

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2014年5月1日(01.05.2014)



(10) 国際公開番号

WO 2014/064803 A1

- (51) 国際特許分類:
G06Q 50/04 (2012.01) G06F 17/50 (2006.01)
G06F 17/30 (2006.01)
- (21) 国際出願番号: PCT/JP2012/077614
- (22) 国際出願日: 2012年10月25日(25.10.2012)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人: 株式会社日立製作所 (HITACHI, LTD.)
[JP/JP]; 〒1008280 東京都千代田区丸の内一丁目
6番6号 Tokyo (JP).
- (72) 発明者: 藤尾 正和 (FUJIO Masakazu); 〒1858601
東京都国分寺市東恋ヶ窪一丁目280番地 株
式会社日立製作所 中央研究所内 Tokyo (JP). 永
崎 健 (NAGASAKI Takeshi); 〒1858601 東京都国
分寺市東恋ヶ窪一丁目280番地 株式会社日
立製作所 中央研究所内 Tokyo (JP). 平山 淳一
(HIRAYAMA Junichi); 〒1858601 東京都国分寺市
東恋ヶ窪一丁目280番地 株式会社日立製作
所 中央研究所内 Tokyo (JP). 多田 彰 (TADA

Akira); 〒2440817 神奈川県横浜市戸塚区吉田町
292番地 株式会社日立製作所 横浜研究所
内 Kanagawa (JP). 今沢 慶 (IMAZAWA Kei); 〒
2440817 神奈川県横浜市戸塚区吉田町292番
地 株式会社日立製作所 横浜研究所内
Kanagawa (JP).

(74) 代理人: 平木 祐輔, 外 (HIRAKI Yusuke et al.); 〒
1056232 東京都港区愛宕2丁目5番1号 愛宕
グリーンヒルズMORIタワー32階 Tokyo
(JP).

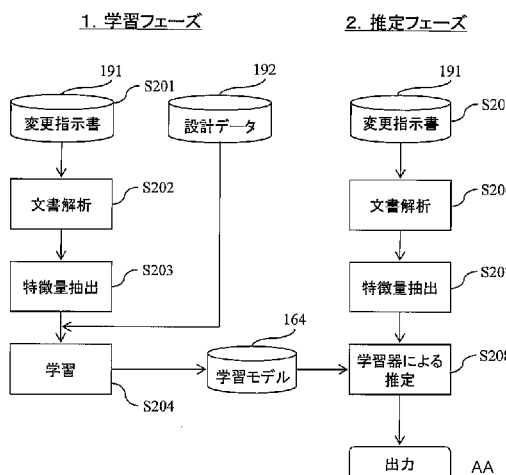
(81) 指定国 (表示のない限り、全ての種類の国内保
護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA,
BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN,
CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES,
FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN,
IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR,
LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX,
MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH,
PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL,
SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG,
US, UZ, VC, VN, ZA, ZM, ZW.

[続葉有]

(54) Title: DOCUMENT PROCESSING PROGRAM, DOCUMENT PROCESSING DEVICE, DOCUMENT PROCESSING SYSTEM, AND DOCUMENT PROCESSING METHOD

(54) 発明の名称: 文書処理プログラム、文書処理装置、文書処理システム、文書処理方法

図2



(57) Abstract: The purpose of the present invention is to estimate the scope of impact of design changes by using change content in which change instructions instructing changes in data such as design data are noted. This document processing program subjects the following to machine learning: the correspondence relationship between change content instructed by document data; and the changed portions of data changed by the change content. On the basis of those results, the document processing program estimates portions for which data will be changed according to new change instructions (see figure 2).

(57) 要約: 本発明は、設計データ等のデータに対する変更を指示する変更指示書が記述している変更内容を用いて、設計変更の影響範囲を推定することを目的とする。本発明に係る文書処理プログラムは、文書データが指示する変更内容と、その変更内容によって変更されたデータの変更部分との間の対応関係を機械学習しておき、その結果に基づき、新たな変更指示によってデータが変更される部分を推定する(図2参照)。

- 1 Learning phase
2 Estimation phase
164 Learning model
192 Design data
S201, S205 Change instructions
S202, S206 Document analysis
S203, S207 Characteristic amount extraction
S204 Learning
S208 Estimation by learning device
AA Output



(84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR),

OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:

— 国際調査報告 (条約第 21 条(3))

明 細 書

発明の名称：

文書処理プログラム、文書処理装置、文書処理システム、文書処理方法
技術分野

[0001] 本発明は、文書を解析する技術に関する。

背景技術

[0002] 製品を設計する際には、新たに設計する新規設計と、既存の設計に基づいて部分的な変更を加える設計変更とがある。例えば、設備プラントに据え付ける製品を考えた場合、設備プラントを現況調査した結果、部品間の干渉等によって当初想定していた据え付け場所に当該製品を置くことができないような場合など、製品の設計変更を余儀なくされる場合がしばしばある。

[0003] 製品を設計変更する場合、機能的に関連する他の部品を付加的に設計変更する必要が生じる場合がある。またその関連部品をどのように設計変更するかなどの派生的な影響も考慮しなければならない。このように、製品の設計変更においては直接・間接の様々な影響を考慮する必要がある。そのため、製品のある部品に対する設計変更による影響範囲（影響が及ぶ他部品の仕様、図面、工程等）やその度合（工数増減、コスト増減等）を迅速に把握することは、設計変更の実施要否を判断する際にきわめて重要である。

[0004] 下記特許文献1は、設計変更の影響を予測する技術として、『変更に関する過去の履歴情報である変更実績データベースを有する変更影響予測装置（請求項1）』を記載している。同文献は、『元の変更内容を考慮することによって影響範囲を精度良く絞り込むことができるパターンを生成する（要約）』ことを目的としている。

先行技術文献

特許文献

[0005] 特許文献1：特開2012-14308号公報

発明の概要

発明が解決しようとする課題

[0006] 実際に設計変更を実施する際には、変更内容を指示する文書データ（変更指示書）によって変更指示がなされる。上記特許文献１に記載されている技術においては、変更実績データベースを利用するに際して、同データベースが格納しているデータ項目などをキーにしてデータベースを検索することが前提になっていると考えられる。したがって、変更指示書から検索キーとするデータ項目などをあらかじめ手作業によって抽出しておく必要があり、相応の作業負担や作業ミスが生じる可能性がある。

[0007] 本発明は、上記のような課題に鑑みてなされたものであり、設計データ等のデータに対する変更を指示する変更指示書が記述している変更内容を用いて、設計変更の影響範囲を推定することを目的とする。

課題を解決するための手段

[0008] 本発明に係る文書処理プログラムは、文書データが指示する変更内容と、その変更内容によって変更されたデータの変更部分との間の対応関係を機械学習しておき、その結果に基づき、新たな変更指示によってデータが変更される部分を推定する。

発明の効果

[0009] 本発明に係る文書処理プログラムによれば、変更指示書が記述している変更内容を用いて変更による影響を推定することができるので、手作業による影響予測に起因する作業負担やミスを軽減し、影響範囲を迅速に予測することができる。

[0010] 上記した以外の課題、構成、および効果は、以下の実施形態の説明により明らかになるであろう。

図面の簡単な説明

[0011] [図１]実施形態１に係る文書処理装置１０の構成を示すブロック図である。

[図２]文書処理装置１０の全体動作を示す処理フロー図である。

[図３]ステップＳ２０２の詳細を説明する処理フロー図である。

[図4]変更指示書 191 のテキスト例とステップ S 2022 の処理イメージを示す図である。

[図5]ステップ S 2022 の処理例を示すフローチャートである。

[図6]ステップ S 2023 の詳細を説明するフローチャートである。

[図7]特殊な書式で記載された変更指示書 191 のテキスト例、およびステップ S 2024 の処理イメージを示す図である。

[図8]ステップ S 2025～S 2027 を省略したと仮定した場合におけるステップ S 203 の処理イメージを示す図である。

[図9]ステップ S 2026 の処理イメージを示す図である。

[図10]ステップ S 2027 の処理イメージを示す図である。

[図11]ステップ S 2026～S 2027 の処理例を示すフローチャートである。

[図12]学習プログラム 154 が使用する類似度ベクトル 1541 の別構成例を示す図である。

[図13]学習プログラム 154 が使用する類似度ベクトル 1541 の別構成例を示す図である。

[図14]学習プログラム 154 が使用する類似度ベクトル 1541 の別構成例を示す図である。

[図15]実施形態 2 に係る文書処理装置 10 の回答書学習フェーズにおける動作を示す処理フロー図である。

[図16]表形式で記述された変更指示書 191 のレイアウト例を示す図である。

[図17]実施形態 3 におけるステップ S 202 の詳細を説明する処理フロー図である。

[図18]実施形態 4 に係る文書処理システム 1000 の構成図である。

[図19]文書処理システム 1000 の全体動作を示す処理フロー図である。

[図20]ステップ S 1901 の処理イメージを示す図である。

[図21]ステップ S 1902 の処理イメージを示す図である。

発明を実施するための形態

[0012] <実施の形態 1：装置構成>

図 1 は、本発明の実施形態 1 に係る文書処理装置 10 の構成を示すブロック図である。文書処理装置 10 は、CAD (Computer Aided Design) データなどのデータ（以下では CAD データとする）に対する変更内容を指示する変更指示書を解析し、その変更内容が影響を及ぼす範囲を推定する装置である。文書処理装置 10 は、一般的なコンピュータを用いて構成することができ、例えば入力装置 11、表示装置 12、CPU (Central Processing Unit) 13、印刷装置 14、メモリ 15、記憶部 16 を備える。

[0013] 入力装置 11 は、ユーザからの操作指示等の入力を受け付ける装置であり、例えばキーボード、マウス、タッチパネルを用いて構成することができる。表示装置 12 は、ユーザに種々の情報を提示する装置であり、例えば液晶ディスプレイのような画面表示装置を用いて構成することができる。印刷装置 14 は、ユーザに提供する種々の情報を必要に応じて印刷する。

[0014] CPU 13 は、メモリ 15 に格納されているプログラムを実行することによって種々の機能を実現する演算装置である。以下では記載の便宜上、各プログラムを動作主体として説明する場合があるが、実際にこれらプログラムを実行するのは CPU 13 であることを付言しておく。

[0015] メモリ 15 は、CPU 13 が実行するプログラムを格納する記憶装置であり、OS (Operating System) 151、通信プログラム 152、文書解析プログラム 153、学習プログラム 154、変更対象推定プログラム 155 を格納する。各プログラムの詳細については後述する。これら各プログラムは、本実施形態 1 における「文書処理プログラム」に相当する。同等の機能を、回路デバイスなどのハードウェアを用いて実現することもできる。メモリ 15 はさらに他のプログラムを格納してもよいし、それらのプログラムを CPU 13 が実行するときに参照されるデータを格納してもよいし、CPU 13 が実行した処理の結果を格納してもよい。

[0016] 記憶部 16 は、CPU 13 が各プログラムの記述にしたがって種々の処理を実行する際に参照する情報を格納する記憶装置であり、変化単語辞書 161、一般用語辞書 162、テンプレート辞書 163、学習モデル 164 を格納する。記憶部 16 はさらに他の情報を格納してもよい。

[0017] 典型的には、メモリ 15 は DRAM (Dynamic Random Access Memory) のような高速かつ揮発性の記憶装置であり、記憶部 16 は HDD (Hard Disk Drive) またはフラッシュメモリのよう大容量かつ不揮発性の記憶装置であるが、その他の種類の記憶装置でもよい。各プログラムは、あらかじめ記憶部 16 に格納しておき、CPU 13 が実行するときにメモリ 15 へコピーするようにしてもよい。記憶部 16 が格納しているデータの少なくとも一部を、必要に応じて一時的にメモリ 15 へコピーするようにしてもよい。

[0018] 文書処理装置 10 は、通信ネットワーク 18 を介してファイルサーバ 19 と接続することができる。ファイルサーバ 19 は、通信ネットワーク 18 に接続された計算機であり、1 台以上存在する。

[0019] 図 1 には文書処理装置 10 が 1 つのコンピュータによって実現される例を示したが、同様の機能を複数の計算機によって実現することもできる。例えば、記憶部 161 が格納している各辞書をいずれかのファイルサーバ 19 に格納し、通信プログラム 152 と通信ネットワーク 18 を介して送受信するようにしてもよい。あるいは、変更対象となる CAD データや変更指示書をいずれかのファイルサーバ 19 に格納し、通信プログラム 152 と通信ネットワーク 18 を介して送受信するようにしてもよい。以下では説明の簡易のため各文書データはファイルサーバ 19 上に格納されているものとする。

[0020] <実施の形態 1：動作手順>

図 2 は、文書処理装置 10 の全体動作を示す処理フロー図である。文書処理装置 10 の動作は、学習フェーズと推定フェーズに分かれる。各フェーズの概略動作について以下に説明する。

[0021] (図 2：学習フェーズ：ステップ S201)

文書処理装置 10 は、はじめに学習フェーズを実施する。学習フェーズにおいて、文書処理装置 10 は、過去に CAD データに対する変更内容を指示した変更指示書 191 を例えば通信プログラム 152 によってファイルサーバ 19 から取得する。通信プログラム 152 は、本実施形態 1 における「文書取得プログラム」「文書取得部」に相当する。

[0022] (図 2 : 学習フェーズ : ステップ S 202)

文書解析プログラム 153 は、ステップ S 201 で取得した変更指示書 191 を解析して単語や文節の係り受け関係を抽出する。本ステップの詳細については後述の図 3 で改めて説明する。

[0023] (図 2 : 学習フェーズ : ステップ S 203)

学習プログラム 154 は、ステップ S 202 における文書解析の結果に基づき、ステップ S 201 で取得した変更指示書 191 の特徴量を算出する。本ステップにおいて算出する特徴量の例については、後述の図 8 ~ 図 10、図 12 ~ 図 14 で改めて説明する。

[0024] (図 2 : 学習フェーズ : ステップ S 204)

文書処理装置 10 は、設計データ 192 (例えば CAD データ) を例えば通信プログラム 152 によってファイルサーバ 19 から取得する。学習プログラム 154 は、ステップ S 203 で抽出した変更指示書 191 の特徴量を用いて、変更指示書 191 が指示する変更内容とその変更内容によって変更された設計データ 192 の変更部分との間の対応関係を学習する。学習結果は学習モデル 164 に格納される。

[0025] (図 2 : 推定フェーズ : ステップ S 205 ~ S 207)

文書処理装置 10 は、未学習の新たな変更内容を記述した変更指示書 191 について、ステップ S 201 ~ S 203 と同様の処理を実施する。

[0026] (図 2 : 推定フェーズ : ステップ S 208)

変更対象推定プログラム 155 は、学習モデル 164 として蓄積されている学習結果を用いて、新たな変更指示書 191 が指示する変更内容によって設計データ 192 が影響を受ける範囲を推定する。例えば、新たな変更指示

書 1 9 1 の特徴量ベクトルと最も距離が近い特徴量ベクトルを特定し、その特徴量ベクトルに関連する過去の変更指示書 1 9 1 が変更指示した設計パラメータが新たな変更指示書 1 9 1 によって変更されると推定することができる。推定結果は、例えば表示装置 1 2 や印刷装置 1 4 などの出力部を介して出力される。

