

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6213019号
(P6213019)

(45) 発行日 平成29年10月18日(2017.10.18)

(24) 登録日 平成29年9月29日(2017.9.29)

(51) Int.Cl.

F 1

G06F 17/50 (2006.01)
G06F 17/27 (2006.01)G06F 17/50 G06F 17/27 G06F 17/50
6 7 2 Z 6 6 5 6 6 4 Z

請求項の数 5 (全 15 頁)

(21) 出願番号 特願2013-157782 (P2013-157782)
 (22) 出願日 平成25年7月30日 (2013.7.30)
 (65) 公開番号 特開2015-28701 (P2015-28701A)
 (43) 公開日 平成27年2月12日 (2015.2.12)
 審査請求日 平成28年4月5日 (2016.4.5)

(73) 特許権者 000005223
 富士通株式会社
 神奈川県川崎市中原区上小田中4丁目1番
 1号
 (74) 代理人 100105142
 弁理士 下田 憲次
 (72) 発明者 山本 達也
 神奈川県川崎市中原区上小田中4丁目1番
 1号 富士通株式会社内
 審査官 松浦 功

最終頁に続く

(54) 【発明の名称】シーケンス抽出方法、シーケンス抽出プログラムおよびシーケンス抽出装置

(57) 【特許請求の範囲】

【請求項 1】

コンピュータを用いて文の順序を抽出する方法であって、

前記コンピュータが有するプロセッサが、解析対象の文章に対し、前記文章を構成する文を処理手順の遷移方向をしめす時間条件語と句読点に基づいて節に分節し、前記節ごとに前記節と前記節を示す第1の識別子と前記節を包含する文を示す第2の識別子と、更に前記節が時間条件語を含む場合は、前記時間条件語と、を含む構造データを前記コンピュータが有する記憶部に記憶させる構造データ生成ステップと、

前記プロセッサが、前記構造データから時間条件語を抽出し、前記時間条件語と、前記時間条件語に対応する遷移方向と、前記時間条件語に対応する第1の識別子と、を含む時間条件データを作成する時間条件抽出ステップと、

前記プロセッサが、前記時間条件データから時間条件語を抽出し、前記構造データから前記時間条件語に対応する第2の識別子と同一の第2の識別子を持つ節に対応する第1の識別子を抽出し、前記第2の識別子について、前記時間条件語に対応する第1の識別子と、前記第2の識別子を有する節の前記第1の識別子と、前記時間条件語に対応する時間方向と、を含む動作シーケンスデータを作成するシーケンス抽出ステップと、

を有する順序抽出方法。

【請求項 2】

請求項1に記載の順序抽出方法であって、

さらに、

前記プロセッサが、前記第2の識別子を同一とする節間の状態遷移について、予め状態遷移方向を定めた文字列（状態遷移語）を抽出し、前記状態遷移語を含む節について、第1の識別子と第2の識別子と、前記状態遷移方向とを含む状態遷移ID対応データを作成する状態遷移抽出ステップと、

を有する順序抽出方法。

【請求項3】

コンピュータに、

解析対象の文章に対し、前記文章を構成する文を処理手順の遷移方向をしめす時間条件語と句読点に基づいて節に分節し、前記節ごとに前記節と前記節を示す第1の識別子と前記節を包含する文を示す第2の識別子と、更に前記節が時間条件語を含む場合は、前記時間条件語と、を含む構造データを作成する構造解析ステップと、

10

前記構造データから時間条件語を抽出し、前記時間条件語と、前記時間条件語に対応する遷移方向と、前記時間条件語に対応する第1の識別子と、を含む時間条件データを作成する時間条件抽出ステップと、

前記時間条件データから時間条件語を抽出し、前記構造データから前記時間条件語に対応する第2の識別子と同一の第2の識別子を持つ節に対応する第1の識別子を抽出し、前記第2の識別子について、前記時間条件語に対応する第1の識別子と、前記第2の識別子を有する節の前記第1の識別子と、前記時間条件語に対応する時間方向と、を含む動作シーケンスデータを作成する順序抽出ステップと、

を実行させるための順序抽出プログラム。

20

【請求項4】

請求項3に記載の順序抽出プログラムであって、

前記コンピュータに、さらに、前記第2の識別子を同一とする節間の状態遷移について、予め状態遷移方向を定めた文字列（状態遷移語）を抽出し、前記状態遷移語を含む節について、第1の識別子と第2の識別子と、前記状態遷移方向とを含む状態遷移ID対応データを作成する状態遷移抽出ステップと、
を実行させるための順序抽出プログラム。

【請求項5】

解析対象の文章に対し、前記文章を構成する文を処理手順の遷移方向をしめす時間条件語と句読点に基づいて節に分節し、前記節ごとに前記節と前記節を示す第1の識別子と前記節を包含する文を示す第2の識別子と、更に前記節が時間条件語を含む場合は、前記時間条件語と、を含む構造データを作成する構造データ生成部と、

30

前記構造データから時間条件語を抽出し、前記時間条件語と、前記時間条件語に対応する遷移方向と、前記時間条件語に対応する第1の識別子と、を含む時間条件データを作成する時間条件抽出部と、

前記時間条件データから時間条件語を抽出し、前記構造データから前記時間条件語に対応する第2の識別子と同一の第2の識別子を持つ節に対応する第1の識別子を抽出し、前記第2の識別子について、前記時間条件語に対応する第1の識別子と、前記第2の識別子を有する節の前記第1の識別子と、前記時間条件語に対応する時間方向と、を含む動作シーケンスデータを作成する順序抽出部と、

40

を有する順序抽出装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、シーケンス抽出にかかる方法、プログラム及び装置に関する。

