文書要素の子供の要素として、他の全ての文書要素が接続されている。

【０００８】

図６８の文書９１と図６９の文書９２から、参考文献エントリに対応する文書要素（以下、エントリ要素）を検索する処理を考える。なお、参考文献エントリとは、図６８の"1.momma."や"1.numata..." に対応する文書要素、及び図６９の" ［1 ］門馬．．．．" や" ［2］沼田．．．．" に対応する文書要素を指すものとする。

【０００９】

文書９１と文書９２のエントリ要素を比較すると、さまざまな相違点が見出される。たとえば、文書９１のエントリ要素の名前は"ITEM"であるのに対して文書９２のエントリ要素の名前は"PARA"である。また、文書９１のエントリ要素はリストを意味するLIST要素でまとめられているのに対し文書９２のエントリ要素は文書全体を意味するDOC 要素の直下に位置している。

【００１０】

このようにインターネット上の構造化文書の論理構造では、同一の用途で用いられるべき文書内容が、多様な構造で表現されるようになる。

構造化文書を対象とした検索では、文書要素の名前、属性や文書要素に対応付けられた内

容に関する条件だけでなく、文書要素間の接続関係も検索条件に利用することにより、再現率を維持したまま適合率を向上させることができる。例えば特開平７－２２５７７０号公報記載のデータ検索装置では、検索対象となる構造化文書の文書要素間の祖孫関係に関する索引をあらかじめ作成しておくことにより、文書要素間の接続関係を利用した高速な検索を可能としている。

【００１１】

【発明が解決しようとする課題】

しかし、上述の技術は、検索対象の文書構造の類似度が高い場合には大きな効果を発揮するが、前述の文書９１と文書９２のように、文書構造上の類似度の比較的低い文書を一括して検索する場合には、文書要素間の祖孫関係に関する条件を利用するのが困難となる。たとえば、文書９１と文書９２のエントリ要素を一括して検索する場合、両者のエントリ要素に共通する祖孫関係は、エントリ要素が文書全体を意味するDOC 要素の子孫であるという関係のみである。上記の技術によりこれらの文書を同時に検索するには、DOC 要素の子孫として特定のエントリ要素が存在するという検索条件を入力しなければならない。この場合、DOC 要素は文書全体を表しているに過ぎないため、文書要素間の接続関係を利用せずに検索した場合と同様の検索結果しか得ることができず、適合率の向上が図れない。

【００１２】

しかも、過去に一度参照した文書の検索を行う場合、利用者がその文書の構造を詳細に記憶していることは少ない。そのため、文書要素間の接続関係を用いた検索を行う場合においても、曖昧な記憶に基づいて検索条件が入力される。従来の技術では、検索条件において指定された構造に適合した場合にのみ検出されるため、曖昧な記憶に基づいた検索条件では利用者が探している文書が検出されない可能性がある。

【００１３】

したがって、利用者が文書を閲覧した際に同等の文書と認識される範囲内の一定の処理を施すことで検索条件に適合するような文書であれば、その文書が検出できるような文書検索処理を行えることが望まれている。

【００１４】

ところで、有向順序木で表現されるのは構造化文書だけではない。オブジェクト指向プログラミングにおけるプログラムの関係も有向順序木で表現することができる。従って、有向順序木で表現される各種情報においても同様に、一定の処理を行うことで検索条件に適合する論理構造を、処理対象となる情報の中から抽出できることが望まれる。

【００１５】

本発明はこのような点に鑑みてなされたものであり、有向序木で表現された情報に対してあいまいな検索条件が入力されても、検索条件に合致する部分構造を抽出することができるデータ処理装置を提供することを目的とする。

【００１６】

また、本発明の第２の目的は、構造化文書に一定の処理を施すことにより、利用者により指定された論理構造に合致する部分構造が生成される場合には、その合致した部分構造を抽出することができる文書処理装置を提供することである。

【００１７】

また、本発明の第３の目的は、有向順序木で表現された情報に一定の処理を施すことにより、利用者により指定された論理構造に合致する部分構造が生成される場合には、その合致した部分構造を抽出するような処理をコンピュータに行わせることができるデータ処理プログラムを記録したコンピュータ読み取り可能な記録媒体を提供することである。

【００１８】

また、本発明の第４の目的は、構造化文書に一定の処理を施すことにより、利用者により指定された論理構造に合致する部分構造が生成される場合には、その合致した部分構造を抽出するような処理をコンピュータに行わせることができる文書処理プログラムを記録したコンピュータ読み取り可能な記録媒体を提供することである。

【００１９】

【課題を解決するための手段】
本発明では上記課題を解決するために、有向順序木で表現された情報に対する処理を行うデータ処理装置において、複数のノードと該ノード間の接続関係とを有向順序木で表現した処理対象情報を格納する情報格納手段と、有向順序木を構成するノードの内容に関する条件を示す述語と、該述語の条件を満たすノード間の接続関係に関する条件を記述したパターン指定とを含む検索条件が入力され、該検索条件のいずれかのパターン指定において接続関係における下位のノードの置き換えを許容することが示されている場合は、前記情報格納手段に格納された前記処理対象情報の有向順序木に対し、該パターン指定で示される接続関係における上位ノードの直下の処理対象のノードを、該処理対象のノードの直下のノード列に置換する操作を行った結果得られる均質化有向順序木を処理対象として、前記均質化有向順序木内の前記検索条件に適合する有向順序木を抽出するマッチング手段と、を有することを特徴とするデータ処理装置が提供される。

【００２０】
このようなデータ処理装置によれば、有向順序木を構成するノードの内容に関する条件を示す述語と、該述語の条件を満たすノード間の接続関係に関する条件を記述したパターン指定とを含む検索条件が入力され、該検索条件のいずれかのパターン指定において接続関係における下位のノードの置き換えを許容することが示されている場合は、マッチング手段により、情報格納手段に格納された処理対象情報の有向順序木に対し、該パターン指定で示される接続関係における上位ノードの直下の処理対象のノードを、該処理対象のノードの直下のノード列に置換する操作を行った結果得られる均質化有向順序木を処理対象として、均質化有向順序木内の検索条件に適合する有向順序木の抽出処理が行われる。

【００２３】
また、上記課題を解決するために、有向順序木で表現された構造化文書に対する処理を行う文書処理装置において、複数の文書要素と該文書要素間の接続関係とを有向順序木で表現した処理対象構造化文書を格納する情報格納手段と、有向順序木を構成する文書要素の内容に関する条件を示す述語と、該述語の条件を満たす文書要素間の接続関係に関する条件を記述したパターン指定とを含む検索条件が入力され、該検索条件のいずれかのパターン指定において接続関係における下位の文書要素の置き換えを許容することが示されている場合は、前記情報格納手段に格納された前記処理対象情報の有向順序木に対し、該パターン指定で示される接続関係における上位文書要素の直下の処理対象の文書要素を、該処理対象の文書要素の直下の文書要素列に置換する操作を行った結果得られる均質化有向順序木を処理対象として、前記均質化有向順序木内の前記検索条件に適合する有向順序木を抽出するマッチング手段と、前記マッチング手段が抽出した有向順序木を論理構造とする構造化文書に対して既定の処理を実行する文書処理手段と、を有することを特徴とする文書処理装置が提供される。

【００２４】
このような文書処理装置によれば、有向順序木を構成する文書要素の内容に関する条件を示す述語と、該述語の条件を満たす文書要素間の接続関係に関する条件を記述したパターン指定とを含む検索条件が入力され、該検索条件のいずれかのパターン指定において接続関係における下位の文書要素の置き換えを許容することが示されている場合は、マッチング手段により、情報格納手段に格納された処理対象構造化文書の有向順序木に対し、該パターン指定で示される接続関係における上位文書要素の直下の処理対象の文書要素を、該処理対象の文書要素の直下の文書要素列に置換する操作を行った結果得られる均質化有向順序木を処理対象として、均質化有向順序木内の検索条件に適合する有向順序木の抽出処理が行われる。すると、文書処理手段により、マッチング手段が抽出した有向順序木を論理構造とする構造化文書に対して既定の処理が行われる。

【００２７】
また、上記課題を解決するために、有向順序木で表現された情報に対する処理を行うデータ処理プログラムを記録したコンピュータ読み取り可能な記録媒体において、複数のノードと該ノード間の接続関係とを有向順序木で表現した処理対象情報を格納する情報格納

手段、有向順序木を構成するノードの内容に関する条件を示す述語と、該述語の条件を満たすノード間の接続関係に関する条件を記述したパターン指定とを含む検索条件が入力され、該検索条件のいずれかのパターン指定において接続関係における下位のノードの置き換えを許容することが示されている場合は、前記情報格納手段に格納された前記処理対象情報の有向順序木に対し、該パターン指定で示される接続関係における上位ノードの直下の処理対象のノードを、該処理対象のノードの直下のノード列に置換する操作を行った結果得られる均質化有向順序木を処理対象として、前記均質化有向順序木内の前記検索条件に適合する有向順序木を抽出するマッチング手段、としてコンピュータを機能させることを特徴とするデータ処理プログラムを記録したコンピュータ読み取り可能な記録媒体が提供される。

【００２８】

このような記録媒体に記録されたデータ処理プログラムをコンピュータに実行させれば、有向順序木を構成するノードの内容に関する条件を示す述語と、該述語の条件を満たすノード間の接続関係に関する条件を記述したパターン指定とを含む検索条件が入力され、該検索条件のいずれかのパターン指定において接続関係における下位のノードの置き換えを許容することが示されている場合は、情報格納手段に格納された処理対象情報の有向順序木に対し、該パターン指定で示される接続関係における上位ノードの直下の処理対象のノードを、該処理対象のノードの直下のノード列に置換する操作を行った結果得られる均質化有向順序木を処理対象として、均質化有向順序木内の検索条件に適合する有向順序木を抽出するような処理機能がコンピュータ上に構築される。

【００２９】

また、上記課題を解決するために、有向順序木で表現された構造化文書に対する処理を行う文書処理プログラムを記録したコンピュータ読み取り可能な記録媒体において、複数の文書要素と該文書要素間の接続関係とを有向順序木で表現した処理対象構造化文書を格納する情報格納手段、有向順序木を構成する文書要素の内容に関する条件を示す述語と、該述語の条件を満たす文書要素間の接続関係に関する条件を記述したパターン指定とを含む検索条件が入力され、該検索条件のいずれかのパターン指定において接続関係における下位の文書要素の置き換えを許容することが示されている場合は、前記情報格納手段に格納された前記処理対象構造化文書の有向順序木に対し、該パターン指定で示される接続関係における上位文書要素の直下の処理対象の文書要素を、該処理対象の文書要素の直下の文書要素列に置換する操作を行った結果得られる均質化有向順序木を処理対象として、前記均質化有向順序木内の前記検索条件に適合する有向順序木を抽出するマッチング手段、前記マッチング手段が抽出した有向順序木を論理構造とする構造化文書に対して既定の処理を実行する文書処理手段、としてコンピュータを機能させることを特徴とする文書処理プログラムを記録したコンピュータ読み取り可能な記録媒体が提供される。

【００３０】

このような記録媒体に記録された文書処理プログラムをコンピュータに実行させれば、有向順序木を構成する文書要素の内容に関する条件を示す述語と、該述語の条件を満たす文書要素間の接続関係に関する条件を記述したパターン指定とを含む検索条件が入力され、該検索条件のいずれかのパターン指定において接続関係における下位の文書要素の置き換えを許容することが示されている場合は、情報格納手段に格納された処理対象構造化文書の有向順序木に対し、該パターン指定で示される接続関係における上位文書要素の直下の処理対象の文書要素を、該処理対象の文書要素の直下の文書要素列に置換する操作を行った結果得られる均質化有向順序木を処理対象として、均質化有向順序木内の検索条件に適合する有向順序木が抽出され、文書処理手段により、マッチング手段が抽出した有向順序木を論理構造とする構造化文書に対して既定の処理を実行するような処理機能がコンピュータ上に構築される。

【００３１】

【発明の実施の形態】

以下、本発明の実施の形態を図面を参照して説明する。

図１は、本発明の原理構成図である。本発明のデータ処理装置は、情報格納手段１とマッチング手段２とからなる。

【００３２】

情報格納手段１は、有向順序木で表現された処理対象情報１ａを格納する。

マッチング手段２は、有向順序木を構成するノードの内容及びノード間の接続関係に関する条件を記述した検索条件３が入力されると、情報格納手段１に格納された処理対象情報１ａの有向順序木の中間ノードを削除し、中間ノードのあった位置に中間ノード直下のノード列を配置する操作を行った結果得られる均質化有向順序木を生成する。この均質化有向順序木は複数生成される。そして、生成された均質化有向順序木を処理対象として、均質化有向順序木内の検索条件に適合する有向順序木４を抽出する。

【００３３】

このようなデータ処理装置によれば、検索条件３によって指定したノード間の接続関係が曖昧であっても、処理対象情報１ａの中間ノードを削除し、その下位のノード列を削除した中間ノードのあった位置に配置することで検索条件に適合する構造が生成できれば、その構造を示す有向順序木が抽出される。したがって、曖昧な記憶に基づいて検索条件を定義しても、利用者の意図した構造の情報を処理対象情報１ａ内から抽出可能となる。

【００３４】

ところで、有向順序木で表現される情報の主なものとして構造化文書がある。構造化文書に対して本発明に係るデータ処理を行えば、文書構造を検索条件に用いて、比較的類似度の低い構造化文書群に対する検索処理を有効に行うことができる。そこで、このような文書検索を行うことができる文書処理装置を、第１の実施の形態として以下に説明する。

【００３５】

図２は、文書処理システムの構成を示すブロック図である。文書処理システムは、文書処理装置１０と入出力装置２０とからなる。文書処理装置１０は、文書保持部１１、検索条件保持部１２、階層オートマトン生成部１３、マッチング部１４、及び文書処理部１５で構成される。

【００３６】

文書保持部１１は、処理対象の構造化文書を保持する。検索条件保持部１２は、検索条件を保持する。本実施の形態における検索条件は、有向順序グラフとして表現される。階層オートマトン生成部１３は、入出力装置２０を介した利用者３０からの検索指令によって、検索条件保持部１２中の検索条件を入力として階層オートマトンを生成する。生成した階層オートマトンは、マッチング部１４に入力される。マッチング部１４は、階層オートマトンと文書保持部１１中の構造化文書を入力として、論理構造のマッチング処理を行う。マッチングの結果として得られる文書要素セット（文書要素の集合）の集合は、文書処理部１５に出力される。文書処理部１５は、文書要素セットの集合を入力として、文書の自動生成や表示、印刷などの文書処理を実行し、入出力装置２０を介して処理結果を利用者３０に提供する。

【００３７】

入出力装置２０は、キーボードやマウスなどの入力装置と、ＣＲＴ(Cathode Ray Tube)、ＬＣＤ(Liquid Crystal Display)などの表示装置、及びプリンタなどの出力装置からなる。利用者３０が入出力装置２０から指示を入力すると、その指示は文書処置装置に入力される。また、文書処置装置の処置結果が表示装置の画面を通じて文書の利用者３０に通知される。あるいは、処理結果が出力装置を介して紙に印刷される。

【００３８】

なお、図２の構成と図１に示した原理図との対応関係は次のとおりである。すなわち、図１の情報格納手段１は、文書保持部１１に対応する。マッチング手段２は、マッチング部１４に対応する。

【００３９】

図２のような構成のシステムにおいて以下のような処理が行われる。

図３は、文書処理装置の処理手順を示すフローチャートである。以下の処理をステップ番

号に沿って説明する（後述する他のフローチャートにおいても同様）。なお、この処理は、利用者３０によって検索指令が入力された際に開始される。検索指令には、検索条件保持部１２内のどの検索条件によって検索を行うのかが指定されている。
［Ｓ１］階層オートマトン生成部１３が、利用者３０によって指定された検索条件を検索条件保持部１２から抽出し、その検索条件に基づいて階層オートマトンを生成する。
［Ｓ２］マッチング部１４が、文書保持部１１中の構造化文書と、階層オートマトン生成部１３が生成した階層オートマトンとのマッチングを行う。
［Ｓ３］文書処理部１５が、マッチング結果の文書要素の集合を対象とした文書処理を行う。

【００４０】

次に、図３の各ステップの処理内容を、具体例を交えながら詳細に説明する。この例では、文書保持部１１内に、図６８に示した文書９１に加えて、以下のような文書が保持されているものとする。

【００４１】

図４は、構造化文書の第３の例を示す図である。この文書９３は、図６８の文書９１と同様の内容を有しているが、論理構造が異なる。文書９１では、参考文献エントリの並びを意味するLIST要素（識別子「007 」）が参考文献全体を意味するSECT要素（識別子「005 」）直下の子供として表現されている。いっぽう、文書９３では、参考文献エントリの並びを意味する要素は存在せず、参考文献エントリはエントリ数が１のLISTリスト要素（識別子「007 」「008 」）として表現されている。

【００４２】

なお、以下の説明では、文書要素を指示するときに論理構造上の文書要素間の接続関係を用いることがある。たとえば、「図６８の文書９１の識別子「002 」の文書要素の長子」とは、「図６８の文書９１の識別子が「002 」である文書要素の先頭の子供要素」、すなわち識別子が「003 」の文書要素を意味するものとする。また、「図６８の文書９１の識別子が「002 」である文書要素の弟要素」とは「図６８の文書の識別子「002 」の文書要素と親要素を共有する要素で、識別子「002 」の文書要素より後に出現する要素のうち最初に出現する要素」すなわち識別子が「005 」の文書要素を意味するものとする。

【００４３】

また、検索条件保持部１２には、以下のような検索条件が保持されているものとする。
図５は、検索条件の例を示す図である。この検索条件４０は、破線で示された矩形（検索条件ノード）からなる有向順序グラフとして表現される。この検索条件４０は、図６８の文書９１と図４の文書９３に含まれている参考文献エントリを一括して得るための検索条件である。

【００４４】

図中実線で示した矩形は、文書要素を引数とする述語４１，４３，４５，４７である。この例では、文書要素に対応づけられたテキスト内容がマッチすべきパターン、及び論理演算子が矩形中に示されている。たとえば、述語４３に示す「" 参考文献"or"References" 」という記述は、「" 参考文献" または「References" という文字列を含むテキスト内容に対応づけられた文書要素が引数に与えられた場合、真を返す」という条件を意味する。
また、述語４５に示す「" ［（数字）＊］"or"（数字）＊．" 」という記述は、「" ［１］、［２］、・・・" または" １．、２．、・・・" という文字列を含むテキスト内容に対応づけられた文書要素が引数に与えられた場合、真を返す」という条件を意味する。

【００４５】

述語に対応する文書要素の下部構造に関する条件を設定するために、パターン指定４２，４４，４６を述語の直下に記述することができる。パターン指定４２，４４，４６は、直上の述語に対応する文書要素直下の文書要素列のパターンを指定する。図５では、パターン指定は楕円で表現されている。図５におけるパターン指定４２，４４，４６中の文字列はパターンの種類を意味する。たとえば、"SEQ" は、直下の文書要素列が図で示された順序（左側が上位）で出現することを意味する。また、"OPTREP"は、直下の文書要素が０回

以上任意の回数出現することを意味する。

【００４６】

図５に示したとおり、パターン指定の直下には述語またはパターン指定の列が出現する。検索条件では、パターン指定と述語の関係を階層的に記述できることから、検索条件を階層的なパターン指定と見なすことができる。さらに図５では用いられていないが、直下の文書要素列のいずれか１つが一度だけ出現することを意味する"CHO" をパターン指定中の文字列("SEQ"および"OPTREP") に追加すれば、検索条件によって、前述の述語からなる任意の階層的な正規表現を表現することが可能となる。

【００４７】

また、パターン指定には特別な記号"+r"を追加することができる。"+r"は、パターン指定直下の文書要素列を決定する処理において、処理対象の文書要素をその文書要素直下の文書要素列と置き換えて処理を進めてもよいこと意味する。たとえば、処理対象の文書要素が図６８のLIST要素（識別子「007 」）であった場合、LIST要素と２つのITEM要素（識別子「008 」、「009 」）からなる要素列を置き換えて処理を進めることができる。もちろん、LIST要素をそのまま用いて処理を進めることもできる。

【００４８】

以上のような文書及び検索条件が保持されている状態で、利用者が図５に示した検索条件４０に適合する構造を有する文書を検索する場合を考える。その場合、まず階層オートマトン生成部１３により階層オートマトンが生成される。以下に、図５の検索条件４０から生成される階層オートマトンを示す。

【００４９】

図６は、階層オートマトンの第１の例を示す図である。この階層オートマトン１００は、４つのオートマトン１１０，１２０，１３０，１４０から構成されている。これらのオートマトンの最上位に位置するオートマトン１１０を、根オートマトンと呼ぶ。

【００５０】

オートマトン１１０，１２０，１３０，１４０の中で、○で示されているのが状態である。状態には、それぞれ「11」、「12」、「21」、「22」、「23」、「31」、「32」、「33」、「41」、「42」、「43」、「44」の識別子が割り振られている。そのうち、始点のない矢印で指される状態「11」、「21」、「31」、「41」が初期状態である。また、◎で示されている状態「12」、「23」、「33」、「44」が受理状態である。さらに、矢印で示されているのが状態遷移であり、状態遷移には、それぞれ「1a」、「2a」、「2b」、「3a」、「3b」、「4a」、「4b」、「4c」、「4d」、「4e」の識別子が割り振られている。状態遷移は、矢印の前の状態で所定の条件を満たす入力があれば、先の状態へ移ることを意味する。

【００５１】

オートマトンの状態遷移には、中央に矩形が記されているものと、矩形の記されていないものとがある。前者は述語を伴う検索条件ノードに対応する状態遷移を意味し、後者は述語を伴わない検索条件ノードに対応する状態遷移を意味する。また、矩形の記述は、状態遷移と文書要素とを対応付けるための条件である。この条件の意味は、検索条件ノードの記述の意味と同一である。図中のオートマトン１３０，１４０には、右上に"+r"が記されている。この記号は、オートマトン上で文書要素と直下の文書要素列との置き換えが許されることを意味する。

