

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号  
特許第7216680号  
(P7216680)

(45)発行日 令和5年2月1日(2023.2.1)

(24)登録日 令和5年1月24日(2023.1.24)

(51)国際特許分類 F I  
 G 0 6 F 40/177 (2020.01) G 0 6 F 40/177  
 G 0 6 F 40/169 (2020.01) G 0 6 F 40/169

請求項の数 11 (全17頁)

(21)出願番号	特願2020-46704(P2020-46704)	(73)特許権者	000003078 株式会社東芝 東京都港区芝浦一丁目1番1号
(22)出願日	令和2年3月17日(2020.3.17)	(74)代理人	100091487 弁理士 中村 行孝
(65)公開番号	特開2021-149306(P2021-149306 A)	(74)代理人	100105153 弁理士 朝倉 悟
(43)公開日	令和3年9月27日(2021.9.27)	(74)代理人	100107582 弁理士 関根 毅
審査請求日	令和3年9月10日(2021.9.10)	(74)代理人	100118876 弁理士 鈴木 順生
		(74)代理人	100206243 弁理士 片桐 貴士
		(72)発明者	長野 伸一 東京都港区芝浦一丁目1番1号 株式会 最終頁に続く

(54)【発明の名称】 情報処理装置、情報処理方法、およびプログラム

(57)【特許請求の範囲】

【請求項1】

第1の表および第2の表に示されている共通属性を検出する検出部と、  
前記第1の表および前記第2の表の記載に基づき、前記共通属性に関する情報を生成する情報生成部

を備え、

前記情報生成部は、

前記第1の表の前記共通属性に係る行または列に存在する第1セルが示すテキストと、前記第2の表の前記共通属性に係る行または列に存在する第2セルが示すテキストと、に基づき、前記共通属性に関する情報を生成し、

前記第1セルが示すテキストを第1制約式に変換し、

前記第2セルが示すテキストを第2制約式に変換し、

前記共通属性に関する情報として第3制約式を、前記第1制約式および前記第2制約式に基づいて生成する

情報処理装置。

【請求項2】

前記検出部は、前記第1の表と参照関係にある表、または、前記第1の表と参照関係にある表を含む第2電子文書を、前記第1の表を含む第1電子文書内のテキストに基づいて検出する

請求項1に記載の情報処理装置。

## 【請求項 3】

前記情報生成部は、

前記第 1 制約式が成り立つならば前記第 2 制約式が必ず成り立つという第 4 制約式の真偽を判定し、

前記第 4 制約式が偽と判定された場合に、前記第 3 制約式を生成する

請求項 1 又は 2 に記載の情報処理装置。

## 【請求項 4】

前記情報生成部は、前記第 4 制約式を否定した第 5 制約式が真となるような前記第 5 制約式の変数への値の割り当てが存在することを確認した場合に、前記第 4 制約式を偽と判定する

請求項 3 に記載の情報処理装置。

## 【請求項 5】

前記情報生成部は、前記第 1 制約式が前記第 2 制約式に包括される場合において、前記第 1 制約式には包括されないが前記第 2 制約式に包括される第 3 制約式を生成する

請求項 1 ないし 4 のいずれか一項に記載の情報処理装置。

## 【請求項 6】

前記情報生成部は、前記第 1 制約式が前記第 2 制約式に包括されない場合に、前記第 2 制約式に包括される第 3 制約式を生成する

請求項 1 ないし 4 のいずれか一項に記載の情報処理装置。

## 【請求項 7】

前記情報生成部は、前記第 1 制約式および前記第 2 制約式を満たす第 3 制約式を生成する

請求項 1 ないし 4 のいずれか一項に記載の情報処理装置。

## 【請求項 8】

前記第 1 の表に含まれるレコードの主キーと、前記第 2 の表に含まれるレコードの主キーと、を特定するテーブル構造解析部をさらに備え、

前記第 1 セルに関するレコードの主キーと、前記第 2 セルに関するレコードの主キーと、が一致する

請求項 1 ないし 7 のいずれか一項に記載の情報処理装置。

## 【請求項 9】

前記検出部は、前記第 1 の表の属性名と、前記第 2 の表の属性名と、の少なくともいずれかを変換することにより、前記共通属性を検出する

請求項 1 ないし 8 のいずれか一項に記載の情報処理装置。

## 【請求項 10】

第 1 の表および第 2 の表に示されている共通属性を検出するステップと、

前記第 1 の表および前記第 2 の表の記載に基づき、前記共通属性に関する情報を生成するステップと、

を備え、

前記情報を生成するステップは、

前記第 1 の表の前記共通属性に係る行または列に存在する第 1 セルが示すテキストと、前記第 2 の表の前記共通属性に係る行または列に存在する第 2 セルが示すテキストと、に基づき、前記共通属性に関する情報を生成し、

前記第 1 セルが示すテキストを第 1 制約式に変換し、

前記第 2 セルが示すテキストを第 2 制約式に変換し、

前記共通属性に関する情報として第 3 制約式を、前記第 1 制約式および前記第 2 制約式に基づいて生成する

情報処理方法。

## 【請求項 11】

コンピュータに、

第 1 の表および第 2 の表に示されている共通属性を検出するステップと、

前記第 1 の表および前記第 2 の表の記載に基づき、前記共通属性に関する情報を生成す

10

20

30

40

50

るステップと、

を実行させるプログラムであって、

前記情報を生成するステップは、

前記第 1 の表の前記共通属性に係る行または列に存在する第 1 セルが示すテキストと、前

記第 2 の表の前記共通属性に係る行または列に存在する第 2 セルが示すテキストと、に基

づき、前記共通属性に関する情報を生成し、

前記第 1 セルが示すテキストを第 1 制約式に変換し、

前記第 2 セルが示すテキストを第 2 制約式に変換し、

前記共通属性に関する情報として第 3 制約式を、前記第 1 制約式および前記第 2 制約式に

基づいて生成する

プログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明の実施形態は、情報処理装置、情報処理方法、およびプログラムに関する。

【背景技術】

【0002】

製品の寸法、品質、形状、材料など決定する際には、様々な技術標準が参照される。例えば、国際標準規格、国内標準規格などといった公的な規格に加え、社内、部門などにおいても独自の規格が設けられていることも多い。製品の設計者は、全ての規格を満たすよ

うにするため、これら複数の技術標準を比較して設計業務を行うこともある。

【0003】

また、これらの技術標準は、他の技術標準を参照して作成されている場合もある。その場合、参照されている技術標準が更新されたときに、参照している技術標準も更新される必要がある。例えば、第 1 の技術標準の属性 A に関する数値範囲が更新された場合、技術標準の管理者は、第 1 の技術標準を参照する第 2 の技術標準に関する文書において属性 A が記載された場所を検出し、当該場所に表示された値を、第 1 の技術標準が示す範囲内の数値に更新するといった処理を行う必要が生じる。