【背景技術】

【0002】

LSI (Large Scale Integration) の開発・設計において、設計者は LSI を IC (Integrated Circuit) の規格や仕様の通りに

50

設計する。検証者は設計者の設計が規格や仕様の通りに設計されているかを検証する。近年技術開発の高度化に伴い、LSIの多機能化に伴いICの規格や仕様が複雑化している。一方で、LSIの開発から製造までの期間は短くなっている。設計者は、短期間で膨大な量の規格や仕様が書かれた規格書や仕様書を読み込み規格や仕様を満たす設計をする。検証者は、短期間で規格や仕様を満たすか検証するため検証環境やテストパターンを構築する。しかしながら、前述の規格や仕様の複雑化により、設計者が規格や仕様を分析して機能を抽出し、完全に理解した上で設計を行うことが困難になっている。また、検証者は、設計者の設計が規格や仕様を満たすか否かを検証するための検証環境やテストパターンを構築することが非常に困難になっている。

【0003】

10

設計の基礎となる規格書や仕様書は自然言語で書かれることが一般的である。自然言語で書かれた文章は、例えばLSIの内部の動作の条件や順番など動作シーケンスを文字で表現する。設計者や検証者は動作の条件や順番を膨大な文章の中から把握しなければならず、見落としが発生することが多い。そこで、動作の条件や順番を網羅的に抽出することが望まれている。

【0004】

20

動作の条件や順番などの情報を自然言語で書かれた文章から抽出するために、自然言語解析が適用される。例えば特許文献1には、自然言語解析により、入力された複数の文の相互の意味的な対応関係を自動的に解釈する技術が開示されている。また、特許文献1には、複数の文の時制を解釈することで時系列関係を解析する技術が開示されている。さらに、非特許文献1には、自然言語で書かれたWebページなどの文章からキーワードを抜き出し、文章と時間帯との関連性を抽出する技術が開示されている。

【先行技術文献】

【特許文献】

【0005】

【特許文献1】特開2006-190101

【非特許文献】

【0006】

【非特許文献1】野呂太一、外3名、「テキスト中のイベントの生起時間帯判定」、情報処理学会論文誌、2007年10月、Vol.48 No.10、p.3405 341

4

30

【発明の概要】

【発明が解決しようとする課題】

【0007】

しかしながら、上述した従来技術では、文章内のイベントの時系列は抽出出来ても、規格書や仕様書に表されるLSI等の動作シーケンスや状態遷移を抽出することは困難である。

【0008】

40

本発明は、上述した従来技術による問題点を解消するため、文章内の動作シーケンスや状態遷移を抽出し、設計者や検証者に図示する方法、プログラム及び装置を提供することを目的とする。

【課題を解決するための手段】

【0009】

上述した課題を解決し、目的を達成するため、本発明の一側面によれば、解析対象の文章に対し、前記文章を構成する文を、時間や条件の要素をもつ節を導く時間条件語と句読点とに基づいて節に分節し、前記節ごとに前記節と前記節を示す第1の識別子と前記節を包含する文を示す第2の識別子と、更に前記節が時間条件語を含む場合は、前記時間条件語と、を含む構造データを作成する構造データ生成ステップと、前記構造データから時間条件語を抽出し、前記時間条件語と、前記時間条件語に対応する遷移方向と、前記時間条件語に対応する第1の識別子と、を含む時間条件データを作成する時間条件抽出ステップ

50

と、前記時間条件データから時間条件語を抽出し、前記構造データから前記時間条件語に対応する第2の識別子と同一の第2の識別子を持つ節に対応する第1の識別子を抽出し、前記第2の識別子について、前記時間条件語に対応する第1の識別子と、前記第2の識別子を有する節の前記第1の識別子と、前記時間条件語に対応する時間方向と、を含む動作シーケンスデータを作成するシーケンス抽出ステップと、を有する順序抽出方法、プログラム及び装置が提案される。

【発明の効果】

【0010】

本発明の一側面によれば、文章内の動作シーケンスや状態遷移を抽出し、設計者や検証者に図示することができる。設計者は、図示された動作シーケンスや状態遷移を参照することにより、規格や仕様を完全に理解した上で設計することができる。また、検証者は、図示された動作シーケンスや状態遷移を参照することにより、設計者の設計が規格や仕様を満たすか否かを検証するための検証環境やテストパターンを構築することが容易となるという効果を奏する。10

【図面の簡単な説明】

【0011】

【図1】図1は、本実施の形態にかかる順序抽出装置の構成を示す説明図である。

【図2】図2は、既存の構造解析ツールによる出力イメージ図である。

【図3】図3は、本実施の形態にかかる入出力のイメージ図である。

【図4】図4は、順序抽出装置のハードウェア構成の一例を示すブロック図である。20

【図5】図5は、ファイルを読み込んでから構造データを生成するまでの全体のフローチャートである。

【図6】図6は、構造データを生成するためのフローチャートである。

【図7】図7は、構造データを生成するために文を節に分割するためのフローチャートである。

【図8】図8は、構造データを生成するために文を節に分割するためのフローチャートである。

【図9】図9は、シーケンスを抽出するためのフローチャートである。

【図10】図10は、状態遷移を抽出するためのフローチャートである。

【図11】図11は、関連する節を抽出するためのフローチャートである。30

【図12】図12は、特定の単語の時間条件方向を表すTC辞書のデータである。

【図13】図13は、文DB111の内容を示す一例である。

【図14】図14は、状態遷移データの内容を示す一例である。

【図15】図14は、状態遷移を抽出し、検証者に図示するイメージ図である。

【発明を実施するための形態】

【0012】

以下に添付図面を参照して、開示の順序抽出装置、方法、およびプログラムの実施の形態を詳細に説明する。

【0013】

図1は、本実施の形態にかかる順序抽出装置の構成を示す説明図である。本実施の形態にかかる順序抽出装置100は、文章の中から順序を抽出し視覚化する装置である。順序抽出装置100は、構造データ生成部101、時間・情報解析部102、時間条件抽出部102、状態遷移抽出部104を含み、データベースとしてTC辞書110、文DB111、状態遷移ID対応データ112を有する。抽出の対象となる自然言語で書かれた規格書や仕様書は、文章DB120に保管されている。40