【００５２】

状態遷移からオートマトンへの参照（オートマトン間の矢印）は、マッチング部１４の処理において状態遷移がマッチング対象の構造化文書の文書要素とマッチしたときに、マッチした文書要素を頂点とする部分木のマッチングに用いるオートマトンへの参照を意味する。部分木のマッチングの結果、参照したオートマトンにおいて受理状態に達すれば、参照した矢印の元に位置する状態遷移の先の状態へ移る。

【００５３】

オートマトン１４０には、「ε」のラベルが付与された状態遷移がある。これはε遷移で

あり、空列を読んで次の状態へ移ることを示している。すなわち、入力が無くても次の状態へ遷移できることを意味する。

【００５４】

以下に、階層オートマトン生成部１３の処理の詳細を説明する。

図７は、階層オートマトン生成処理部が行う処理を示すフローチャートである。この処理は、すべて階層オートマトン生成部１３が行う。

［Ｓ１１］検索条件の根ノードを引数にして、根オートマトン生成処理を呼び出す。根オートマトン生成処理の結果、変数Ｔに出力された階層オートマトンが返される。この処理の詳細は、図８に示す。

［Ｓ１２］変数Ｔに出力された階層オートマトンをマッチング部１４へ出力し、処理を終了する。

【００５５】

図８は、根オートマトン生成処理手順を示すフローチャートである。この処理は、全て階層オートマトン生成部１３が行う処理である。この処理の入力は、検索条件の根ノードであり、出力は変数Ｔに対して出力されたオートマトンである。

［Ｓ２１］初期状態と受理状態とを１つずつ有するオートマトンを生成し、変数Ｔに格納する。

［Ｓ２２］検索条件の根ノードに保持されている条件を伴う状態遷移を生成する。そして、生成した状態遷移の開始端をステップＳ２１で格納したオートマトンの初期状態とし、生成した状態遷移の終了端をステップＳ２１で格納したオートマトンの受理状態とする。

［Ｓ２３］検索条件の根ノードを含む検索条件ノードと状態遷移を引数にして、オートマトン生成処理を呼び出す。この処理の詳細は、図９に示す。

【００５６】

図９は、オートマトン生成処理手順を示すフローチャートの前半である。この処理は、すべて階層オートマトン生成部１３が行う。この処理の入力は検索条件ノードと状態遷移であり、出力は変数Ｔへ出力されるオートマトンである。

［Ｓ３１］検索条件ノードに対応するオートマトンが、変数Ｔに含まれるか否かを判断する。含まれる場合は処理をステップＳ３２に進め、含まれない場合は処理をステップＳ３３に進める。

［Ｓ３２］入力の状態遷移から検索条件ノードに対応するオートマトンへの参照を生成し、処理を終了する。

［Ｓ３３］オートマトンを生成し、入力の状態遷移から生成したオートマトンへの参照を生成する。

［Ｓ３４］入力の検索条件ノードのパターン指定に"+r"が設定されているか否かを判断する。"+r"が設定されていれば処理をステップＳ３５に進め、設定されていなければ処理をステップＳ３６に進める。

［Ｓ３５］生成したオートマトンに"+r"を設定する。

［Ｓ３６］生成したオートマトン上に、初期状態と受理状態を生成する。

［Ｓ３７］検索条件ノードのパターン指定の種類を判断する。パターン指定が"SEQ" なら処理をステップＳ３８に進め、パターン指定が"CHO" なら処理をステップＳ３９に進め、パターン指定が"OPTREP"なら処理をステップＳ４０に進める。

【００５７】

図１０は、オートマトン生成処理手順を示すフローチャートの後半である。

［Ｓ３８］生成したオートマトンを引数にしてSEQ パターン指定生成処理を呼び出す。その後、処理をステップＳ４１に進める。SEQ パターン指定生成処理の詳細は、図１１、図１２において説明する。

［Ｓ３９］生成したオートマトンを引数にしてCHO パターン指定生成処理を呼び出す。その後、処理をステップＳ４１に進める。CHO パターン指定生成処理の詳細は、図１３において説明する。

［Ｓ４０］生成したオートマトンを引数にしてOPTREPパターン指定生成処理を呼び出す。

その後、処理をステップＳ４１に進める。OPTREPパターン指定生成処理の詳細は、図１４において説明する。
［Ｓ４１］ステップＳ３８～Ｓ４０の各処理の返り値のオートマトンを変数Ｔに格納する。
【００５８】
図１１は、SEQ パターン生成処理手順を示すフローチャートの前半である。この処理は、全て階層オートマトン生成部１３が行う処理である。この処理の入力はオートマトンと検索条件ノードであり、出力はオートマトンである。
［Ｓ５１］SEQ パターン指定直下のすべての検索条件ノードを処理したか否かを判断する。すべての検索条件ノードを処理したのであれば、処理をステップＳ６１に進め、そうでなければ処理をステップＳ５２に進める。
［Ｓ５２］SEQ パターン指定直下の未処理の検索条件ノードのうち、先頭の検索条件ノードを処理対象とする。
［Ｓ５３］処理対象の検索条件ノードの位置を判断する。先頭の検索条件ノードであれば処理をステップＳ５４に進め、末尾の検索条件ノードであれば処理をステップＳ５５に進め、それらのいずれでもない検索条件ノードであれば処理をステップＳ５６に進める。
［Ｓ５４］状態を１つ生成する。そして、初期状態を開始端とし、このステップで生成した状態を終了端とした状態遷移を生成する。その後、処理をステップＳ５７に進める。
［Ｓ５５］直前に作成された状態を開始端とし、受理状態を終了端とする状態遷移を生成する。その後、処理をステップＳ５７に進める。
［Ｓ５６］状態を１つ生成する。そして、直前に生成された状態を開始端として、このステップで生成された状態を終了端とする状態遷移を生成する。
［Ｓ５７］処理対象の検索条件ノードに述語が含まれているか否かを判断する。述語が含まれている場合には処理をステップＳ５８に進め、そうでない場合には処理をステップＳ５９に進める。
［Ｓ５８］直前に生成された状態遷移に、処理対象の検索条件ノードの述語を対応付ける。
【００５９】
図１２は、SEQ パターン生成処理手順を示すフローチャートの後半である。
［Ｓ５９］処理対象の検索条件ノードにパターン指定が含まれるか否かを判断する。パターン指定が含まれる場合には処理をステップＳ６０に進め、そうでない場合には処理をステップＳ５１に進める。
［Ｓ６０］処理対象の検索条件ノードと直前に生成された状態遷移とを引数として、オートマトン生成処理（図９、図１０に示す）を呼び出す。その後、処理をステップＳ５１に進める。
［Ｓ６１（図１１に示す）］オートマトンを出力し、処理を終了する。
【００６０】
図１３は、CHO パターン生成処理手順を示すフローチャートである。この処理は、すべて階層オートマトン生成部１３が行う。この処理の入力はオートマトンと検索条件ノードであり、出力はオートマトンである。
［Ｓ７１］CHO パターン指定直下のすべての検索条件ノードを処理したか否かを判断する。すべての検索条件ノードを処理していれば処理をステップＳ７８に進め、そうでなければ処理をステップＳ７２に進める。
［Ｓ７２］CHO パターン指定直下の未処理の検索条件ノードのうち、先頭の検索条件ノードを処理対象とする。
［Ｓ７３］初期状態を開始端とし、受理状態を終了端とする状態遷移を生成する。
［Ｓ７４］処理対象の検索条件ノードに述語が含まれているか否かを判断する。述語が含まれている場合には処理をステップＳ７５に進め、述語が含まれていない場合には処理をステップＳ７６に進める。
［Ｓ７５］直前に生成された状態遷移に処理対象の検索条件ノードの述語を対応付ける。

［Ｓ７６］処理対象の検索条件ノードにパターン指定が含まれるか否かを判断する。パターン指定が含まれる場合には処理をステップＳ７７に進め、パターン指定が含まれない場合には処理をステップＳ７１に進める。
［Ｓ７７］処理対象の検索条件ノードと、直前に生成された状態遷移とを引数としてオートマトン生成処理（図９、図１０に示す）を呼び出す。その後、処理をステップＳ７１に進める。
［Ｓ７８］オートマトンを出力し、処理を終了する。
【００６１】
図１４は、OPTREPパターン生成処理手順を示すフローチャートである。この処理は、すべて階層オートマトン生成部１３が行う。この処理の入力はオートマトンと検索条件ノードであり、出力はオートマトンである。
［Ｓ８１］オートマトン上に２つの状態（第１の状態、第２の状態）を生成する。
［Ｓ８２］初期状態を開始端として第１の状態を終了端とするε遷移、初期状態を開始端として受理状態を終了端とするε遷移、第２の状態を開始端として第１の状態を終了端とするε遷移、及び第２の状態を開始端として受理状態を終了端とするε遷移を生成する。
［Ｓ８３］第１の状態を開始端として第２の状態を終了端とする状態遷移を生成する。
［Ｓ８４］処理対象の検索条件ノードに述語が含まれるか否かを判断する。述語が含まれる場合には処理をステップＳ８５に進め、述語が含まれない場合には処理をステップＳ８６に進める。
［Ｓ８５］ステップＳ８３で生成された状態遷移に、処理対象の検索条件ノードの述語を対応付ける。
［Ｓ８６］処理対象の検索条件ノードにパターン指定が含まれるか否かを判断する。パターン指定が含まれる場合には処理をステップＳ８７に進め、パターン指定が含まれない場合には処理をステップＳ８８に進める。
［Ｓ８７］処理対象の検索条件ノードと、ステップＳ８３で生成された状態遷移とを引数としてオートマトン生成処理（図９、図１０に示す）を呼び出す。
［Ｓ８８］オートマトンを出力する。
【００６２】
このような処理を階層オートマトン生成部１３が実行することにより、図５に示す検索条件４０から図６に示す階層オートマトン１００が生成される。生成された階層オートマトン１００は、マッチング部１４に渡される。
【００６３】
次に、マッチング部１４が行う処理の詳細を説明する。
図１５は、マッチング部の処理手順を示すフローチャートである。この処理は、すべてマッチング部１４によって行われる。また、この処理の入力は階層オートマトンと構造化文書であり、出力は入力の構造化文書の文書要素セット（文書要素の集合）の集合である。なお、マッチング部１４では、以下の処理で利用される変数Ｖを保持する。変数Ｖには、任意の数の文書要素セットが保持されるものとする。また、以下のステップＳ１０３の処理で利用されているスタックは、文書要素と状態との組を要素とする。
［Ｓ１０１］処理対象の文書要素を構造化文書の根文書要素とする。
［Ｓ１０２］階層オートマトンの根オートマトンの初期状態、処置対象の文書要素、空の文書要素セット、及び空のスタックを引数にしてオートマトンマッチング処理を呼び出す。この処理の詳細は、図１６に示す。
［Ｓ１０３］構造化文書の論理構造を前から順に走査した場合に、処理対象の文書要素の次の文書要素が存在するか否かを判断する。次の文書要素が存在する場合には処理をステップＳ１０４に進め、存在しない場合には処理をステップＳ１０５に進める。なお、論理構造上の文書要素の順番とは、親子の文書要素間では親の方が先の順である。兄弟の文書要素間では兄の方が先の順である。さらに、弟の文書要素は、兄の文書要素の子孫にあたる文書要素の最後尾の文書要素の次になる。文書要素の識別子は、論理構造上の順番にしたがって振られている。

［Ｓ１０４］処理対象の文書要素の次の文書要素を処理対象とする。
［Ｓ１０５］変数Ｖの値である文書要素セットを出力する。
【００６４】
図１６は、オートマトンマッチング処理手順を示す第１のフローチャートである。この処理は、すべてマッチング部１４が行う。また、この処理の入力は階層オートマトン中のオートマトンの状態、構造化文書の文書要素、文書要素セット、及びスタックであり、出力は変数Ｖに書き込まれる。なお、入力の文書要素としてNULL（空の値）を設定することも可能である。
［Ｓ１１１］入力の文書要素がスタックの先頭の文書要素であるか否かを判断する。先頭の文書要素であれば処理をステップＳ１１２に進め、そうでない場合は処理をステップＳ１１６に進める。
［Ｓ１１２］入力の状態とスタック先頭の状態を引数にした到達可能状態特定処理の返り値である状態の集合にスタック先頭の状態が含まれるか否かを判断する。含まれる場合は処理をステップＳ１１３に進め、そうでない場合は処理を終了する。なお、到達可能状態特定処理の詳細は、図１９に示す。
［Ｓ１１３］スタック先頭の文書要素と状態の組をpop し、入力の状態をpop した状態と置き換える。
［Ｓ１１４］スタックが空であるか否かを判定する。スタックが空の場合は処理をステップＳ１１５へ、そうでない場合は処理をステップＳ１１６へ進める。
［Ｓ１１５］変数Ｖに入力の文書要素セットを書き出す。
［Ｓ１１６］入力の状態を引数にしてマッチング可能状態遷移特定処理を起動した結果の返り値の状態遷移を変数Ｘに格納する。マッチング可能状態遷移特定処理の詳細は、図２２に示す。
［Ｓ１１７］変数Ｘのすべての状態遷移が処理済みか否かを判断する。すべての状態遷移が処理済の場合は処理を終了し、そうでない場合は処理をステップＳ１１８に進める。
【００６５】
図１７は、オートマトンマッチング処理手順を示す第２のフローチャートである。
［Ｓ１１８］変数Ｘの未処理の状態遷移の１つを処理対象とする。
［Ｓ１１９］処理対象の状態遷移と入力の文書要素を引数として文書要素マッチング処理を呼び出し、返り値が真か否かを判断する。戻り値が真なら処理をステップＳ１２０に進め、偽なら処理をＳ１２７に進める。なお、文書要素マッチング処理の詳細は、図２４に示す。
［Ｓ１２０］入力の文書要素セットのコピーを作成し、文書要素セットのコピーに入力の文書要素を追加する。
［Ｓ１２１］入力の文書要素を引数にして次要素特定処理を起動した結果の返り値が入力の文書要素の長子であるか否かを判定する。長子である場合は処理をステップＳ１２２に進め、そうでない場合は処理をステップＳ１２６に進める。なお、次要素特定処理の詳細は、図２５に示す。
［Ｓ１２２］処理対象の状態遷移がオートマトンを参照しているか否か判定する。オートマトンを参照している場合は処理をステップＳ１２３へ進め、そうでない場合は処理をステップＳ１２６へ進める。
［Ｓ１２３］スタックのコピーを生成する。
［Ｓ１２４］スタックのコピーに、処理対象の状態遷移の終了端の状態と、入力の文書要素を引数とした弟要素特定処理の返り値との組をpushする。
【００６６】
図１８は、オートマトンマッチング処理手順を示す第３のフローチャートである。
［Ｓ１２５］処理対象の状態遷移が参照するオートマトンの初期状態を引数として到達可能状態特定処理を呼び出した場合の返り値の状態ごとに、入力の文書要素の長子、返り値の状態、ステップＳ１２０で作成した文書要素セットのコピー、及びステップＳ１２４で作成したスタックのコピーを入力としてオートマトンマッチング処理を呼び出し、処理を

ステップＳ１２７に進める。
［Ｓ１２６］処理対象の状態遷移の終了端の状態を引数として到達可能状態特定処理を呼び出した場合の返り値の状態ごとに、入力の文書要素を引数として弟要素特定処理を起動した結果の返り値、返り値の状態、ステップＳ１２０で作成した文書要素セットのコピー、及び入力のスタックのコピーを入力としてオートマトンマッチング処理を呼び出す。
［Ｓ１２７］処理対象の状態遷移が属するオートマトンに"+r"が記されているか否か判定する。"+r"が記されている場合は処理をステップＳ１２８へ進め、そうでない場合は処理をステップＳ１１７に進める。
［Ｓ１２８］入力の文書要素を引数にして次要素特定処理を起動した結果の返り値が入力の文書要素の長子であるか否か判定する。入力の文書要素の長子である場合は処理をステップＳ１２９に、そうでない場合は処理をステップＳ１１７に進める。
［Ｓ１２９］入力の状態、入力の文書要素の長子の文書要素、入力の文書要素セットのコピー、及び入力のスタックのコピーを入力としてオートマトンマッチング処理を呼び出す。その後、処理をステップＳ１１７に進める。
【００６７】
次に、ステップＳ１１２で行われる到達可能状態特定処理の手順を説明する。
図１９は、到達可能状態特定処理の手順を示すフローチャートである。この処理は、すべてマッチング部１４で行われる。また、この処理の入力は階層オートマトン中のオートマトンの状態の組（第１の状態と第２の状態）であり、出力は階層オートマトン中のオートマトンの状態の集合である。本処理および本処理で呼び出される処理では、任意の数の状態を保持する共通の変数Ｙを保持する。
［Ｓ１３１］入力の第１の状態と第２の状態とを引数にして状態チェック処理を呼び出す。状態チェック処理の詳細は、図２０に示す。
［Ｓ１３２］変数Ｙを出力する。
【００６８】
図２０は、状態チェック処理のフローチャートの前半である。この処理は、すべてマッチング部１４で行われる。また、この処理の入力はオートマトンの状態の組（第１の状態と第２の状態）であり、出力は変数Ｙに反映される。
［Ｓ１４１］入力の第１の状態が変数Ｙに含まれているか否か判定する。変数Ｙに含まれている場合には処理を終了し、含まれていない場合には処理をステップＳ１４２に進める。
［Ｓ１４２］入力の第１の状態を変数Ｙに追加する。
［Ｓ１４３］入力の第１の状態と第２の状態が同一であるか否か判定する。これらが同一である場合は処理を終了し、そうでない場合は処理をステップＳ１４４に進める。
［Ｓ１４４］入力の第１の状態を開始端とするすべての状態遷移を処理したか否か判定する。未処理の状態遷移がある場合は処理をステップＳ１４５へ進め、すべての状態遷移が処理済みの場合は処理を終了する。
［Ｓ１４５］未処理の状態遷移を１つ選択し、処理対象とする。
［Ｓ１４６］処理対象の状態遷移がε遷移の場合は処理をステップＳ１４７へ進め、述語を伴う状態遷移の場合は処理をステップＳ１４４に進め、いずれでもない場合は処理をステップＳ１４８へ進める。
［Ｓ１４７］処理対象の状態遷移の終了端の状態と第２の状態とを引数にして状態チェック処理を呼び出し、処理をステップＳ１４９に進める。
［Ｓ１４８］処理対象の状態遷移が参照するオートマトンの初期状態を引数にして状態チェック処理を呼び出し、処理をステップＳ１４９に進める。
【００６９】
図２１は、状態チェック処理のフローチャートの後半である。
［Ｓ１４９］入力の第１の状態が受理状態である場合には処理をステップＳ１５０に進め、そうでない場合は処理をステップＳ１４４に進める。
［Ｓ１５０］第１の状態を含むオートマトンを参照している状態遷移の終了端の状態と第

２の状態とを引数にして状態チェック処理を呼び出す。その後、処理をステップＳ１４４に進める。

【００７０】

図２２は、マッチング可能状態遷移特定処理の手順を示すフローチャートである。この処理は、すべてマッチング部１４が行う。また、この処理の入力はオートマトンの状態の組（第１の状態と第２の状態）であり、出力は状態遷移の集合である。この処理およびこの処理から呼び出される処理では、任意の数の状態を保持する共通の変数Ｚと、任意の数の状態遷移を保持する共通の状態Ｗを保持する。

［Ｓ１６１］入力の第１の状態と第２の状態を引数にして状態遷移チェック処理を呼び出す。状態遷移チェック処理の詳細は、図２３に示す。

［Ｓ１６２］変数Ｗを出力する。

【００７１】

図２３は、状態遷移チェック処理の手順を示すフローチャートである。この処理は、すべてマッチング部１４が行う。また、この処理の入力はオートマトンの状態であり、出力は変数Ｚおよび変数Ｗに反映される。

［Ｓ１７１］入力の第１の状態が変数Ｚに含まれているか否か判定する。変数Ｚに含まれている場合は処理を終了し、含まれていない場合には処理をステップＳ１７２に進める。

［Ｓ１７２］入力の第１の状態を変数Ｚに追加する。

［Ｓ１７３］入力の第１の状態と第２の状態が同一であるか否か判定する。同一である場合は処理を終了し、そうでない場合は処理をステップＳ１７４に進める。

［Ｓ１７４］入力の第１の状態を開始端とするすべての状態遷移を処理したか否か判定し、処理済みの場合は処理を終了し、そうでない場合は処理をステップＳ１７５へ進める。

［Ｓ１７５］未処理の状態遷移を１つ選択し、処理対象とする。

［Ｓ１７６］処理対象の状態遷移の種別を判別する。状態遷移がε遷移の場合は処理をステップＳ１７７へ進め、条件を伴う状態遷移の場合は処理をステップＳ１７９へ進め、いずれでもない場合は処理をステップＳ１７８へ進める。

［Ｓ１７７］処理対象の状態遷移の終了端の状態と第２の状態を引数にして状態遷移チェック処理を呼び出し、処理をステップＳ１８０に進める。

［Ｓ１７８］処理対象の状態遷移が参照するオートマトンの初期状態と第１の状態を引数にして状態遷移チェック処理を呼び出し、処理をステップＳ１８０に進める。

［Ｓ１７９］処理対象の状態遷移を変数Ｗに追加し、処理をステップＳ１８０に進める。

［Ｓ１８０］入力の第１の状態が受理状態であるか否かを判断し、受理状態の場合には処理をステップＳ１８１に進め、そうでない場合は処理をステップＳ１７４に進める。

［Ｓ１８１］第１の状態を含むオートマトンを参照している状態遷移の終了端の状態と第２の状態を引数にして状態遷移チェック処理を呼び出す。その後、処理をステップＳ１７４に進める。

【００７２】

なお、到達可能状態特定処理およびマッチング可能状態遷移特定処理はマッチング部に入力される構造化文書とは無関係に実行できるので、これらの処理をあらかじめ一度だけ実行して処理結果を保持し、オートマトンマッチング処理では保持されている処理結果を適宜利用することにより、同一の引数に対する到達可能状態特定処理およびマッチング可能状態遷移特定処理を繰り返し実行することによるオーバヘッドを未然に防ぐように構成することも可能である。

【００７３】

図２４は、文書要素マッチング処理の手順を示すフローチャートである。この処理は、すべてマッチング部１４が行う。また、この処理の入力は文書要素と状態遷移であり、出力は真偽値である。