【0004】

このような状況は、設計者、管理者などに余計な負担を与えるため、技術標準に関する運用を効率化する技術が求められている。技術標準の設計属性は、通常、表によって示されている。しかし、表の構造は様々であり、参照関係を有する表同士の間同じ位置に存在するセル同士が、同一属性に関する値を示すとは限らない。そのため、表同士を単純に比較することが難しい。また、一般的に、技術文書には複数の表が示されており、技術文書同士が参照関係にあることが判明していても、当該技術文書同士のいずれの表が参照関係にあるかは判明していない場合もあり得る。

【先行技術文献】

【特許文献】

【0005】

【文献】国際公開 2018/025707号

特開 2019-032704号公報

国際公開 2015/162889号

【発明の概要】

【発明が解決しようとする課題】

【0006】

本発明の一実施形態は、異なる表に共通して示されている属性を検出し、当該属性に関する情報を生成する情報処理装置を提供する。

【課題を解決するための手段】

【0007】

本発明の一実施形態としての情報処理装置は、検出部と、情報生成部と、を備える。前

10

20

30

40

50

記検出部は、第 1 の表および第 2 の表に示されている共通属性を検出する。前記情報生成部は、前記第 1 の表および前記第 2 の表の記載に基づき、前記共通属性に関する情報を生成する。

【図面の簡単な説明】

【0008】

【図 1】本発明の一実施形態に係る情報処理装置の一例を示すブロック図。

【図 2】参照関係にある表の一例を示す図。

【図 3】解析部の処理のフローの一例を示す図。

【図 4】参照関係にある表の他の一例を示す図。

【図 5】参照関係検出部の処理のフローの一例を示す図。

10

【図 6】検索部の処理のフローの一例を示す図。

【図 7】情報生成部の処理のフローの一例を示す図。

【図 8】SAT ツールを用いた判定スクリプトを示す図。

【図 9】本発明の一実施形態におけるハードウェア構成の一例を示すブロック図。

【発明を実施するための形態】

【0009】

以下、図面を参照しながら、本発明の実施形態について説明する。

【0010】

(本発明の一実施形態)

図 1 は、本発明の一実施形態に係る情報処理装置の一例を示すブロック図である。本実施形態に関する情報処理装置 1 は、入出力部 11 と、解析部 12 と、参照関係検出部 13 と、検索部 14 と、情報生成部 15 と、記憶部 16 と、を備える。解析部 12 は、文書構造解析部 121 と、テキスト解析部 122 と、テーブル構造解析部 123 と、を備える。参照関係検出部 13 は、参照テーブル検出部 131 と、共通属性検出部 132 と、を備える。検索部 14 は、指定情報抽出部 141 と、参照情報抽出部 142 と、を備える。情報生成部 15 は、制約式変換部 151 と、制約式充足部 152 と、制約式生成部 153 と、を備える。

20

【0011】

情報処理装置 1 は、異なる表に共通して示されている属性を検出することにより、参照関係にある表を特定する。また、参照関係にある表が示す記載に基づいて、当該属性に関する情報を生成する。以降、当該属性を共通属性と記載する。

30

【0012】

図 2 は、参照関係にある表の一例を示す図である。図 2 に示された第 1 の表と第 2 の表は、「XXX」という鋼材と、「YYY」という鋼材と、についての規定を示したものである。例えば、「XXX」という鋼材を、両方の表が示す規定を満たすように製造する場合、厚さを 13 mm 以上、長さを 180 mm 以下にすべきである。

【0013】

しかし、参照する表が一目で分かるように配置されているとは限らない。参照する表が異なる文書に記載されていることもある。また、同一文書に記載されていても、異なるページに記載されていることもある。そのため、人ではなく装置によって、参照する表の管理および上記のような条件を算出させることが好ましい。

40

【0014】

しかし、上記のような条件を自動的に算出させようとする、図 2 の例では問題が生じる。図 2 の例では、第 1 の表の列見出しに示された属性名と、第 2 の表の列見出しに示された属性名と、が完全には一致していない。そのため、第 1 の表が示す属性と、第 2 の表が示す属性と、が同じであると認識することができない。例えば、第 1 の表では第 1 の列に示された属性名は「種類記号」であり、第 2 の表では第 1 の列に示された属性名は「材料記号」である。そのため、第 1 の表の第 1 の列と、第 2 の表の第 1 の列と、が、鋼材の種別という同一の情報を示すものと認識されない。また、第 1 の表では mm といった単位が属性名に含まれているが、第 2 の表では属性名に含まれていない。

50

## 【 0 0 1 5 】

そこで、本実施形態の情報処理装置 1 は、与えられた電子文書および当該電子文書内の表を解析し、解析結果に基づいて共通属性を検出する。

## 【 0 0 1 6 】

なお、解析される電子文書は、HTML (Hyper Text Markup Language) ファイル、PDF (Portable Document Format) ファイルなどといった解析可能なものであればよい。例えば、インターネット上のサイトのページであってもよい。なお、電子文書が、図 2 示したような技術文書に限られるわけではない。また、表が示す情報も、製品に関するものでなくともよい。また、生成される情報も条件式に限られない。

10

## 【 0 0 1 7 】

情報処理装置 1 の内部構成について説明する。なお、図 1 に示された内部構成は例であり、図 1 に示されていない構成要素が情報処理装置 1 内に存在してもよい。また、図 1 に示した情報処理装置 1 の各構成要素が、細分化されてもよいし、集約されてもよい。また、図 1 に示された各構成要素が、情報処理装置 1 とは別の装置に含まれていてもよい。例えば、解析部 1 2、参照関係検出部 1 3、検索部 1 4、および情報生成部 1 5 がそれぞれ個別の装置であってもよい。また、例えば、ネットワークエリアストレージなどといった記憶装置に情報処理装置 1 の処理に関するデータが格納されていてもよい。すなわち、記憶部 1 6 が情報処理装置 1 の外部に存在していてもよい。

## 【 0 0 1 8 】

入出力部 1 1 は、情報処理装置 1 とは別の装置などから、処理に必要な情報を取得し、処理結果を出力する。なお、入出力される情報は特に限られるものではない。例えば、情報処理装置 1 によって解析される電子文書、検索対象とする表を識別するための識別子などを取得する。

20

## 【 0 0 1 9 】

解析部 1 2 は、電子文書に含まれる文章および表を解析する。図 3 は、解析部 1 2 の処理のフローの一例を示す図である。本フローは、電子文書が入出力部 1 1 に入力された場合に開始されることを想定する。