TC辞書110は、時間条件語(TC語)と、時間条件方向を組みにしたデータベースである。時間条件語とは、“After”や“When”など、時間(Time)や条件(Condition)の要素を持った単語のことである。時間条件語は、文法用語で言えば、「時の副詞節を導く接続詞」である。解析対象が規格書や仕様書である文章の例にあてはめると、時間条件語は、LSIの内部の動作の実行時期や実行条件を記述する副詞節

を導く接続詞である。

【 0 0 1 4 】

図2は、既存の構造解析ツールによって仕様書に書かれた文章を解析した出力イメージ図である。仕様書に書かれた文章として、例えば式1の文章を入力する。

【 0 0 1 5 】

【数1】

After the Configure Endpoint Command and SET_CONFIGURATION request complete successfully, software may schedule TDs on any enabled endpoint Transfer Ring. ... (1)

10

【 0 0 1 6 】

既存の構造解析ツールは、式 1 の構造を解析して図 2 を得る。図 2 では、S (s e n t e n c e) 2 0 1 や 2 0 3 は節を示し、P N (P u n c t u a t i o n) 2 0 2 は読点を示す。式 2 は読点 2 0 2 で 2 つの節に分割される。既存の構造解析ツールは、節 2 0 1、2 0 2 について、それぞれ、単語レベルに至るまで構造を解析する。

図3は、順序抽出装置が仕様書にかかれた文章を解析し、順序を抽出するイメージ図である。式1の文章に対し、順序抽出装置の動作を説明する。図3では、図解の視認性向上のために「Configure Endpoint Command」を「CEC」と表示し、「SET_CONFIGURATION」を「SC」と表示する。

20

[0 0 1 7]

構造データ生成部 101 は、式 1 の文章の形態素解析や係り受け解析を行う。形態素解析や係り受け解析は既存の技術であるため詳細を省略する。係り受け解析の結果、単語 “request” は CEC301 と SC302 にかかっていることが判明する。グループ 304 は CEC301 と SC302 についての状態遷移の要件をまとめた。

[0 0 1 8]

グループ307はグループ304の条件が満たされたときに遷移する先である。グループ304の時間条件語“After”303が示すとおり、グループ304が完了(complete)したら、グループ307に遷移する。グループ307は動作主体であるsoftware305と動作対象であるTDS306にかかる単語softwareとTDSを保存する。

30

(0 0 1 9)

図4は、順序抽出装置100のハードウェア構成の一例を示すブロック図である。図4において、順序抽出装置100は、Central Processing Unit(CPU)401と、Read Only Memory(ROM)402と、Random Access Memory(RAM)403と、を含む。また、順序抽出装置100は、記憶装置404と、インターフェイス408と、を含む。また、順序抽出装置100は、ディスプレイ305と、キーボード406と、マウス407と、を含む。また、CPU401～インターフェイス408はバス400によってそれぞれ接続されている。

40

[0 0 2 0]

CPU401は、順序抽出装置400の全体の制御を司る演算処理装置である。ROM402は、ブートプログラムなどのプログラムを記憶する不揮発性メモリである。RAM403は、CPU401のワークエリアとして使用される揮発性メモリである。

[0 0 2 1]

記憶装置404は、記憶媒体を内部に保持する内部記憶装置409と記憶媒体が外部に取り出せる外部記憶装置410を含んでいる。また、記憶装置404はCPU401の制御にしたがって内部記憶媒体411または外部記憶媒体412に対するデータのリードお

50

およびライトを制御する制御装置である。内部記憶装置 409 には、内部記憶媒体 411 の種類に応じて磁気ディスクドライブやソリッドステートドライブ (SSD) を選択することができる。外部記憶装置 410 は、たとえば、磁気テープ装置、磁気ディスクドライブ、光磁気ディスクドライブ、光ディスクドライブ、SSD などである。外部記憶媒体 412 は、外部記憶装置 410 の制御で書き込まれたデータを記憶する不揮発性メモリである。たとえば外部記憶装置 410 が磁気ディスクドライブである場合、外部記憶媒体 412 は、磁気ディスクである例えはフレキシブルディスク (FD) である。また、外部記憶装置 410 が光ディスクドライブである場合、外部記憶媒体 412 は、光ディスクである例えは Compact Disc Recordable (CD-R) である。また、外部記憶装置が SSD である場合、外部記憶媒体は、半導体素子メモリである。

10

【0022】

ディスプレイ 405 は、カーソル、アイコンあるいはツールボックスをはじめ、文書、画像、機能情報などのデータ表示する装置である。ディスプレイ 405 は、たとえば、Cathode Ray Tube (CRT)、Thin Film Transistor (TFT) 液晶ディスプレイ、プラズマディスプレイ、有機エレクトロルミネッセンスディスプレイなどである。

【0023】

キーボード 406 は、文字、数字、各種指示などの入力のためのキーを有し、データの入力を行う装置である。また、キーボード 406 は、タッチパネル式の入力パッドやテンキーなどであってもよい。マウス 407 は、カーソルの移動や範囲選択、あるいはウィンドウの移動やサイズの変更などを行う装置である。マウス 407 は、ポインティングデバイスとして同様に機能を有するものであれば、トラックボールやジョイスティックなどであってもよい。