［Ｓ１９１］文書要素に対応付けられた文書内容の文字列が、状態遷移に対応付けられた文字列パターンにマッチするか否かを判断する。文字列パターンにマッチする場合は処理をステップＳ１９２へ進め、そうでない場合は処理をステップＳ１９３に進める。

［Ｓ１９２］真を返し、処理を終了する。
［Ｓ１９３］偽を返し、処理を終了する。
【００７４】
なお、本実施の形態では、状態遷移に対応付けられる条件として文書内容の文字列パターンのみが利用可能と想定したが、状態遷移に対応付けられる条件は、文書要素を入力として真偽値を返す述語であればどのようなものでも構わない。このような条件には、文書要素の名称の完全一致・部分一致・パターンマッチ、文書要素の属性値の完全一致・部分一致・パターンマッチ、属性値の範囲指定、属性値と他の状態遷移にマッチした文書要素の属性値との同値関係・大小関係、および、これらの条件を論理結合子で結合したものが含まれる。
【００７５】
図２５は、次要素特定処理の手順を示すフローチャートである。この処理は、すべてマッチング部１４が行う。また、本処理の入力は構造化文書中の文書要素であり、出力は入力の文書要素と同一の構造化文書中の文書要素である。
［Ｓ２０１］入力の文書要素に子供がいるか否か判定する。子供がいる場合は処理をステップＳ２０２へ進め、子供がいない場合は処理をＳ２０３へ進める。
［Ｓ２０２］入力の文書要素の長子を出力し、処理を終了する。
［Ｓ２０３］入力の文書要素を引数にして弟要素特定処理を呼び出す。弟要素特定処理の詳細は、図２６に示す。
［Ｓ２０４］弟処理特定処理の返り値を出力し、処理を終了する。
【００７６】
図２６は、弟要素特定処理の手順を示すフローチャートである。この処理は、すべてマッチング部１４が行う。また、この処理の入力は構造化文書中の文書要素であり、出力は入力の文書要素と同一の構造化文書中の文書要素である。
［Ｓ２１１］入力の文書要素に弟がいるか否か判定する。弟がいる場合は処理をステップＳ２１２へ進め、弟がいない場合は処理をステップＳ２１３へ進める。
［Ｓ２１２］入力の文書要素の弟要素を出力し、処理を終了する。
［Ｓ２１３］入力の文書要素に親がいるか否か判定する。親がいる場合は処理をステップＳ２１４へ進め、親がいない場合は処理をステップＳ２１６へ進める。
［Ｓ２１４］入力の文書要素の親要素を引数にして弟要素特定処理を呼び出す。
［Ｓ２１５］ステップＳ２１４で呼び出した弟要素特定処理の返り値を出力し、処理を終了する。
［Ｓ２１６］空（NULL値）を出力し、処理を終了する。
【００７７】
以上の処理をマッチング部が行うことにより、文書保持部１１に格納されている文書の中で、階層オートマトン生成部１３が生成した階層オートマトンに適合する文書を抽出することができる。たとえば、図６の階層オートマトン１００を入力とするオートマトンマッチング処理では、図６８の文書９１と図４の文書９３のどちらを入力の文書としても、参考文献エントリに対応する文書要素を得ることができる。以下に、図６８の文書９１と図４の文書９３を対象としたオートマトンマッチング処理の呼び出し関係について説明する。
【００７８】
図６８の文書９１に対する処理では、図６の識別子「１ａ」の状態遷移と識別子「005 」の文書要素がマッチングする場合にのみ参考文献エントリが得られる。識別子「１ａ」の状態遷移の開始端の状態である識別子「11」の状態と識別子「005 」の文書要素との組を入力とするオートマトンマッチング処理、および、この処理から直接または間接的に呼び出されるオートマトンマッチング処理の呼び出し関係を以下に示す。
【００７９】
図２７は、図６８に示した文書における呼び出し関係を示す第１の図である。また、図２８は、図６８に示した文書における呼び出し関係を示す第２の図である。

【００８０】
図２７、図２８のノード２０１～２１５はオートマトンマッチング処理を意味し、ノード間の実線の矢印は処理の間の直接的な呼び出し関係、ノード間の破線は処理の間の間接的な呼び出し関係を意味する。また、○印は文書要素セットを変数Ｖに書き込めることを意味し、×印は文書要素セットを変数Ｖに書き込めないことを意味する。ノード中の４つ組は、入力の状態の識別子、入力の文書要素の識別子、入力の文書要素セット、及び入力のスタックを意味する。
【００８１】
オートマトンマッチング処理を実行した場合の呼び出し関係は一般には有向木となるが、オートマトンマッチング処理起動の履歴を内部的に保持し、この履歴を利用して同一の引数に対する処理呼び出しをただ一１度に抑えることができる。図２７、図２８の有向グラフは、同一の引数に対する処理呼び出しをただ一度に抑えた場合の呼び出し関係を表現している。
【００８２】
図２７、図２８では、ノード２０４の処理が実行されるオートマトン（図６のオートマトン１３０）に記号"+r"が設定されているためノード２０５の処理が呼び出される。この呼び出しにより、識別子「007 」の文書要素を読み飛ばし、識別子「008 」の文書要素や識別子「009 」の文書要素を対象としたオートマトンマッチング処理を呼び出すことを可能としている。この結果、識別子「008 」の文書要素と識別子「009 」の文書要素を参考文献エントリとして得ることができる。
【００８３】
図４の文書９３に対する処理でも、図６の状態遷移１ａと識別子「005 」の文書要素がマッチングする場合にのみ参考文献が得られる。状態遷移１ａの開始端の状態である識別子「11」の状態と識別子「005 」の文書要素の組を入力とするオートマトンマッチング処理、および、この処理から直接または間接的に呼び出されるオートマトンマッチング処理の呼び出し関係を以下に示す。
【００８４】
図２９は、図４に示した文書における呼び出し関係を示す第１の図である。また、図３０は、図４に示した文書における呼び出し関係を示す第２の図である。図２９、図３０の表記方法は図２７、図２８と同一であり、各ノード３０１～３１９がオートマトンマッチング処理を示している。
【００８５】
図２９、図３０では、図２７、図２８と同様、ノード３０４の処理からノード３０５の処理を呼び出すことによって識別子「007 」の文書要素を読み飛ばすことが可能となっている。さらに、ノード３０６～３０９の処理からノード３１０～ノード３１３の処理を呼び出すことによって識別子「009 」の文書要素を読み飛ばすことを可能としている。この結果、識別子「008 」の文書要素と識別子「010 」の文書要素を参考文献エントリとして得ることができる。
【００８６】
図３１は、図６の階層オートマトンと図６８の文書をマッチング部に入力した結果得られる文書要素セットを示す図である。また、図３２は、図６の階層オートマトンと図４の文書をマッチング部に入力した結果得られる文書要素セットを示す図である。なお、本実施の形態で示した処理の流れでは、同一の文書要素セットが複数出力されることがあるが、簡単にするため図３１、図３２では同一の文書要素セットを１つだけ示している。
【００８７】
次に、文書処理部１５の処理について説明する。
本実施の形態における文書処理部１５は、マッチング部１４で得られた文書要素セットごとに、処理対象の文書の論理構造を縮退した結果得られる構造化文書から可視化データを生成し、入出力装置２０のＣＲＴディスプレイに出力するものとする。
【００８８】

図３３は、図６８の構造化文書を処理対象としたときにＣＲＴディスプレイに表示される画面を示す図である。この画面４１０は、部分構造表示部４１１と適合内容表示部４１２とがある。部分構造表示部４１１には、マッチング部１４で得られた文書要素セットに含まれる文書要素の構造が表示されている。また、適合内容表示部４１２には、部分構造表示部４１１に表示された文書要素の内容が表示されている。

【００８９】

図３４は、図４の構造化文書を処理対象としたときにＣＲＴディスプレイに表示される画面を示す図である。なお、図６８の文書９１と図４の文書９３とは論理構造が異なるだけで文書の内容は同じであるため、この画面４２０の部分構造表示部４２１と適合内容表示部４２２との表示内容は、図３３の画面４１０の表示内容と同じである。

【００９０】

以上で、第１の実施の形態における文書処理装置の処理は終了する。

なお、第１の実施の形態で用いた検索条件ノードおよび階層オートマトンでは、パターン指定やオートマトンに記号"+r"を明示的に表示する表記法を採用したが、他の表記方法を採用してもよい。すなわち、パターン指定直下の文書要素列を決定する処理において、処理対象の文書要素をその文書要素直下の文書要素列と置き換えて処理を進めることができるパターン指定やオートマトンを、他のパターン指定等と区別できればよい。たとえば、上記の例で記号"+r"を伴わないパターン指定やオートマトンに記号"-r"を設定し、記号"+r"を伴うパターン指定やオートマトンには記号を設定しない表記法を採用することもできる。この場合でも、第１の実施の形態で示した処理の流れにおいて記号"+r"を扱う処理を適切に変更することにより、同一の処理の流れを実現することが容易に可能である。

【００９１】

また、第１の実施の形態のオートマトン生成処理（図９、図１０に示す）の流れでは、SEQ パターン指定、CHO パターン指定、またはOPTREPパターン指定に対応するオートマトンが生成される。しかし、オートマトン生成処理を以下のように実現することで、正規表現に対応するオートマトンを生成することも可能である。その例を、第１の実施の形態に関する応用例として以下に説明する。

【００９２】

図３５は、第１の実施の形態に関する応用例のオートマトン生成処理手順を示すフローチャートである。この処理は、すべて階層オートマトン生成部１３が行う。

［Ｓ３０１］検索条件ノードに対応するオートマトンが変数Ｔに含まれる場合には処理をステップＳ３０２へ進め、そうでない場合は処理をステップＳ３０３へ進める。

［Ｓ３０２］入力の状態遷移から、検索条件ノードに対応するオートマトンへの参照を生成し、処理を終了する。

［Ｓ３０３］オートマトンを生成し、入力の状態遷移からオートマトンへの参照を生成する。

［Ｓ３０４］入力の検索条件ノードのパターン指定に"+r"が設定されているか否か判定し、設定されている場合には処理をステップＳ３０５へ進め、そうでない場合には処理をステップＳ３０６へ進める。

［Ｓ３０５］生成したオートマトンに"+r"を設定する。

［Ｓ３０６］オートマトンに初期状態と受理状態を１つずつ生成する。

［Ｓ３０７］検索条件ノードの正規表現とオートマトンを引数にして属性変換文法評価処理を呼び出す。この処理の詳細は、図３６に示す。

［Ｓ３０８］ステップＳ３０７で生成されたオートマトンを変数Ｔに出力する。

【００９３】

図３６は、属性変換文法評価処理の手順を示すフローチャートである。この処理は、すべて階層オートマトン生成部１３で行われる。また、この処理の入力は正規表現とオートマトンであり、出力はオートマトンである。

［Ｓ３１１］「プログラミング言語処理系」（佐々政孝著、岩波書店）６５ページ図３．５記載の属性変換文法で入力の正規表現を解析し、構文木を構成する。

［Ｓ３１２］構文木の根を処理対象とする。
［Ｓ３１３］処理対象に対応する生成規則を上記属性変換文法から特定し、特定した生成規則に対応する操作を行う。
［Ｓ３１４］ステップＳ３１３で特定された生成規則が生成規則(2) であり、かつ、ステップＳ３１３で状態遷移に対応付けられた述語と同一の検索条件ノードに正規表現が含まれる場合には処理をステップＳ３１５へ進め、そうでない場合には処理をステップＳ３１６へ進める。
［Ｓ３１５］ステップＳ３１３で生成された状態遷移と、ステップＳ３１３で状態遷移に対応付けられた述語と同一の検索条件ノードに含まれる正規表現を引数にして、図３５のオートマトン生成処理を呼び出す。
［Ｓ３１６］構文木を前順に走査したときに処理対象の次に走査される記号が存在すれば処理をステップＳ３１７へ進め、そうでなければステップＳ３１８に進む。
［Ｓ３１７］ステップＳ３１６の記号を処理対象とし、処理をステップＳ３１３に進める。
［Ｓ３１８］オートマトンを出力し、処理を終了する。
【００９４】
このようなオートマトン生成処理を用いることで、検索条件ノードのパターン指定として任意の正規表現を用いることができる。ただし、この場合の検索条件ノードには常に述語が含まれる。
【００９５】
さらに、図３６に示した属性変換文法評価処理を以下のように実現することにより、検索条件ノードのパターン指定として任意の正規表現を用いることができるだけでなく、検索条件中に述語を含まない検索条件ノードを用いることが可能となる。
【００９６】
図３７は、属性変換文法評価処理の変形例を示す図である。この処理は、図３６の処理に代えて行われ、すべての処理が階層オートマトン生成部１３で行われる。また、この処理の入力は正規表現とオートマトンであり、出力はオートマトンである。
［Ｓ３２１］「プログラミング言語処理系」（佐々政孝著、岩波書店）６５ページ図３．５記載の属性変換文法で入力の正規表現を解析し、構文木を構成する。
［Ｓ３２２］構文木の根を処理対象とする。
［Ｓ３２３］処理対象に対応する生成規則を上記属性変換文法から特定し、特定した生成規則に対応する操作を行う。
［Ｓ３２４］ステップＳ３２３で状態遷移に対応付けられた検索条件ノードに正規表現が含まれる場合には処理をステップＳ３２５へ進め、そうでない場合には処理をステップＳ３２６へ進める。
［Ｓ３２５］ステップＳ３２３で生成された状態遷移と、ステップＳ３２３で状態遷移に対応付けられた述語と同一の検索条件ノードに含まれる正規表現を引数にして、図３５のオートマトン生成処理を呼び出す。
［Ｓ３２６］構文木を前順に走査したときに処理対象の次に走査される記号が存在するか否かを判断する。存在すれば処理をステップＳ３２７へ進め、そうでなければ処理をステップＳ３２８に進める。
［Ｓ３２７］ステップＳ３２６の記号を処理対象とし、処理をステップＳ３２３に進める。
［Ｓ３２８］オートマトンを出力し、処理を終了する。
【００９７】
次に、第２の実施の形態について説明する。第２の実施の形態における処理対象の構造化文書は、図６９の構造化文書である。第１の実施の形態で示した文書処理装置に図５の検索条件と図６９の文書９２を入力しても、識別子「002 」の文書要素が検索条件ノード中の述語にマッチしないために参考文献エントリを得ることができない。
【００９８】

そこで、図６９の文書９２に対しても所望の結果を得るため、本実施の形態の文書処理装置では、検索条件のパターン指定に"+p"を設定することができるようにする。"+p"は、このパターン指定と文書要素とのマッチング中に、マッチング対象の文書要素を読み飛ばしてもよいことを意味する。

【００９９】

第２の実施の形態で用いる検索条件を以下に示す。

図３８は、"+p"を用いて検索条件の例を示す図である。この検索条件５０は、図５の検索条件４０のパターン指定４２に記号"+p"を追加し、パターン指定４４およびパターン指定４６から記号"+r"を取り除いたものであり、文書要素を引数とする述語５１，５３，５５，５７と、パターン指定５２，５４，５６とで表されている。

【０１００】

"+p"の追加により、述語５３にマッチする文書要素の前に出現する文書要素群、述語５３にマッチする文書要素とパターン指定５４にマッチする文書要素列の間の文書要素群、及びパターン指定５４にマッチする文書要素列の後に出現する文書要素群を読み飛ばしてマッチングを行うことが可能となる。具体的には、図６９の文書に対する処理では識別子が「002 」、「003 」、「007 」、「008 」の文書要素を読み飛ばして所望の文書要素を得ることができる。

【０１０１】

第２の実施の形態に係る文書処理装置を実現するための構成要素は、第１の実施の形態と同様である。そのため、図２に示した構成を用いて第２の実施の形態を説明する。第２の実施の形態では、"+p"を含む検索条件を処理するため、本実施の形態の文書処理装置では、第１の実施の形態の文書処理装置のオートマトン生成処理とオートマトンマッチング処理に以下の変更を施した処理が実行される。

【０１０２】

図３９は、第２の実施の形態におけるオートマトン生成処理の手順を示すフローチャートである。これは、第１の実施の形態の図９、図１０の処理に代えて階層オートマトン生成部１３で行われる。

［Ｓ４０１］検索条件ノードに対応するオートマトンが変数Ｔに含まれるか否かを判断する。変数Ｔに含まれる場合は処理をステップＳ４０２へ進め、そうでない場合は処理をステップＳ４０３へ進める。

［Ｓ４０２］入力の状態遷移から、検索条件ノードに対応するオートマトンへの参照を生成し、処理を終了する。

［Ｓ４０３］オートマトンを生成し、入力の状態遷移からオートマトンへの参照を生成する。

［Ｓ４０４］入力の検索条件ノードのパターン指定に"+p"が設定されているか否か判定し、設定されている場合には処理をステップＳ４０５へ進め、そうでない場合には処理をステップＳ４０６へ進める。

［Ｓ４０５］生成したオートマトンに"+p"を設定する。

［Ｓ４０６］オートマトン上に初期状態と受理状態を生成する。

［Ｓ４０７］検索条件ノードのパターン指定の種別を判別する。パターン指定が"SEQ" なら処理をステップＳ４０８へ進め、"CHO" なら処理をステップＳ４０９へ進め、"OPTREP" なら処理をステップＳ４１０へ進める。

［Ｓ４０８］オートマトンを引数にしてSEQ パターン指定生成処理を呼び出す。その後、処理をステップＳ４１１に進める。なお、SEQ パターン指定生成処理の詳細は、図１１、図１２に示した処理と同様である。

［Ｓ４０９］オートマトンを引数にしてCHO パターン指定生成処理を呼び出す。その後、処理をステップＳ４１１に進める。なお、CHO パターン指定生成処理の詳細は、図１３に示した処理と同様である。

［Ｓ４１０］オートマトンを引数にしてOPTREPパターン指定生成処理を呼び出す。その後、処理をステップＳ４１１に進める。なお、OPTREPパターン指定生成処理の詳細は、図１

４に示した処理と同様である。
［Ｓ４１１］各生成処理の返り値のオートマトンを変数Ｔに格納する。
【０１０３】
図３８の検索条件５０を第２の実施の形態のオートマトン生成処理に入力した結果出力される階層オートマトンを以下に示す。
図４０は、第２の実施の形態の文書処理装置により生成された階層オートマトンを示す図である。階層オートマトン５００は、４つのオートマトン５１０，５２０，５３０，５４０からなる。各オートマトン５１０，５２０，５３０，５４０の状態と、状態遷移とには、図６の例と同様の識別子が割り振られている。
【０１０４】
この階層オートマトン５００は、オートマトン５２０の右上に"P" が設定されている。この記号は、オートマトン上でマッチング対象の文書要素を読み飛ばしてもよいことを意味する。
【０１０５】
第２の実施の形態におけるオートマトンマッチング処理は、以下のように変更される。
図４１は、第２の実施の形態におけるオートマトンマッチング処理の手順を示すフローチャートの前半である。この処理は、第１の実施の形態の図１６～図１８の処理に代えて行われる。
［Ｓ４２１］入力の文書要素がスタック先頭の文書要素であるか否かを判断する。スタック先頭の文書要素であれば処理をステップＳ４２２へ進め、そうでない場合は処理をステップＳ４２６に進める。
［Ｓ４２２］入力の状態とスタック先頭の状態を引数にした到達可能状態特定処理の返り値である状態の集合に、スタック先頭の状態が含まれるか否かを判断する。スタック先頭の状態が含まれる場合は処理をステップＳ４２３に進め、そうでない場合は処理を終了する。
［Ｓ４２３］スタック先頭の文書要素と状態の組をpop し、入力の状態をpop した状態と置き換える。
［Ｓ４２４］スタックが空であるか否かを判定し、空の場合は処理をステップＳ４２５へ進め、そうでない場合は処理をステップＳ４２６へ進める。
［Ｓ４２５］変数Ｖに入力の文書要素セットを書き出す。
［Ｓ４２６］入力の状態を引数にしてマッチング可能状態遷移特定処理を起動した結果の返り値の状態遷移を変数Ｘに格納する。マッチング可能状態遷移特定処理の詳細は、図２２に示す通りである。
［Ｓ４２７］変数Ｘのすべての状態遷移が処理済みの場合は処理を終了し、そうでない場合は処理をステップＳ４２８に進める。
［Ｓ４２８］変数Ｘの未処理の状態遷移の１つを処理対象とする。
［Ｓ４２９］処理対象の状態遷移と入力の文書要素を引数として文書要素マッチング処理を呼び出し、返り値の値を判断する。返り値が真なら処理をステップＳ４３０に進め、返り値が偽なら処理をステップＳ４３７に進める。なお、文書要素マッチング処理の詳細は、図２４に示す通りである。
［Ｓ４３０］入力の文書要素セットのコピーを作成し、文書要素セットのコピーに入力の文書要素を追加する。
【０１０６】
図４２は、第２の実施の形態におけるオートマトンマッチング処理の手順を示すフローチャートの後半である。
［Ｓ４３１］入力の文書要素を引数にして次要素特定処理を起動した結果の返り値が入力の文書要素の長子であるか否かを判定する。長子である場合は処理をステップＳ４３２に進め、そうでない場合は処理をステップＳ４３６に進める。なお、次要素特定処理の詳細は、図２５に示す通りである。
［Ｓ４３２］処理対象の状態遷移がオートマトンを参照しているか否か判定し、参照して