## 【 0 0 2 0 】

解析部 1 2 が入出力部 1 1 を介して、電子文書を取得する (S 1 0 1)。解析部 1 2 の文書構造解析部 1 2 1 は、当該電子文書から、文章と表を分離する (S 1 0 2)。分離の方法は、公知の手法でよい。例えば、電子文書が PDF ファイルの場合、文書または表を抽出する様々な抽出ソフトおよびコマンドが公開されており、それらを用いてもよい。

30

## 【 0 0 2 1 】

解析部 1 2 のテキスト解析部 1 2 2 は、電子文書から分離された表および文章それぞれに対し、テキスト解析を行う (S 1 0 3)。例えば、所定の文字列が、表および文章に含まれているかを検出する。例えば、表のレコードにおいて主キーとなる可能性が高い名称を定めておく。また当該文字列が正規表現で表されてもよい。

## 【 0 0 2 2 】

解析部 1 2 のテーブル構造解析部 1 2 3 は、抽出された表の構造を特定する (S 1 0 4)。例えば、主キーとなる可能性が高い属性の名称として、「記号」という文字列が定められていたとする。そして、図 2 に示した表がテキスト解析部 1 2 2 により解析された場合、テーブル構造解析部 1 2 3 は、図 2 の表に「記号」という文字列を含む属性があることを認識することができる。そして、図 2 の表に、主キーとなる可能性が高い他の名称が含まれていなければ、「記号」という文字列を含む属性を主キーと定める。すなわち、第 1 の表の主キーは「種類記号」であり、第 2 の表の主キーは「材料記号」とであると決定する。このようにして、テーブル構造解析部 1 2 3 は、いわゆる「表記ゆれ」があっても、レコードの主キーを特定することができる。

40

## 【 0 0 2 3 】

また、HTML、PDF などの電子文書では、ソースコードに、表のタイトル、見出し

50

などを示すタブなどが含まれている。そのため、テーブル構造解析部 1 2 3 は、電子文書のソースコードに基づいて、表内の見出しの位置などの構造を解析してもよい。

#### 【 0 0 2 4 】

また、本来、レコードは表の行に参照するが、複数のレコードを 1 行で表している表も存在する。図 4 は、参照関係にある表の一例を示す図である。図 4 に示された第 3 の表および第 4 の表は、複数のセルが結合されている。本来、第 3 の表および第 4 の表には、SM 4 0 0 A のレコードと、SM 4 0 0 B のレコードと、が存在するが、「SM 4 0 0 A SM 4 0 0 B」というレコードが存在するようにも解釈できる。

#### 【 0 0 2 5 】

テーブル構造解析部 1 2 3 は、表が図 4 のような構造の場合に、本来のレコードを特定する。例えば、製品の名称などが、1 レコードとして扱われる可能性が名称として、予め決めておく。例えば、SM 4 0 0 A、SM 4 0 0 B、および SM 4 0 0 C が定められている。テーブル構造解析部 1 2 3 は、解析されたテキストに基づいて、「SM 4 0 0 A SM 4 0 0 B」という文字列を認識し、さらに、当該文字列に二つの名称が含まれていることを認識する。認識は、正規表現検索によって行うことが可能である。

10

#### 【 0 0 2 6 】

テーブル構造解析部 1 2 3 は、複数の特定名称が認識された場合に、当該文字列に係る行を、特定名称ごとの行に分割してもよい。つまり、各属性の値が同じであるレコードが特定名称ごとに作成されてもよい。あるいは、セルに改行を意味する改行コードが含まれていると認識された場合に、当該セルに係る行は複数のレコードを含むとみなし、当該行を改行コードごとに分割してもよい。

20

#### 【 0 0 2 7 】

分割の際に行および列のいずれを増やすかは、主体および属性が行および列のいずれに記載されているかにより判断されてもよい。例えば、図 4 においては、「SM 4 0 0 A」と「SM 4 0 0 B」と同じく製品の名称である「SM 4 0 0 C」が、図 4 の前述のセルと同じ列に記載されている。ゆえに、図 4 の前述のセルを、「SM 4 0 0 A」を示すセルと、「SM 4 0 0 B」を示すセルと、に分割する場合は、行が増やされる。

#### 【 0 0 2 8 】

また、図 4 に示すように、列見出しが複数の行にわたって記載されている場合もあり得る。図 4 の例では、「降伏点又は体力  $N/mm^2$ 」と、「厚さ mm」と、「16 以下」と、それぞれ別のセルに記載されているが、これは、本来、1 つのセルに記載されるものである。テーブル構造解析部 1 2 3 は、このような表に対して、列見出しの行数を認識し、列見出しが 1 行で示された場合における属性名に変換してもよい。列見出しの行数は、前述の通り、電子文書のソースコードに基づいて、認識することが可能である。例えば、図 4 の例では、「降伏点又は体力  $N/mm^2$ 、厚さ mm、16 以下」のような属性名に変換する。このようにして、列見出しに係る行数が異なっている表同士でも、属性名を比較可能にする。

30

#### 【 0 0 2 9 】

このようにして、テーブル構造解析部 1 2 3 は、レコードおよびその属性を明確化して、共通属性を見つけやすくする。

40

#### 【 0 0 3 0 】

解析部 1 2 が処理結果を記憶部 1 6 に格納し (S 1 0 5)、フローは終了する。例えば、解析されたテキストに係るテキスト情報 1 6 1 が格納される。また、抽出された表およびその構造を示すテーブル情報 1 6 2 が電子文書と対応付けて記憶されてもよい。このようにして、電子文書が入出力部 1 1 に入力される度に本フローが実施されて、記憶部 1 6 に情報が蓄積されていく。

#### 【 0 0 3 1 】

参照関係検出部 1 3 は、電子文書、電子文書に含まれる表、表に含まれる属性などに対し、それらが参照するものを検出する。図 5 は、参照関係検出部 1 3 の処理のフローの一例を示す図である。本フローは、解析部 1 2 の処理が行われた後に実施される。なお、解

50

析部 1 2 の処理の実施の度に行われてもよいし、解析部 1 2 の処理が複数回行われた後にまとめて実行されてもよい。

【 0 0 3 2 】

参照関係検出部 1 3 の参照テーブル検出部 1 3 1 は、電子文書の解析されたテキストを取得する ( S 2 0 1 ) 。当該テキストは解析部 1 2 から取得してもよいし、記憶部 1 6 のテキスト情報を用いてもよい。参照テーブル検出部 1 3 1 は、当該テキストに基づいて、当該電子文書または当該電子文書内の表と参照する表を検出する ( S 2 0 2 ) 。