20

【0024】

インターフェイス 408 は、ネットワーク 220 と接続し、外部装置からのデータの入出力を制御する制御装置である。インターフェイス 408 は、接続するネットワーク 420 の種類に対応する。インターフェイス 408 は、例えは有線 LAN や無線 LAN の Network Interface Card (NIC) である。

【0025】

図 5 は本願発明の全体フローである。構造データ生成部 101 は、文章 DB に格納されたファイルを読み込む (ステップ S501)。構造データ生成部 101 は、ファイルの中の文章のうち、文がいくつあるか数を取得する (ステップ S502)。構造データ生成部 101 は、ファイルの中の文章の最初の文から最後の文それぞれについて、ステップ S504 ~ S507 の処理を行う (ステップ S503 乃至ステップ S508)。構造データ生成部 101 は、文に対し一意の識別子 (文 ID) を付与する (ステップ S504)。次に構造データ生成部 101 は、文について形態素解析 (ステップ S505)、構造解析 (ステップ S506) を行う。ここで形態素解析 (ステップ S505)、構造解析 (ステップ S506) は既存の自然言語解析技術を利用することができる。続いて構造データ生成部 101 は、構造データを生成する (ステップ S507)。構造データについては後述する。構造データ生成部 101 は、最後の文が終了するまでステップ S504 ~ S507 の処理を行う (ステップ S508)。

30

【0026】

図 6 は構造データ生成 (ステップ S507) のフローである。最初に構造データ生成部 101 は、文についてカンマが含まれるか判定する (ステップ S601)。文にカンマが含まれる場合、構造データ生成部 101 は、文頭からカンマ前までを第 1 の節、カンマからピリオド前までを第 2 の節として抽出する (ステップ S602)。次に構造データ生成部 101 は、第 1 の節、第 2 の節両方に対してステップ S604、ステップ S605 の処理を行う (ステップ S603 乃至ステップ S606)。構造データ生成部 101 は、節に対し、一意の識別子 (節 ID) を付与する (ステップ S604)。節 ID は、文 ID を接頭辞として、節の順番を子番号として付すことができる。一例として、文 ID を “AAA”

40

50

としたときに、第1の節の節IDを“AAA.1”、第2の節の節IDを“AAA.2”と付すことができる。その後、構造データ生成部101は、節を細分化する節細分化1を行う(ステップS605)。構造データ生成部101は、最後の節が終わるまで、ステップS604、S605の処理を行う(ステップS606)。

【0027】

構造データ生成部101は、文にカンマが含まれないと判定したときは、次に文に時間条件語があるか否かを判定する(ステップS610)。構造データ生成部101は、文に時間条件語が含まれないと判断した時は終了する。構造データ生成部101は、文に時間条件語が含まれると判断した時は、かかる時間条件語が文頭にあるか否かを判定する(ステップS611)。構造データ生成部101は、文頭に時間条件語がある時は終了する。構造データ生成部101は、文頭に時間条件語がない場合、文頭から時間条件語前までを第3の節、時間条件語からピリオド前までを第4の節として抽出する(ステップS612)。次に構造データ生成部101は、第3の節と第4の節に対してそれぞれステップS614の処理を行う(ステップS613、ステップS615)。構造データ生成部101は、節に対し、一意の識別子を付与する(ステップS614)。かかる識別子は、ステップS604で説明した方法で付すことができる。構造データ生成部101は、最後の節が終わるまで、ステップS614の処理を行う(ステップS615)。

【0028】

図7はステップS605で述べた、文を節に細分化する節細分化1のフローである。構造データ生成部101は、節の中に時間条件語があるか否かを判定する(ステップS701)。構造データ生成部101は、節に時間条件語が含まれないと判断した時は終了する。構造データ生成部101は、節に時間条件語が含まれると判断した時は、かかる時間条件語が節の頭にあるか否かを判定する(ステップS702)。構造データ生成部101は、節の頭に時間条件語がないと判定した時は、節の頭から時間条件語の前までと、時間条件語から節の末尾までを新たな節として分割する(ステップS703)。構造データ生成部101は、節の頭に時間条件語があるときは、節に文頭を表すフラグを付与する(ステップS704)。構造データ生成部101は、文頭の時間条件語から節の末尾までを節として抽出し(ステップS705)、節ID、時間条件語、文頭フラグを文DB111に保存する(ステップS706)。

【0029】

図8は式1の文801に対して、図5～7のフローを経た文と節の関係を具体例で示したものである。構造データ生成部101は、式1の文801に文ID“AAA”を付与し(ステップS504)、形態素解析(ステップS505)、構造解析(ステップS506)を行った後、図6の構造データ生成フローに記載する処理を行う。

【0030】

構造データ生成部101は、文801についてカンマが含まれるか判定する(ステップS601)。ここで式1にはカンマが含まれるため、構造データ生成部101は、文頭からカンマ前までの節802と、カンマからピリオドまでの節803をそれぞれ抽出する(ステップS602)。次に、構造データ生成部101は、節802と節803の両方に対し以下の処理を行う(ステップS603～ステップS606)。構造データ生成部101は、節802に文801の文ID“AAA”を接頭辞とする節ID“AAA.1”を付与する(ステップS604)。ここで、「.1」なる子番号は、文頭から近い順に、昇順で付与される。次に構造データ生成部101は、節802の細分化を試みる(ステップS505)。

【0031】

構造データ生成部101は、節802に時間条件語が含まれるか判定する(ステップS701)。節802には時間条件語である“A f t e r”が存在するため、次に、構造データ生成部101は、節802の節頭に時間条件語があるか判定する(ステップS702)。時間条件語“A f t e r”は節802の節頭に存在するため、構造データ生成部101は、節802に文頭フラグを付与する(ステップS704)。次に、構造データ生成部