いる場合は処理をステップＳ４３３へ進め、そうでない場合は処理をステップＳ４３６へ進める。
［Ｓ４３３］スタックのコピーを作成する。
［Ｓ４３４］スタックのコピーに、処理対象の状態遷移の終了端の状態と、入力の文書要素を引数とした弟要素特定処理の返り値の組をpushする。
［Ｓ４３５］処理対象の状態遷移が参照するオートマトンの初期状態を引数として到達可能状態特定処理を呼び出した場合の返り値の状態ごとに、入力の文書要素の長子と、返り値の状態と、ステップＳ４３０で作成した文書要素セットのコピーと、ステップＳ４３４で作成したスタックのコピーを入力としてオートマトンマッチング処理を呼び出す。その後、処理をステップＳ４３７に進める。
［Ｓ４３６］処理対象の状態遷移の終了端の状態を引数として到達可能状態特定処理を呼び出した場合の返り値の状態ごとに、入力の文書要素を引数として弟要素特定処理を起動した結果の返り値と、返り値の状態と、ステップＳ４３０で作成した文書要素セットのコピーと、入力のスタックのコピーを入力としてオートマトンマッチング処理を呼び出す。
［Ｓ４３７］処理対象の状態遷移が属するオートマトンに"+p"が記されているか否か判定し、記されている場合は処理をステップＳ４３８へ進め、そうでない場合は処理をステップＳ４２７に進める。
［Ｓ４３８］入力の状態、入力の文書要素を弟要素特定処理に入力した場合の返り値の文書要素、入力の文書要素セットのコピー、入力のスタックのコピーを入力としてオートマトンマッチング処理を呼び出す。その後、処理をステップＳ４２７に進める。
【０１０７】
第２の実施の形態のマッチング部１４に図６９の構造化文書と図４０の階層オートマトンを入力したときのオートマトンマッチング処理の呼び出し関係を以下に示す。
【０１０８】
図４３は、第２の実施の形態におけるオートマッチング処理の呼び出し関係を示す第１の図である。また、図４４は、第２の実施の形態におけるオートマッチング処理の呼び出し関係を示す第２の図である。図４３、図４４の表記法は図２７、図２８の表記法と同一であり、ノード６０１～６１４はオートマトンマッチング処理を意味する。なお、簡単のため、図４３、図４４では文書要素セット(001,004,005,006) の出力に至る呼び出し関係のみを示し、文書要素セット(001,004) および文書要素セット(001,004,005) の出力に至る呼び出し関係は省略してある。
【０１０９】
図４３、図４４では、ノード６０２の処理が実行されるオートマトン（図４０のオートマトン５２０）に記号"+p"が設定されているため、ノード６０３の処理およびノード６０４の処理が呼び出される。この呼び出しにより、識別子が「002 」の文書要素および識別子が「003 」の文書要素を読み飛ばし、識別子が「004 」、「005 」、「006 」の文書要素を対象としたオートマトンマッチング処理を呼び出すことを可能としている。さらに、ノード６１１の処理が実行されているオートマトン（図４０のオートマトン５２０）に記号"+p"が設定されているため、ノード６１２の処理およびノード６１４の処理が呼び出される。この呼び出しにより、識別子が「007 」の文書要素と「008 」の文書要素を読み飛ばすことができる。この結果、識別子が「005 」の文書要素と「006 」の文書要素を参考文献エントリとして得ることができる。
【０１１０】
図４５は、図６９の構造化文書と図３８の検索条件を第２の実施の形態の文書処理装置に入力した結果得られるマッチング部の出力を示す図である。このように、第２の実施の形態によれば、図６９のような文書の中からも、図３８の検索条件に適合する構造を抽出することができる。
また、第１の実施の形態の応用例（図３５に示す）と同様の変更を、第２の実施の形態に対して行うこともできる。そのような、第２の実施の形態に対する応用例を以下に示す。
【０１１１】

図４６は、第２の実施の形態の応用例に係るオートマトン生成処理のフローチャートである。これは、図３９の処理に代えて階層オートマトン生成部１３で行われる処理である。
［Ｓ５０１］検索条件ノードに対応するオートマトンが変数Ｔに含まれるか否かを判断する。変数Ｔに含まれる場合には処理をステップＳ５０２へ進め、そうでない場合は処理をステップＳ５０３へ進める。
［Ｓ５０２］入力の状態遷移から、検索条件ノードに対応するオートマトンへの参照を生成し、処理を終了する。
［Ｓ５０３］オートマトンを生成し、入力の状態遷移からオートマトンへの参照を生成する。
［Ｓ５０４］入力の検索条件ノードのパターン指定に"+p"が設定されているか否か判定する。"+p"が設定されている場合には処理をステップＳ５０５へ進め、そうでない場合には処理をステップＳ５０６へ進める。
［Ｓ５０５］生成したオートマトンに"+p"を設定する。
［Ｓ５０６］オートマトンに初期状態と受理状態を１つずつ生成する。
［Ｓ５０７］検索条件ノードの正規表現とオートマトンを引数にして属性変換文法評価処理を呼び出す。
［Ｓ５０８］ステップＳ５０７で生成されたオートマトンを変数Ｔに出力し、処理を終了する。

【０１１２】

このようにして、第１の実施の形態の応用例と同様の変更を、第２の実施の形態に対して行うことができる。
次に、第３の実施の形態について説明する。第３の実施の形態は、第１の実施の形態の機能と、第２の実施の形態の機能とを併せ持った文書処理装置である。

【０１１３】

第３の実施の形態において処理対象として想定している構造化文書は、次のような文書である。
図４７は、構造化文書の第４の例を示す図である。第１の実施の形態で示した文書処理装置に図５の検索条件４０と図４７の文書９４を入力しても、識別子「007 」のPARA要素が検索条件ノード中の述語にマッチしないために参考文献エントリを得ることができない。また、第２の実施の形態で示した文書処理装置に図５の検索条件４０と図４７の文書９４を入力しても、識別子「008 」のLIST要素が検索条件ノード中の述語にマッチしないために参考文献エントリを得ることができない。

【０１１４】

そこで、図４７の文書９４に対しても所望の結果を得るため、第３の実施の形態の文書処理装置では、検索条件のパターン指定に"+r"と"+p"の両方を設定することができるようにする。

【０１１５】

図４８は、"+r"と"+p"とを用いた検索条件を示す図である。図４８の検索条件６０は、図５の検索条件４０のパターン指定４２に記号"+p"を追加したものであり、文書要素を引数とする述語６１，６３，６５，６７と、パターン指定６２，６４，６６とで表されている。この追加により、述語６３にマッチする文書要素の前に出現する文書要素群、述語６３にマッチする文書要素とパターン指定６４にマッチする文書要素列の間の文書要素群、及びパターン指定６４にマッチする文書要素列の後に出現する文書要素群を読み飛ばしてマッチングを行うことが可能となる。具体的には、図４７の文書９４に対する処理では識別子「007 」「009 」の文書要素を読み飛ばして所望の文書要素を得ることができる。

【０１１６】

ここで、"+r"と"+p"を含む検索条件を処理するため、第３の実施の形態の文書処理装置では、第１の実施の形態の文書処理装置のオートマトン生成処理とオートマトンマッチング処理に以下の変更を施した処理が実行される。なお、第３の実施の形態に係る文書処理装置に必要な構成要素は、第１の実施の形態と同様であるため、図２に示した構成を用いて

第３の実施の形態を説明する。

【０１１７】

図４９は、第３の実施の形態におけるオートマトン生成処理手順を示すフローチャートである。この処理は、第１の実施の形態のオートマトン生成処理（図９、図１０に示す）に代えて行われる処理であり、すべて階層オートマトン生成部１３で行われる。

［Ｓ６０１］検索条件ノードに対応するオートマトンが変数Ｔに含まれるか否かを判断し、変数Ｔに含まれる場合は処理をステップＳ６０２へ進め、そうでない場合は処理をステップＳ６０３へ進める。

［Ｓ６０２］入力の状態遷移から、検索条件ノードに対応するオートマトンへの参照を生成し、処理を終了する。

［Ｓ６０３］オートマトンを生成し、入力の状態遷移からオートマトンへの参照を生成する。

［Ｓ６０４］入力の検索条件ノードのパターン指定に"+r"が設定されているか否か判定し、設定されている場合には処理をステップＳ６０５へ進め、そうでない場合には処理をステップＳ６０６へ進める。

［Ｓ６０５］生成したオートマトンに"+r"を設定する。

［Ｓ６０６］入力の検索条件ノードのパターン指定に"+p"が設定されているか否か判定し、設定されている場合には処理をステップＳ６０７へ進め、そうでない場合には処理をステップＳ６０８へ進める。

［Ｓ６０７］生成したオートマトンに"+p"を設定する。

［Ｓ６０８］オートマトン上に初期状態と受理状態を生成する。

［Ｓ６０９］検索条件ノードのパターン指定の種別を判別する。パターン指定が"SEQ" なら処理をステップＳ６１０へ進め、"CHO" なら処理をステップＳ６１１へ進め、"OPTREP" なら処理をステップＳ６１２へ進める。

［Ｓ６１０］オートマトンを引数にしてSEQ パターン指定生成処理を呼び出す。その後、処理をステップＳ６１３に進める。なお、SEQ パターン指定生成処理の詳細は、図１１、図１２に示した処理と同様である。

［Ｓ６１１］オートマトンを引数にしてCHO パターン指定生成処理を呼び出す。その後、処理をステップＳ６１３に進める。なお、CHO パターン指定生成処理の詳細は、図１３に示した処理と同様である。

［Ｓ６１２］オートマトンを引数にしてOPTREPパターン指定生成処理を呼び出す。なお、OPTREPパターン指定生成処理の詳細は、図１４に示した処理と同様である。

［Ｓ６１３］各生成処理の返り値のオートマトンを変数Ｔに格納する。

【０１１８】

このようにして、"+r"と"+p"との両方を使用した検索条件に対する階層オートマトンが生成される。

図５０は、第３の実施の形態の文書処理装置により生成された階層オートマトンを示す図である。これは、図４８の検索条件を第３の実施の形態のオートマトン生成処理に入力した結果出力される階層オートマトン７００である。この階層オートマトン７００は、４つのオートマトン７１０，７２０，７３０，７４０で構成される。オートマトン７２０には"+p"の記号が付与されており、オートマトン７３０，７４０には"+r"の記号が付与されている。

【０１１９】

このような階層オートマトン７００を用いてオートマトンマッチング処理が行われる。オートマトンマッチング処理は以下のように変更される。

図５１は、第３の実施の形態におけるオートマトンマッチング処理の手順を示すフローチャートの前半である。この処理は、第１の実施の形態のオートマトンマッチング処理（図１６～図１８）に代えて、マッチング部１４によって行われる処理である。

［Ｓ６２１］入力の文書要素がスタック先頭の文書要素であるか否かを判断する。スタック先頭の文書要素であれば処理をステップＳ６２２へ進め、そうでない場合は処理をステ

ップＳ６２６に進める。
［Ｓ６２２］入力の状態とスタック先頭の状態を引数にした到達可能状態特定処理の返り値である状態の集合に、スタック先頭の状態が含まれるか否かを判断する。含まれる場合は処理をステップＳ６２３に進め、そうでない場合は処理を終了する。なお、到達可能状態特定処理の詳細は、図１９に示す通りである。
［Ｓ６２３］スタック先頭の文書要素と状態の組をpop し、入力の状態をpop した状態と置き換える。
［Ｓ６２４］スタックが空であるか否かを判定し、空の場合は処理をステップＳ６２５へ進め、そうでない場合は処理をステップＳ６２６へ進める。
［Ｓ６２５］変数Ｖに入力の文書要素セットを書き出す。
［Ｓ６２６］入力の状態を引数にしてマッチング可能状態遷移特定処理を起動した結果の返り値の状態遷移を変数Ｘに格納する。マッチング可能状態遷移特定処理の詳細は、図２２に示す通りである。
［Ｓ６２７］変数Ｘのすべての状態遷移が処理済みであるか否かを判断し、処理済みの場合は処理を終了し、そうでない場合は処理をステップＳ６２８に進める。
［Ｓ８２７］変数Ｘの未処理の状態遷移の１つを処理対象とする。
［Ｓ６２９］処理対象の状態遷移と入力の文書要素を引数として文書要素マッチング処理を呼び出し、返り値を判断する。返り値が真なら処理をステップＳ６３０に進め、偽なら処理をステップＳ６３７に進める。なお、文書要素マッチング処理の詳細は、図２４に示す通りである。
［Ｓ６３０］入力の文書要素セットのコピーを作成し、文書要素セットのコピーに入力の文書要素を追加する。
【０１２０】
図５２は、第３の実施の形態におけるオートマトンマッチング処理の手順を示すフローチャートの後半である。
［Ｓ６３１］入力の文書要素を引数にして次要素特定処理を起動した結果の返り値が入力の文書要素の長子であるか否かを判定する。返り値が長子である場合は処理をステップＳ６３２に進め、そうでない場合は処理をステップＳ６３６に進める。なお、次要素特定処理の詳細は、図２５に示す通りである。
［Ｓ６３２］処理対象の状態遷移がオートマトンを参照しているか否か判定し、参照している場合は処理をステップＳ６３３へ進め、そうでない場合は処理をステップＳ６３６へ進める。
［Ｓ６３３］スタックのコピーを作成する。
［Ｓ６３４］スタックのコピーに、処理対象の状態遷移の終了端の状態と、入力の文書要素を引数とした弟要素特定処理の返り値の組をpushする。
［Ｓ６３５］処理対象の状態遷移が参照するオートマトンの初期状態を引数として到達可能状態特定処理を呼び出した場合の返り値の状態ごとに、入力の文書要素の長子と、返り値の状態と、ステップＳ６３０で作成した文書要素セットのコピーと、ステップＳ６３４で作成したスタックのコピーとを入力としてオートマトンマッチング処理を呼び出し、処理をステップＳ６３７に進める。
［Ｓ６３６］処理対象の状態遷移の終了端の状態を引数として到達可能状態特定処理を呼び出した場合の返り値の状態ごとに、入力の文書要素を引数として弟要素特定処理を起動した結果の返り値と、返り値の状態と、ステップＳ６３０で作成した文書要素セットのコピーと、入力のスタックのコピーを入力としてオートマトンマッチング処理を呼び出す。
［Ｓ６３７］処理対象の状態遷移が属するオートマトンに"+r"が記されているか否か判定し、記されている場合は処理をステップＳ６３８へ進め、そうでない場合は処理をステップＳ６４０に進める。
［Ｓ６３８］入力の文書要素を引数にして次要素特定処理を起動した結果の返り値が入力の文書要素の長子であるか否か判定し、長子である場合は処理をステップＳ６３９に、そうでない場合は処理をステップＳ６４０に進める。

［Ｓ６３９］入力の状態、入力の文書要素の長子の文書要素、入力の文書要素セットのコピー、入力のスタックのコピーを入力としてオートマトンマッチング処理を呼び出す。
［Ｓ６４０］処理対象の状態遷移が属するオートマトンに"+p"が記されているか否か判定し、記されている場合はステップＳ６４１へ進め、そうでない場合は処理をＳ６２７に進める。
［Ｓ６４１］入力の状態、入力の文書要素を弟要素特定処理に入力した場合の返り値の文書要素、入力の文書要素セットのコピー、入力のスタックのコピーを入力としてオートマトンマッチング処理を呼び出す。その後、処理をステップＳ６２７に進める。
【０１２１】
以下に、第３の実施の形態のマッチング部１４に図４７の文書９４と図５０の階層オートマトン７００を入力したときの、オートマトンマッチング処理の呼び出し関係を示す。
【０１２２】
図５３は、第３の実施の形態におけるオートマトンマッチング処理の呼び出し関係を示す第１の図である。また、図５４は、第３の実施の形態におけるオートマッチング処理の呼び出し関係を示す第２の図である。図５３、図４３の表記法は図２７、図２８の表記法と同一であり、ノード８０１～８１３はオートマトンマッチング処理を意味する。なお、簡単のため、図５３、図５４では、オートマトンマッチング処理に４つ組(11,005,NULL,NULL)が入力された時点からの呼び出し関係のみを示す。さらに、文書要素セット(005,006,009,010) の出力に至る呼び出し関係のみを示し、文書要素セット(005,006) および文書要素セット(005,006,009) の出力に至る呼び出し関係は省略してある。
【０１２３】
図５３、図５４では、ノード８０２の処理が実行されるオートマトン（図５０のオートマトン７２０）に記号"+p"が設定されているため、識別子「007 」の文書要素を読み飛ばしてノード８０３の処理を呼び出すことができる。また、ノード８０５の処理が実行されるオートマトン（図５０のオートマトン７３０）に記号"+r が設定されているため、識別子「008 」の文書要素を読み飛ばしてノード８０６の処理を呼び出すことができる。さらに、ノード８１１の処理が実行されるオートマトン（図５０のオートマトン７２０）に記号"+p"が設定されているため、識別子「011 」の文書要素を読み飛ばしてノード８１２の処理を呼び出し、結果的に識別子「009 」の文書要素と識別子「010 」の文書要素を参考文献エントリとして得ることができる。
【０１２４】
図５５は、第４の実施の形態におけるマッチング部の出力例を示す図である。これは、図４７の構造化文書と図４８の検索条件を本実施例の文書処理装置に入力した結果得られるマッチング部の出力結果である。
【０１２５】
このように、第１の実施の形態の機能と第２の実施の形態の機能とを兼ね備えることで、様々な構造化文書の中から、利用者の意図に合った部分構造をマッチング処理により検出することができる。
【０１２６】
ところで、第１の実施の形態において説明した応用例（図３５に示す）と同様の応用を、第３の実施の形態に対して行うこともできる。その場合図３５に示したオートマトン生成処理は以下のように変更される。
【０１２７】
図５６は、第３の実施の形態のオートマトン生成処理に関する応用例を示すフローチャートである。この処理は、図４９の処理に代えて階層オートマトン生成部１３で行われる処理である。
［Ｓ７０１］検索条件ノードに対応するオートマトンが変数Ｔに含まれるか否かを判断し、変数Ｔに含まれる場合には処理をステップＳ７０２へ進め、そうでない場合は処理をステップＳ７０３へ進める。
［Ｓ７０２］入力の状態遷移から、検索条件ノードに対応するオートマトンへの参照を生

成し、処理を終了する。
［Ｓ７０３］オートマトンを生成し、入力の状態遷移からオートマトンへの参照を生成する。
［Ｓ７０４］入力の検索条件ノードのパターン指定に"+r"が設定されているか否か判定し、設定されている場合には処理をステップＳ７０５へ進め、そうでない場合には処理をステップＳ７０６へ進める。
［Ｓ７０５］生成したオートマトンに"+r"を設定する。
［Ｓ７０６］入力の検索条件ノードのパターン指定に"+p"が設定されているか否か判定し、設定されている場合には処理をステップＳ７０７へ進め、そうでない場合には処理をステップＳ７０８へ進める。
［Ｓ７０７］生成したオートマトンに"+p"を設定する。
［Ｓ７０８］オートマトンに初期状態と受理状態を１つずつ生成する。
［Ｓ７０９］検索条件ノードの正規表現とオートマトンを引数にして属性変換文法評価処理を呼び出す。
［Ｓ７１０］ステップＳ７０７で生成されたオートマトンを、変数Ｔに出力する。
【０１２８】
次に、第４の実施の形態について説明する。第４の実施の形態は、要素に対して属性が与えられている場合に、その属性を、該当する要素の子孫に対して継承させながらマッチングを行うものである。
【０１２９】
第４の実施の形態では、以下のような構造化文書を処理対象として想定している。
図５７は、構造化文書の第５の例を示す図である。この文書９５の一部の文書要素には、名前だけでなく属性も設定されている。たとえば識別子「003 」の文書要素では、要素名として"LIST"が設定されているだけでなく、属性として"ALIGN="C"" が設定されている。本実施例では、文書要素の属性ALIGN はその文書要素を頂点とする部分木に対応する文書内容を表示するときの配置を規定するものとする。そして、属性値"C" は、文書内容がセンタリングされて表示されることを意味するものとする。
【０１３０】
図５７の構造化文書から参考文献エントリを得るための検索条件を以下に示す。
図５８は、第４の実施の形態に用いる検索条件を示す図である。この検索条件７０は、文書要素を引数とする述語７１，７３，７５，７７と、パターン指定７２，７４，７６とで表されている。検索条件７０の述語７３，７５，７７には、"CONTENT" と"ATTR"という２種類の条件が示されている。"CONTENT" は文書要素の内容に関する条件を意味し、"ATTR" は文書要素の属性に関する条件を意味する。たとえば図５８の検索条件の述語７３に示されている条件は「" 参考文献" または"References"を内容に含み、属性ALIGN の値が"C" である文書要素」となる。
【０１３１】
図５９は、属性を指定した検索条件から生成される階層オートマトンを示す図である。この階層オートマトン８００は、４つのオートマトン８１０，８２０，８３０，８４０からなる。
【０１３２】
なお、第４の実施の形態の文書処理装置の装置構成および処理の流れは、基本的には第３の実施の形態で示したものと同一である。ただし、図４８に示した検索条件を処理できるようにするため、図５１、図５２のオートマトンマッチング処理は次のように変更される。
【０１３３】
図６０は、第４の実施の形態のオートマトンマッチング処理手順を示す第１のフローチャートである。
［Ｓ８０１］入力の文書要素がスタック先頭の文書要素であるか否かを判断し、スタック先頭の文書要素である場合は処理をステップＳ８０２へ進め、そうでない場合は処理をス