[0027] <実施の形態 1 : 動作手順の詳細>

図 3 は、ステップ S 2 0 2 の詳細を説明する処理フロー図である。以下、図 3 の各ステップについて説明する。

[0028] (図 3 : ステップ S 2 0 2 1 ~ S 2 0 2 2)

文書解析プログラム 1 5 3 は、変更指示書 1 9 1 からテキスト (文字列) 部分を抽出する (S 2 0 2 1)。文書解析プログラム 1 5 3 は、後述の図 4 で説明する手順にしたがって、そのテキストを変更内容に関する部分とそれ以外の部分 (例えば経過を説明する前段部分) に振り分ける (S 2 0 2 2)。

[0029] (図 3 : ステップ S 2 0 2 3)

文書解析プログラム 1 5 3 は、ステップ S 2 0 2 2 において抽出した、変更内容を指示するテキストに対して、後述の図 6 で説明する言語解析を実施する。

[0030] (図 3 : ステップ S 2 0 2 4)

文書解析プログラム 1 5 3 は、ステップ S 2 0 2 2 において抽出した、変更内容を指示するテキストから、テンプレート辞書 1 6 3 を用いることにより単語を抽出する。本ステップは、ステップ S 2 0 2 3 の言語解析によって抽出することが難しい書式で記述されたテキストから単語を抽出するためのものである。本ステップの詳細は後述の図 7 で改めて説明する。

[0031] (図 3 : ステップ S 2 0 2 5)

文書解析プログラム 1 5 3 は、変化単語辞書 1 6 1 を用いて、設計データ 1 9 2 に変更箇所を生じさせる指示を記載していると想定される単語を、キーワードとして抽出する。例えば、「変更する」「訂正する」といった単語

は、設計データ 192 に対する設計変更を指示している可能性が高いと考えられるので、本ステップにおいてこれら語句をキーワードとして抽出する。本ステップは、各単語の特徴量を定める際の準備としての意義もあるし、次のステップ S 2026 における処理開始箇所を定める意義もある。

[0032] (図 3 : ステップ S 2026)

文書解析プログラム 153 は、ステップ S 2025 で抽出したキーワードを起点として所定範囲内でテキストを前方または後方に探索し、キーワードの対象になっていると想定される語句を抽出する。本ステップの詳細は、後述の図 9 で改めて説明する。

[0033] (図 3 : ステップ S 2027)

文書解析プログラム 153 は、一般用語辞書 162 を用いて、設計データ 192 に対する変更箇所に影響しないと想定される単語を、ステップ S 2022 において抽出したテキストから除外する。本ステップの詳細は後述の図 10 で改めて説明する。

[0034] (図 3 : ステップ S 2025 ~ S 2027 : 補足)

ステップ S 2025 ~ S 2027 は、ステップ S 2022 において抽出したテキストと設計データ 192 の変更箇所との間の相関度を高めるために実施する処理であるため、省略することもできる。これらステップを省略する場合の例については、後述の図 8 で改めて説明する。文書解析プログラム 153 は、以上のステップによって抽出されたテキストを、次のステップ S 203 に対して出力する。

[0035] 図 4 は、変更指示書 191 のテキスト例とステップ S 2022 の処理イメージを示す図である。変更指示書 191 は、必ずしも設計データ 192 に対する変更内容のみを記載しているわけではなく、その他のテキストが含まれるのがむしろ通常である。ステップ S 2022 において文書解析プログラム 153 は、変更指示書 191 を変更内容部分 1912 とその他部分 1911 に振り分ける。振り分けの基準は、例えば以下のような手法が考えられる。

[0036] (図 4 : 判定手法の例その 1)

変更内容またはそれ以外の部分を記述する文章において典型的に含まれる語句が、今回ステップS 2 0 2 2において処理する文章内に存在するか否かに基づき判定する。例えば「～ので」「～に伴い」などの語句は、日本語においては因果関係を示唆する語句であるため、その後方には設計データ 1 9 2の変更箇所に関連する語句が存在していると想定される。したがってこれらの語句を境界にして、以後のテキストは変更内容を記述していると推定することができる。同様の手法は後述の図 9においても用いている。

[0037] (図 4 : 判定手法の例その 2)

過去の変更指示書 1 9 1において変更内容またはそれ以外の部分を記述した文章と、今回ステップS 2 0 2 2において処理する文章との間で、文章パターンマッチングなどの類否判定を実施し、その結果得られた類似度に基づきテキストの内容を判定する。

[0038] 図 5 は、ステップS 2 0 2 2の処理例を示すフローチャートである。ここでは図 4 で説明した例その 1 を前提とする。ステップS 2 0 2 2 1～S 2 0 2 2 3は後述の図 6 におけるステップS 2 0 2 3 1～S 2 0 2 3 3と同様である。ステップS 2 0 2 2 4において、文書解析プログラム 1 5 3は、変更内容またはそれ以外の部分を記述する文章において典型的に含まれる語句を登録したキーワード辞書を参照し、ステップS 2 0 2 2 1～S 2 0 2 2 3で抽出した語句がそれらに該当するか否かを判定する。ステップS 2 0 2 2 5において、文書解析プログラム 1 5 3は、ステップS 2 0 2 2 4の判定結果にしたがってテキストを振り分ける。

[0039] 図 6 は、ステップS 2 0 2 3の詳細を説明するフローチャートである。ステップS 2 0 2 3 1において、文書解析プログラム 1 5 3は、形態素辞書を用いて、入力されたテキストの形態素解析を実施する。ステップS 2 0 2 3 2において、文書解析プログラム 1 5 3は、係り受け辞書を用いて、入力されたテキストの係り受けを実施する。ステップS 2 0 2 3 3において、文書解析プログラム 1 5 3は、ストップワード辞書を用いて、入力されたテキストのストップワードを判定する。

- [0040] 以上の処理により、文書解析プログラム 153 は、入力されたテキストから、単語、文節、およびそれらの係り受け関係を抽出し、その結果を出力する。各辞書は、例えば記憶部 16 にあらかじめ格納しておくことができる。
- [0041] 図 7 は、特殊な書式で記載された変更指示書 191 のテキスト例、およびステップ S2024 の処理イメージを示す図である。図 7 に示す変更指示書 191 は、括弧「()」、コロン「:」、矢印「→」を用いて記述されている。
- [0042] これら特殊記号を用いて記述されたテキスト内の語句は、言語解析によって抽出することが難しい一方で、経験則上、変更内容に係る事項を記述している場合が多いと考えられる。そこで文書解析プログラム 153 は、これら特殊書式をあらかじめ定義したテンプレート辞書 163 を用いてその特殊書式を認識し、そのなかに含まれる単語を抽出することとした。
- [0043] 図 7 に示すテンプレート辞書 163 は、2つのテンプレートを例示している。テンプレート 1 は、テキストをコロン直前の単語とその後のテキストに分け、さらに後者は矢印の前後に記述された単語に分けるべき旨を定義している。テンプレート 2 は、テキストを開始括弧直前の単語と括弧内のテキストに分け、さらに後者は矢印の前後に記述された単語に分けるべき旨を定義している。
- [0044] 例えばテンプレート 1 については、コロン直前の単語を抽出すべき旨を定義し、テンプレート 2 については、開始括弧直前の単語と括弧内の矢印直前の単語を抽出すべき旨を定義しているものと仮定する。この場合、図 7 に示す変更指示書 191 から、同図の符号 191' に示すように各単語を抽出することができる。
- [0045] なお、図 7 に示したテンプレート辞書 163 は例示であり、文字や記号などのテキストパターンによって定義することができるものであれば任意のテンプレートを定義することができる。
- [0046] 図 8 は、ステップ S2025～S2027 を省略したと仮定した場合におけるステップ S203 の処理イメージを示す図である。変更指示書 191 は

、ステップS 2 0 2の結果、図8の符合1 9 1' 'に示すように、各単語に分割される。学習プログラム1 5 4は、各単語に適当な特徴量を付与し、各単語の個数を次元数とする特徴量ベクトル1 5 4 1として出力する。例えば図8に例示する文章は、2 1個の単語からなるので2 1次元の特徴量ベクトルとしてその特徴を表現することができる。

[0047] 特徴量ベクトル1 5 4 1における各単語の特徴量の具体的な数値は、例えばその単語によって設計データ1 9 2が変更される度合いを、その単語と当該変更箇所の相関度などによって数値化したものを用いることができる。この相関度は、経験則に基づいて定めてもよいし、過去実績の統計などに基づいて定めてもよい。

[0048] 図9は、ステップS 2 0 2 6の処理イメージを示す図である。変化単語辞書1 6 1に登録されているキーワードの前後に存在する語句は、設計データ1 9 2の変更箇所との間の相関が高いと想定される。そこで文書解析プログラム1 5 3は、変更指示書1 9 1を各単語に分割した後、キーワードを起点として前方または後方に単語を検索する。学習プログラム1 5 4は、その範囲内にある単語にはその他の単語よりも大きな特徴量を付与する。これにより、設計データ1 9 2の変更箇所と変更指示書1 9 1内の各単語との間の相関度を、特徴量ベクトル1 5 4 1としてより正確に表現することができる。

[0049] 図9に示す変化単語辞書1 6 1は、2種類のキーワードを例示している。キーワード1は、日本語において因果関係を示唆する語句であるため、その後方には設計データ1 9 2の変更箇所に関連する語句が存在していると想定される。キーワード2は、日本語において変更を示唆する語句であるため、その前方には変更箇所に関連する語句が存在していると想定される。このように、キーワードの前方と後方のいずれを探索するかは、キーワードの種類によって異なるので、キーワードとセットにして定義しておけばよい。

[0050] 図1 0は、ステップS 2 0 2 7の処理イメージを示す図である。文章を構成するために必要な一般用語（例えば助詞、当該分野における慣用語など）は、設計データ1 9 2の変更箇所との間の相関が低いと想定される。かかる

一般用語を用いて特徴量ベクトル 1 5 4 1 を構成しても、その特徴量ベクトル 1 5 4 1 は変更指示書 1 9 1 と設計データ 1 9 2 の変更箇所との間の対応関係を十分に表現できないと考えられる。

[0051] そこで文書解析プログラム 1 5 3 は、一般用語辞書 1 6 2 に登録されている一般用語を、変更指示書 1 9 1 からあらかじめ削除する。学習プログラム 1 5 4 は、一般用語を削除した後のテキストを用いて特徴量ベクトル 1 5 4 1 を構成する。これにより、特徴量ベクトル 1 5 4 1 と変更箇所との間の相関度を高め、両者の対応関係をより適切に表現することができる。

[0052] 図 1 1 は、ステップ S 2 0 2 6 ～ S 2 0 2 7 の処理例を示すフローチャートである。ステップ S 2 0 2 6 1 ～ S 2 0 2 6 3 はステップ S 2 0 2 3 1 ～ S 2 0 2 3 3 と同様であるため、これらステップの結果を流用して省略することもできる。ステップ S 2 0 2 6 4 ～ S 2 0 2 6 5 において、文書解析プログラム 1 5 3 は、図 9 ～ 図 1 0 で説明した処理を実施する。

[0053] 図 1 2 は、学習プログラム 1 5 4 が使用する類似度ベクトル 1 5 4 1 の別構成例を示す図である。類似度ベクトル 1 5 4 1 は、変更指示書 1 9 1 が指示する設計データ 1 9 2 に対する変更内容と、その変更内容によって変更される設計データ 1 9 2 の変更部分との間の対応関係を学習するために使用するものであるから、同様の機能を有する限り、図 8 ～ 図 1 0 で例示した以外の構成を採用することができる。

[0054] 図 1 2 に示す例においては、変更指示書 1 9 1 が変更指示する設計パラメータ（例えば C A D パラメータ）と、過去に各設計パラメータを変更するよう指示した変更指示書内の文章との間の類似度を、過去の各文章について求め、これを類似度ベクトル 1 5 4 1 のベクトル要素とした。

[0055] フィールド 1 5 4 1 1 は、変更指示書 1 9 1 内において設計データ 1 9 2 を変更するよう指示している文章を保持する。フィールド 1 5 4 1 2 は、フィールド 1 5 4 1 1 の文章が変更するよう指示する設計パラメータを保持する。フィールド 1 5 4 1 3 は、フィールド 1 5 4 1 1 の文章と、過去の変更指示書 1 9 1 内の文章のうち各設計パラメータ（設計パラメータ 1 ～ n を例

示した)を変更するよう指示するものとの間の類似度を保持する。

[0056] 図12に例示する類似度ベクトル1541の構成例によれば、変更すべき設計パラメータを媒介にして他の設計パラメータとの間の相関度を学習することができる。以下に説明する図13における構成例についても同様である。

[0057] 図13は、学習プログラム154が使用する類似度ベクトル1541の別構成例を示す図である。図12に示す例においては、変更指示書191とは関係しない設計パラメータを変更するよう指示する過去の文章についても、設計パラメータ毎に変更指示書191との間の類似度を求めることとした。図13に示す例においては、変更指示書191とは関係しない設計パラメータについては、例えば各過去文章と変更指示書191との間の類似度の平均値などを用いて、1つの類似度に集約することとした。

[0058] フィールド15414は、フィールド15411の文章と、過去の変更指示書191内の文章のうち同じ設計パラメータ（すなわちフィールド15412の設計パラメータ）を変更するよう指示するものとの間の類似度を保持する。フィールド15415は、フィールド15411の文章と、過去の変更指示書191内の文章のうちその他の設計パラメータ（すなわちフィールド15412以外の設計パラメータ）を変更するよう指示するものとの間の類似度を保持する。

[0059] 図14は、学習プログラム154が使用する類似度ベクトル1541の別構成例を示す図である。フィールド15416は、フィールド15411の文章と、過去の変更指示書191内の文章のうち同じ設計パラメータ（すなわちフィールド15412の設計パラメータ）を変更するよう指示するものとの間の類似度を保持する。ただし、類似度を算出する手法については複数のそれぞれ異なるものを採用し、各手法を用いて求めた類似度をベクトル要素とした。

[0060] <実施の形態1：まとめ>

以上のように、本実施形態1に係る文書処理装置10は、変更指示書19

1 が指示する変更内容と、その変更内容によって変更される設計データ 1 9 2 の変更箇所との間の対応関係をあらかじめ機械学習しておき、その学習結果に基づき、新たな変更指示書 1 9 1 によって設計データ 1 9 2 が変更される箇所を推定する。これにより、文章として記述された変更指示書 1 9 1 のみを用いて、変更指示による影響範囲を予測することができる。したがって、変更指示書 1 9 1 から設計パラメータを手作業で抽出するなどの作業が不要になり、影響範囲を予測する際の作業負担やコストを軽減することができる。

[0061] <実施の形態 2>

実施形態 1 では、過去の変更指示書 1 9 1 を用いて設計データ 1 9 2 に対する影響範囲を学習する動作例を説明したが、実際の設計作業においては、変更指示書 1 9 1 によって指示された変更内容を実現できないなどの理由により、その変更内容が設計データ 1 9 2 へ反映されない場合がある。設計変更の可否は、一般に変更指示書 1 9 1 に対する回答書として記述される。そこで本発明の実施形態 2 では、回答書の内容を追加的に学習する動作例を説明する。