10

20

30

40

50

101は、時間条件語から節末までを節として抽出する（ステップS705）。言い換えると、文頭の“After”から節末まで、すなわち節802そのものを節として抽出する。構造データ生成部101は、節802について節IDである“AAA.1”及び時間条件語“After”、文頭フラグ関連付けて文DB111に保存する（ステップS706）。

【0032】

図9は時間条件を抽出するフローである。時間条件時間条件抽出部102は、文DB111から一行を選択する（ステップS901）。時間条件抽出部102は、選択された行の中から時間条件語を選択する（ステップS902）。次に、時間条件抽出部102は、選択された行の中に文頭フラグが付されているか判定する（ステップS903）。文頭フラグが付されていないと判定した場合、時間条件抽出部102は、TC辞書を探索し、選択した時間条件語の方向を取得し（ステップS904）、終了する。文頭フラグが付されていると判定した場合、時間条件抽出部102は、TC辞書を探索して選択した時間条件語が文頭に存在する場合の方向を取得し（ステップS905）、文DB111の選択した行に方向を記録する（ステップS906）。

【0033】

時間条件語は文頭に存在する場合と文頭以外に存在する場合がある。式1の文801では“After”という時間条件語が文頭に存在する。この場合は、時間条件語“After”を含む節802の内容の後に後半の節803の内容が発生する。一方、時間条件語“After”が文頭にない式2の場合を説明する。

【0034】

【数2】

An endpoint is in the Stopped state after the CEC request complete successfully. ・・・(2)

【0035】

式2の文に対し、図5～7の処理を行うと、文頭から時間条件語“After”的前までの節と、時間条件語“After”を含む節とに分割される。時間条件語“After”を含む節の内容が先にあり、文頭から時間条件語“After”的前の節の内容が後に発生する。即ち、同じ単語で表される時間条件語でも、文頭に存在するか否かで文の構造の時間・条件の方向が変化する。文DB111に文頭フラグがある場合は、時間条件語が文頭に存在する場合の方向を文DB111へ格納する（ステップS906）。

【0036】

図10は状態遷移を抽出するフローである。状態遷移抽出部104は、文DB111のすべての行に対し以下の処理を行う（ステップS1001、ステップS1006）。状態遷移抽出部104は、文DB111の1行を選択し、当該行に保存されている節を第一の実体とする（ステップS1002）。状態遷移抽出部は、第1の実体と一致する節があるか否か文DB111の検索を行う（ステップS1003）。一致する節がある場合、状態遷移抽出部104は、関連情報を抽出し（ステップS1004）、状態遷移データに保存する（ステップS1005）。すべての行に対し処理が終わったら、状態遷移抽出を終了する（ステップS1006）。

【0037】

図11は、関連情報の抽出についてのフローである。状態遷移抽出部104は、前述の第1の実体と一致した節（第2の実体）の属する行の節IDから、文IDと子番号を付した回数による階層情報を取得する（ステップS1101）。次に、状態遷移抽出部は、階層情報が1以上であるか判定する（ステップS1102）。階層情報が1以上であれば、状態遷移抽出部は、文DBから同一の文IDで同一の階層情報を持つ行の節（第3の実体

10

20

30

40

50

)を探索する(S 1 1 0 3)。次に状態遷移抽出部は、第2の実体と第3の実体が時間条件語で接続されるか判定する(ステップS 1 1 0 4)。具体的には、第2の実体の行に属する「方向」と、第2の実体の節IDと第3の実体の節IDの先後関係とで判断する(ステップS 1 1 0 4)。第2の実体の行に属する「方向」が、「右」であり、かつ第3の実体の節IDが第2の実体の節IDの後であるとき、第2の実体の節IDをポイント先として接続されると判断し、第3の実体を取得する(ステップS 1 1 0 5)。第2の実体の行に属する「方向」が、「左」であり、かつ第3の実体の節IDが第2の実体の節IDの前であるとき、第2の実体の節IDをポイント先として接続されると判断し、第3の実体を取得する(ステップS 1 1 0 6)。これら以外のときはそのまま終了する。

【0038】

10

図12は時間条件語と、時間条件方向を組みにしたデータベースであるTC辞書1200の一例を示したものである。例えば、時間条件語1202毎に時間の経過方向や、条件の成立による関係を示す時間条件方向1201を記録する。時間条件語1202が“After(文頭)”である行1203に着目すると、時間条件方向は「右」となっている。文頭に時間条件語“After”がある節は、右側の節、即ちその節に続く節との間に関係を有していることを示す。一方で、時間条件語1202が“after”である行1204に着目すると、時間条件方向は「左」となっている。afterがある節は、左側の節、即ちその節の前の節との関係を有していることを示す。

【0039】

20

図13は文DB111の内容を表す例である。文DB1300は、構成要素として節ID1301、時間条件語1302、文頭フラグ1303、節の実体1304及び方向1305を有する。

行1306に着目すると、節ID1301として、“AAA.1”が記録されている。これから行1306の節は、文IDが“AAA”である文の一部であることが分かる。時間条件語1302には、“After”が記録されている。そして、方向1305には右と記録され、文頭フラグ1303には1が記録されている。これらのことから、行1306の節は、文頭に時間条件語“After”が存在し、右側の節、即ち節IDが“AAA.1”的節に続く節との間に関係を有している節であることが分かる。節の実体1304には「After the CEC and SC request complete successfully」が記録されている。

【0040】

30

行1307に着目すると、時間条件語1302“after”が記録されている。そして、方向1305には左と記録され、文頭フラグ1303には0が記録されている。これらのことから、行1307の節は、文頭ではない時間条件語“after”が存在し、左側の節、即ち節IDが“BBB.2”的節の前の“BBB.1”的節との間に関係を有している節であることが分かる。節の実体1304には「after the CEC request complete successfully」が記録されている。