テップＳ８０６に進める。
［Ｓ８０２］入力の状態とスタック先頭の状態を引数にした到達可能状態特定処理の返り値である状態の集合に、スタック先頭の状態が含まれるか否かを判断する。スタック先頭の状態が含まれる場合は処理をステップＳ８０３に進め、そうでない場合は処理を終了する。
［Ｓ８０３］スタック先頭の文書要素と状態の組をpop し、入力の状態をpop した状態と置き換える。
［Ｓ８０４］スタックが空であるか否かを判定し、空の場合は処理をステップＳ８０５へ進め、そうでない場合は処理をステップＳ８０６へ進める。
［Ｓ８０５］変数Ｖに入力の文書要素セットを書き出す。
［Ｓ８０６］入力の状態を引数にしてマッチング可能状態遷移特定処理を起動した結果の返り値の状態遷移を変数Ｘに格納する。なお、マッチング可能状態遷移特定処理の詳細は、図２２に示す通りである。
［Ｓ８０７］変数Ｘのすべての状態遷移が処理済みか否かを判断し、処理済みの場合は処理を終了し、そうでない場合は処理をステップＳ８０８に進める。
［Ｓ８０８］変数Ｘの未処理の状態遷移の１つを処理対象とする。
［Ｓ８０９］処理対象の状態遷移と入力の文書要素を引数として文書要素マッチング処理を呼び出し、返り値が真なら処理をステップＳ８１０に進め、偽なら処理をステップＳ８１７に進める。文書要素マッチング処理の詳細は、図６３に示す。
【０１３４】
図６１は、第４の実施の形態のオートマトンマッチング処理手順を示す第２のフローチャートである。
［Ｓ８１０］入力の文書要素セットのコピーを作成し、文書要素セットのコピーに入力の文書要素を追加する。
［Ｓ８１１］入力の文書要素を引数にして次要素特定処理を起動した結果の返り値が入力の文書要素の長子であるか否かを判定し、長子である場合は処理をステップＳ８１２に進め、そうでない場合は処理をステップＳ８１６に進める。
［Ｓ８１２］処理対象の状態遷移がオートマトンを参照しているか否か判定し、参照している場合は処理をステップＳ８１３へ進め、そうでない場合は処理をステップＳ８１６へ進める。
［Ｓ８１３］スタックのコピーを作成する。
［Ｓ８１４］スタックのコピーに、処理対象の状態遷移の終了端の状態と、入力の文書要素を引数とした弟要素特定処理の返り値の組をpushする。
［Ｓ８１５］処理対象の状態遷移が参照するオートマトンの初期状態を引数として到達可能状態特定処理を呼び出した場合の返り値の状態ごとに、入力の文書要素の長子と、返り値の状態と、ステップＳ８１０で作成した文書要素セットのコピーと、ステップＳ８１４で作成したスタックのコピーを入力としてオートマトンマッチング処理を呼び出し、処理をステップＳ８１７に進める。
［Ｓ８１６］処理対象の状態遷移の終了端の状態を引数として到達可能状態特定処理を呼び出した場合の返り値の状態ごとに、入力の文書要素を引数として弟要素特定処理を起動した結果の返り値と、返り値の状態と、ステップＳ８１０で作成した文書要素セットのコピーと、入力のスタックのコピーを入力としてオートマトンマッチング処理を呼び出す。
［Ｓ８１７］処理対象の状態遷移が属するオートマトンに"+r"が記されているか否か判定し、記されている場合は処理をステップＳ８１８へ進め、そうでない場合は処理を終了する。
［Ｓ８１８］入力の文書要素を引数にして次要素特定処理を起動した結果の返り値が入力の文書要素の長子であるか否か判定し、長子である場合は処理をステップＳ８１９に進め、そうでない場合は処理をステップＳ８２１に進める。
［Ｓ８１９］入力の文書要素に設定されている属性を入力の文書要素の子供の文書要素に設定する。

【０１３５】
図６２は、第４の実施の形態のオートマトンマッチング処理手順を示す第３のフローチャートである。
［Ｓ８２０］入力の状態、入力の文書要素の長子の文書要素、入力の文書要素セットのコピー、入力のスタックのコピーを入力としてオートマトンマッチング処理を呼び出す。
［Ｓ８２１］処理対象の状態遷移が属するオートマトンに"+p"が記されているか否か判定し、記されている場合は処理をステップＳ８２２へ進め、そうでない場合は処理をステップＳ８０７に進める。
［Ｓ８２２］入力の状態、入力の文書要素を弟要素特定処理に入力した場合の返り値の文書要素、入力の文書要素セットのコピー、入力のスタックのコピーを入力としてオートマトンマッチング処理を呼び出す。その後、処理をステップＳ８０７に進める。
【０１３６】
図６０～図６２の処理の流れは、図５１、図５２の処理の流れにステップＳ８１９の処理が追加されたものである。また、図２４の文書要素マッチング処理は以下のように変更される。
【０１３７】
図６３は、第４の実施の形態における文書要素マッチング処理の手順を示すフローチャートである。この処理は、すべてマッチング部１４によって行われる。
［Ｓ８３１］文書要素に対応付けられた文書内容の文字列が、状態遷移に対応付けられた述語のCONTENT エントリの文字列パターンにマッチするか否かを判断する。マッチする場合は処理をステップＳ８３２へ進め、そうでない場合は処理をステップＳ８３４に進める。
［Ｓ８３２］状態遷移に対応付けられた属性の名前と値の組が、文書要素に対応付けられた述語のATTRエントリの属性の名前と値の組の中に存在するか否かを判断する。存在する場合は処理をステップＳ８３３へ進め、そうでない場合は処理をステップＳ８３４に進める。
［Ｓ８３３］真を返し、処理を終了する。
［Ｓ８３４］偽を返し、処理を終了する。
【０１３８】
ここで、第４の実施の形態のマッチング部１４に図５７に示す文書９５と図５９に示す階層オートマトン８００とを入力した際の、オートマトンマッチング処理の呼び出し関係を以下に示す。
【０１３９】
図６４は、第４の実施の形態による呼び出し関係を示す図である。図６４の表記法は図２７、図２８などの表記法と同一であり、ノード９０１～９０５はオートマトンマッチング処理を意味する。なお、図６４では、本実施の形態の文書処理装置の特徴的な処理の流れを説明するのに必要な呼び出し関係のみを示している。
【０１４０】
図６４では、第３の実施の形態と同様、ノード９０２の処理が実行されるオートマトン（図５９のオートマトン８２０）に記号"+p"が設定されているため、識別子「003 」の文書要素から識別子「009 」の文書要素を読み飛ばし、識別子「010 」文書要素を処理対象とすることができる（ノード９０３の処理）。ここで、第３の実施の形態の文書処理装置では、識別子「001 」の文書要素をマッチングの対象にするか、識別子「010 」の文書要素を読み飛ばして識別子「011 」の文書要素をマッチング対象にすることしかできなかった。そのため、識別子「３ａ」の状態遷移の条件を満足する文書要素を得ることはできない。
【０１４１】
しかし、第４の実施の形態の文書処理装置では、識別子「010 」の文書要素の属性を識別子「011 」の文書要素に追加して得られる文書要素をマッチングの対象とすることができる（ノード９０４の処理）。したがって、識別子「011 」の文書要素を識別子「３ａ」の

状態遷移とマッチさせ、識別子「012 」の文書要素以降を対象とした処理を続けることができる。そして、結果的に、識別子「013 」の文書要素と識別子「014 」の文書要素とを参考文献エントリとして得ることができる。

【０１４２】

次に、第５の実施の形態について説明する。第５の実施の形態は、検索条件に適合した構造の内容を、別の内容に置き換えて出力するようにしたものである。

構造化文書を対象とした処理では、処理対象の文書の部分構造を検索するだけでなく、検索された部分構造の文書要素の名前や属性を変換することにより、文書作成時に想定された用途以外の用途に利用することが行われている。この処理に対応するため、第５の実施の形態の文書処理装置では、第３の実施の形態の文書処理装置の入力となる検索条件の述語に、述語にマッチした文書要素に設定すべき名前や属性を対応付けておく。このような検索条件の例を以下に示す。

【０１４３】

図６５は、第５の実施の形態に用いる検索条件を示す図である。この検索条件８０は、文書要素を引数とする述語８１，８４，８６，８８と、パターン指定８２，８３，８５，８７とで表されている。図６５の述語に対応する矩形には、記号" →" に続いて文字列が記載されている。これは、述語にマッチした文書要素の名前を" →" に続く文字列に置き換えることを意味する。

【０１４４】

第５の実施の形態における文書処理装置の装置構成および処理の流れは、第３の実施の形態における装置構成および処理の流れと基本的に同一である。唯一の差異は、図５１に示したオートマトンマッチング処理のステップＳ６３０が、以下のように変更される点である。

［Ｓ６３０ａ］入力の文書要素セットのコピーを作成し、入力の状態遷移に対応する検索条件ノードに設定された名前や属性を入力の文書要素のコピーに設定して、このコピーを文書要素セットのコピーに追加する。

【０１４５】

第５の実施の形態における文書処理装置に図６５の検索条件と図６９の構造化文書を入力した場合のオートマトンマッチング処理の呼び出し関係を、以下に示す。

【０１４６】

図６６は、第５の実施の形態による呼び出し関係を示す図である。図６６の表記法は図２７、図２８などの表記法と同一であり、ノード１００１～１００８はオートマトンマッチング処理を意味する。図６６の呼び出し関係は図４３、図４４の呼び出し関係と基本的に同一であるが、処理呼び出し時の第３引数の文書要素セットには、文書要素識別子と文書要素に設定された要素名が記号"-" で接続されたものが記されている。たとえば、ノード１００２の処理の第３引数である「001-" 参考文献" 」は、「文書要素セットがマッチング部から出力されるときには、識別子「001 」である文書要素（DOC 要素）の名前を" 参考文献" に変更して出力すること」を意味する。

【０１４７】

また、本実施の形態における文書処理装置に図６５の検索条件と図６９の構造化文書を入力した結果得られる文書要素セット(001,004,005,006) についてディスプレイに表示されるイメージの例を以下に示す。

【０１４８】

図６７は、第５の実施の形態によって表示される画面を示す図である。この画面４３０は図３３と同様に、部分構造表示部４３１と適合内容表示部４３２とで構成されている。そして、部分構造表示部４３１の表示内容は、要素名が置き換えられた状態で表示されている。

【０１４９】

なお、第５の実施の形態の文書処理装置の処理の流れは、第３の実施の形態における文書処理装置との差異を示すことによって説明した。しかし、他の実施例における文書処理装

置に対しても、上記の説明で示した変更と同様の変更を施すことで、図６５の検索条件を処理することが可能となる。

【０１５０】

なお、上記の処理機能は、コンピュータによって実現することができる。その場合、データ処理装置及び文書処理装置が有すべき機能の処理内容は、コンピュータで読み取り可能な記録媒体に記録されたプログラムに記述しておく。そして、このプログラムをコンピュータで実行することにより、上記処理がコンピュータで実現される。コンピュータで読み取り可能な記録媒体としては、磁気記録装置や半導体メモリ等がある。市場に流通させる場合には、ＣＤ－ＲＯＭ(Compact Disk Read Only Memory) やフロッピーディスク等の可搬型記録媒体にプログラムを格納して流通させたり、ネットワークを介して接続されたコンピュータの記憶装置に格納しておき、ネットワークを通じて他のコンピュータに転送することもできる。コンピュータで実行する際には、コンピュータ内のハードディスク装置等にプログラムを格納しておき、メインメモリにロードして実行する。

【０１５１】

【発明の効果】

以上説明したように本発明の第１のデータ処理装置では、有向順序木の中間ノードを削除し、中間ノードのあった位置に中間ノード直下のノード列を配置する操作を行った結果得られる均質化有向順序木を処理対象として、検索条件に対する適合構造の抽出を行うようにしたため、論理構造が異なる多数の情報に対する検索を行う際においても、ノード間の接続関係を利用した検索条件を利用できる。

【０１５２】

また、本発明の第２のデータ処理装置では、有向順序木中のいずれかのノードを頂点とする部分木を削除する操作を行った結果得られる均質化有向順序木を処理対象として、検索条件に対する適合構造の抽出を行うようにしたため、論理構造が異なる多数の情報に対する検索を行う際においても、ノード間の接続関係を利用した検索条件を利用できる。

【０１５３】

また、本発明の第１の文書処理装置では、有向順序木の中間ノードを削除し、中間ノードのあった位置に中間ノード直下のノード列を配置する操作を行った結果得られる均質化有向順序木を処理対象として、検索条件に対する適合構造の抽出を行うようにしたため、論理構造が異なる多数の構造化文書に対する検索を行う際においても、文書要素間の接続関係を利用した検索条件を利用できる。

【０１５４】

また、本発明の第２の文書処理装置では、有向順序木中のいずれかのノードを頂点とする部分木を削除する操作を行った結果得られる均質化有向順序木を処理対象として、検索条件に対する適合構造の抽出を行うようにしたため、論理構造が異なる多数の構造化文書に対する検索を行う際においても、文書要素間の接続関係を利用した検索条件を利用できる。

【０１５５】

また、本発明の第１のデータ処理プログラムを記録したコンピュータ読み取り可能な記録媒体では、有向順序木の中間ノードを削除し、中間ノードのあった位置に中間ノード直下のノード列を配置する操作を行った結果得られる均質化有向順序木を処理対象として、検索条件に対する適合構造の抽出を行うような処理をコンピュータに行わせることができるため、ノード間の接続関係を利用した検索条件を利用して論理構造が異なる多数の情報に対する検索を行う機能を、コンピュータ上に構築することができる。

【０１５６】

また、本発明の第２のデータ処理プログラムを記録したコンピュータ読み取り可能な記録媒体では、有向順序木中のいずれかのノードを頂点とする部分木を削除する操作を行った結果得られる均質化有向順序木を処理対象として、検索条件に対する適合構造の抽出を行うような処理をコンピュータに行わせることができるため、ノード間の接続関係を利用した検索条件を利用して論理構造が異なる多数の情報に対する検索を行う機能を、コンピュ

ータ上に構築することができる。

【０１５７】

また、本発明の第１の文書処理プログラムを記録したコンピュータ読み取り可能な記録媒体では、有向順序木の中間ノードを削除し、中間ノードのあった位置に中間ノード直下のノード列を配置する操作を行った結果得られる均質化有向順序木を処理対象として、検索条件に対する適合構造の抽出を行うような処理をコンピュータに行わせることができるため、ノード間の接続関係を利用した検索条件を利用して論理構造が異なる多数の構造化文書に対する検索を行う機能を、コンピュータ上に構築することができる。

【０１５８】

また、本発明の第２の文書処理プログラムを記録したコンピュータ読み取り可能な記録媒体では、有向順序木中のいずれかのノードを頂点とする部分木を削除する操作を行った結果得られる均質化有向順序木を処理対象として、検索条件に対する適合構造の抽出を行うような処理をコンピュータに行わせることができるため、ノード間の接続関係を利用した検索条件を利用して論理構造が異なる多数の構造化文書に対する検索を行う機能を、コンピュータ上に構築することができる。

【図面の簡単な説明】

【図１】　本発明の原理構成図である。

【図２】　文書処理システムの構成を示すブロック図である。

【図３】　文書処理装置の処理手順を示すフローチャートである。

【図４】　構造化文書の第３の例を示す図である。

【図５】　検索条件の例を示す図である。

【図６】　階層オートマトンの第１の例を示す図である。

【図７】　階層オートマトン生成処理部が行う処理を示すフローチャートである。

【図８】　根オートマトン生成処理手順を示すフローチャートである。

【図９】　オートマトン生成処理手順を示すフローチャートの前半である。

【図１０】　オートマトン生成処理手順を示すフローチャートの後半である。

【図１１】　SEQ パターン生成処理手順を示すフローチャートの前半である。

【図１２】　SEQ パターン生成処理手順を示すフローチャートの後半である。

【図１３】　CHO パターン生成処理手順を示すフローチャートである。

【図１４】　OPTREPパターン生成処理手順を示すフローチャートである。

【図１５】　マッチング部の処理手順を示すフローチャートである。

【図１６】　オートマトンマッチング処理手順を示す第１のフローチャートである。

【図１７】　オートマトンマッチング処理手順を示す第２のフローチャートである。

【図１８】　オートマトンマッチング処理手順を示す第３のフローチャートである。

【図１９】　到達可能状態特定処理の手順を示すフローチャートである。

【図２０】　状態チェック処理のフローチャートの前半である。

【図２１】　状態チェック処理のフローチャートの後半である。

【図２２】　マッチング可能状態遷移特定処理の手順を示すフローチャートである。

【図２３】　状態遷移チェック処理の手順を示すフローチャートである。

【図２４】　文書要素マッチング処理の手順を示すフローチャートである。

【図２５】　次要素特定処理の手順を示すフローチャートである。

【図２６】　弟要素特定処理の手順を示すフローチャートである。

【図２７】　図６８に示した文書における呼び出し関係を示す第１の図である。

【図２８】　図６８に示した文書における呼び出し関係を示す第２の図である。

【図２９】　図４に示した文書における呼び出し関係を示す第１の図である。

【図３０】　図４に示した文書における呼び出し関係を示す第２の図である。

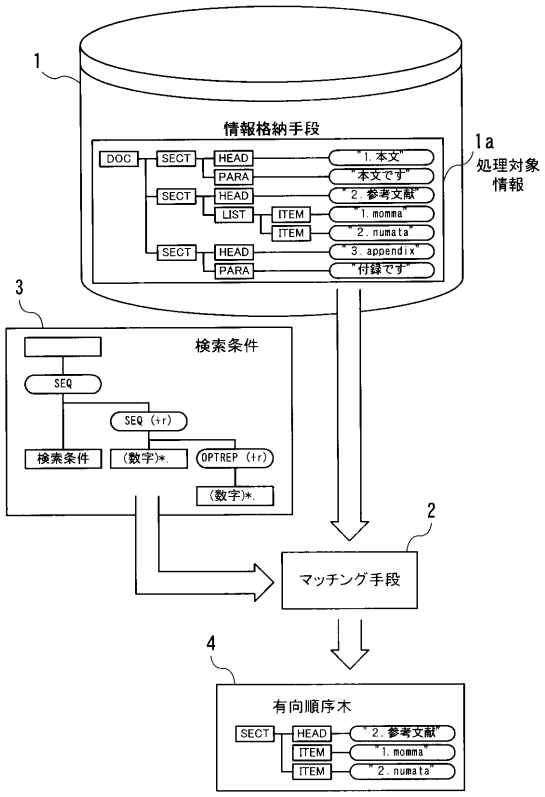
【図３１】　図６の階層オートマトンと図６８の文書をマッチング部に入力した結果得られる文書要素セットを示す図である。

【図３２】　図６の階層オートマトンと図４の文書をマッチング部に入力した結果得られる文書要素セットを示す図である。

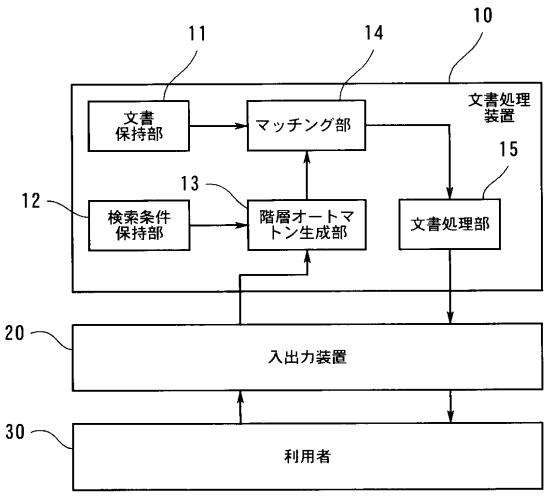
【図３３】　図６８の構造化文書を処理対象としたときにＣＲＴディスプレイに表示される画面を示す図である。
【図３４】　図４の構造化文書を処理対象としたときにＣＲＴディスプレイに表示される画面を示す図である。
【図３５】　第１の実施の形態に関する応用例のオートマトン生成処理手順を示すフローチャートである。
【図３６】　属性変換文法評価処理の手順を示すフローチャートである。
【図３７】　属性変換文法評価処理の変形例を示す図である。
【図３８】　"+p"を用いて検索条件の例を示す図である。
【図３９】　第２の実施の形態におけるオートマトン生成処理の手順を示すフローチャートである。
【図４０】　第２の実施の形態の文書処理装置により生成された階層オートマトンを示す図である。
【図４１】　第２の実施の形態におけるオートマトンマッチング処理の手順を示すフローチャートの前半である。
【図４２】　第２の実施の形態におけるオートマトンマッチング処理の手順を示すフローチャートの後半である。
【図４３】　第２の実施の形態におけるオートマッチング処理の呼び出し関係を示す第１の図である。
【図４４】　第２の実施の形態におけるオートマッチング処理の呼び出し関係を示す第２の図である。
【図４５】　図６９の構造化文書と図３８の検索条件を第２の実施の形態の文書処理装置に入力した結果得られるマッチング部の出力を示す図である。
【図４６】　第２の実施の形態の応用例に係るオートマトン生成処理のフローチャートである。
【図４７】　構造化文書の第４の例を示す図である。
【図４８】　"+r"と"+p"とを用いた検索条件を示す図である。
【図４９】　第３の実施の形態におけるオートマトン生成処理手順を示すフローチャートである。
【図５０】　第３の実施の形態の文書処理装置により生成された階層オートマトンを示す図である。
【図５１】　第３の実施の形態におけるオートマトンマッチング処理の手順を示すフローチャートの前半である。
【図５２】　第３の実施の形態におけるオートマトンマッチング処理の手順を示すフローチャートの後半である。
【図５３】　第３の実施の形態におけるオートマッチング処理の呼び出し関係を示す第１の図である。
【図５４】　第３の実施の形態におけるオートマッチング処理の呼び出し関係を示す第２の図である。
【図５５】　第４の実施の形態におけるマッチング部の出力例を示す図である。
【図５６】　第３の実施の形態のオートマトン生成処理に関する応用例を示すフローチャートである。
【図５７】　構造化文書の第５の例を示す図である。
【図５８】　第４の実施の形態に用いる検索条件を示す図である。
【図５９】　属性を指定した検索条件から生成される階層オートマトンを示す図である。
【図６０】　第４の実施の形態のオートマトンマッチング処理手順を示す第１のフローチャートである。
【図６１】　第４の実施の形態のオートマトンマッチング処理手順を示す第２のフローチャートである。
【図６２】　第４の実施の形態のオートマトンマッチング処理手順を示す第３のフローチ

ャートである。
【図６３】　第４の実施の形態における文書要素マッチング処理の手順を示すフローチャートである。
【図６４】　第４の実施の形態による呼び出し関係を示す図である。
【図６５】　第５の実施の形態に用いる検索条件を示す図である。
【図６６】　第５の実施の形態による呼び出し関係を示す図である。
【図６７】　第５の実施の形態によって表示される画面を示す図である。
【図６８】　構造化文書の第１の例を示す図である。
【図６９】　構造化文書の第２の例を示す図である。
【符号の説明】
1　情報格納手段
2　マッチング手段
3　検索条件
4　有向順序木