[0062] 図 1 5 は、本実施形態 2 に係る文書処理装置 1 0 の回答書学習フェーズにおける動作を示す処理フロー図である。本実施形態 2 において、文書処理装置 1 0 は、学習フェーズと推定フェーズの間において、新たに回答書学習フェーズを実施する。回答書学習フェーズは、変更指示書 1 9 1 に対する回答書 1 9 4 が設計データ 1 9 2 に与えた影響を学習するフェーズであり、学習フェーズと類似しているが、設計変更不可の旨を回答書 1 9 4 が回答している場合は学習フェーズと反対方向に学習を進める点が異なる。以下、学習フェーズとの差異点を中心に説明する。

[0063] (図 1 5 : 回答書学習フェーズ : ステップ S 1 5 0 1)

文書処理装置 1 0 は、過去に変更指示書 1 9 1 に対する回答を記述した回答書 1 9 4 を例えば通信プログラム 1 5 2 によってファイルサーバ 1 9 から取得する。

[0064] (図 15 : 回答書学習フェーズ : ステップ S 1502)

文書解析プログラム 153 は、ステップ S 1501 で取得した回答書 194 を解析して単語や文節の係り受け関係を抽出する。本ステップは概ねステップ S 202 と同様であるが、設計データ 192 に対する変更内容を抽出することに代えて、変更指示書 191 に対する回答内容を抽出する点が異なる。したがって、「変更する」などの変更を示唆する単語に代えて、「変更不可」などの変更内容に対する回答を示唆する単語を抽出する必要がある。変化単語辞書 161 は、これらの単語をあらかじめ格納しておくこととする。回答書 194 の回答対象は設計パラメータであると考えられるので、設計パラメータを示唆する単語を抽出する点は実施形態 1 と同様である。

[0065] (図 15 : 回答書学習フェーズ : ステップ S 1503)

学習プログラム 154 は、ステップ S 1502 における文書解析の結果に基づき、ステップ S 1501 で取得した回答書 194 の特徴量を算出する。本ステップの処理は、ステップ S 203 と同様である。

[0066] (図 15 : 回答書学習フェーズ : ステップ S 1504)

学習プログラム 154 は、ステップ S 1503 で抽出した回答書 191 の特徴量を用いて、回答書 191 が指示する回答内容とその回答内容によって変更された設計データ 192 の変更部分との間の対応関係を学習する。

[0067] (図 15 : 回答書学習フェーズ : ステップ S 1504 : 補足)

設計変更を許可する旨の回答である場合には、変更指示書 191 によって変更内容を学習することができるので、回答書 194 によって新たに学習すべき事項はない。設計変更を拒否する旨の回答である場合には、変更指示書 191 が指示した変更内容は設計データ 192 に反映されなかったことになるので、変更指示書 191 による学習結果をキャンセルする必要がある。そこで学習プログラム 154 は、回答書 194 による学習結果を学習モデル 164 に対して負の方向に反映させる。例えば、特徴量ベクトル 1541 のベクトル要素に -1 を乗算して符号を反転させたものを用いて学習を実施すればよい。

[0068] <実施の形態 2 : まとめ>

以上のように、本実施形態 2 に係る文書処理装置 10 は、変更指示書 191 が指示する変更内容を拒否する旨を回答書 194 が回答している場合は、変更指示書 191 内の変更内容と設計データ 192 の変更箇所との間の対応関係をキャンセルする向きに機械学習を実施する。これにより、設計変更の可否を学習結果に反映し、より精度よく影響範囲を推定することができる。

[0069] <実施の形態 3>

実施形態 1 ~ 2 では、変更指示書 191 が記号や文字列によって記述されている例を説明した。しかし実際の変更指示書 191 は、例えば表形式のように様々な書式で記述されている場合がある。そこで本発明の実施形態 3 では、変更指示書 191 を解析する際に、その書式を解析してテキスト部分を抽出する動作例を説明する。その他の構成は実施形態 1 ~ 2 と同様であるため、以下では変更指示書 191 の書式を解析する動作について中心に説明する。

[0070] 図 16 は、表形式で記述された変更指示書 191 のレイアウト例を示す図である。図 16 のうち最下段に記述されているテキストは実施形態 1 の図 4 で説明したものと同一であるが、その前に他のセルが存在するため、これらセルを除去するかまたはテキスト部分の位置を特定する必要がある。

[0071] 図 17 は、本実施形態 3 におけるステップ S202 の詳細を説明する処理フロー図である。以下、図 17 の各ステップについて説明する。

[0072] (図 17 : ステップ S2028)

文書解析プログラム 153 は、ステップ S2021 に代えて以下のステップ S2028 と S2029 を実施する。ステップ S2028 において文書解析プログラム 153 は、変更指示書 191 を座標付テキストに変換する。座標付テキストとは、変更指示書 191 が記述しているテキスト部分について、例えば文書の左上を原点とする座標を内部的に付与した文書データである。座標付テキストは公知技術であるため、その詳細については省略する。

[0073] (図 17 : ステップ S2029)

文書解析プログラム１５３は、ステップＳ２０２８の結果を用いて、変更指示書１９１のレイアウトを解析する。座標付テキストを用いて文書のレイアウトを解析する手法は公知技術であるため、その詳細については省略する。以下のステップは実施形態１のステップＳ２０２２以降と同様である。

[0074] ＜実施の形態４＞

図１８は、本発明の実施形態４に係る文書処理システム１０００の構成図である。文書処理システム１０００は、実施形態１～３で説明した文書処理装置１０に加え、種々の付加機能を備えたシステムである。文書処理システム１０００は、ストレージ１１００、抽出処理装置（以下では、ＥＴＬ：Extract／Transform／Load）１２００、コンテンツサーバ１３００、検索サーバ１４１０、メタデータサーバ１４２０、実施形態１～３で説明した文書処理装置１０、アプリケーションプログラム１５００を有する。

[0075] ストレージ１１００は、種々のデータを格納する記憶装置であり、実施形態１～３におけるファイルサーバ１９の代替として機能する。例えば文書データ１１０１、ＣＡＤデータ１１０２、メールデータ１１０３などを格納する。変更指示書１９１や回答書１９４もストレージ１１００上に格納される。

[0076] ＥＴＬ１２００は、ストレージ１１００が格納している各データから必要な項目を抽出する装置である。ＥＴＬ１２００は、変更箇所抽出プログラム１２０１、対応付けプログラム１２０２、対応付け規則データ１２０３を、ＨＤＤなどの記憶装置に格納する。これらの動作については後述する。

[0077] コンテンツサーバ１３００は、上位のサーバ群に対して各種データを提供するサーバであり、メタデータ１３０１、コンテンツデータ１３０２を、ＨＤＤなどの記憶装置に格納する。メタデータ１３０１は、コンテンツデータ１３０２の属性情報を記述するデータである。

[0078] 検索サーバ１４１０とメタデータサーバ１４２０は、アプリケーションプログラム１５００からの要求を受けてコンテンツデータ１３０２またはメタ

データ 1301 を検索するサーバである。検索サーバ 1401 は、コンテンツデータ 1302 から抽出したインデックス 1411 を用いてコンテンツデータ 1302 を検索する。メタデータサーバ 1420 は、メタデータ 1301 をインデックス化したデータベース 1421 を用いてメタデータ 1301 を検索する。

[0079] アプリケーションプログラム 1500 は、適当なコンピュータによって実行され、その用途に応じて様々な機能を提供する。例えば、上記検索要求の他、文書処理装置 10 に対して変更指示書 191 による影響範囲を推定するよう指示し、その結果を受け取ってユーザに提示することができる。CAD ツールなどの設計アプリケーションや、変更指示書 191 または回答書 194 を送受信するワークフローアプリケーションなどを含めることもできる。

[0080] 文書処理装置 10 とその他のサーバや装置は、一体的に構成することもできる。例えば ETL 1200 が搭載している各プログラムを文書処理装置 10 上に実装し、文書処理装置 10 がこれらプログラムを実行するようにしてもよい。

[0081] <実施の形態 4 : ETL 1200 の機能>

実施形態 1 ~ 3 では、変更指示書 191 と設計データ 192 の変更箇所との間の対応関係を学習することを説明した。しかし、設計データ 192 が例えば CAD データのように特殊な形式で記述されている場合、過去に設計データ 192 が変更された箇所を容易に特定できない可能性がある。そこで本実施形態 4 において、ETL 1200 は、設計データ 192 の変更箇所を抽出し、作業者の負担を軽減することとした。

[0082] また、変更指示書 191 内の変更内容と設計データ 192 の変更箇所との間の対応関係を学習する前提として、変更指示書 191 がどの設計データ 192 に対して変更指示したのかを特定する必要がある。しかし、変更指示書 191 と設計データ 192 がそれぞれ別のシステムやワークフローを介して作成される場合、設計工程が終了した後になって別の作業者が両者の対応関係を把握することが困難になる可能性がある。そこで本実施形態 4 において

、E T L 1 2 0 0 は、前段処理として、変更指示書 1 9 1 と設計データ 1 9 2 をあらかじめ対応付ける処理を支援し、作業者の負担を軽減することとした。

[0083] 図 1 9 は、文書処理システム 1 0 0 0 の全体動作を示す処理フロー図である。文書処理システム 1 0 0 0 は、実施形態 1 ～ 3 で説明した文書処理装置 1 0 の動作に加えて、新たにステップ S 1 9 0 1 ～ S 1 9 0 2 を実施する。以下、これらのステップについて説明する。

[0084] (図 1 9 : 学習フェーズ : ステップ S 1 9 0 1)

E T L 1 2 0 0 は、C P U によって変更箇所抽出プログラム 1 2 0 1 を実行し、過去の変更指示書 1 9 1 によって設計データ 1 9 2 が変更された箇所を抽出する。本ステップの詳細は後述の図 2 0 で改めて説明する。

[0085] (図 1 9 : 学習フェーズ : ステップ S 1 9 0 2)

E T L 1 2 0 0 は、C P U によって対応付けプログラム 1 2 0 2 を実行し、過去の変更指示書 1 9 1 とこれによって変更された設計データ 1 9 2 を対応付ける。本ステップの詳細は後述の図 2 1 で改めて説明する。

[0086] 図 2 0 は、ステップ S 1 9 0 1 の処理イメージを示す図である。ここでは設計データ 1 9 2 が C A D データである場合を例示した。C A D データは図面データであるため、その変更箇所を特定するためには、テキスト形式でダンプしたログデータなどにいったん変換する必要がある。ここではログデータを表計算ソフト上に取り込んだ様子を例示した。

[0087] 図 2 0 において、設計データ 1 9 2 が変更指示書 1 9 1 によって符号 1 9 1 ' のように変更されたものと仮定する。一般的には、変更箇所を特定するためには両者の差分を取ればよいが、C A D データに関しては必ずしも単純に差分を取れば変更箇所を特定できるとは限らない。部品番号 1 9 2 1 が作業工程に応じて（例えば月毎に）変更される可能性があるからである。

[0088] ただし、部品番号 1 9 2 1 が変更になっても、その並び順は同じ部品であれば同一に保たれるのが通常である。変更箇所抽出プログラム 1 2 0 1 は、この特徴を利用して設計データ 1 9 2 の変更箇所を抽出する。具体的には、

変更前後に係る設計データ 191 と 191' を部品番号 1921 でソートし、その他の列（例えば図 20 の A E 列に示す部品名称）の一致度が最も高い箇所が互に対応しているものと推定する。一致度は例えば D P マッチングなどの公知手法によって求めればよい。変更箇所をより確実に抽出するためには、変更箇所抽出プログラム 1201 が抽出した結果をユーザが目視確認するなどして必要に応じ訂正すればよい。

[0089] 図 21 は、ステップ S 1902 の処理イメージを示す図である。一般に変更指示書 191 と設計データ 192 は異なるワークフローやアプリケーションを介して作成されるのが通常であるため、両者の書式は必ずしも合致せず、これらに対応付けるために相応の作業負担が生じる。

[0090] ただし、設計データ 192 に含まれている部品番号などの設計情報は、変更指示書 191 のなかに何らかの形式で記述されているはずである。C A D データは図面データであるため、変更指示書 191 のなかにおいては図面番号として C A D データが指定されている場合が多いと考えられる。そこで対応付けプログラム 1202 は、変更指示書 191 のなかに記載されている図面番号が含まれる設計データ 192 が、当該変更指示書 191 と対応していると仮定し、その旨をユーザに提示する。ユーザはその提示を目視確認して両者をより正確に対応付けることができる。

[0091] 図面番号は、典型的には変更指示書 191 の件名部分や本文に記載されていると考えられるので、対応付けプログラム 1202 はこれら部分と設計データ 192 の対応関係を優先的に検査するようにしてもよい。また、図面番号が同一であっても日時が離れている場合は両者の関係性が低いと考えられるので、両者の日時（文書内に記載されている日時でもよいし、更新日時でもよい）が所定範囲内（例えば 1 週間以内など）である場合に限り、両者に対応付けるようにしてもよい。

[0092] 変更指示書 191 内のいずれの項目を、設計データ 192 のいずれの項目と対応付けるかについての規則は、対応付け規則データ 1203 としてあらかじめ定義しておくことができる。

[0093] <実施の形態４：まとめ>

以上のように、本実施形態４に係る文書処理システム１０００において、ＥＴＬ１２００は、過去の変更指示書１９１によって設計データ１９２が変更された箇所を抽出する。これにより、設計データ１９２が特殊な書式で記述されている場合であっても、変更指示書１９１とこれによる変更箇所との間の対応関係を効率よく学習することができる。

[0094] また、本実施形態２に係る文書処理システム１０００において、ＥＴＬ１２００は、過去の変更指示書１９１とこれによって変更された設計データ１９２を対応付ける。これにより、両者が個別に作成された場合であっても、作業者がこれらに対応付けるための作業負担を軽減することができる。

[0095] 本発明は上記した実施形態に限定されるものではなく、様々な変形例が含まれる。上記実施形態は本発明を分かりやすく説明するために詳細に説明したものであり、必ずしも説明した全ての構成を備えるものに限定されるものではない。また、ある実施形態の構成の一部を他の実施形態の構成に置き換えることもできる。また、ある実施形態の構成に他の実施形態の構成を加えることもできる。また、各実施形態の構成の一部について、他の構成を追加・削除・置換することもできる。

[0096] 例えば、以上の実施形態ではＣＡＤデータなどの設計データ１９２に対して変更を指示する例を説明したが、その他のデータに対して文書をもって変更指示する場合においても本発明を適用することができる。

[0097] 上記各構成、機能、処理部、処理手段等は、それらの一部や全部を、例えば集積回路で設計する等によりハードウェアで実現してもよい。また、上記の各構成、機能等は、プロセッサがそれぞれの機能を実現するプログラムを解釈し、実行することによりソフトウェアで実現してもよい。各機能を実現するプログラム、テーブル、ファイル等の情報は、メモリ、ハードディスク、ＳＳＤ（Solid State Drive）等の記録装置、ＩＣカード、ＳＤカード、ＤＶＤ等の記録媒体に格納することができる。