【 0 0 3 3 】

例えば、電子文書 A には引用文献として電子文書 B が記載されていると想定する。その場合、電子文書 A の解析されたテキストには電子文書 B の名称が示されている。参照テーブル検出部 1 3 1 は、電子文書 A の解析されたテキストに電子文書 B の名称があることを検出し、電子文書 B を電子文書 A によって参照される電子文書とみなしてもよい。電子文書 B の名称など、引用されると想定される電子文書の名称を予め記憶部 1 6 が記憶しておき、参照テーブル検出部 1 3 1 は、それらの電子文書の名称がテキストに示されているかを検索すればよい。あるいは、解析部 1 2 が入力された電子文書の名称を記憶部 1 6 に記憶し、参照テーブル検出部 1 3 1 は、以前に入力された電子文書の名称がテキストに示されているかを検索してもよい。

10

【 0 0 3 4 】

また、参照テーブル検出部 1 3 1 は、電子文書 A のテキストに、「を参照」などといった参照関係を示す文字列を検出してもよい。当該文字列が検出された場合に、当該文字列の目的語とされる電子文書および表が、参照関係にあるものとみなしてもよい。参照関係を示す文字列は予め記憶部 1 6 が記憶していればよい。また、例えば、電子文書 A の表 3 の前後に記載された文章に「引用文献の表 7 を参照」といったテキストがあるが、電子文書 B には、電子文書 A の記載がなかった場合、電子文書 A が電子文書 B を参照し、電子文書 A の表 3 が、電子文書 B の表 7 を参照するとみなしてもよい。このようにして、参照方向を推定してもよい。

20

【 0 0 3 5 】

なお、上記のような参照関係の特定方法では、誤検出の恐れがある。ゆえに、参照する表であると断定せず、候補とみなし、共通属性検出部 1 3 2 によって参照する属性が検出された場合に、参照する表であると確定されてもよい。

30

【 0 0 3 6 】

共通属性検出部 1 3 2 は、テキスト情報 1 6 1 、テーブル情報 1 6 2 などに基づいて、参照関係にある表に共通する属性 ( 共通属性 ) を検出する ( S 2 0 3 ) 。例えば、共通属性検出部 1 3 2 は、参照関係にあるとみなされた電子文書 A の表および電子文書 B の表を記憶部 1 6 から抽出し、両表の属性を比較して、属性が一致するかを確認し、両表に示されている共通属性を検出する。

【 0 0 3 7 】

なお、属性の名称が完全に一致しなくとも、属性が一致するとみなしてよい。例えば、図 2 の例では、「種類記号」と「材料記号」のように属性の名称が完全に一致していない。しかし、共通属性検出部 1 3 2 は、「種類記号」と「材料記号」を「記号」という名称に変換し、主キーに該当する列の各行に記載された「 X X X 」、「 Y Y Y 」という文字列が一致していることから、属性が一致していると判定してもよい。変換方法は、予め重要とされる文字列を定めておき、当該文字列が含まれる場合は、当該文字列以外の部分を消去してもよい。あるいは、言い換えなどを記憶しておき、「種類記号」と「材料記号」のいずれか一方を他方に変換してもよい。このようにして、第 1 の表の属性名と、第 2 の表の属性名と、の少なくともいずれかを変換することにより、共通属性を検出してもよい。

40

【 0 0 3 8 】

また、参照関係にある属性が 1 対 1 であるとは限らない。例えば、図 4 に示した第 3 の表には、「降伏点又は体力  $N / m m ^ 2$ 、厚さ  $m m$ 、40 を超え 75 以下」という属性と、「降伏点又は体力  $N / m m ^ 2$ 、厚さ  $m m$ 、75 を超え 100 以下」という属性と、が

50

四角い枠で囲まれて示されている。しかし、これらの属性に関する値が「215以上」で同じであるため、第4の表では、これらの属性と参照関係にある属性が、「降伏点又は体力  $N/mm^2$ 、厚さ mm、40を超え100以下」とまとめられて、示されている。このように、参照関係にある表においては、複数の属性が一つにまとめられていることもあり得る。

【0039】

そのような場合に対応するために、共通属性検出部132は、「以上」、「以下」、「未満」、「超え」などといった特定の単語と、数値と、によって構成された属性名を数式に変換する。変換は、変換前の単語と変換後の数学記号との対応関係を示した対応表に基づき行われればよく、記憶部16が当該対応表を予め記憶しておけばよい。

10

【0040】

例えば、第1の表の「降伏点又は体力  $N/mm^2$ 、厚さ mm、40を超え75以下」を、「 $40 <$ 降伏点又は体力  $N/mm^2$ 、厚さ mm 75」と変換する。同様に、第2の表の「降伏点又は体力  $N/mm^2$ 、厚さ mm、40を超え100以下」を「 $40 <$ 降伏点又は体力  $N/mm^2$ 、厚さ mm 100」と変換する。これにより、第1の表の変換された属性は、第2の表の変換された属性に包括されており、参照関係にあると判定することができる。

【0041】

なお、参照関係にある電子文書は判明したが、参照関係にある表が不明である場合は、参照テーブル検出部131は、参照関係にある電子文書内の表ごとに、共通属性があるかを探索して、参照関係にある表を検出してもよい。

20

【0042】

参照関係検出部13は処理結果を記憶部16に格納し(S204)、フローは終了する。例えば、電子文書および表の参照関係を示す参照関係情報163が格納される。また、参照関係にある表の共通属性を示す共通属性情報164が記憶されてもよい。このようにして、入出力部11に入力された電子文書およびそれらに含まれる表の参照関係に関する情報が蓄積されていく。

【0043】

検索部14は、ユーザから指定された検索対象に対する情報を記憶部16から取得する。図6は、検索部14の処理のフローの一例を示す図である。本フローは、指定対象が入出力部11に入力された場合に開始されることを想定する。

30

【0044】

検索部14が、入出力部11を介して、指定対象に係る情報を取得する(S301)。例えば、検索対象として、電子文書およびその表の名称などの識別子が指定されてもよい。また、表の属性が指定されてもよいし、表に含まれるテキストが指定されてもよい。例えば、図2に示したような「XXX」などが指定されてもよい。

【0045】

検索部14の指定情報抽出部141は、記憶部16から指定対象に係る情報を取得する(S302)。例えば、記憶部16に記憶された電子文書および表の全体が抽出されてもよいし、レコード、セルといった表の一部が抽出されてもよい。

40

【0046】

検索部14の参照情報抽出部142は、参照関係情報163、共通属性情報164などに基づいて、記憶部16から指定対象の参照関係を確認し(S303)、テーブル情報162から参照関係にある情報を抽出する(S304)。例えば、図2の第1の表と、「XXX」というテキストと、が指定された場合、参照関係情報163によって図2の第2の表との参照関係が認識できるため、参照情報抽出部142は、図2の第2の表の「XXX」に係るレコードを抽出してもよい。さらに「厚さ」が指定されていた場合、当該「厚さ」は、第2の表における「鋼板の厚さ」に対応することが共通属性情報164によって認識できるため、参照情報抽出部142は、図2の第2の表の「XXX」と「鋼板の厚さ」に係るセル「10mm以上」を抽出してもよい。