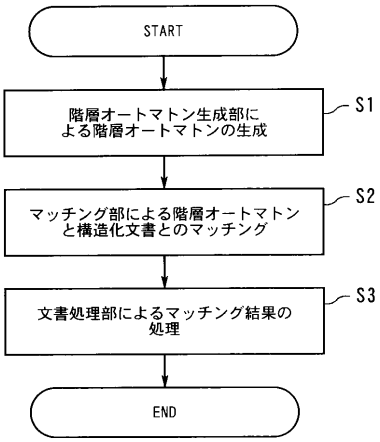
【図１】



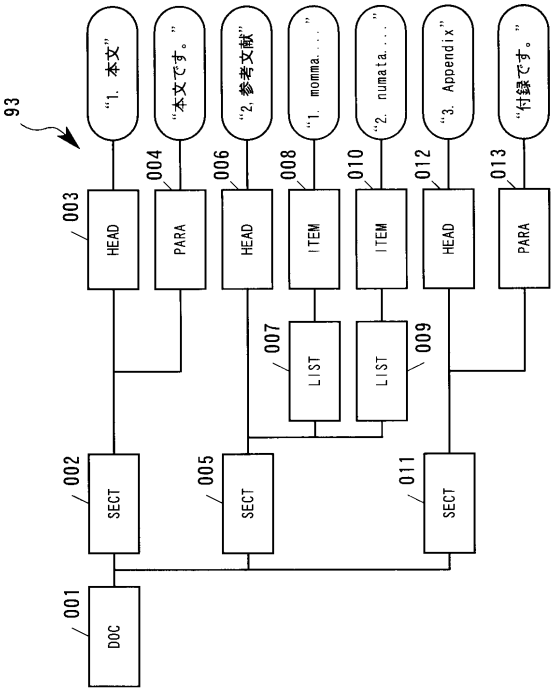
【図２】



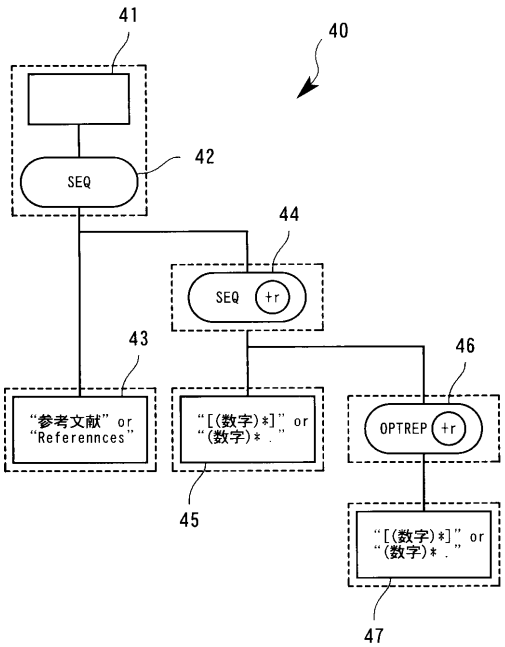
【図３】


START
階層オートマトン生成部による階層オートマトンの生成
S1
マッチング部による階層オートマトンと構造化文書とのマッチング
S2
文書処理部によるマッチング結果の処理
S3
END

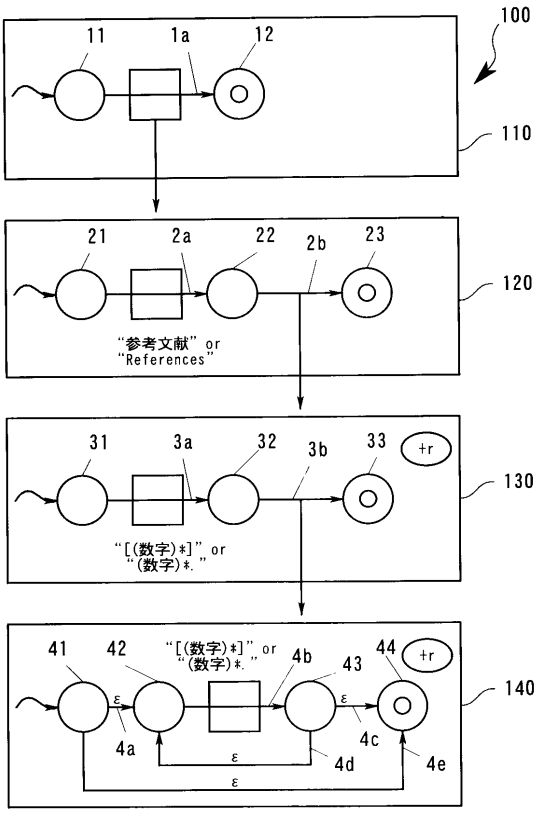
【図４】


93
"1. 本文"
"本文です。"
"2. 参考文献"
"1. momma..."
"2. numata..."
"3. Appendix"
"付録です。"
003 HEAD
004 PARA
006 HEAD
008 ITEM
010 ITEM
012 HEAD
013 PARA
007 LIST
009 LIST
002 SECT
005 SECT
011 SECT
001 DOC

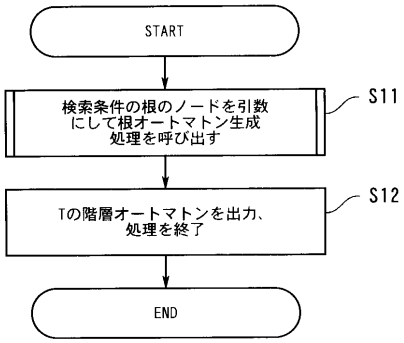
【図５】


41
40
42
SEQ
44
SEQ +r
43
46
"参考文献" or "Referennces"
"[(数字)*]" or "(数字)*."
OPTREP +r
45
"[(数字)*]" or "(数字)*."
47

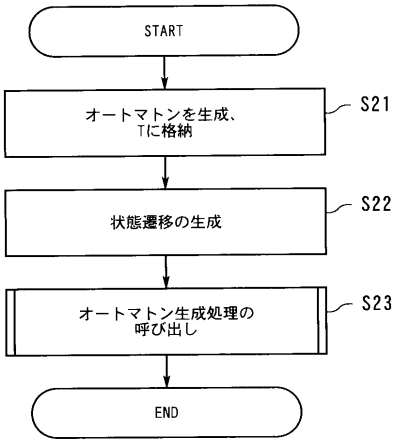
【図６】


100
11
1a
12
110
21
2a
22
2b
23
120
"参考文献" or "References"
31
3a
32
3b
33
+r
130
"[(数字)*]" or "(数字)*."
41
42
"[(数字)*]" or "(数字)*."
4b
43
44
+r
140
ε
4a
4c
4d
4e

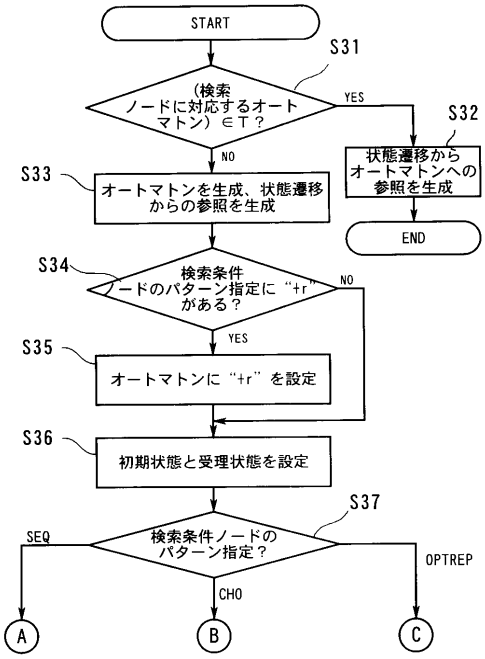
【図７】


START
検索条件の根のノードを引数にして根オートマトン生成処理を呼び出す
S11
Tの階層オートマトンを出力、処理を終了
S12
END

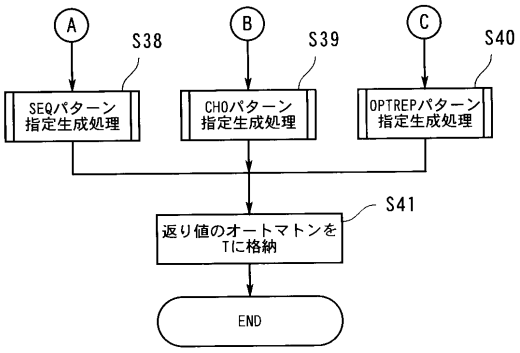
【図８】


START
オートマトンを生成、Tに格納
S21
状態遷移の生成
S22
オートマトン生成処理の呼び出し
S23
END

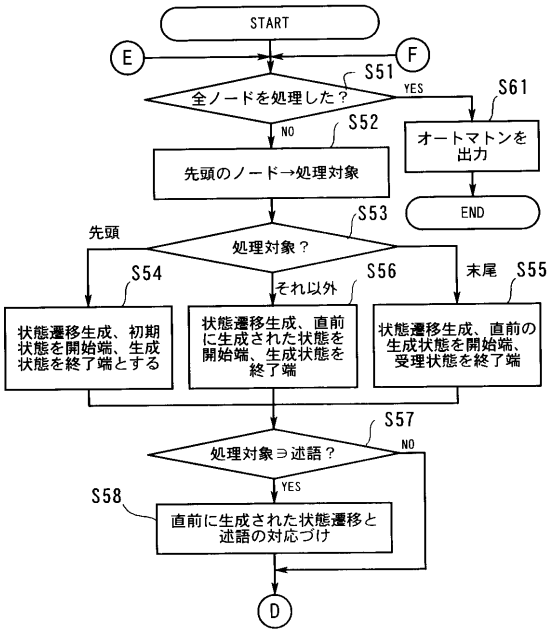
【図９】


START
S31
（検索ノードに対応するオートマトン）∈T？
YES
S32
状態遷移からオートマトンへの参照を生成
END
NO
S33
オートマトンを生成、状態遷移からの参照を生成
S34
検索条件ノードのパターン指定に“+r”がある？
NO
YES
S35
オートマトンに“+r”を設定
S36
初期状態と受理状態を設定
S37
検索条件ノードのパターン指定？
SEQ
CHO
OPTREP
A
B
C

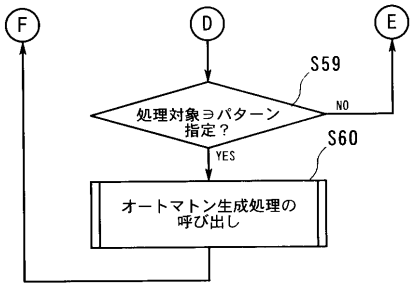
【図１０】


A
S38
SEQパターン指定生成処理
B
S39
CHOパターン指定生成処理
C
S40
OPTREPパターン指定生成処理
S41
返り値のオートマトンをTに格納
END