符号の説明

[0098] 10 : 文書処理装置、11 : 入力装置、12 : 表示装置、13 : CPU、14 : 印刷装置、15 : メモリ、151 : OS、152 : 通信プログラム、153 : 文書解析プログラム、154 : 学習プログラム、155 : 変更対象推定プログラム、16 : 記憶部、161 : 変化単語辞書、162 : 一般用語辞書、163 : テンプレート辞書、164 : 学習モデル、18 : 通信ネットワーク、19 : ファイルサーバ、1000 : 文書処理システム、1100 : ストレージ、1200 : 抽出処理装置、1201 : 変更箇所抽出プログラム、1202 : 対応付けプログラム、1203 : 対応付け規則データ、1300 : コンテンツサーバ、1410 : 検索サーバ、1420 : メタデータサーバ、1500 : アプリケーションプログラム。

請求の範囲

- [請求項1] コンピュータに文書処理させるプログラムであって、前記コンピュータに、
- データに対する変更内容を指示する文書データを取得する文書取得ステップ、
- 前記文書データが指示する前記変更内容と、前記変更内容によって変更された前記データの変更部分との間の対応関係を機械学習する学習ステップ、
- 前記機械学習の結果に基づき、新たな前記文書データが指示する前記データに対する変更内容によって前記データが変更される部分を推定する推定ステップ、
- を実行させることを特徴とする文書処理プログラム。
- [請求項2] 前記文書処理プログラムはさらに、前記コンピュータに、
- 前記文書データが記述しているテキストの単語または文節係受関係のうち少なくともいずれかを抽出する文書解析ステップを実行させ、
- 前記学習ステップでは、前記コンピュータに、
- 前記文書解析ステップにおいて抽出した前記単語または前記文節係受関係のうち少なくともいずれかと、前記文書データによって変更された前記データの変更部分との間の相関度を求めることにより、前記対応関係を機械学習するステップ、
- を実行させることを特徴とする請求項1記載の文書処理プログラム。
- [請求項3] 前記文書処理プログラムはさらに、前記コンピュータに、
- 前記文書データが指示する前記データに対する変更内容に対する回答を記述した回答文書を取得する回答文書取得ステップを実行させ、
- 前記文書解析ステップでは、前記コンピュータに、
- 前記回答文書が記述しているテキストのうち前記変更内容を許可

しない旨の前記回答を記述している部分を抽出させ、

前記学習ステップでは、前記コンピュータに、

前記変更内容を許可しない旨の前記回答を記述しているテキストが前記回答文書内に存在する場合は、前記対応関係を弱める向きに前記機械学習を実施させる

ことを特徴とする請求項2記載の文書処理プログラム。

[請求項4]

前記文書処理プログラムはさらに、前記コンピュータに、

前記データの変更部分を生じさせる単語として定義されたキーワードを登録した単語辞書を参照することにより、前記文書データ内に含まれる前記キーワードを抽出するステップを実行させ、

前記学習ステップでは、前記コンピュータに、

前記文書データから抽出された前記キーワードと、前記文書データによって変更された前記データの変更部分との間の相関度を求めることにより、前記対応関係を機械学習させる

ことを特徴とする請求項2記載の文書処理プログラム。

[請求項5]

前記文書処理プログラムはさらに、前記コンピュータに、

前記データの変更部分を生じさせる単語として定義されたキーワードの周辺に存在するテキストパターンを記述するテンプレートを前記文書データに適用することにより、前記文書データ内に含まれる前記キーワードを抽出するステップを実行させ、

前記学習ステップでは、前記コンピュータに、

前記テンプレートを用いて前記文書データから抽出された前記キーワードと、前記文書データによって変更された前記データの変更部分との間の相関度を求めることにより、前記対応関係を機械学習させる

ことを特徴とする請求項2記載の文書処理プログラム。

[請求項6]

前記文書処理プログラムはさらに、前記コンピュータに、

前記データの変更部分を生じさせる単語として定義されたキーワー

ドを前記文書データから抽出するステップを実行させ、

前記学習ステップでは、前記コンピュータに、

前記文書データ内に含まれるテキストのうち前記キーワードから所定範囲内に存在する単語と、前記文書データによって変更された前記データの変更部分との間の相関度を求めることにより、前記対応関係を機械学習させる

ことを特徴とする請求項2記載の文書処理プログラム。

[請求項7]

前記文書処理プログラムはさらに、前記コンピュータに、

前記データの変更部分を生じさせない単語として定義された一般用語を登録した一般用語辞書を参照することにより、前記文書データ内に含まれる前記一般用語を特定するステップを実行させ、

前記学習ステップでは、前記コンピュータに、

前記文書データが記述しているテキストから前記一般用語を除いたものと、前記文書データによって変更された前記データの変更部分との間の相関度を求めることにより、前記対応関係を機械学習させる

ことを特徴とする請求項2記載の文書処理プログラム。

[請求項8]

前記文書処理プログラムはさらに、前記コンピュータに、

前記文書データのレイアウトを解析し、前記解析によって得られた前記レイアウトのうち前記変更内容を記述した部分を特定するレイアウト解析ステップを実行させ、

前記学習ステップでは、前記コンピュータに、

前記レイアウト解析ステップにおいて特定した前記変更内容と、前記文書データによって変更された前記データの変更部分との間の相関度を求めることにより、前記対応関係を機械学習させる

ことを特徴とする請求項2記載の文書処理プログラム。

[請求項9]

前記文書処理プログラムはさらに、前記コンピュータに、

前記文書データに含まれるテキストのうち前記変更内容を記述した部分とそれ以前の部分との間の境界部分に配置される区切り語句を

抽出するステップを実行させ、

前記学習ステップでは、前記コンピュータに、

前記文書データが記述しているテキストから前記区切り語句以前の部分を除いたものと、前記文書データによって変更された前記データの変更部分との間の相関度を求めることにより、前記対応関係を機械学習させる

ことを特徴とする請求項 2 記載の文書処理プログラム。

[請求項10]

前記学習ステップでは、前記コンピュータに、

前記文書データの前記変更内容が指示する前記変更部分と、同じ前記変更部分を変更するよう指示する変更内容を含む他の文書データとの間の類似度を、複数の前記文書データについてそれぞれ記述することによって形成された特徴ベクトルを用いて、前記機械学習を実施させる

ことを特徴とする請求項 1 記載の文書処理プログラム。

[請求項11]

前記文書取得ステップでは、前記コンピュータに、

C A D データに対する変更内容を指示する前記文書データを取得させ、

前記学習ステップでは、前記コンピュータに、

前記変更内容と、前記変更内容によって変更された前記 C A D データの変更部分との間の対応関係を機械学習させ、

前記推定ステップでは、前記コンピュータに、

前記機械学習の結果に基づき、新たな前記文書データが指示する前記 C A D データに対する変更内容によって前記 C A D データが変更される部分を推定させる

ことを特徴とする請求項 1 記載の文書処理プログラム。

[請求項12]

文書进行处理する装置であって、

データに対する変更内容を指示する文書データを取得する文書取得部と、

前記文書データが指示する前記変更内容と、前記変更内容によって変更された前記データの変更部分との間の対応関係を機械学習する学習部と、

前記機械学習の結果に基づき、新たな前記文書データが指示する前記データに対する変更内容によって前記データが変更される部分を推定する推定部と、

を備えることを特徴とする文書処理装置。

[請求項13]

請求項12記載の文書処理装置と、

前記文書データが指示する前記変更内容によって前記データが変更された箇所を抽出する抽出処理装置と、

を有し、

前記データは、

物品を構成する部品の設計情報を記述したCADデータとして構成されており、

前記抽出処理装置は、

前記CADデータが前記文書データによって変更された前後に係る前記部品の部品番号とその並び順を変更レコードとして取り出し、

前記部品番号が前記変更の前後において一致しない場合には、前記変更レコード内の記録のうち、前記部品の並び順が最も類似するものを特定することにより、前記CADデータが変更された箇所を抽出する

ことを特徴とする文書処理システム。

[請求項14]

前記抽出処理装置は、

前記文書データが記述している件名、図面番号、および前記CADデータの更新日時と前記文書データの更新日時が所定範囲内にあるか否かに基づき、前記文書データと前記CADデータを対応付け、

前記文書処理装置は、

前記抽出処理装置が対応付けた前記文書データと前記CADデー

タを用いて前記対応関係を学習する

ことを特徴とする請求項 13 記載の文書処理システム。

[請求項15]

文書进行处理する方法であって、

データに対する変更内容を指示する文書データを取得する文書取得ステップ、

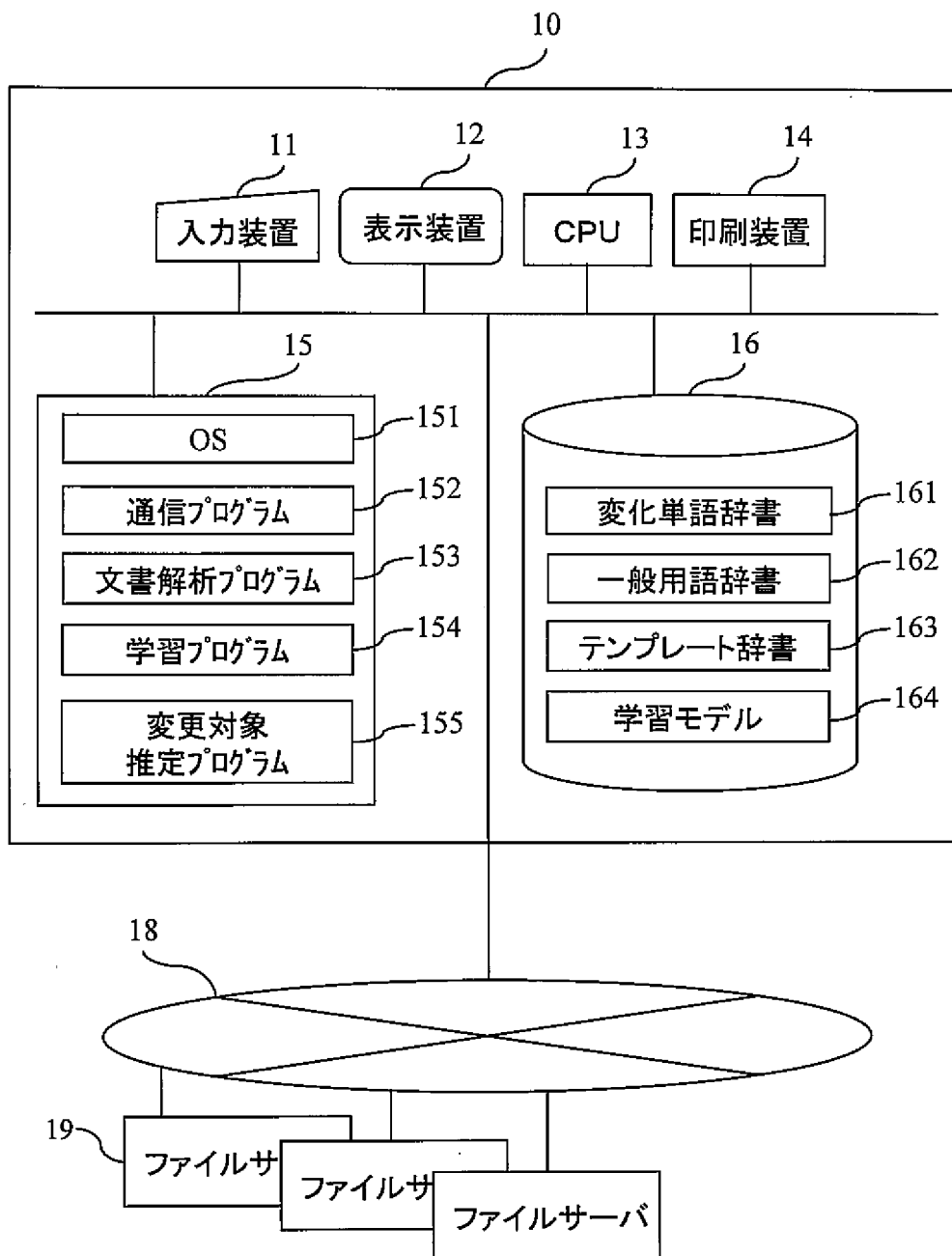
前記文書データが指示する前記変更内容と、前記変更内容によって変更された前記データの変更部分との間の対応関係を機械学習する学習ステップ、

前記機械学習の結果に基づき、新たな前記文書データが指示する前記データに対する変更内容によって前記データが変更される部分を推定する推定ステップ、

を有することを特徴とする文書処理方法。

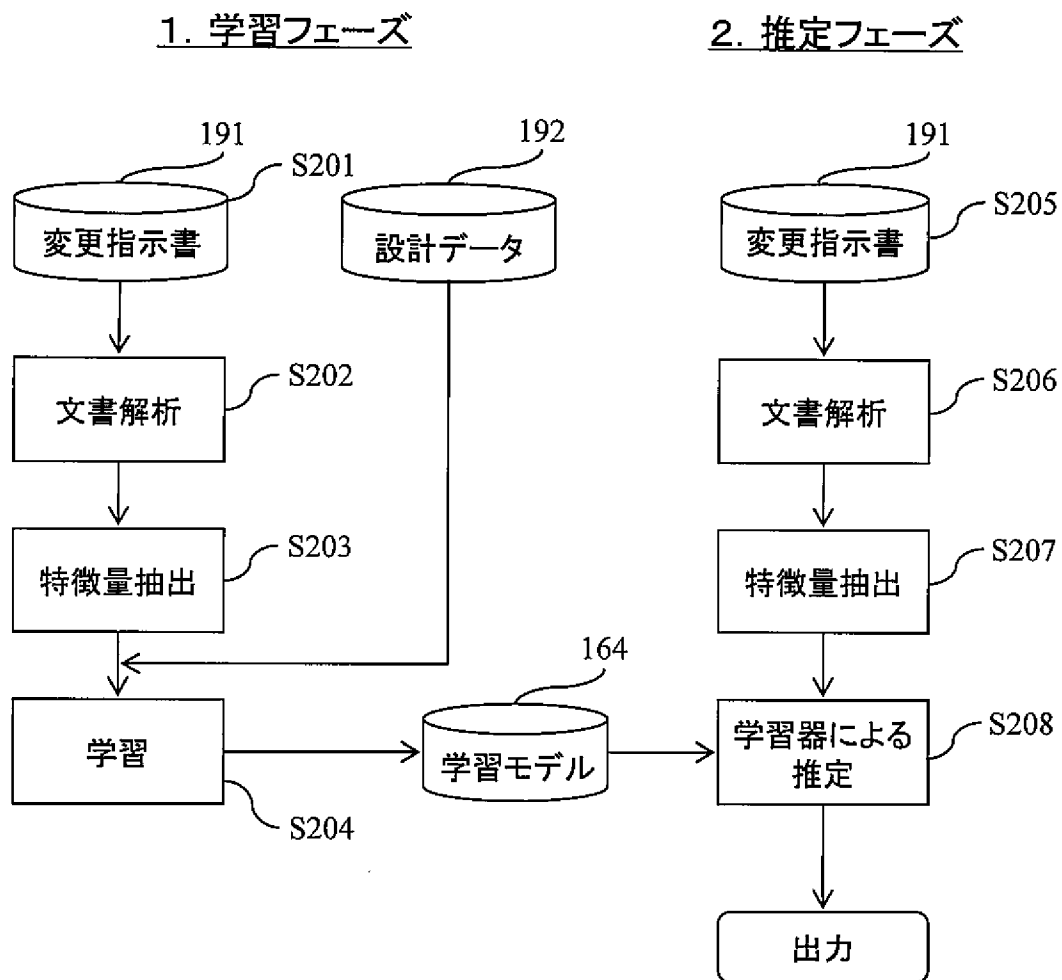
[図1]