【0041】

図14は状態遷移データの例である。状態遷移データ1400には、節ID1401、節の実体1402、ポイント先の節ID1403、時間条件語1404が記録されている。ポイント先節ID1403は、今の節がどの節の実体に遷移するか遷移先を示したものである。時間条件語1404には、遷移に関係する時間条件語を保存する。

40

行1405の節に着目すると、節ID1401は“AAA.1”を記録する。節の実体1402には「After the CEC and SC request complete successfully」が記録されている。時間条件語1404には、時間条件語“After”が記録されている。ポイント先節ID1403には、節の実体が“AAA.2”的節の実体に遷移することを示している。

【0042】

行1406の節に着目すると、節ID1401は“BBB.2”を記録する。節の実体1402には「After the CEC request complete successfully」が記録されている。時間条件語1404には、時間条件語“after”が記録されている。ポイント先節ID1403には、節の実体が“BBB.1”的節の実体に遷移することを示している。時間条

50

件語“*a f t e r*”は“*B B B . 1*”の節の実体への遷移に影響する。

【0043】

図15は本願発明における実施に際し、表示される実行結果のイメージ図である。出力イメージ1500は、式1の状態遷移を視覚化したものである。オブジェクト1501は式1の節802の実体である。オブジェクト1502は、式1の節803の実体である。方向1503はオブジェクト1501とオブジェクト1502との遷移の関係を示す。

【0044】

このように、本発明の第一の実施の形態について、式1の文801の遷移関係を表示するまでを説明したが、この開示の一部をなす論述及び図面は本発明を限定するものではない。

10

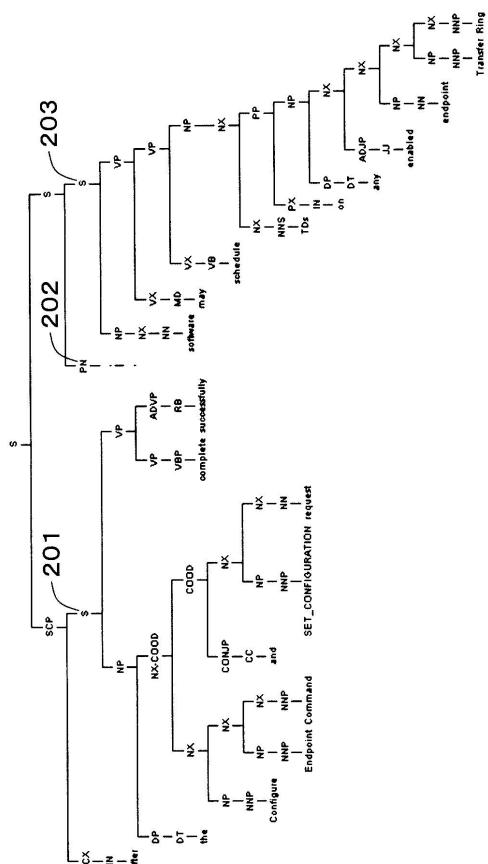
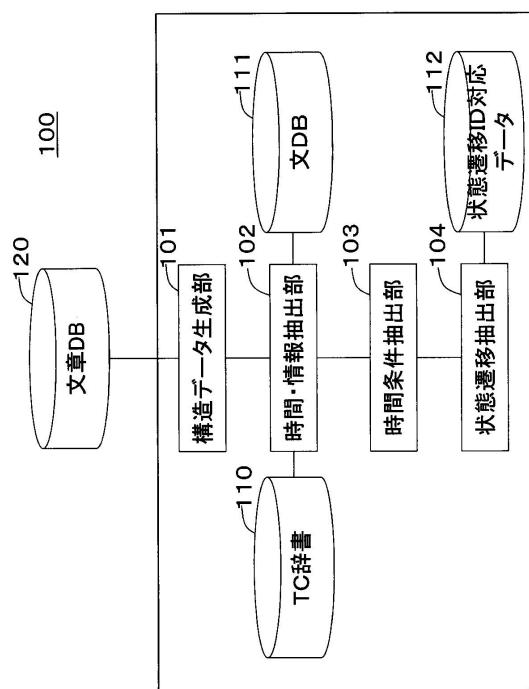
【符号の説明】

【0045】

- 100 設計支援装置
- 101 構造データ生成部
- 102 時間・情報抽出部
- 103 時間条件抽出部
- 104 状態遷移抽出部
- 110 TC辞書
- 111 文DB
- 112 状態遷移ID対応データ
- 120 文章DB
- 1200 TC辞書
- 1300 文DB構成
- 1400 状態遷移データ例
- 1500 実行結果イメージ