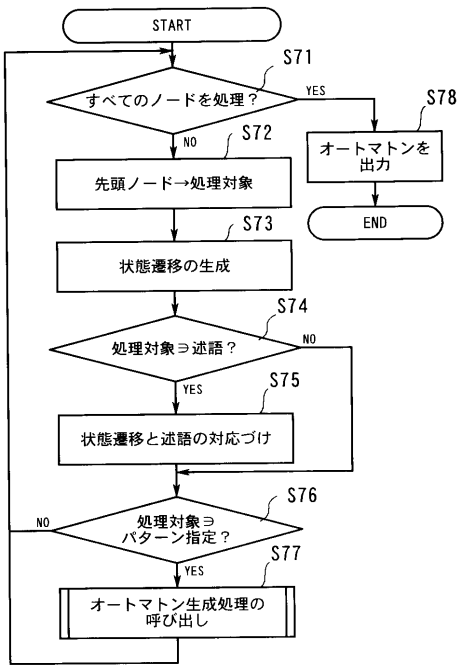
【図１１】



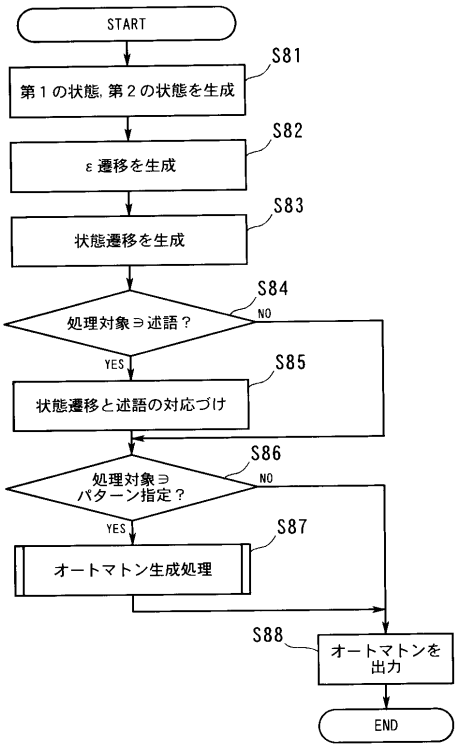
【図１２】



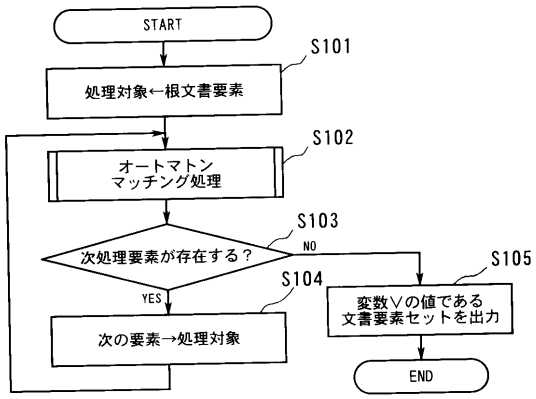
【図１３】



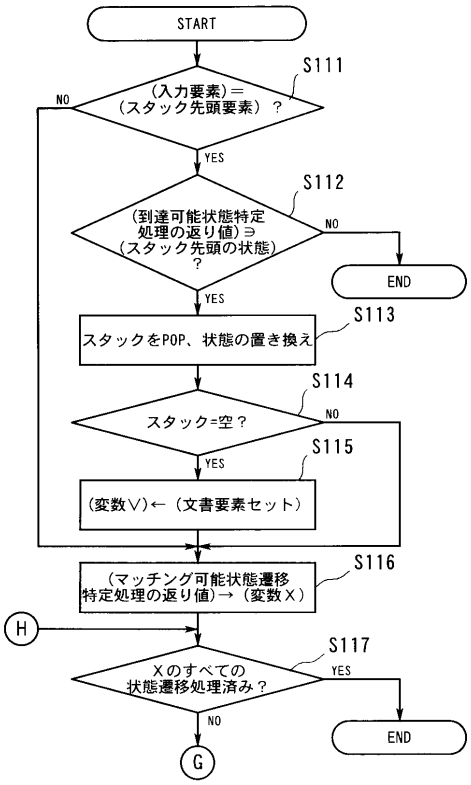
【図１４】



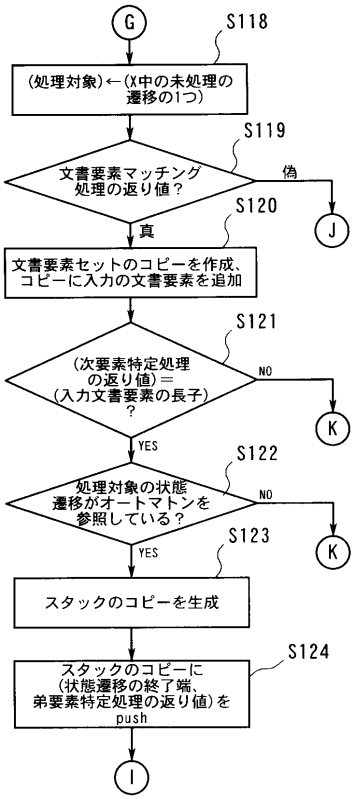
【図１５】



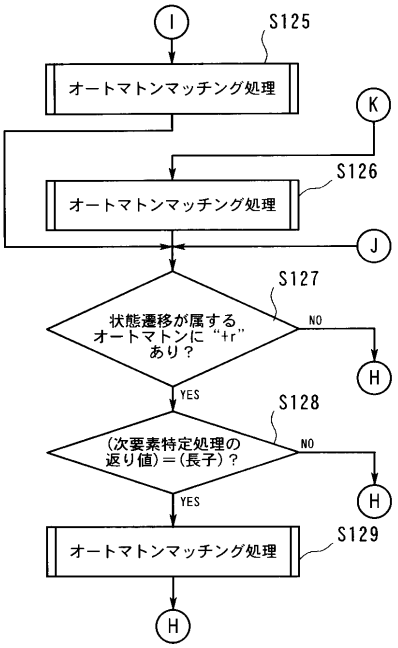
【図１６】



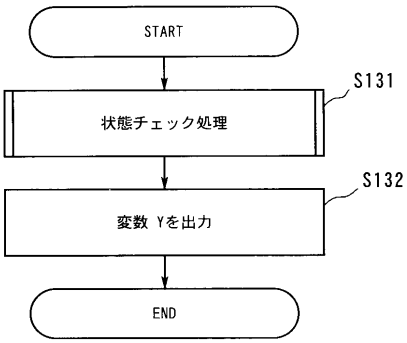
【図１７】



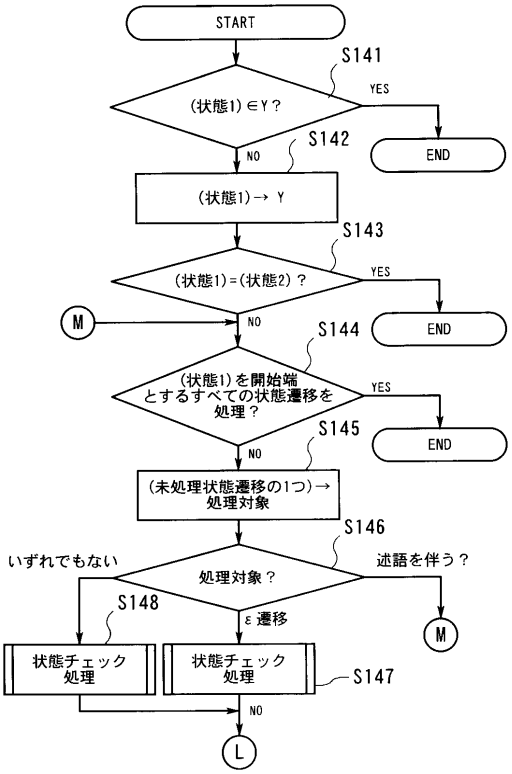
【図１８】



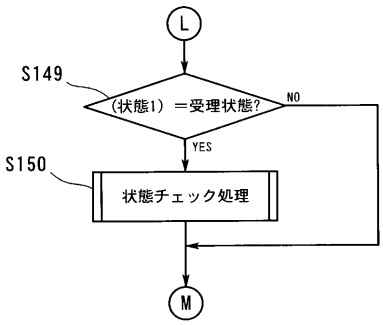
【図１９】



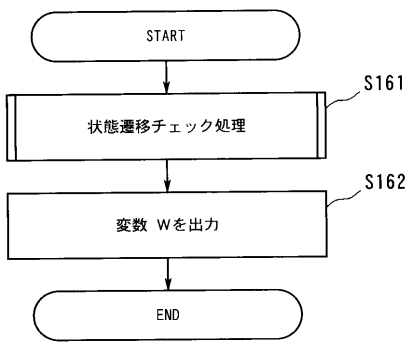
【図２０】



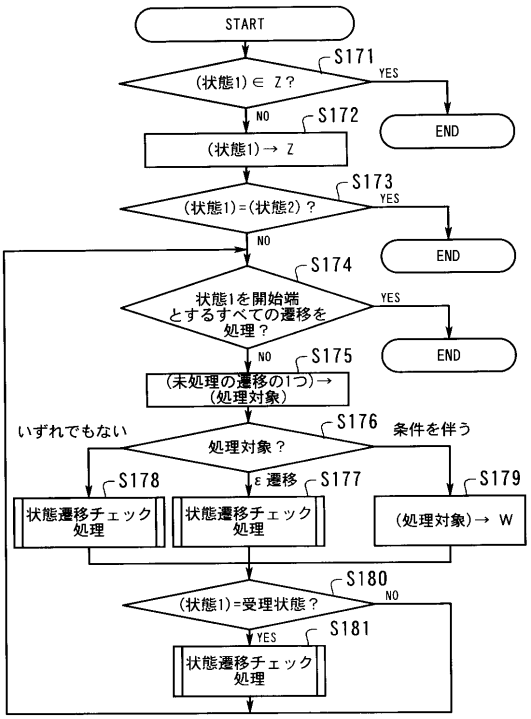
【図２１】



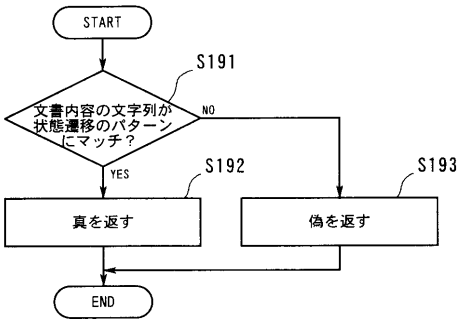
【図２２】



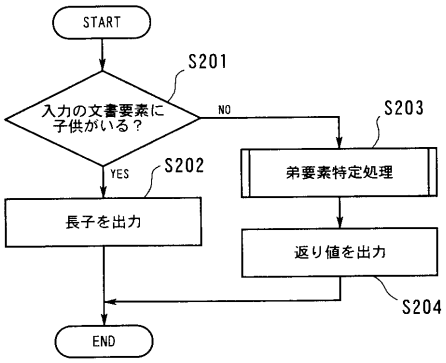
【図２３】



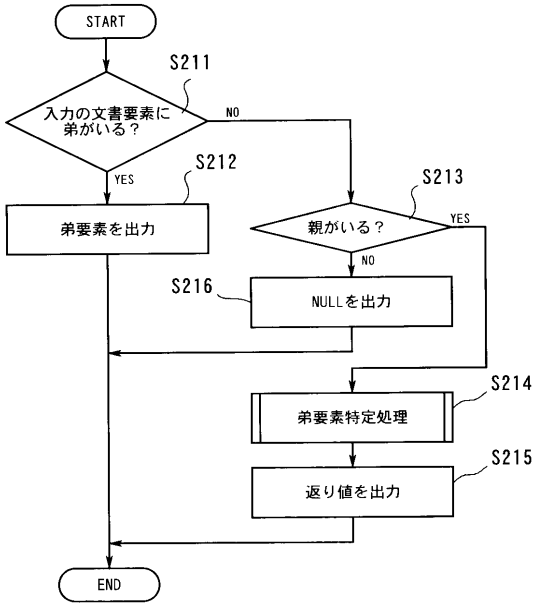
【図２４】



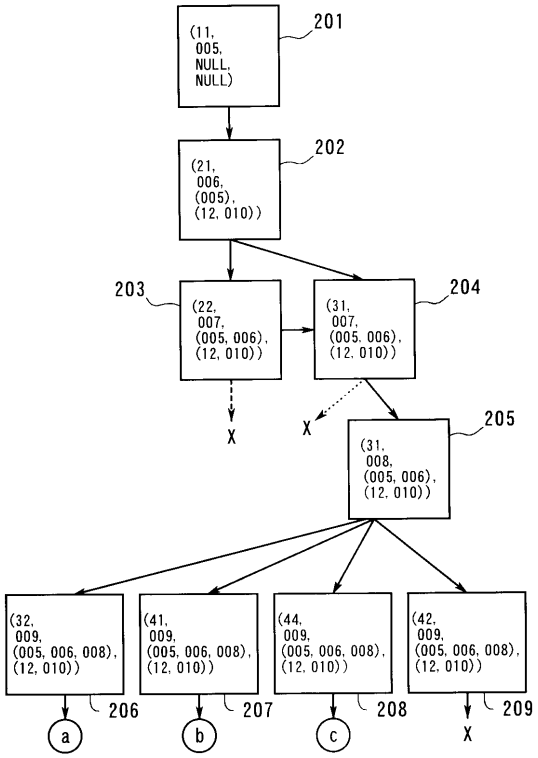
【図２５】



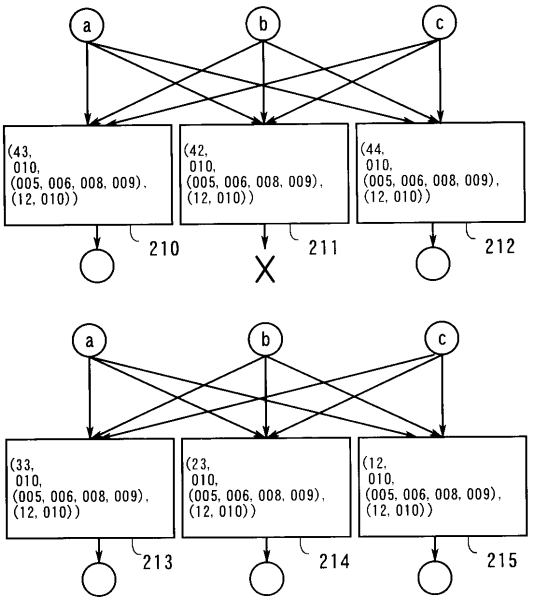
【図２６】



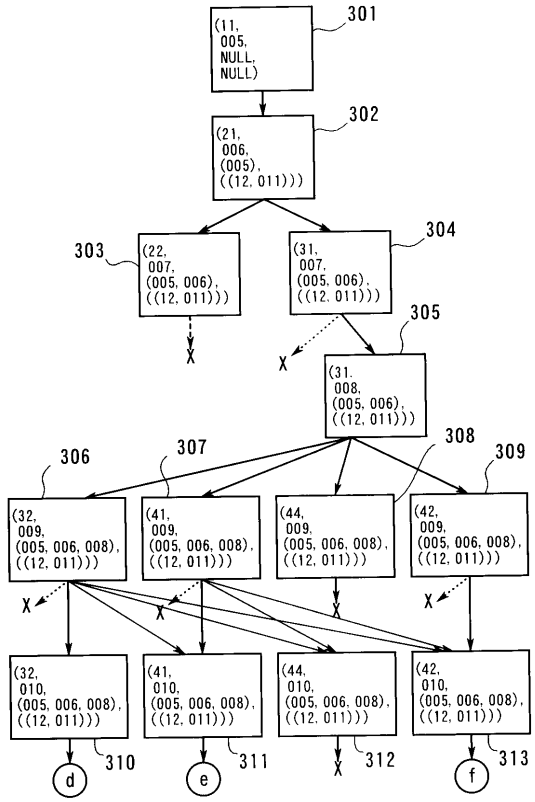
【図２７】



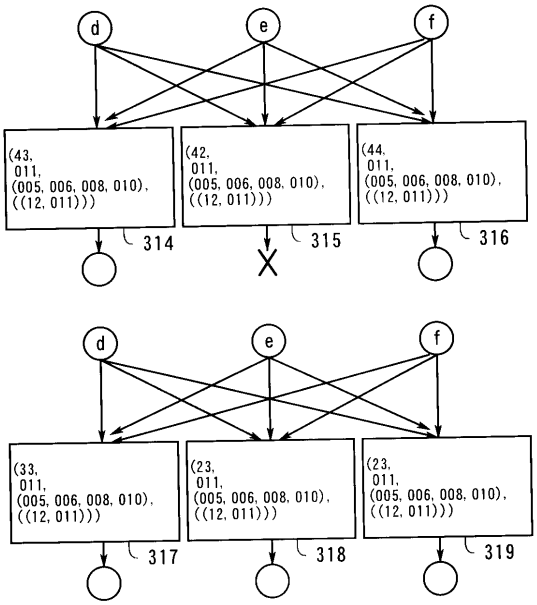
【図２８】



【図２９】



【図３０】



【図３１】

(005, 006, 008)

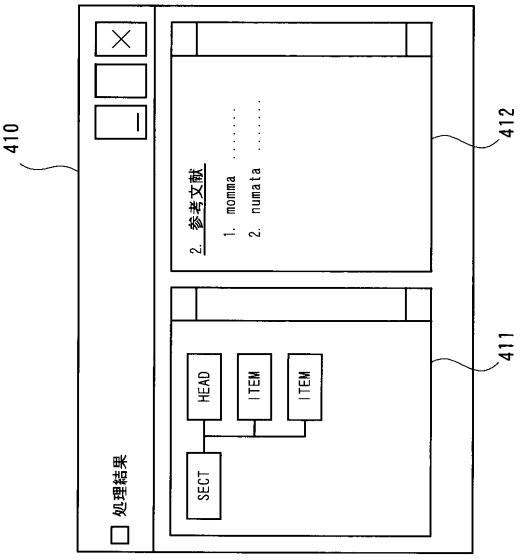
(005, 006, 008, 009)

【図３２】

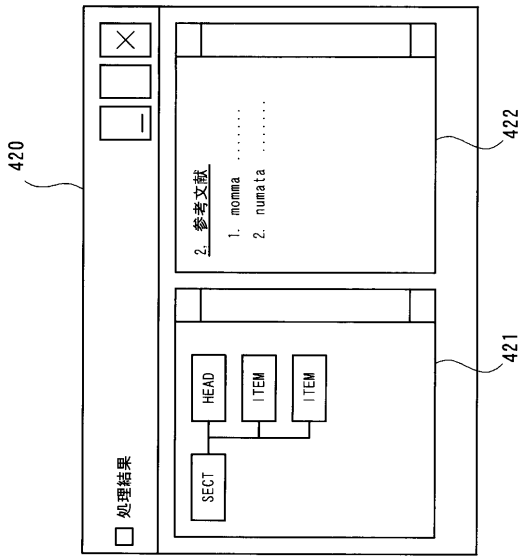
(005, 006, 008)

(005, 006, 008, 010)

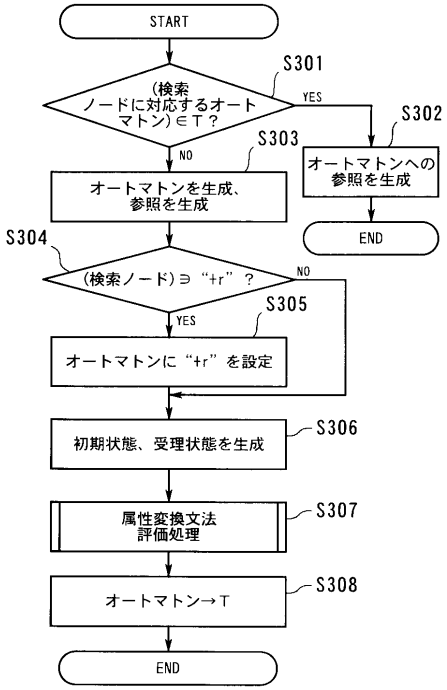
【図３３】



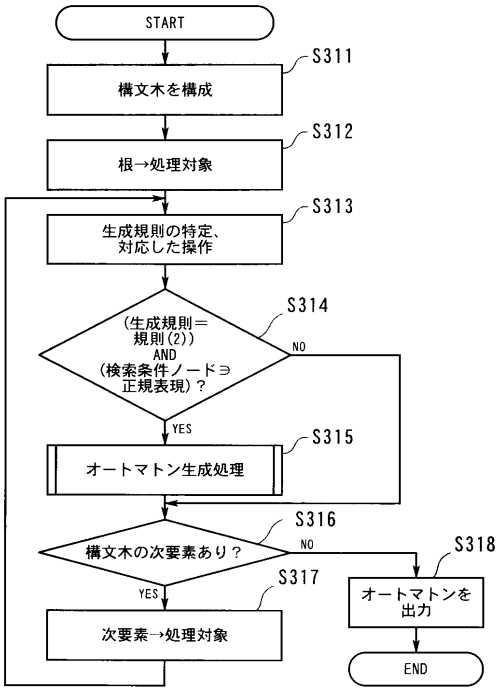
【図３４】



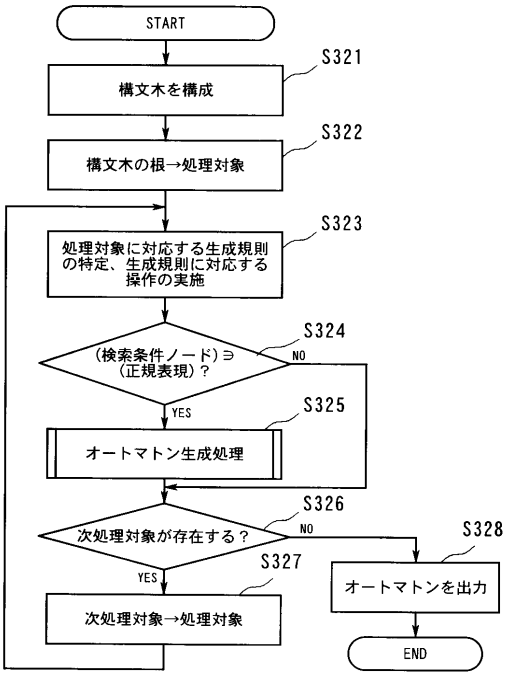
【図３５】



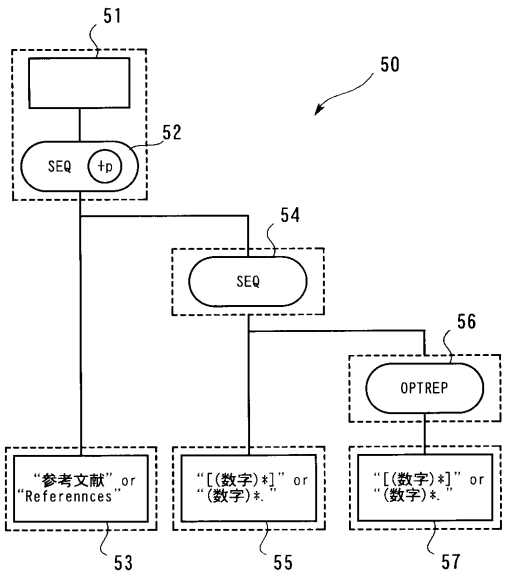
【図３６】



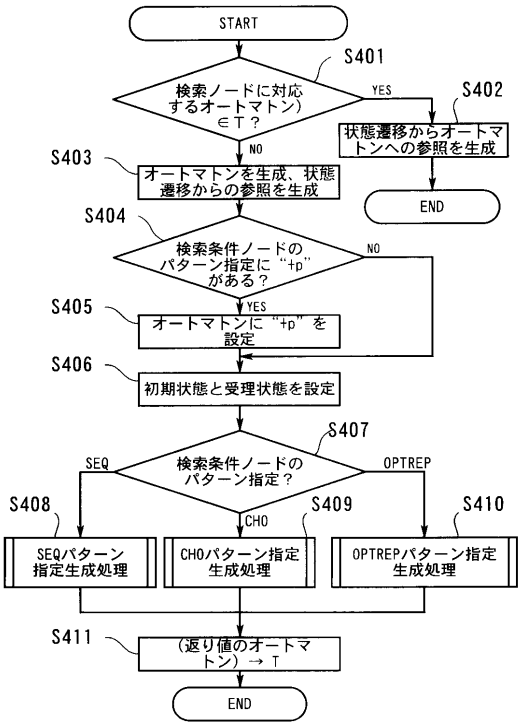
【図３７】



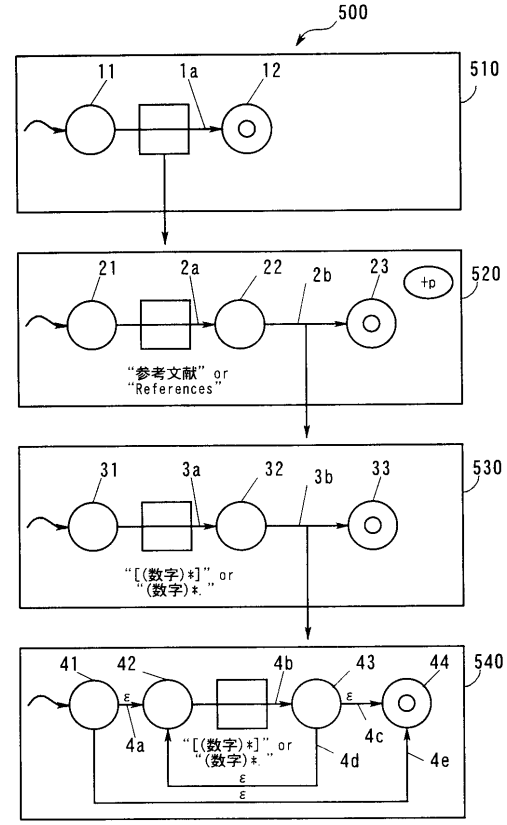
【図３８】



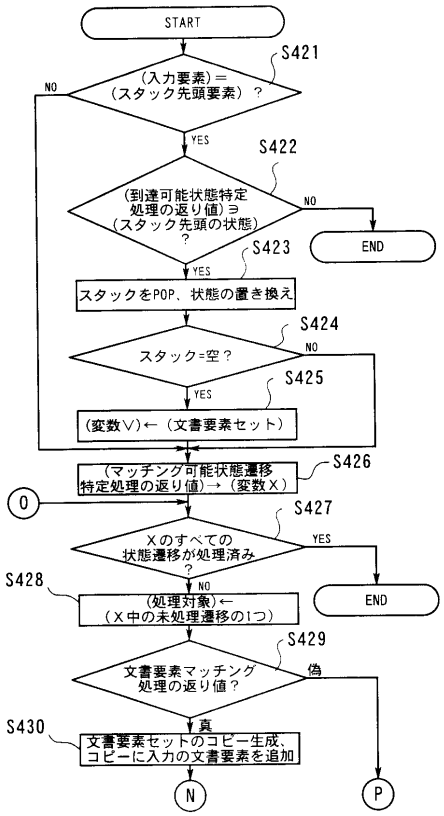
【図３９】



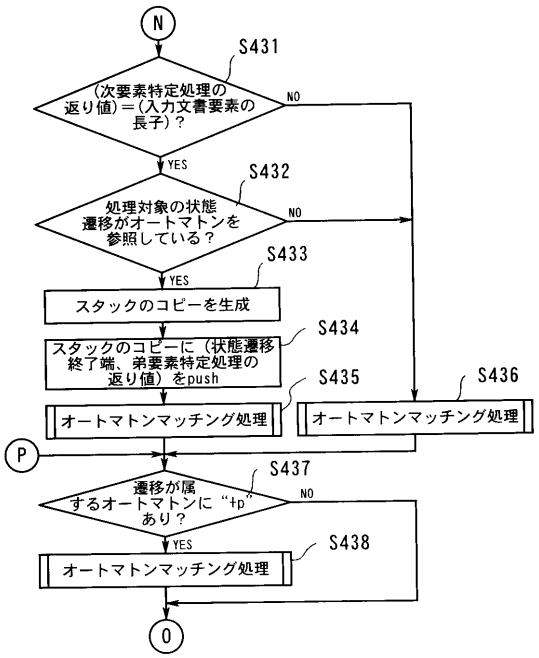
【図４０】



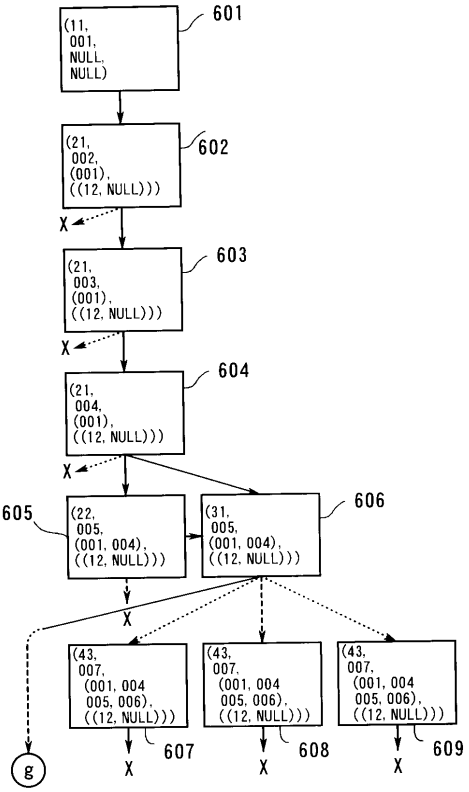
【図４１】



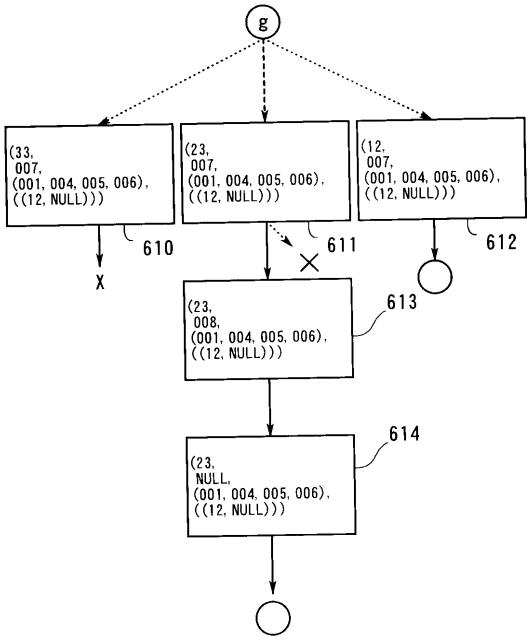
【図４２】



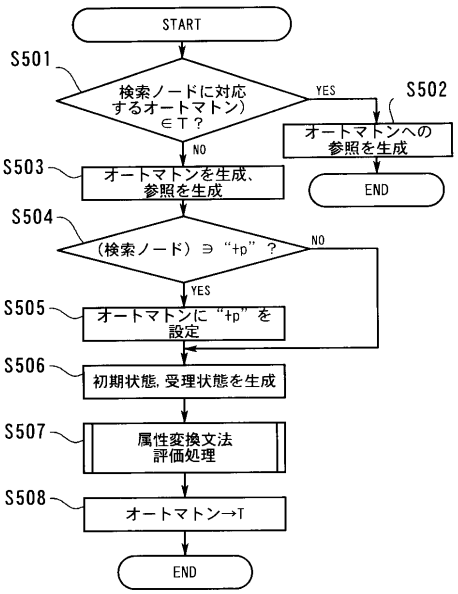
【図４３】



【図４４】



【図４６】



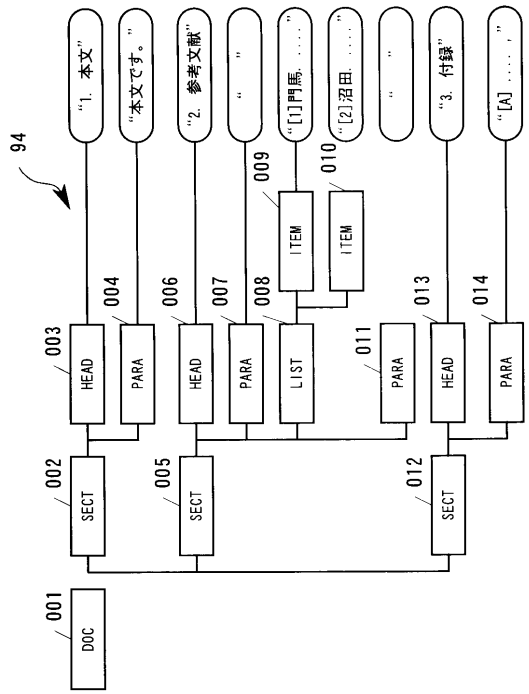
【図４５】

(001, 004,)

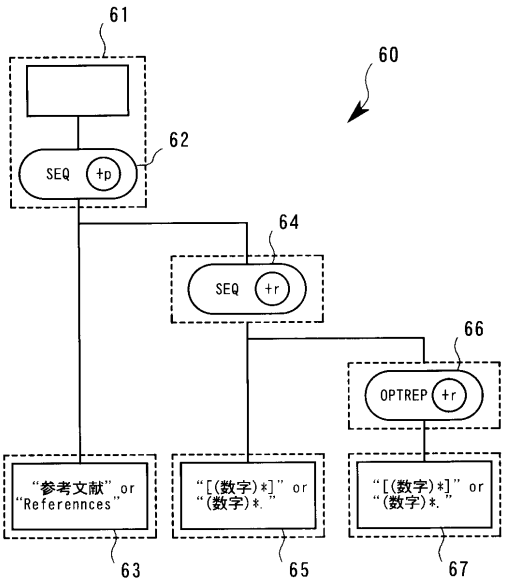
(001, 004, 005)

(001, 004, 005, 006)