50



## 【 0 0 4 7 】

検索部 1 4 は、指定対象の情報と、当該情報と参照関係にある情報と、を送信する（ S 3 0 5 ）。送信先は、入出力部 1 1 であってもよい。その場合、ユーザが指定レコードおよび参照レコードの内容を確認することになる。また、送信先は、情報生成部 1 5 であってもよい。あるいは、それらの情報が記憶部 1 6 に記憶され、記憶部 1 6 を介して、情報生成部 1 5 に提供されてもよい。それらの場合、情報生成部 1 5 がそれらの情報に基づいて新たな情報を生成することになる。

## 【 0 0 4 8 】

情報生成部 1 5 は、指定対象に係る情報を生成して提供する。例えば、情報生成部 1 5 は、指定された表と、当該表に参照する表と、の記載に基づき、共通属性に関する情報を生成する。具体的には、指定された表の共通属性に係る列に存在するセルが示すテキストと、参照する表の共通属性に係る列に存在するセルが示すテキストと、に基づき、共通属性に関する情報を生成する。なお、見出しが列ではなく行に記載されている場合は、「共通属性に係る列」は「共通属性に係る行」と読み替えられる。

10

## 【 0 0 4 9 】

なお、生成される情報は様々であるが、本実施形態では、指定された表に数値または数値範囲を示す制約式が示されている場合を想定する。図 7 は、情報生成部 1 5 の処理のフローの一例を示す図である。本フローは、検索部 1 4 からの検索結果を受信した場合に開始されることを想定する。

## 【 0 0 5 0 】

情報生成部 1 5 が、指定対象の情報と、当該情報と参照関係にある情報と、を受信する（ S 4 0 1 ）。ここでは、図 2 に示した第 1 の表と、それに参照される第 2 の表と、を受信したとする。

20

## 【 0 0 5 1 】

ここで、例えば、図 2 に示した第 1 の表が社内技術標準を規定したもので、第 2 の表が国内技術標準を規定したものであるならば、第 1 の表の各レコードに関する規定内容は、第 2 の表の当該レコードに関する規定内容を満たしているかどうかを判定する必要がある。本実施形態では、この判定処理を、数理論理の分野における命題論理式の充足可能性問題（ S A T ）とみなす。命題論理式とは、制約条件を表す数学的な表現である。以降では単に制約式と呼ぶ。情報生成部 1 5 は、 S A T ツールと呼ばれる判定ツールを利用して、充足可能性問題を解く。充足可能性問題および S A T ツールは、コンピュータサイエンス、人工知能などの分野において広く利用されており、そのような S A T ツールを用いればよい。

30

## 【 0 0 5 2 】

情報生成部 1 5 の制約式変換部 1 5 1 が、受信した情報を制約式に変換する（ S 4 0 2 ）。具体的には、受信した両表の各セルが示すテキストを制約式に変換する。なお、共通属性に係るセルが示すテキストのみを制約式に変換してもよい。制約式変換部 1 5 1 は、前述の共通属性検出部 1 3 2 と同様、「以上」、「以下」、「未満」、「超え」などの単語を所定の変換法則に基づいて数学記号に変換する。

## 【 0 0 5 3 】

例えば、第 1 の表の「 X X X 」のレコードは、「厚さ 1 3 長さ 1 8 0」という制約式に変換される。ここで「 $\wedge$ 」は AND を表し、2 つの制約式「厚さ 1 3」と「長さ 1 8 0」が同時に成立する場合に真になることを意味する。同様に、第 2 の表の「 X X X 」のレコードは、「厚さ 1 0 長さ 2 0 0」という制約式に変換される。

40

## 【 0 0 5 4 】

次に、「 X X X 」のレコードについて、第 1 の表で規定されている制約式「厚さ 1 3 長さ 1 8 0」が、第 2 の表で規定されている制約式「厚さ 1 0 長さ 2 0 0」に含意されている必要がある。このことは、「厚さ 1 0 長さ 2 0 0 厚さ 1 0 長さ 2 0 0」という制約式で表される。ここで「 $\supseteq$ 」は含意関係を表し、「厚さ 1 0 長さ 2 0 0」が成り立つならば、制約式「厚さ 1 0 長

50

さ 200」が成り立つ』場合に真になることを意味する。

【0055】

情報生成部15の制約式充足部152は、これらの制約式の充足可否（つまり、成立するかどうか）を判定する（S403）。充足しない場合（S404のNO）、制約式生成部153は、これらの制約式に基づいて、新たな制約式を生成する（S405）。

【0056】

例えば、第1の表の各レコードに関する規定内容が、第2の表の当該レコードに関する規定内容を満たすためには、厚さおよび長さを変数と見なしたとき、両変数に任意の値を割り当てたととしても、前述の含意関係を表す制約式が常に真となる（恒真と呼ぶ）必要がある。なぜならば、含意関係が成立しないような変数への値の割り当てが存在することは、第1の表の規定内容が第2の表の規定内容を満たさない場合が存在することを意味するからである。

10

【0057】

例えば、第1の表の「XXX」のレコードに関して、厚さおよび長さの両変数に任意の値を割り当てても、第1の表で規定されている制約式「厚さ 13 長さ 180」と、第2の表で規定されている制約式「厚さ 10 長さ 200」が同時かつ常に真となるので、含意関係を表す制約式「厚さ 10 長さ 200 厚さ 10 長さ 200」は常に真となる（恒真となる）。

【0058】

一方、「YYY」のレコードに関して、例えば、厚さが13で長さが230の場合、第1の表で規定されている制約式「厚さ 13 長さ 230」は真となるが、第2の表で規定されている制約式「厚さ 15 長さ 250」は偽となるので、含意関係を表す制約式「厚さ 13 長さ 230 厚さ 15 長さ 250」は偽となり、成立しない。

20

【0059】

このように、含意関係を表す制約式が恒真であるかどうかは、各変数へ値を順に割り当てた場合に、真および偽のどちらになるかを確認することにより、判定可能である。しかしながら、各変数が取りうる全ての値の組み合わせを網羅的に調べることは、膨大な時間を要する。

【0060】

30

そこで、本実施形態では、組み合わせを網羅的に調べずに、判定を行う。ある制約式Pが恒真であるとは、Pに対する否定である $\neg P$ に対して、真となるような変数への値の割り当てが存在しないことと等価である。逆に、 $\neg P$ に対して真となるような変数への値の割り当てが1つでも存在すれば、Pを偽にするような変数への値の割り当てが存在するため、Pは恒真ではない。本実施形態では、この性質を利用して、含意関係を表す制約式が恒真であることを判定するために、含意関係を表す制約式を否定した制約式を生成し、真となるような変数への値の割り当てが存在しないことを確認する。