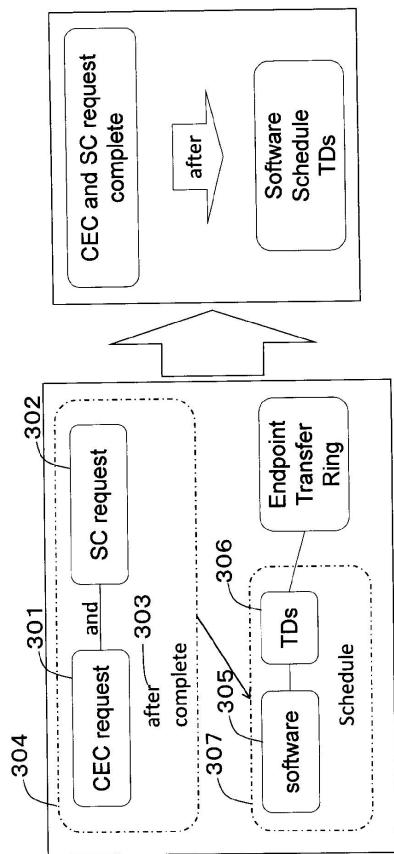
【図1】

【図2】

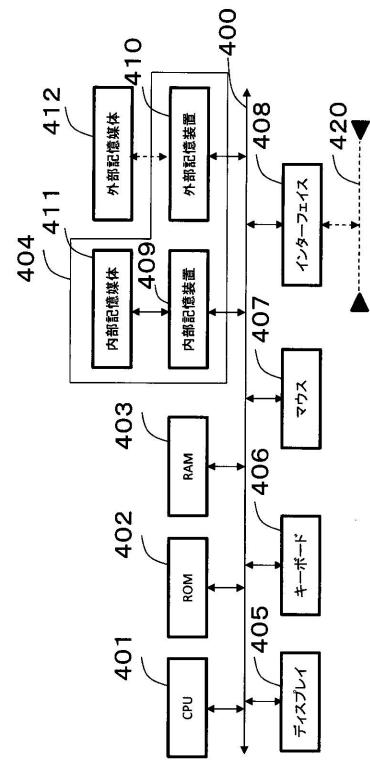


20

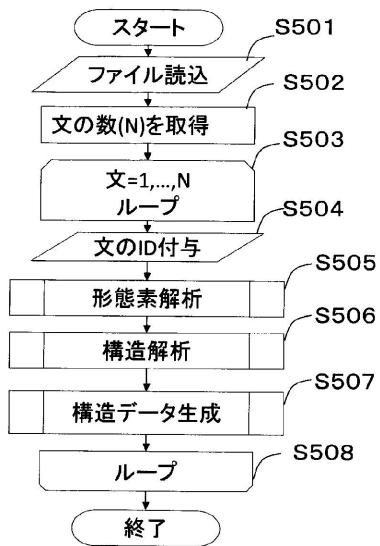
【図3】



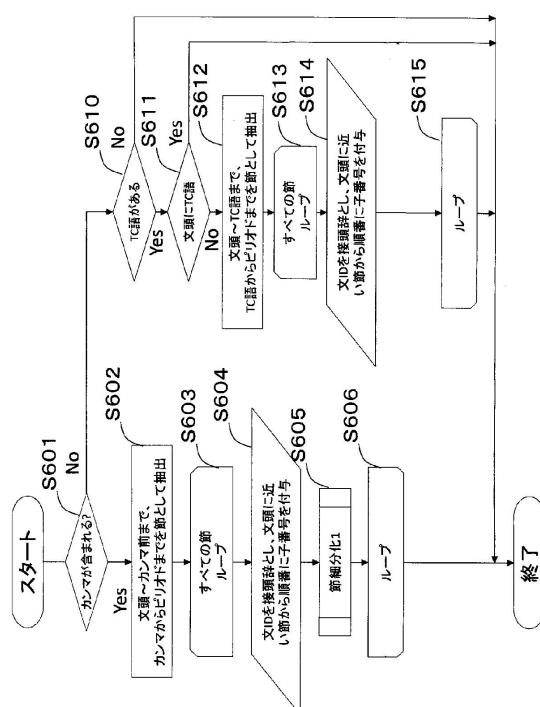
【図4】



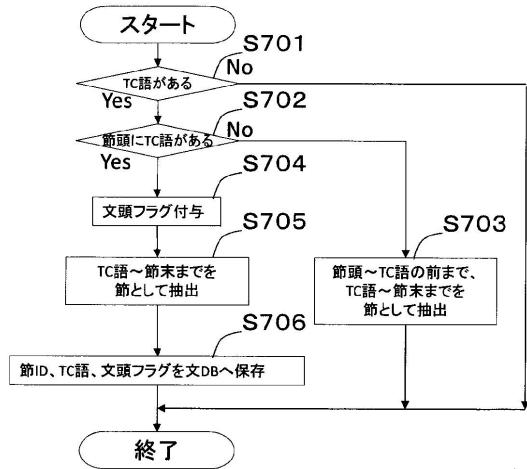
【図5】



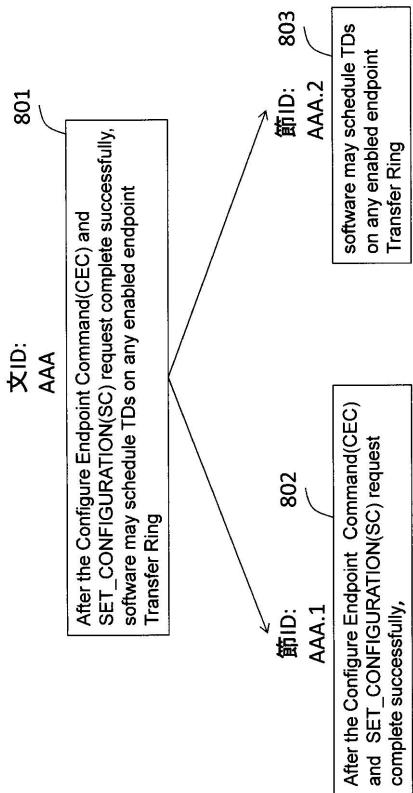
【図6】



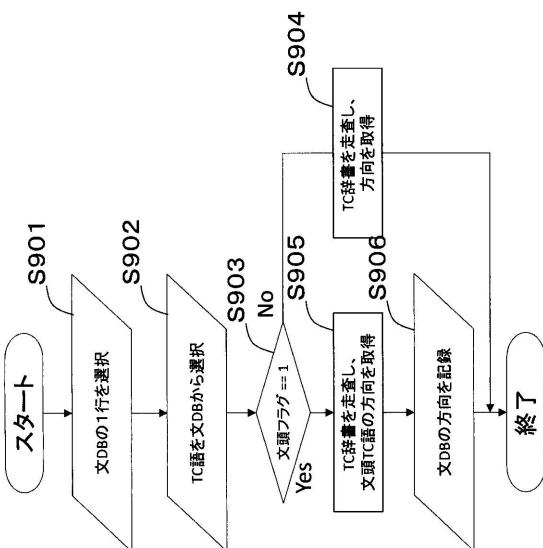
【図7】



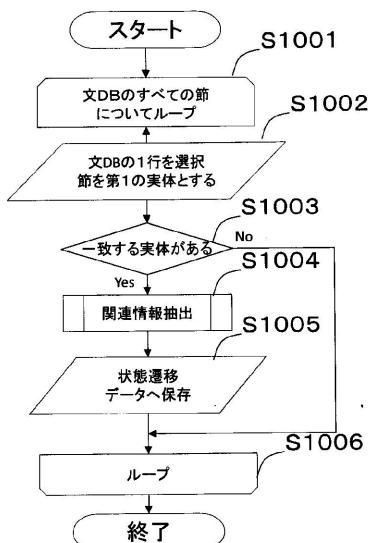
【図8】



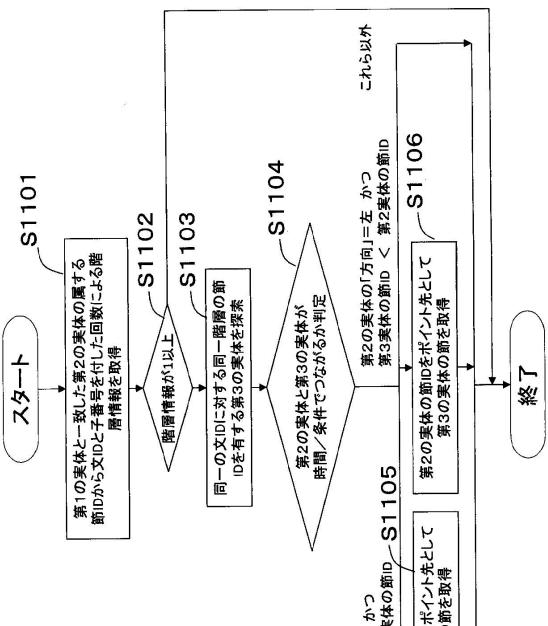
【図9】



【図10】



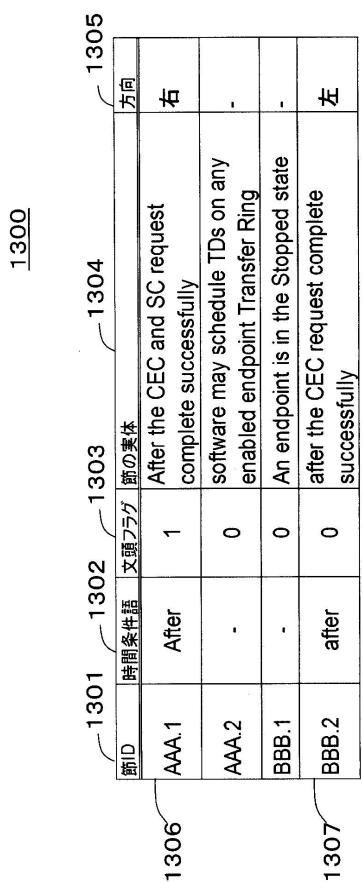