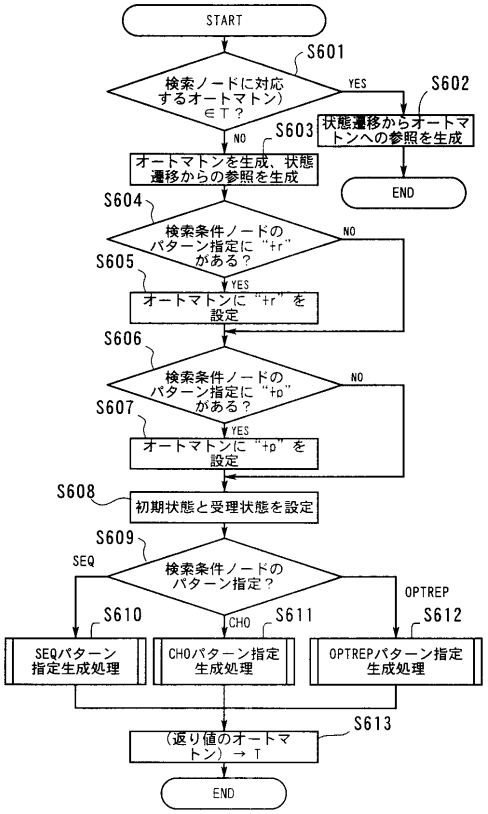
【図４７】



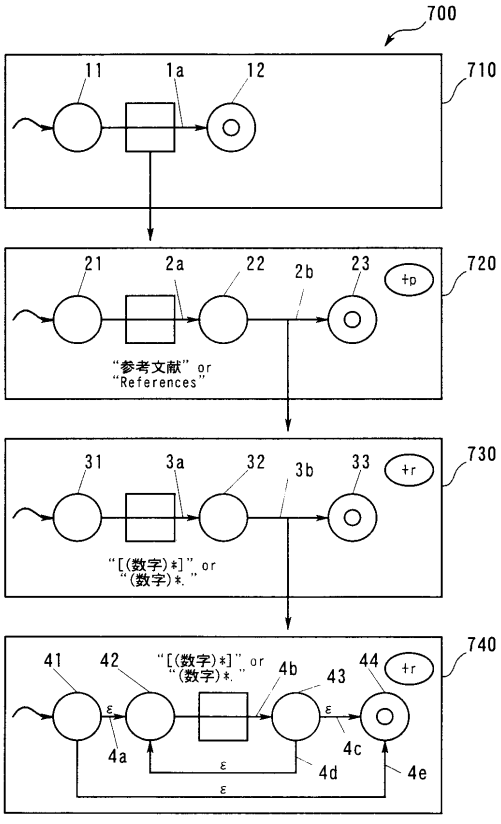
【図４８】



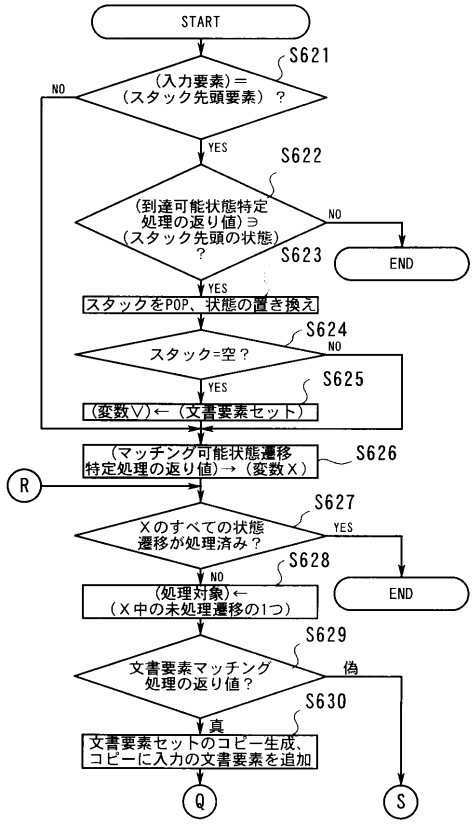
【図４９】



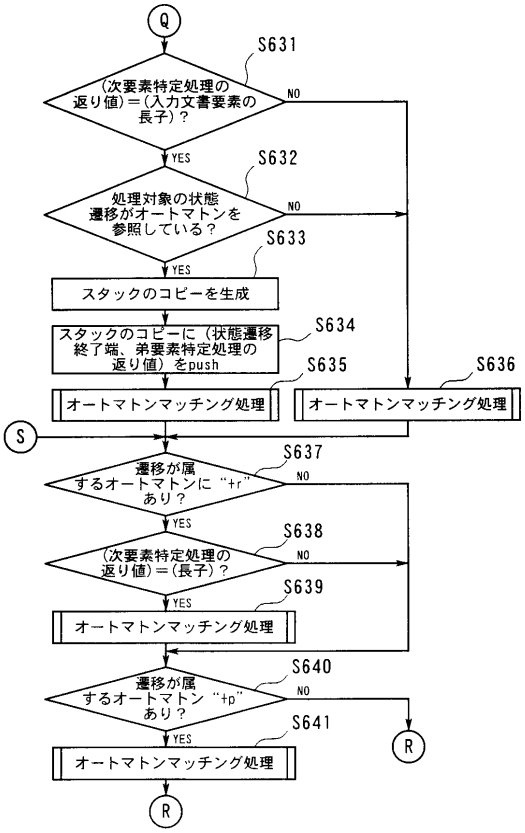
【図５０】



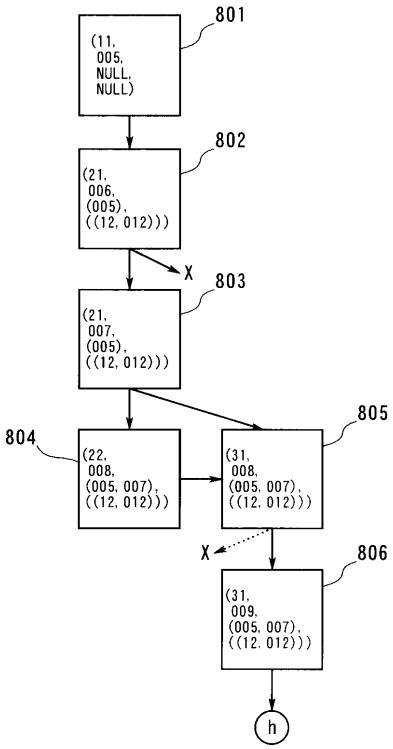
【図５１】



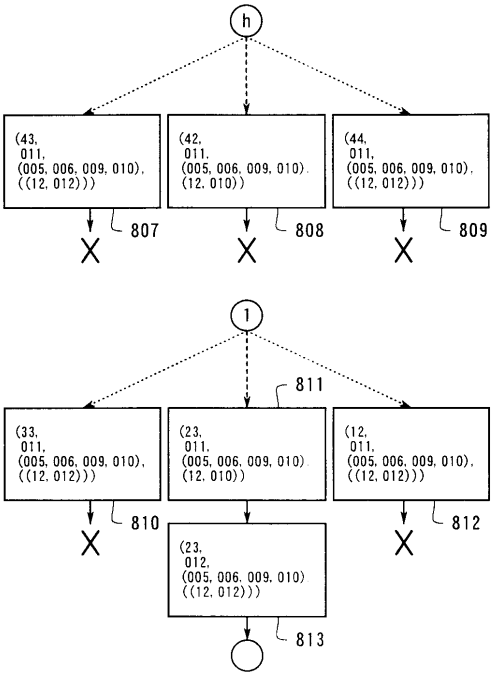
【図５２】



【図５３】



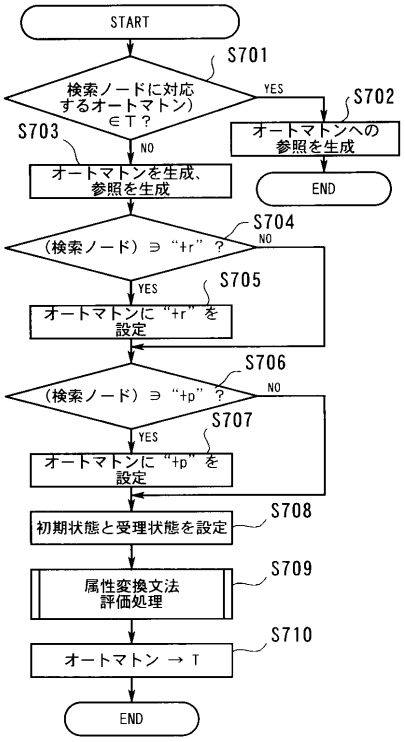
【図５４】



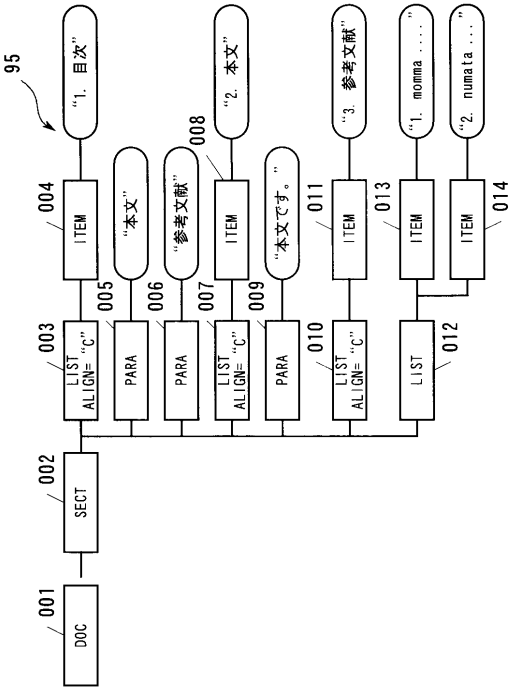
【図５５】

(005, 006)
(005, 006, 009)
(005, 006, 009, 010)

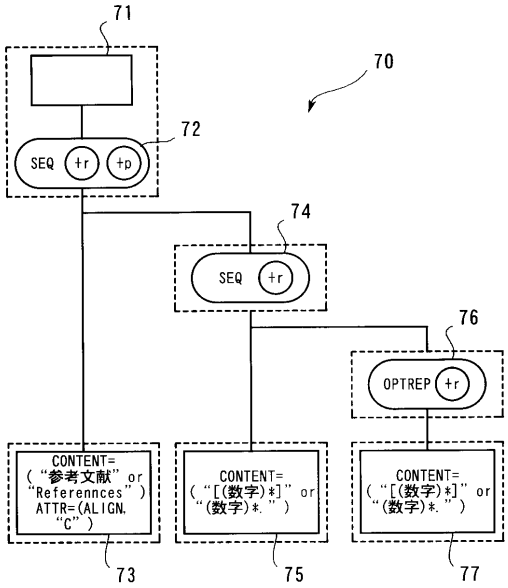
【図５６】



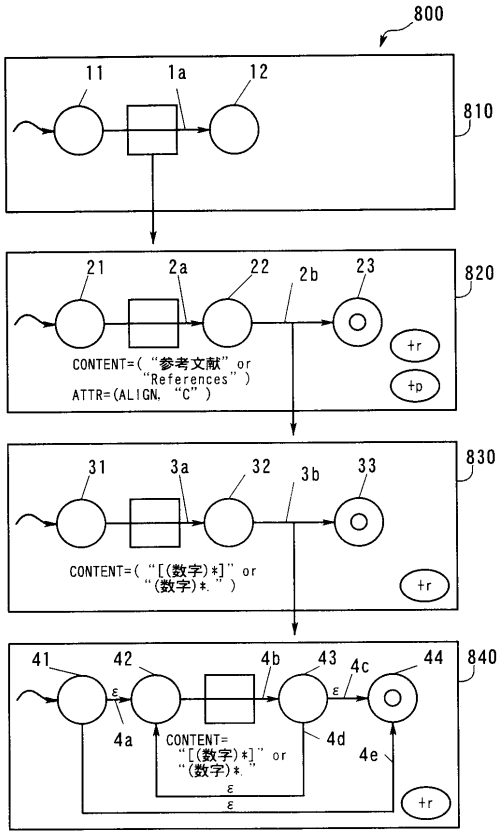
【図５７】



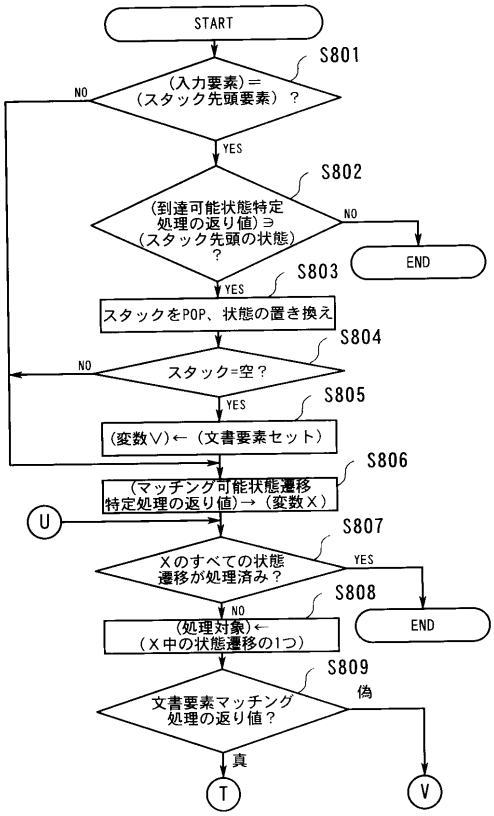
【図５８】



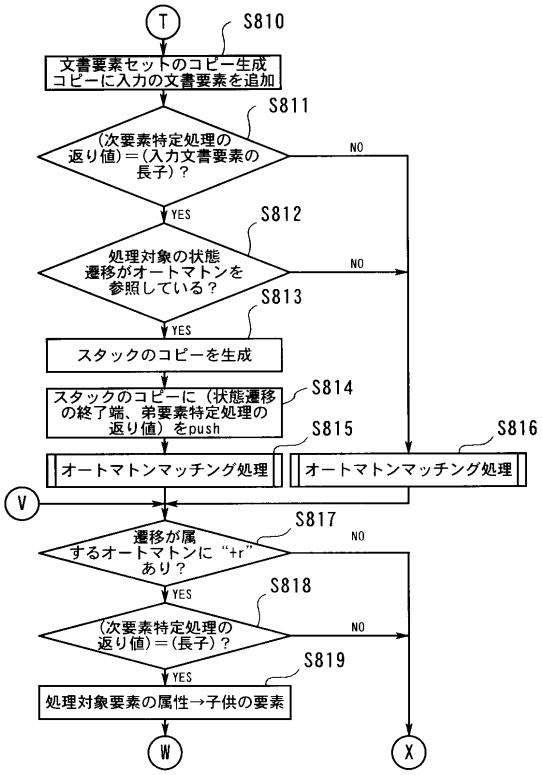
【図５９】



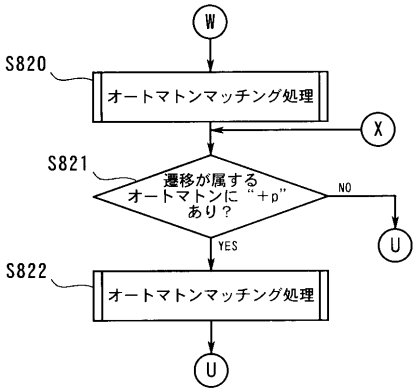
【図６０】



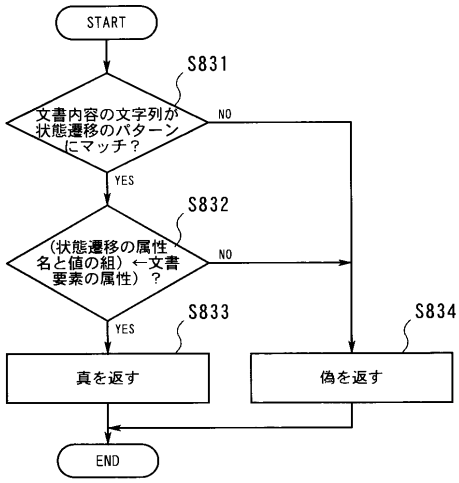
【図６１】



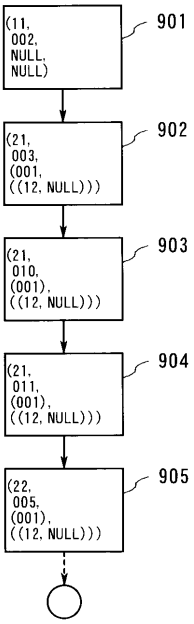
【図６２】



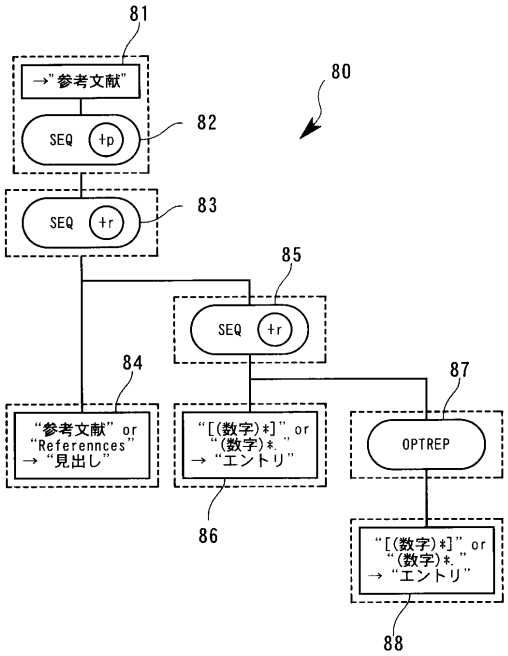
【図６３】



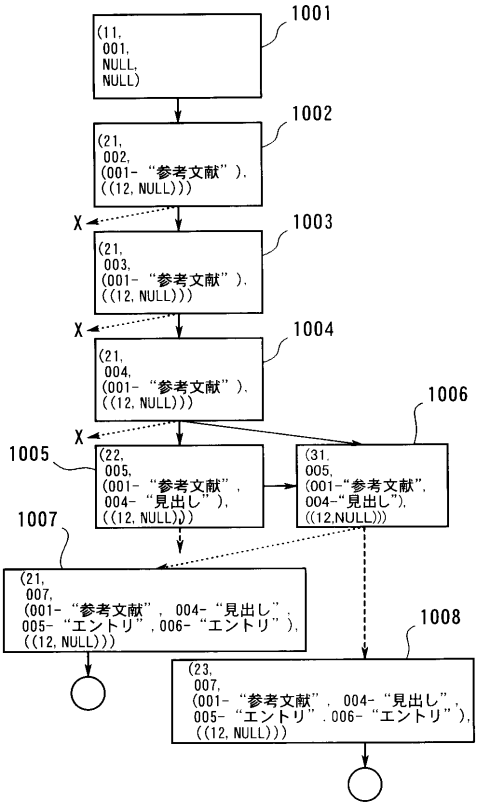
【図６４】



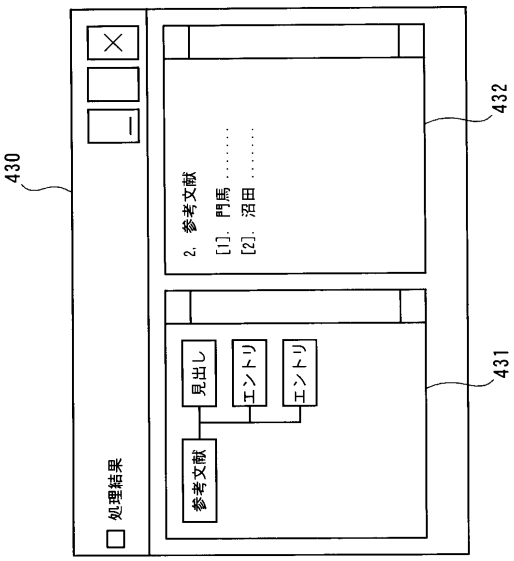
【図６５】



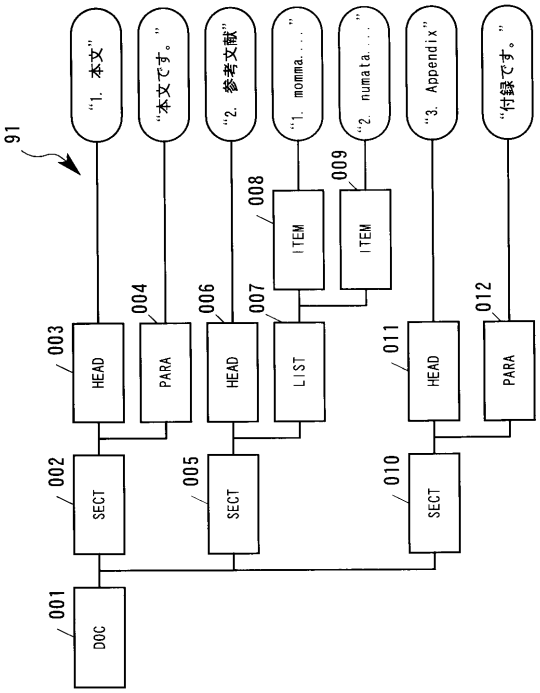
【図６６】



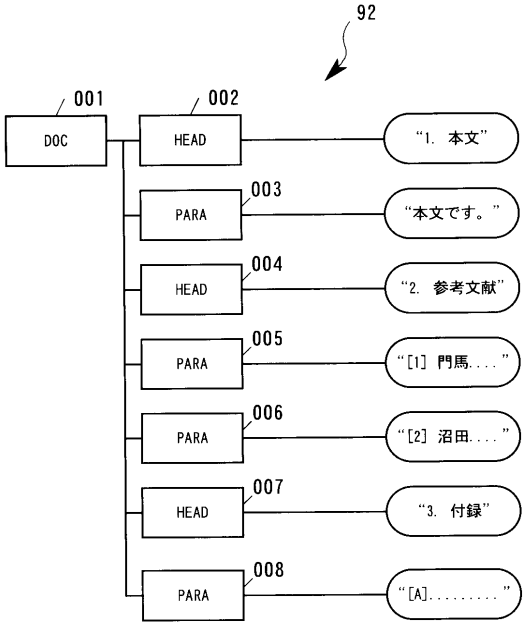
【図６７】



【図６８】


91
"1. 本文"
"本文です。"
"2. 参考文献"
"1. momma....."
"2. numata....."
"3. Appendix"
"付録です。"
001
DOC
002
SECT
003
HEAD
004
PARA
005
SECT
006
HEAD
007
LIST
008
ITEM
009
ITEM
010
SECT
011
HEAD
012
PARA

【図６９】


92
001
DOC
002
HEAD
"1. 本文"
003
PARA
"本文です。"
004
HEAD
"2. 参考文献"
005
PARA
"[1] 門馬...."
006
PARA
"[2] 沼田...."
007
HEAD
"3. 付録"
008
PARA
"[A]....."

フロントページの続き

(58)調査した分野(Int.Cl., ＤＢ名)
G06F 17/30

(54)【発明の名称】データ処理装置、文書処理装置、データ処理プログラムを記録したコンピュータ読み取り可能な記録媒体、文書処理プログラムを記録したコンピュータ読み取り可能な記録媒体、データ処理方法、および文書処理方法