【0061】

40

当該確認には、SATツールを利用する。SATツールにより、命題論理式の真とするような変数の割り当てが存在するか否かを自動的に判定することができる。

【0062】

図8は、SATツールを用いた判定スクリプトを示す図である。図8(A)には、「XXX」のレコードに関する判定スクリプトの一例が示されている。図8(B)には、「YYY」のレコードに関する判定スクリプトの一例が示されている。これらのスクリプトは、代表的なSATツールの一つであるZ3のpython実装を利用して生成されている。また、判定結果が図8(A)および8(B)の一番下に、「unsat」または「sat」で表されている。

【0063】

50

図8(A)に示すように、「XXX」のレコードに関する判定結果は「unsat」であり、すなわち前記否定を取った制約式を真にするような、変数への値の割り当てが存在しないことを意味する。したがって、否定を取った含意関係を表す制約式は恒真となるため、「XXX」のレコードに関して、第1の表の規定内容は第2の表の規定内容を満たしていると言える。

【0064】

「YYY」のレコードに関する判定結果は「sat」、すなわち否定を取った制約式を真にするような、変数への値の割り当てが存在することを意味する。したがって、否定を取った含意関係を表す制約式は恒真ではないため、「YYY」のレコードに関して、第1の表の規定内容は第2の表の規定内容を満たしていると言えない。上記において、厚さが13で長さが230の場合、含意関係を表す制約式「厚さ 13 長さ 230 厚さ 15 長さ 250」は偽となることを示したが、その通りの結果が得られている。

10

【0065】

判定結果が偽であった場合、制約式充足部152は、新たな制約式を生成する。どのような制約式を生成するかは、情報処理装置1の用途に応じて、様々であってよい。例えば、第1の表の「XXX」のレコードの制約式を緩和することが考えられる。第1の表の「XXX」に対する「厚さ 13、長さ 230」という制約式は、「厚さ」は10まで、「長さ」は200まで緩和することができる。ゆえに、制約式生成部153は、「XXX」の「厚さ」が10以上で13以下の新たな制約式を生成してもよいし、「10 調整可能な厚さ 13」といった緩和可能な範囲を示してもよい。このように、情報生成部15は、第1の表の制約式が第2の表の制約式に包括される場合に、第1の表の制約式には包括されないが第2の表の制約式に包括されるための情報を生成してもよい。

20

【0066】

また、例えば、第1の表の「YYY」のレコードの制約式を厳格化することが考えられる。前述の例では、第1の表の「YYY」に対する「厚さ 13、長さ 230」という制約式は、「厚さ」を15以上にしなければ第2の表の制約式を満たすことができない。そこで、制約式生成部153は、「厚さ」が15以上の新たな制約式を作成してもよいし、「満たすべき厚さ 15」といった満たすべき範囲を示してもよい。このように、情報生成部15は、第1の表の制約式が第2の表の制約式に包括されない場合に、第2の表の制約式に包括されるための情報を算出してもよい。

30

【0067】

また、制約式生成部153は、第1の表の制約式と、第2の表の制約式と、の両方を満たす新たな制約式を生成してもよい。あるいは、両方とも満たさない制約式を生成してもよい。

【0068】

このような新たな制約式などといった情報を、情報生成部15が、入出力部11を介して、送信する(S405)。これを受けてユーザは、参照関係にある表に矛盾があるかどうかを認識し、共通属性の緩和可能な、または、満たすべき、数値または範囲を認識することができる。

40

【0069】

以上のように、本実施形態の情報処理装置1は、電子文書および当該電子文書に含まれる表の記載などに基づいて、当該表と参照関係にある表を検出する。また、参照関係にある表が示すテキスト(例えば、属性名、セルの記載)を解析し、解析結果に基づいて、参照関係にある表の共通属性を検出する。これにより、属性名称が完全に一致しない場合であっても、同一の属性であると認識することができる。そして、検出された共通属性に対するテキストに基づいて情報が生成されるため、複数の文書および表を一つ一つ参照する手間をなくすことができる。

【0070】

また、文書の参照関係、表の参照関係を検出して記憶することができる。これにより、

50

例えば、表および文書の参照関係を記した文書を別に用意して管理するといった運用の手間をなくし、管理を効率化することができる。また、例えば、ある表が示す値が更新される場合に、その他の表の値も更新する必要があるかどうかを容易に認識することができる。

【0071】

なお、上記の実施形態の少なくとも一部は、プロセッサ、メモリなどを実装しているIC(Integrated Circuit:集積回路)などの専用の電子回路(すなわちハードウェア)により実現されてもよい。また、上記の実施形態の少なくとも一部は、ソフトウェア(プログラム)を実行することにより、実現されてもよい。例えば、汎用のコンピュータ装置を基本ハードウェアとして用い、コンピュータ装置に搭載されたCPUなどのプロセッサにプログラムを実行させることにより、上記の実施形態の処理を実現することが可能である。

10

【0072】

例えば、コンピュータが読み取り可能な記憶媒体に記憶された専用のソフトウェアをコンピュータが読み出すことにより、コンピュータを上記の実施形態の装置とすることができる。記憶媒体の種類は特に限定されるものではない。また、通信ネットワークを介してダウンロードされた専用のソフトウェアをコンピュータがインストールすることにより、コンピュータを上記の実施形態の装置とすることができる。こうして、ソフトウェアによる情報処理が、ハードウェア資源を用いて、具体的に実装される。

【0073】

図9は、本発明の一実施形態におけるハードウェア構成の一例を示すブロック図である。情報処理装置1は、プロセッサ21と、主記憶装置22と、補助記憶装置23と、ネットワークインタフェース24と、デバイスインタフェース25と、を備え、これらがバス26を介して接続されたコンピュータ装置2として実現できる。記憶部16は、主記憶装置22または補助記憶装置23により実現可能であり、その他の構成要素は、プロセッサ21により実現可能である。

20

【0074】

なお、図9のコンピュータ装置2は、各構成要素を一つ備えているが、同じ構成要素を複数備えていてもよい。また、図9では、1台のコンピュータ装置2が示されているが、ソフトウェアが複数のコンピュータ装置にインストールされて、当該複数のコンピュータ装置それぞれがソフトウェアの異なる一部の処理を実行してもよい。

30

【0075】

プロセッサ21は、コンピュータの制御装置および演算装置を含む電子回路である。プロセッサ21は、コンピュータ装置2の内部構成の各装置などから入力されたデータやプログラムに基づいて演算処理を行い、演算結果や制御信号を各装置などに出力する。具体的には、プロセッサ21は、コンピュータ装置2のOS(オペレーティングシステム)や、アプリケーションなどを実行し、コンピュータ装置2を構成する各装置を制御する。プロセッサ21は、上記の処理を行うことができれば特に限られるものではない。