図 1



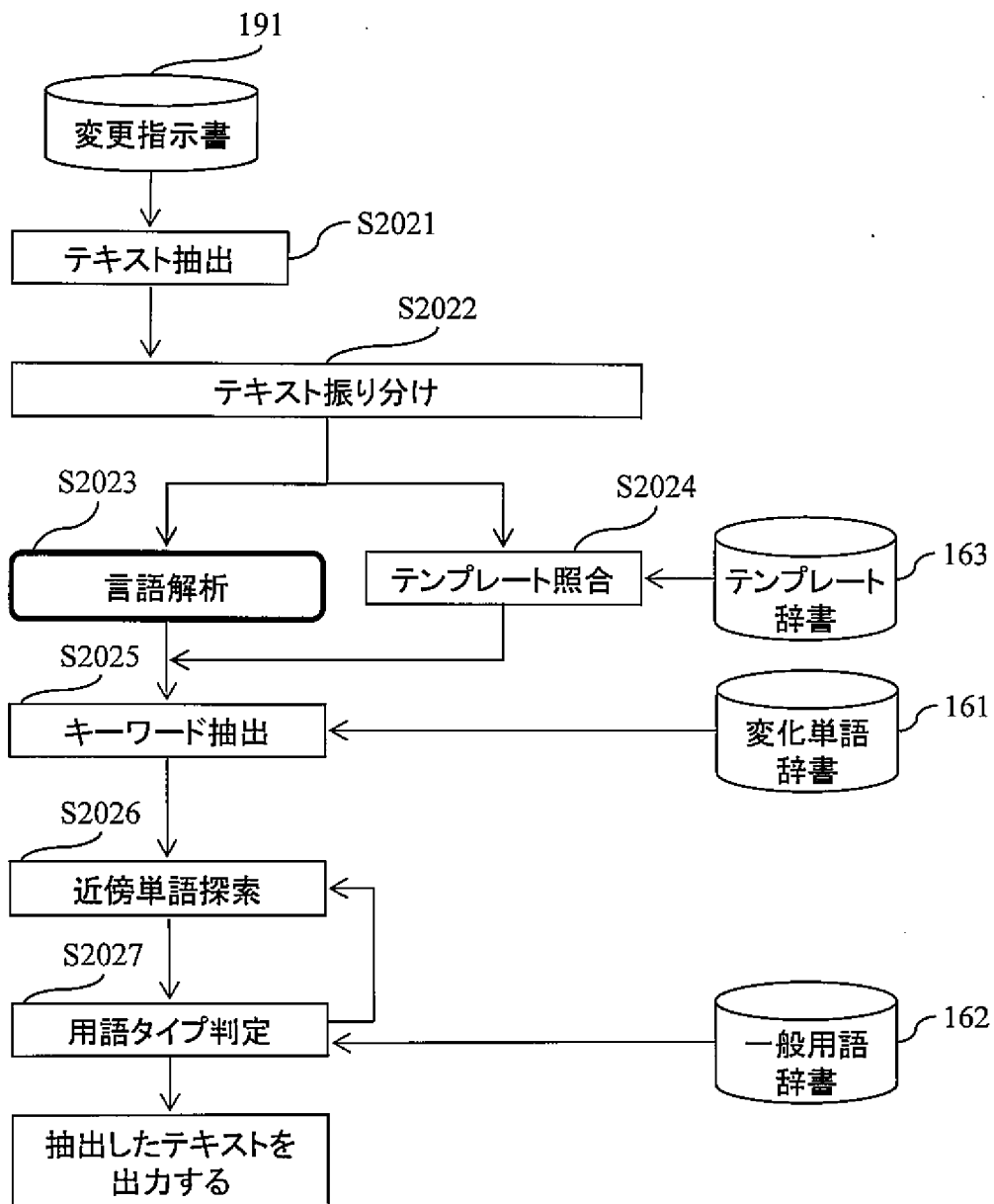
[図2]

図 2



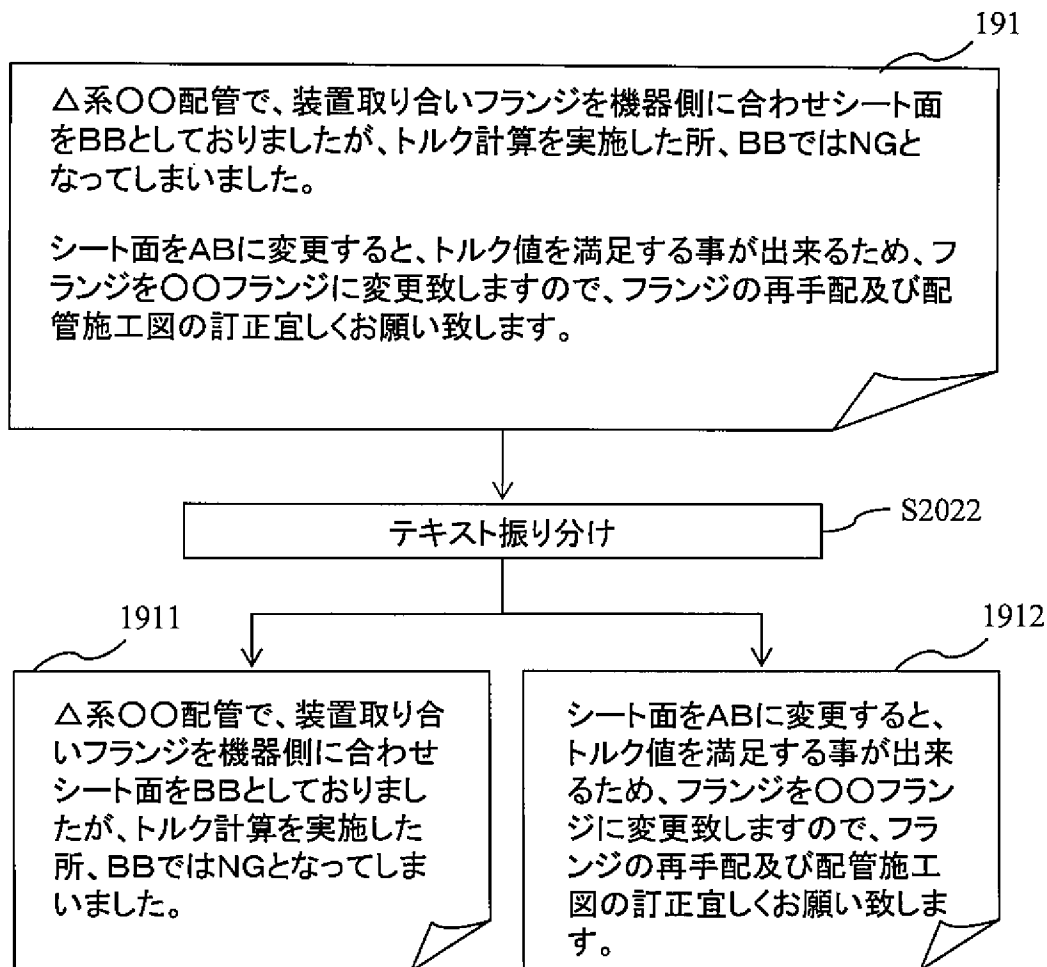
[図3]

図 3



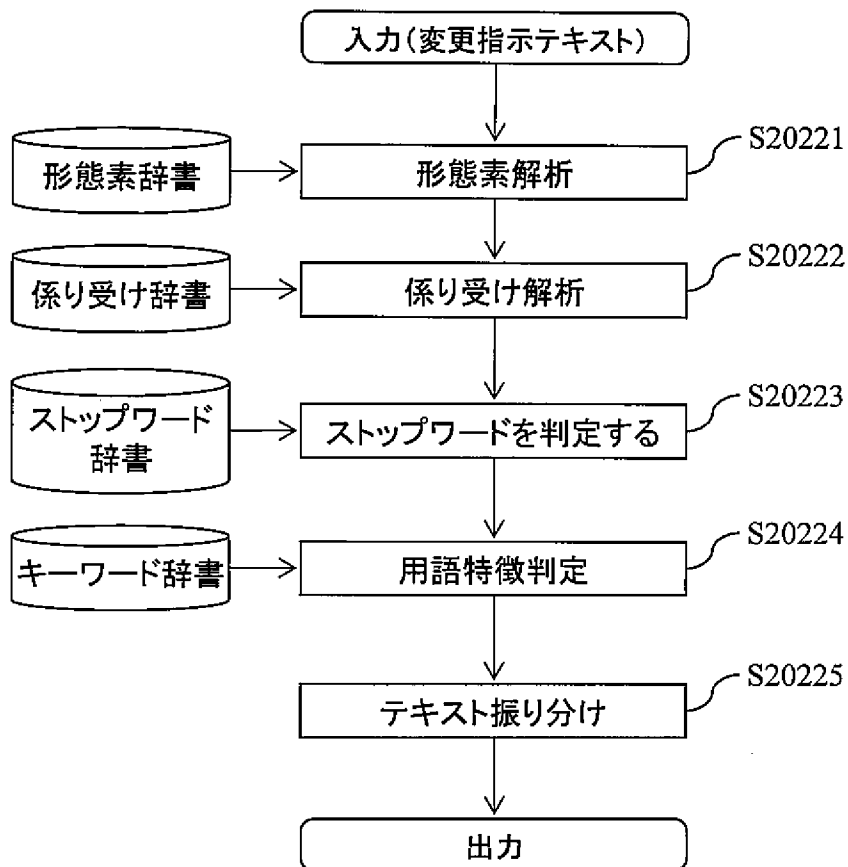
[図4]

図 4



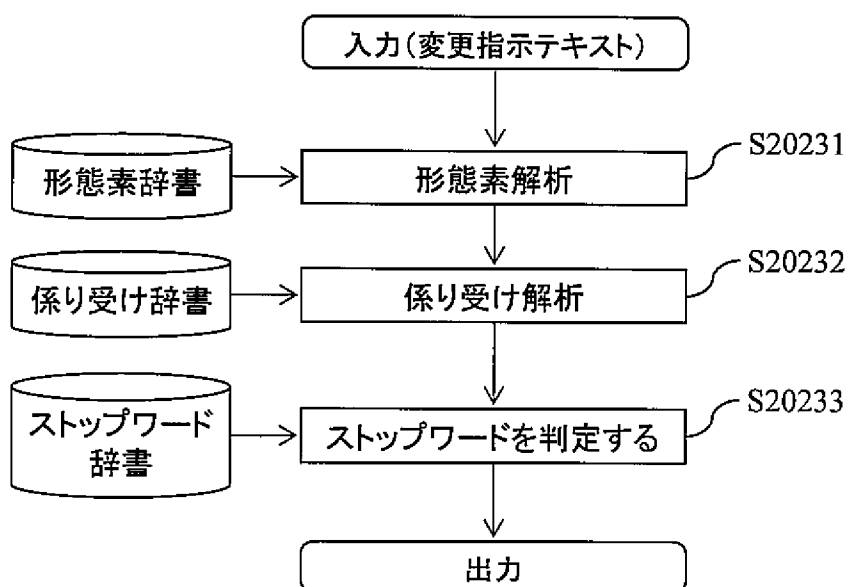
[図5]

図 5



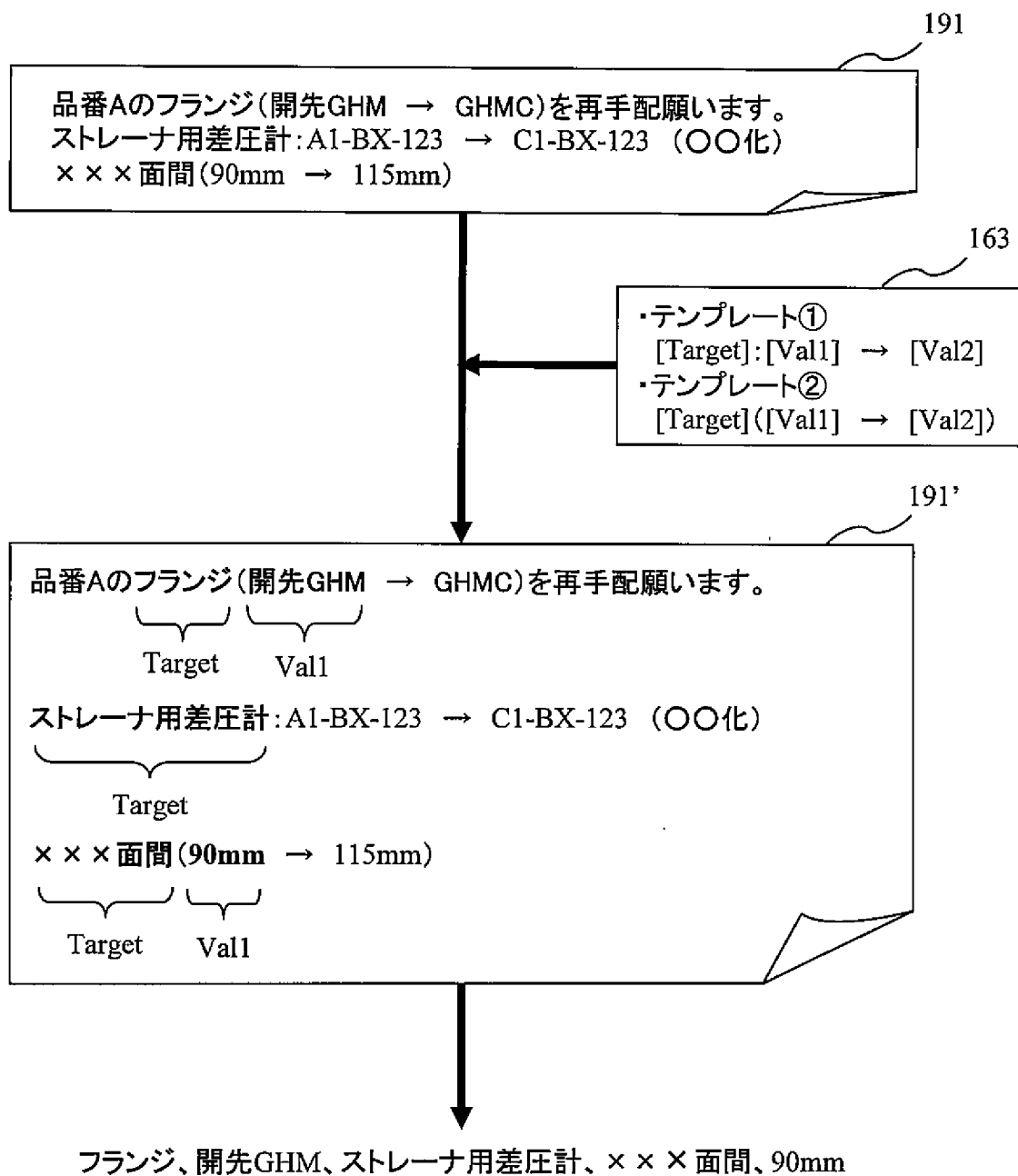
[図6]