【図 1 1】



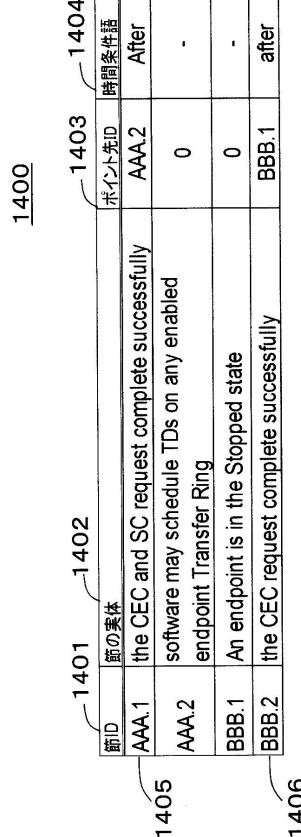
【図 1 2】



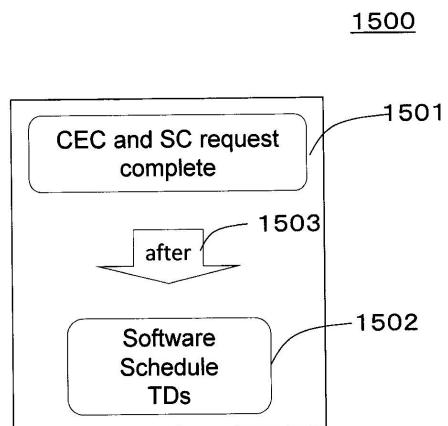
【図 1 3】



【図 1 4】



【図 15】



フロントページの続き

(56)参考文献 特開平09-006781(JP,A)
特開平04-264972(JP,A)
特開2007-257291(JP,A)
特開2009-289020(JP,A)
米国特許出願公開第2006/0161414(US,A1)
田中克己,自然言語による仕様記述文解析のための知識の取り扱いについて,情報処理学会研究
報告,社団法人情報処理学会,1992年9月10日,Vol.92 No.70,pp.8
1-89

(58)調査した分野(Int.Cl., DB名)

G 06 F 17 / 50
G 06 F 17 / 27 - 17 / 28
G 06 F 9 / 44 - 9 / 45
G 06 F 11 / 22