【0076】

主記憶装置22は、プロセッサ21が実行する命令および各種データなどを記憶する記憶装置であり、主記憶装置22に記憶された情報がプロセッサ21により直接読み出される。補助記憶装置23は、主記憶装置22以外の記憶装置である。なお、これらの記憶装置は、電子情報を格納可能な任意の電子部品を意味するものとし、メモリでもストレージでもよい。また、メモリには、揮発性メモリと、不揮発性メモリがあるが、いずれでもよい。

40

【0077】

ネットワークインタフェース24は、無線または有線により、通信ネットワーク3に接続するためのインタフェースである。ネットワークインタフェース24は、既存の通信規格に適合したものをいれればよい。ネットワークインタフェース24により、通信ネットワーク3を介して通信接続された外部装置4Aと情報のやり取りが行われてもよい。

【0078】

50

デバイスインタフェース 25 は、外部装置 4 B と直接接続する USB などのインタフェースである。外部装置 4 B は、外部記憶媒体でもよいし、データベースなどのストレージ装置でもよい。

【0079】

外部装置 4 A および 4 B は出力装置でもよい。出力装置は、例えば、画像を表示するための表示装置でもよいし、音声などを出力する装置などでもよい。例えば、LCD (Liquid Crystal Display)、CRT (Cathode Ray Tube)、PDP (Plasma Display Panel)、スピーカなどがあるが、これらに限られるものではない。

【0080】

なお、外部装置 4 A および 4 B は入力装置でもよい。入力装置は、キーボード、マウス、タッチパネルなどのデバイスを備え、これらのデバイスにより入力された情報をコンピュータ装置 2 に与える。入力装置からの信号はプロセッサ 21 に出力される。

【0081】

上記に、本発明の一実施形態を説明したが、これらの実施形態は、例として提示したものであり、発明の範囲を限定することは意図していない。これら新規な実施形態は、その他の様々な形態で実施されることが可能であり、発明の要旨を逸脱しない範囲で、種々の省略、置き換え、移行を行うことができる。これら実施形態やその変形は、発明の範囲や要旨に含まれるとともに、特許請求の範囲に記載された発明とその均等の範囲に含まれる。

【符号の説明】

【0082】

- 1 情報処理装置
- 11 入出力部
- 12 解析部
  - 121 文書構造解析部
  - 122 テキスト解析部
  - 123 テーブル構造解析部
- 13 参照関係検出部
  - 131 参照テーブル検出部
  - 132 共通属性検出部
- 14 検索部
  - 141 指定情報抽出部
  - 142 参照情報抽出部
- 15 情報生成部
  - 151 制約式変換部
  - 152 制約式充足部
  - 153 制約式生成部
- 16 記憶部
  - 161 テキスト情報
  - 162 テーブル情報
  - 163 参照関係情報
  - 164 共通属性情報
- 2 コンピュータ装置
  - 21 プロセッサ
  - 22 主記憶装置
  - 23 補助記憶装置
  - 24 ネットワークインタフェース
  - 25 デバイスインタフェース
  - 26 バス
- 3 通信ネットワーク

10

20

30

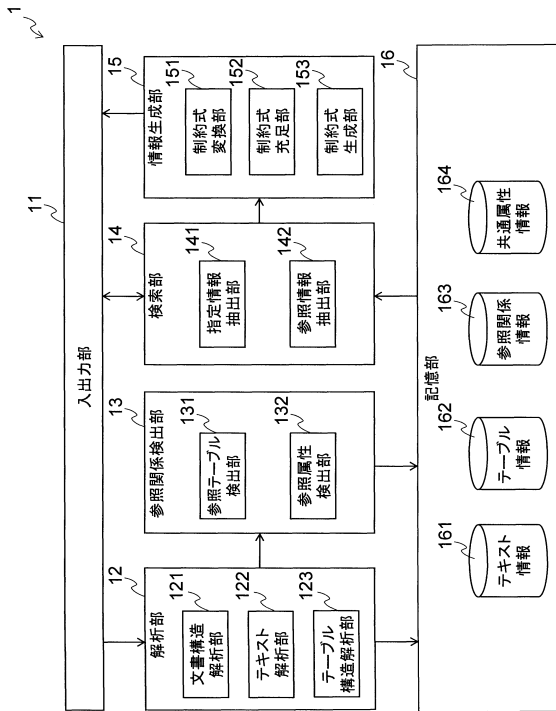
40

50

### 4 A および 4 B 外部装置

【図面】

【図 1】



【図 2】

第1の表

種類記号	厚さ(mm)	長さ(mm)
XXX	13 以上	180 以下
YYY	13 以上	230 以下

参照関係  
(第1の表が第2の表を参照)

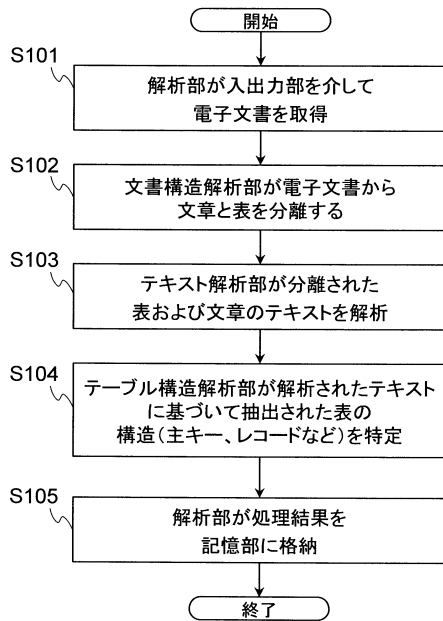
第2の表

材料記号	鋼板の厚さ	鋼板の長さ
XXX	10mm 以上	200mm 以下
YYY	15mm 以上	250mm 以下

10

20

【図 3】



【図 4】

第3の表

種類の記号	降伏点又は耐力 N/mm <sup>2</sup>		伸び	
	厚さ mm	厚さ mm	試験片	%
SM400A SM400B SM400C	16 以下	245 以上	5 以下	23 以上
	16 を超え 40 以下	235 以上	5 以下	18 以上
	40 を超え 100 以下	215 以上	5 以下	22 以上
SM400A SM400B SM400C	100 以下	205 以上	5 以下	18 以上
	100 を超え 160 以下	195 以上	5 以下	22 以上
	160 以下	205 以上	5 以下	24 以上
SM400A SM400B SM400C	100 を超え 160 以下	160 を超え 200 以下	5 以下	23 以上
	160 以下	200 以下	5 以下	18 以上
	160 を超え 200 以下	200 以下	5 以下	22 以上
SM400A SM400B SM400C	16 以下	400~510	5 以下	23 以上
	16 を超え 40 以下	400~510	5 以下	18 以上
	40 を超え 100 以下	400~510	5 以下	22 以上