図 6



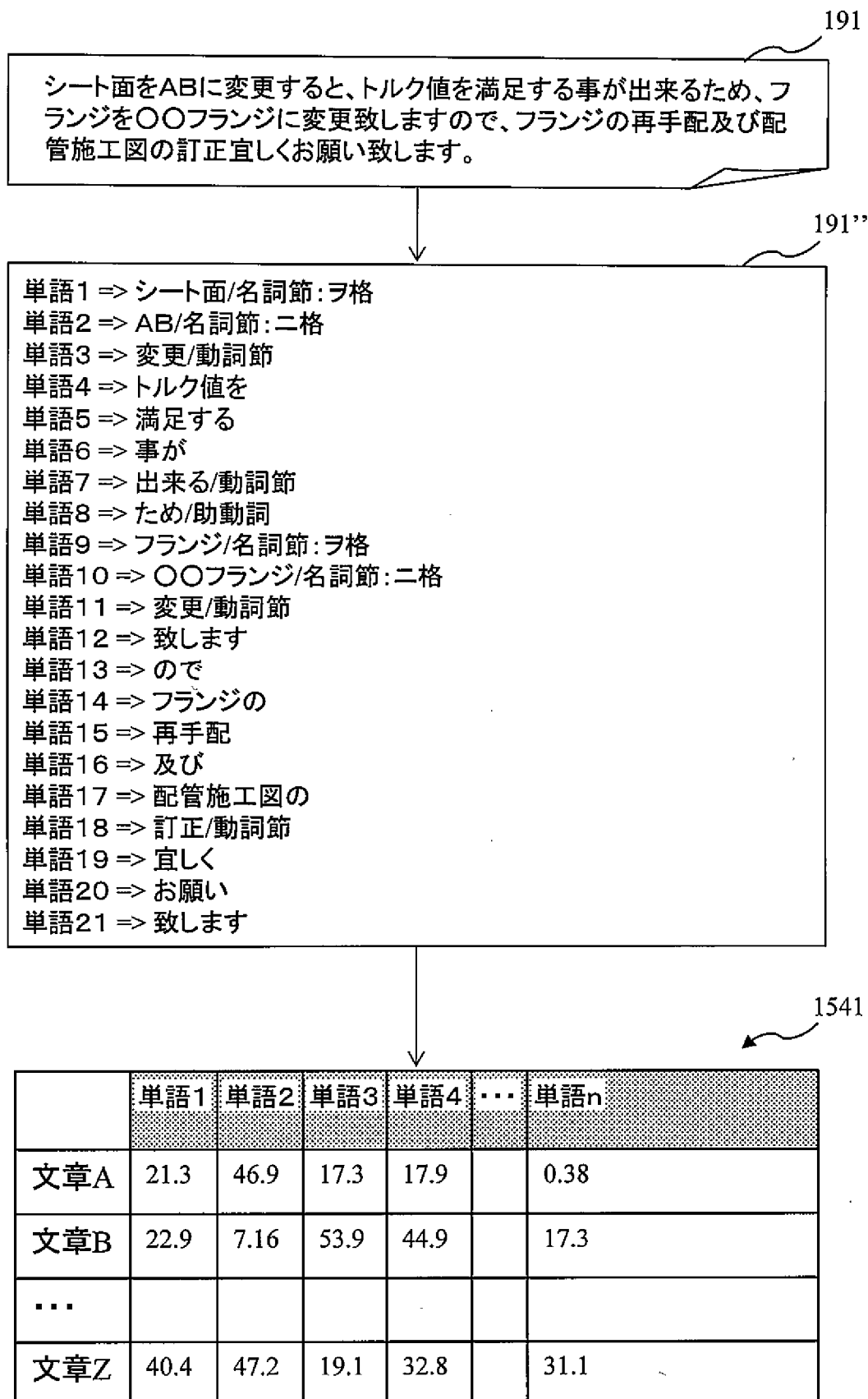
[図7]

図 7



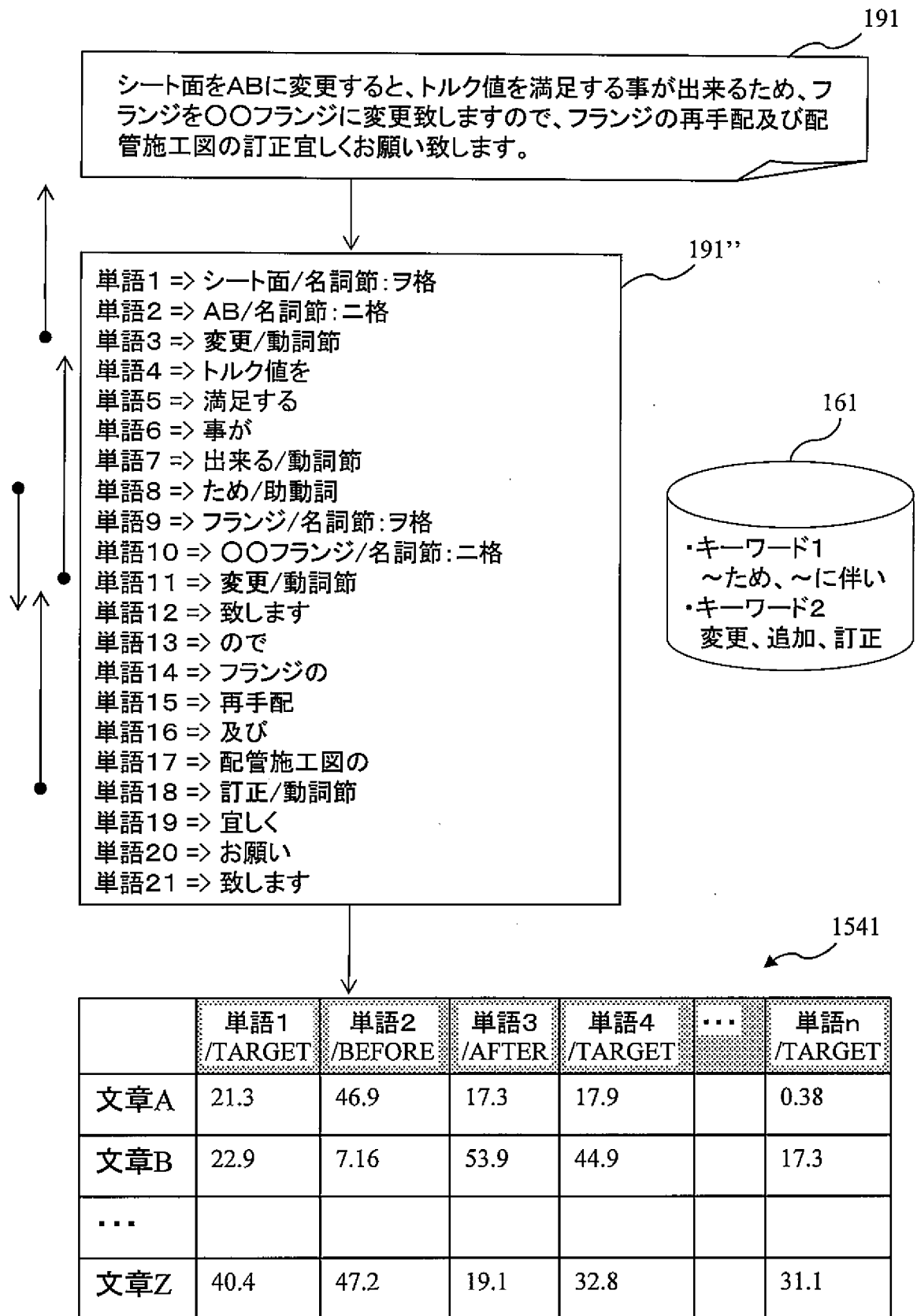
[図8]

図 8



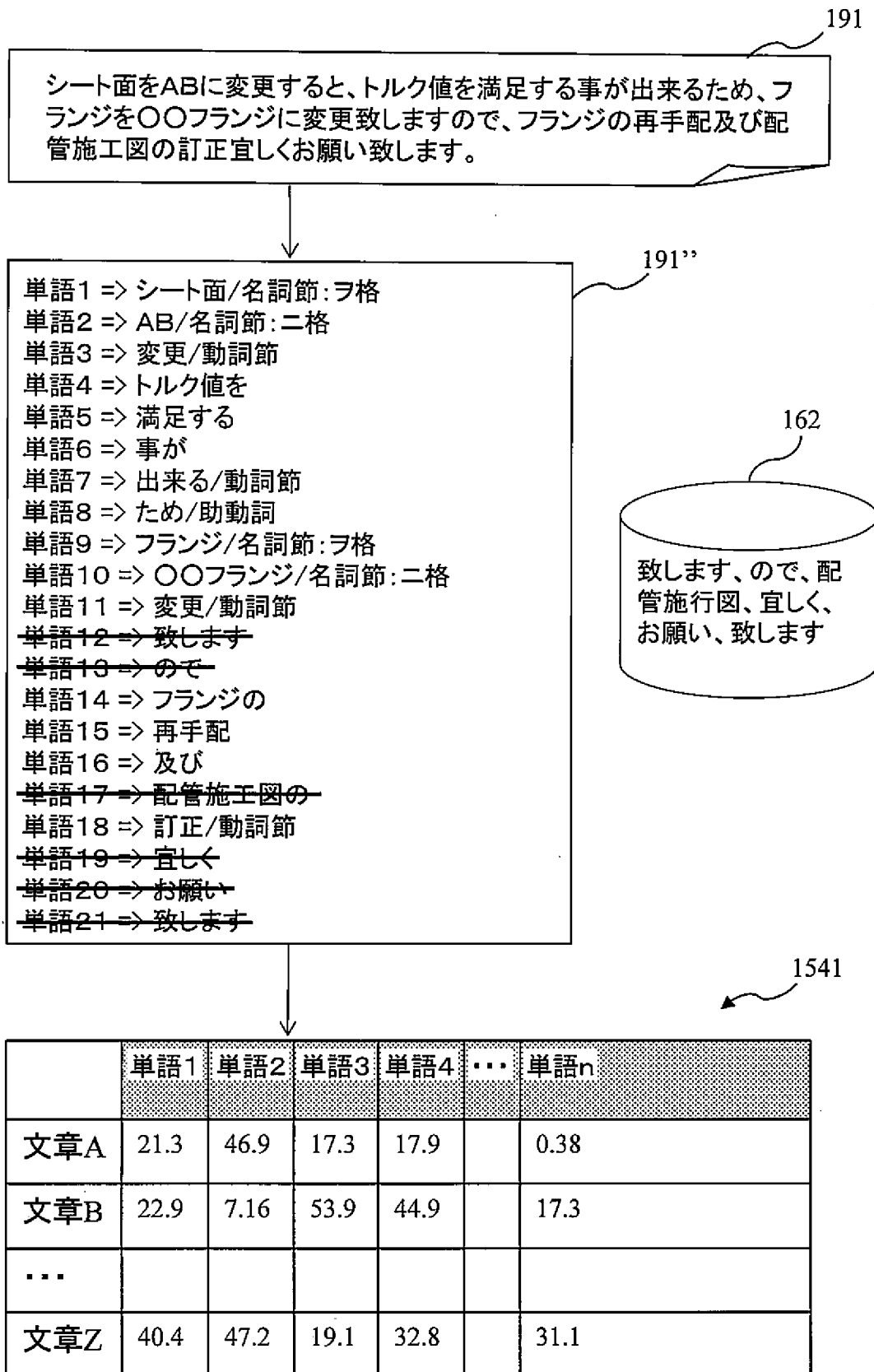
[図9]

図 9



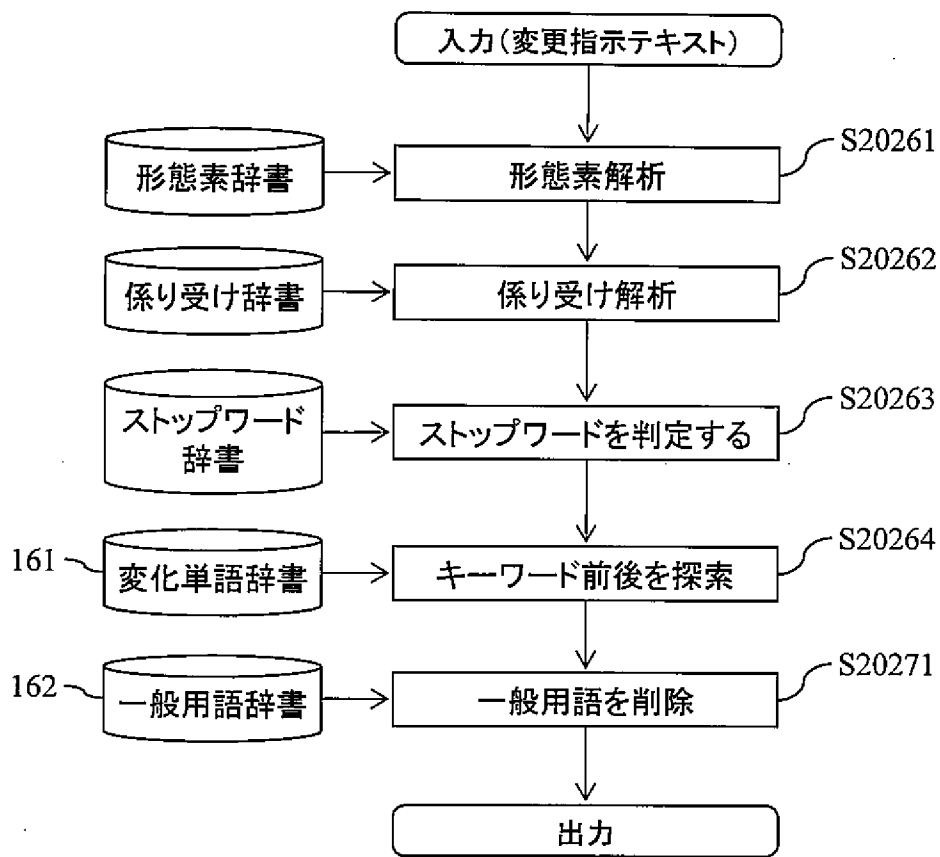
[図10]

図 10



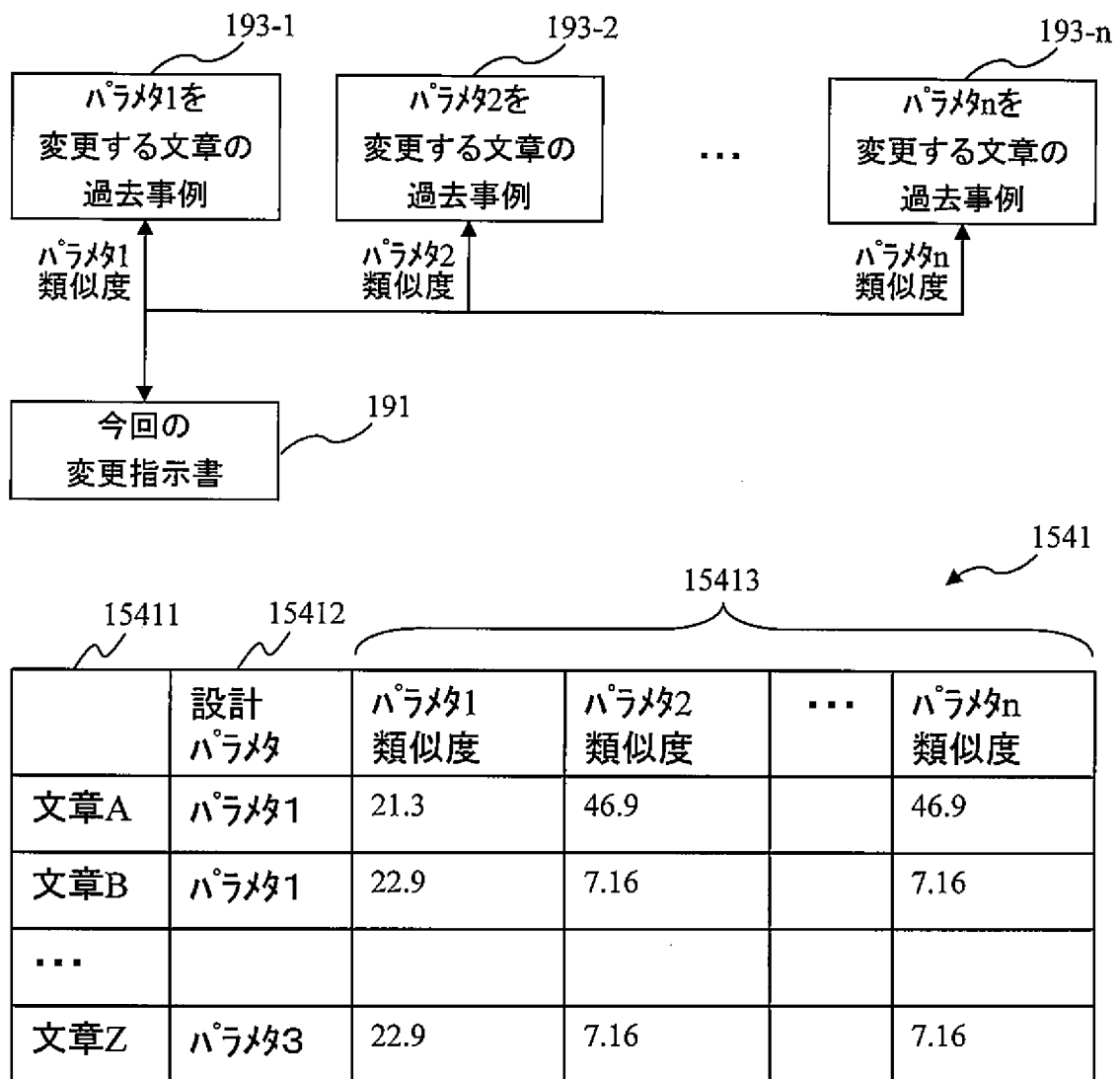
[図11]

図 1 1



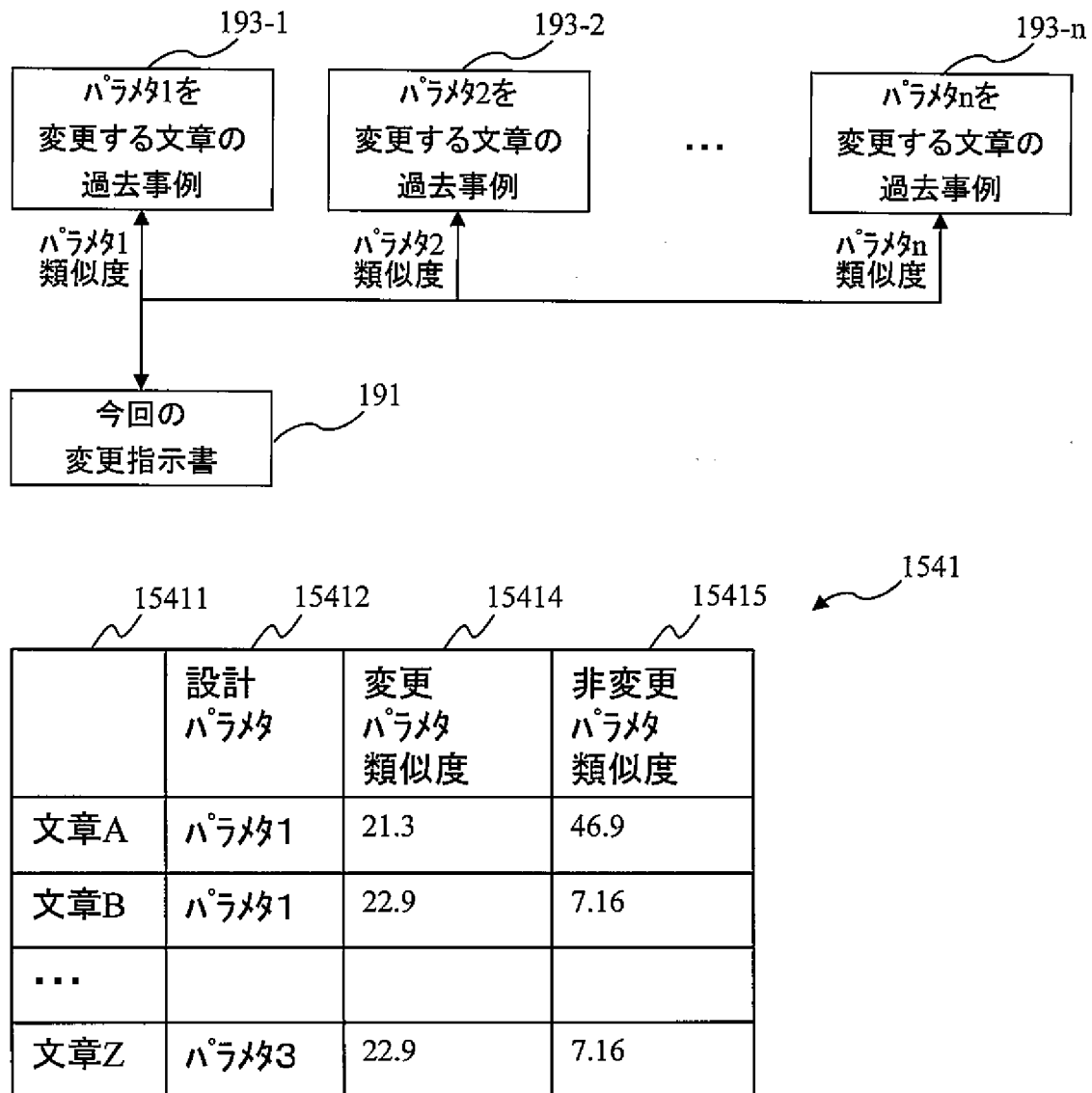
[図12]

図 1 2



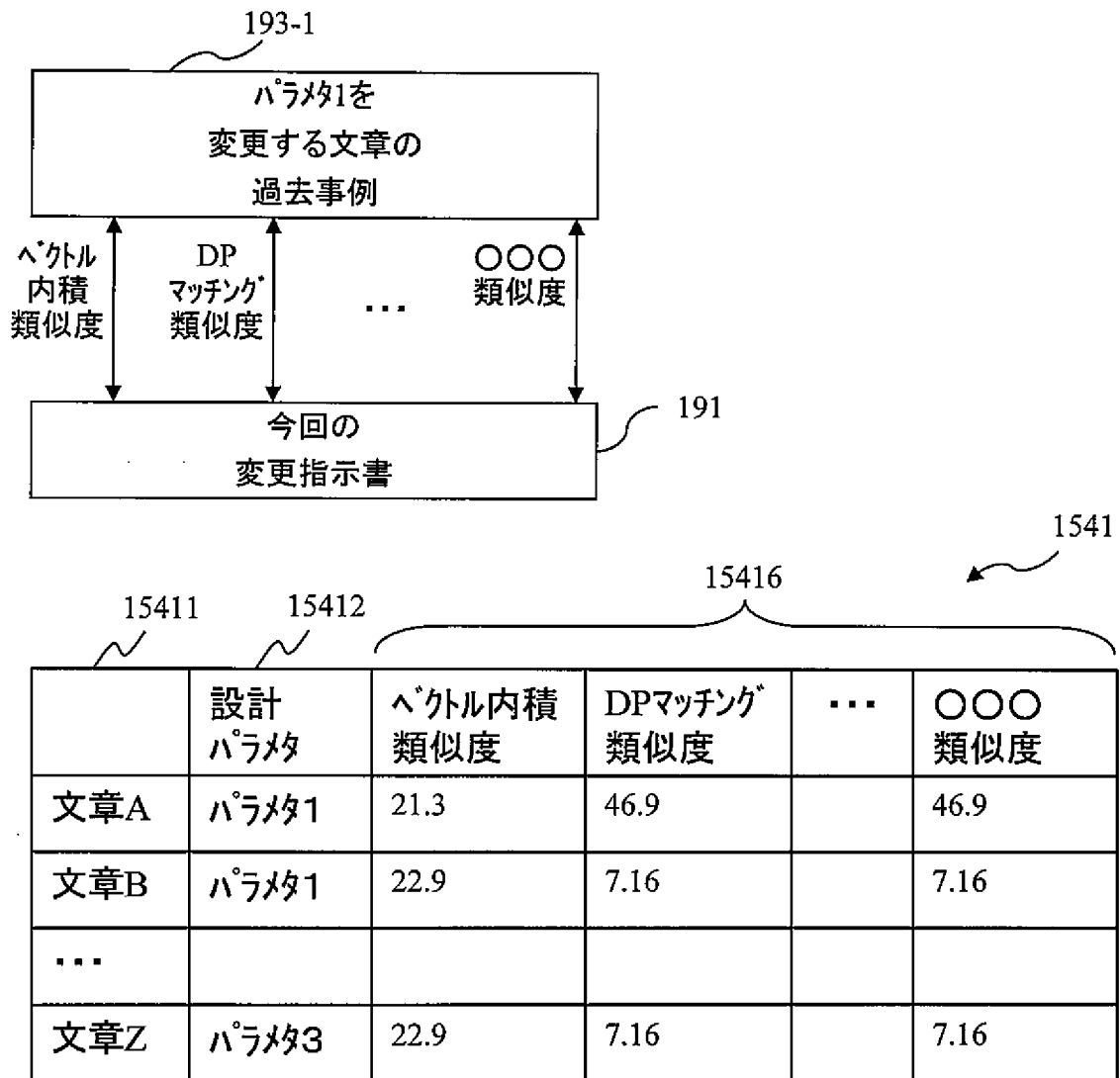
[図13]

図 1 3



[図14]

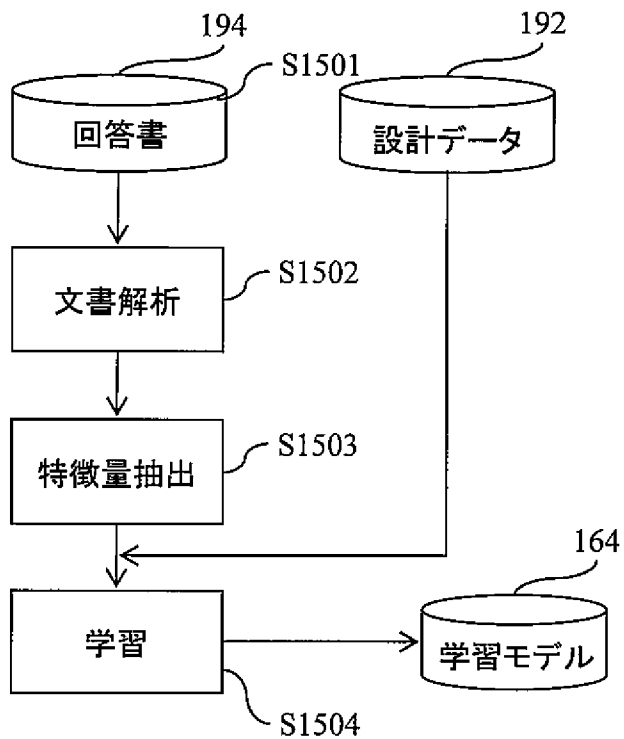
図 1 4



[図15]

図 1 5

3. 回答書学習フェーズ



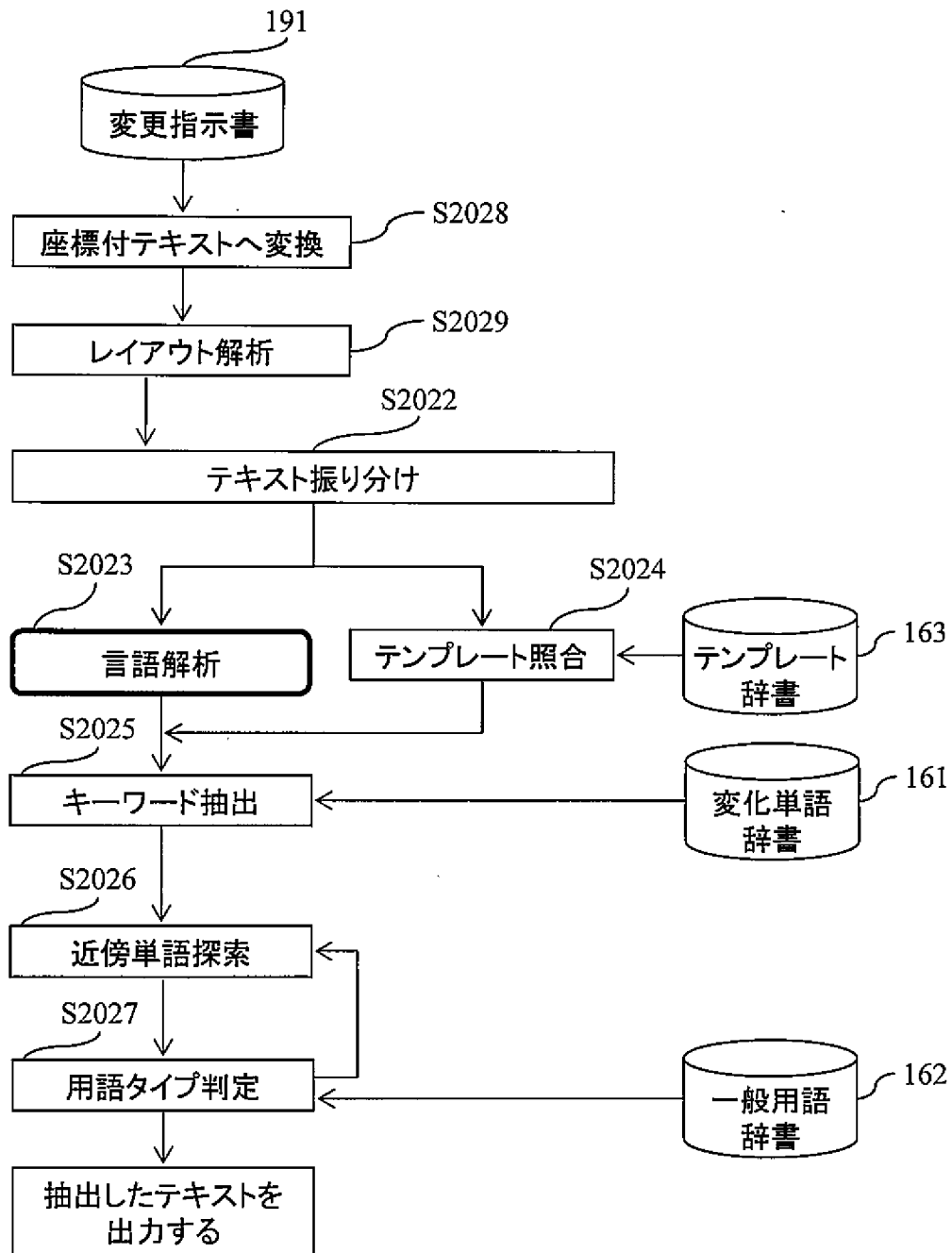
[図16]