第4の表

材料記号	降伏点又は耐力 N/mm <sup>2</sup>		伸び	
	鉄板の厚さ mm	鉄板の厚さ mm	試験片	%
SM400A SM400B SM400C	16 以下	205 以上	5 以下	23 以上
	16 を超え 40 以下	195 以上	5 以下	18 以上
	40 を超え 100 以下	205 以上	5 以下	22 以上
SM400A SM400B SM400C	100 以下	160 を超え 200 以下	5 以下	18 以上
	160 以下	200 以下	5 以下	22 以上
	160 を超え 200 以下	200 以下	5 以下	24 以上
SM400A SM400B SM400C	16 以下	400~510	5 以下	23 以上
	16 を超え 40 以下	400~510	5 以下	18 以上
	40 を超え 100 以下	400~510	5 以下	22 以上

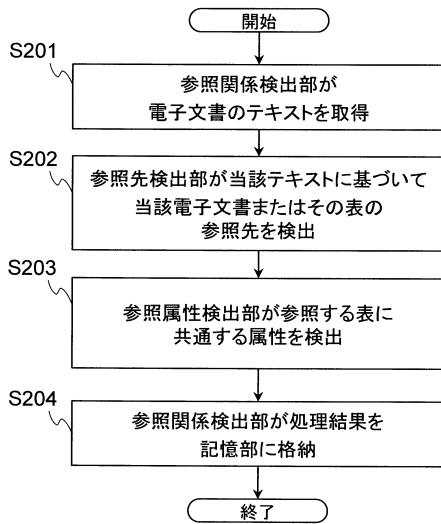
参照関係(第4の表が第3の表を参照)

30

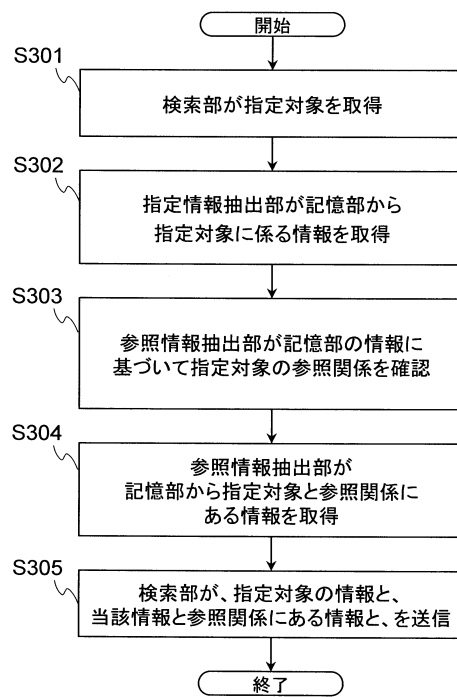
40

50

【図 5】



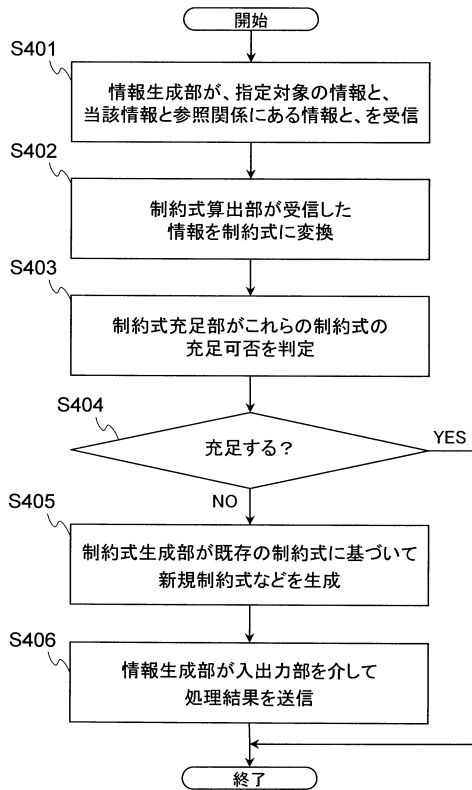
【図 6】



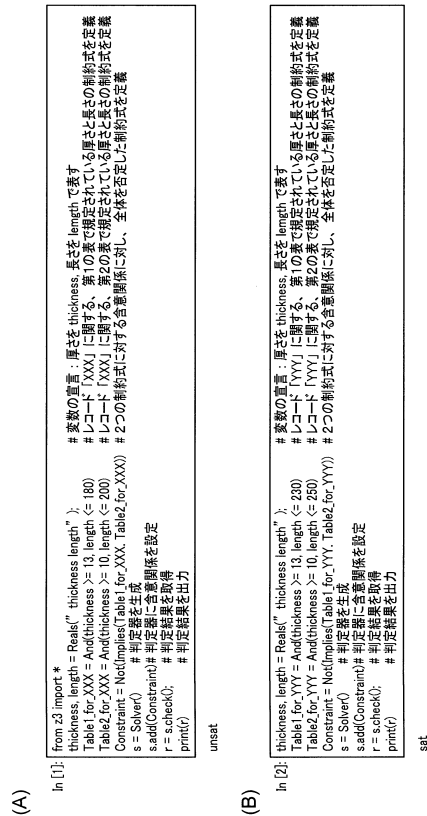
10

20

【図 7】



【図 8】

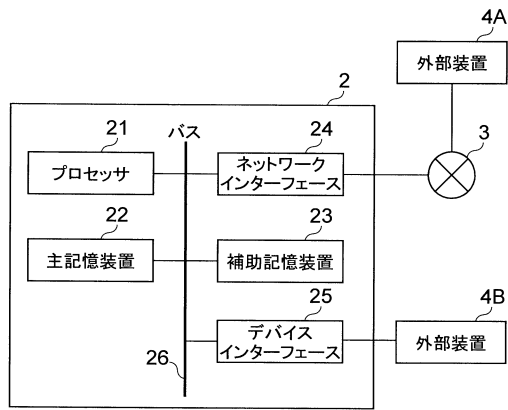


30

40

50

【図 9】



10

20

30

40

50



## フロントページの続き

- 社東芝内  
(72)発明者 張 瑞剛  
東京都港区芝浦一丁目1番1号 株式会社東芝内
- (72)発明者 岩政 幹人  
東京都港区芝浦一丁目1番1号 株式会社東芝内
- 審査官 木村 大吾
- (56)参考文献 特開2006-099236(JP,A)  
特開2004-086782(JP,A)  
米国特許第10248736(US,B1)  
米国特許第09715487(US,B2)
- (58)調査した分野 (Int.Cl., DB名)  
G06F 40/00 - 40/58  
G06F 16/00 - 16/958