図 1 6

項目 1		項目 2		項目 3	
項目 4				項目 5	
文書名				No.	Rev.
				1A3B456-789	0
				1A3B456-123	0
△系〇〇配管で、装置取り合いフランジを機器側に合わせシート面をBBとしておりましたが、トルク計算を実施した所、BBではNGとなってしまいました。 シート面をABに変更すると、トルク値を満足する事が出来るため、フランジを〇〇フランジに変更致しますので、フランジの再手配及び配管施工図の訂正宜しくお願い致します。					

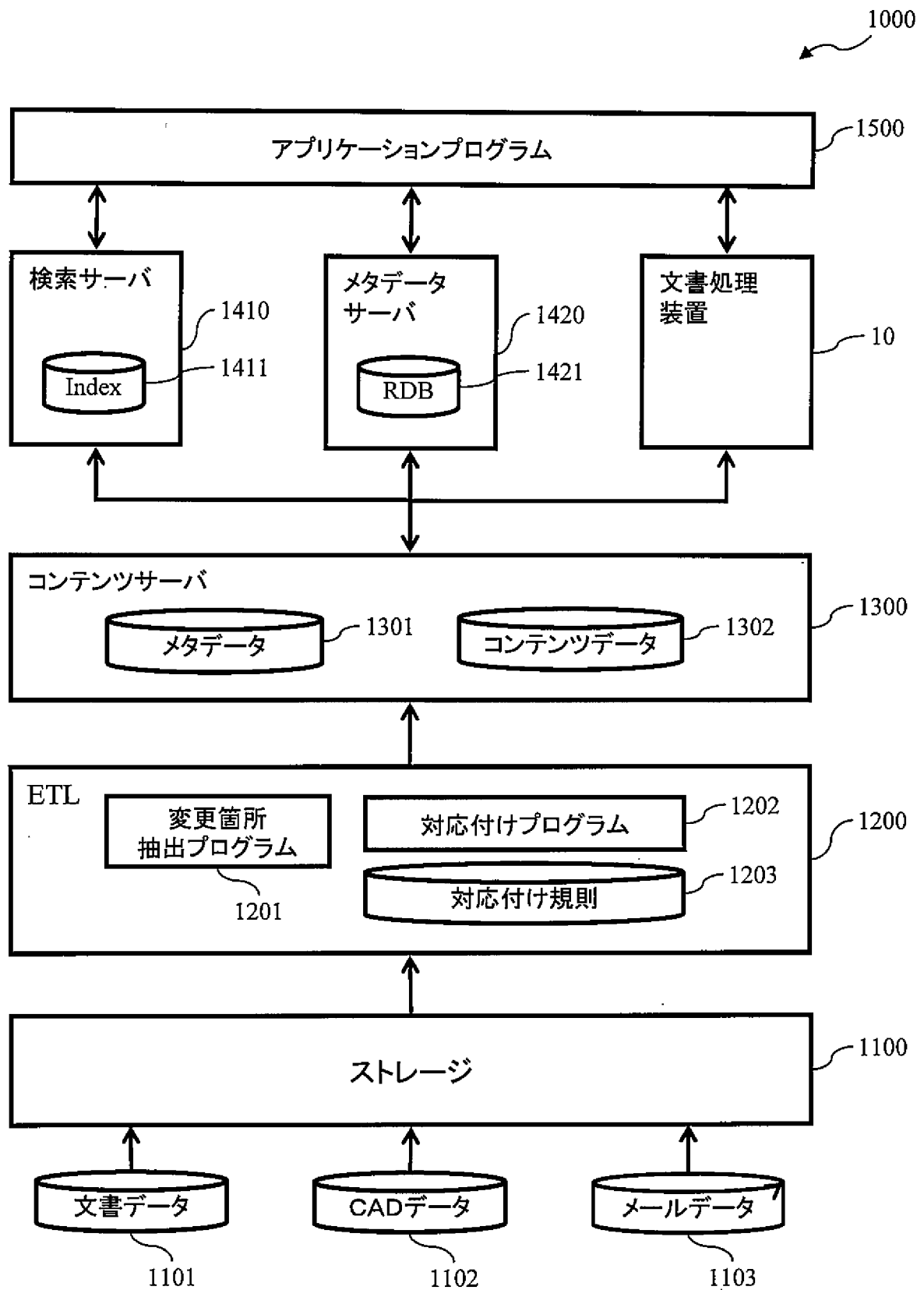
[図17]

図 1 7



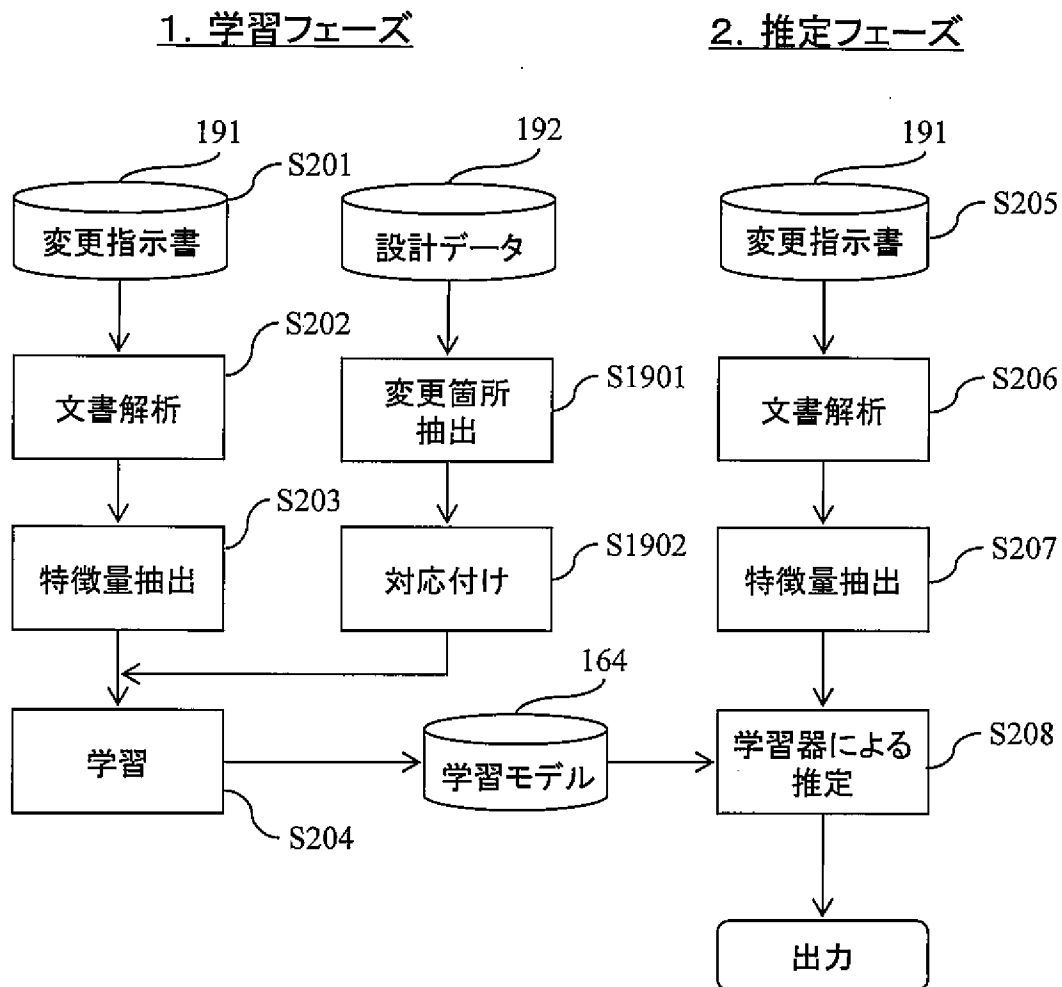
[図18]

図 18



[図19]

図 1 9



[図20]

図 20

1921		1922		192	
A1	A9	AA	AB	AC	AE
2008/3/25	N71504B	1	1		フランジ
2008/3/25	N71504B	1	2		ボルト
2008/3/25	N71504B	1	3		ナット
2008/3/25	N71504B	1	4	CR	パッキン
2008/3/25	N71504B	1	5	SPCC	ワッシャ
2008/3/25	N71504B	1	6	FRP	
2008/3/25	N71504B	27	7	FRP	

↓

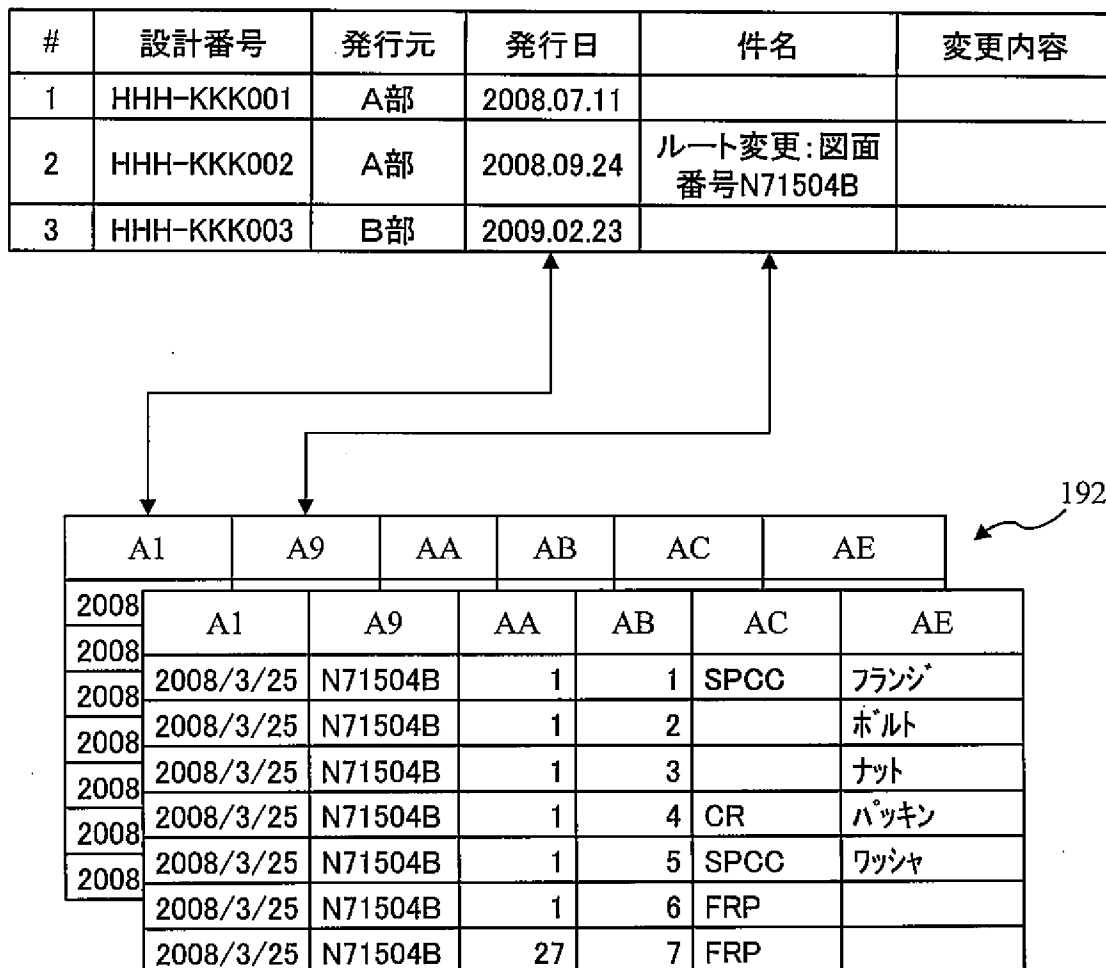
A1	A9	AA	AB	AC	AE
2008/3/25	N71504B	1	1	SPCC	フランジ
2008/3/25	N71504B	1	2		ボルト
2008/3/25	N71504B	1	3		ナット
2008/3/25	N71504B	1	4	CR	パッキン
2008/3/25	N71504B	1	5	SPCC	ワッシャ
2008/3/25	N71504B	1	6	FRP	
2008/3/25	N71504B	27	7	FRP	

192'

材質変更

[図21]

図 2 1



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2012/077614

A. CLASSIFICATION OF SUBJECT MATTER

G06Q50/04(2012.01)i, G06F17/30(2006.01)i, G06F17/50(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06Q50/04, G06F17/30, G06F17/50

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2013
Kokai Jitsuyo Shinan Koho	1971-2013	Toroku Jitsuyo Shinan Koho	1994-2013

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2000-047861 A (NEC Corp.), 18 February 2000 (18.02.2000), entire text; all drawings (Family: none)	1-15
A	JP 2012-014308 A (Hitachi-GE Nuclear Energy, Ltd.), 19 January 2012 (19.01.2012), entire text; all drawings (Family: none)	1-15
A	JP 11-184861 A (Hitachi, Ltd.), 09 July 1999 (09.07.1999), entire text; all drawings (Family: none)	1-15



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

21 January, 2013 (21.01.13)

Date of mailing of the international search report

29 January, 2013 (29.01.13)

Name and mailing address of the ISA/

Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2012/077614

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2010-287026 A (EXA Corp.), 24 December 2010 (24.12.2010), entire text; all drawings (Family: none)	1-15
A	JP 2006-344200 A (Hitachi, Ltd.), 21 December 2006 (21.12.2006), entire text; all drawings & US 2007/0078553 A1	1-15

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int.Cl. G06Q50/04(2012.01)i, G06F17/30(2006.01)i, G06F17/50(2006.01)i

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int.Cl. G06Q50/04, G06F17/30, G06F17/50

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2013年
日本国実用新案登録公報	1996-2013年
日本国登録実用新案公報	1994-2013年

国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 2000-047861 A (日本電気株式会社) 2000.02.18, 全文, 全図 (ファミリーなし)	1-15
A	JP 2012-014308 A (日立GEニュークリア・エナジー株式会社) 2012.01.19, 全文, 全図 (ファミリーなし)	1-15
A	JP 11-184861 A (株式会社日立製作所) 1999.07.09, 全文, 全図 (ファミリーなし)	1-15

☒ C欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」 特に関連のある文献ではなく、一般的技術水準を示すもの
「E」 国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」 優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)
「O」 口頭による開示、使用、展示等に言及する文献
「P」 国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」 国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」 特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」 特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」 同一パテントファミリー文献

国際調査を完了した日

21.01.2013

国際調査報告の発送日

29.01.2013

国際調査機関の名称及びあて先

日本国特許庁 (ISA/JP)

郵便番号100-8915

東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

唐橋 拓史

電話番号 03-3581-1101 内線 3562

5L

3575

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 2010-287026 A (株式会社エクサ) 2010.12.24, 全文, 全図 (ファミリーなし)	1-15
A	JP 2006-344200 A (株式会社日立製作所) 2006.12.21, 全文, 全図 & US 2007/0078553 A1	1-15