

(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
2017 年 2 月 2 日(02.02.2017)



(10) 国際公開番号

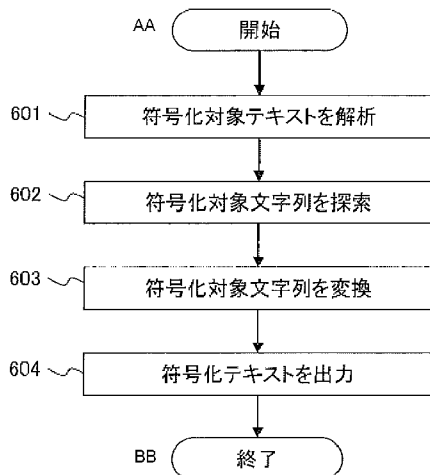
WO 2017/017738 A1

- (51) 国際特許分類:
G06F 17/22 (2006.01)
- (21) 国際出願番号: PCT/JP2015/071171
- (22) 国際出願日: 2015 年 7 月 24 日(24.07.2015)
- (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: 富士通株式会社 (FUJITSU LIMITED) [JP/JP]; 〒2118588 神奈川県川崎市中原区上小田中 4 丁目 1 番 1 号 Kanagawa (JP).
- (72) 発明者: 中村文昭 (NAKAMURA, Fumiaki); 〒2118588 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 Kanagawa (JP). 片岡正弘 (KATAOKA, Masahiro); 〒2118588 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 Kanagawa (JP). 出内将夫 (IDEUCHI, Masao); 〒2118588 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 Kanagawa (JP).
- (74) 代理人: 大菅義之 (OSUGA, Yoshiyuki); 〒1020084 東京都千代田区二番町 8 番地 2 O 二番町ビル 3 F 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, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, 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, SA, 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.
- (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, ST, 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),

[続葉有]

(54) Title: ENCODING PROGRAM, ENCODING DEVICE, AND ENCODING METHOD

(54) 発明の名称: 符号化プログラム、符号化装置、及び符号化方法



601 Analyze text to be encoded
602 Search for character string to be encoded
603 Transform character string to be encoded
604 Output encoded text
AA Start
BB End

(57) Abstract: A computer analyzes a text to be encoded on the basis of morpheme segmentation information which is used in the segmentation of morphemes (step 601). In each of a plurality of records, the morpheme segmentation information includes a header morpheme of a registered character string of each record which is included in character string encoding information, said character string encoding information including a plurality of records in which registered character strings are associated with encodings. Next, the computer searches in the character string encoding information for a character string to be encoded which begins with a given morpheme which is included in the character string to be encoded, according to the identification of the given morpheme which is included in the record of the morpheme segmentation information which corresponds to a given registered character string (step 602). If the character string to be encoded corresponds to the given registered character string, the computer transforms the character string to be encoded into an encoding which is associated in the character string encoding information with the given registered character string (step 603), and outputs an encoded text which includes the encoding which results after the transformation (step 604).

(57) 要約:

[続葉有]

WO 2017/017738 A1



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

添付公開書類:

— 国際調査報告 (条約第 21 条(3))

コンピュータは、形態素の区分に用いられる形態素区分情報に基づいて、符号化対象テキストを解析する (ステップ 601)。形態素区分情報は、登録文字列と符号とを対応付けたレコードを複数有する文字列符号化情報に含まれる各レコードの登録文字列の先頭形態素を複数のレコード各々に含む。次に、コンピュータは、ある登録文字列に対応した形態素区分情報のレコードに含まれるある形態素の識別に応じて、符号化対象テキストに含まれるある形態素から始まる符号化対象文字列を文字列符号化情報において探索する (ステップ 602)。そして、コンピュータは、符号化対象文字列がある登録文字列に対応する場合、符号化対象文字列を文字列符号化情報においてある登録文字列に対応付けられた符号に変換し (ステップ 603)、変換後の符号を含む符号化テキストを出力する (ステップ 604)。

明 細 書

発明の名称： 符号化プログラム、符号化装置、及び符号化方法
技術分野

[0001] 本発明は、符号化プログラム、符号化装置、及び符号化方法に関する。

背景技術

[0002] 形態素解析は、テキストを形態素に分割し、各形態素に対して品詞情報を付与する処理である。形態素解析により得られる形態素は、単語として扱われることもある。

[0003] 文書情報から抽出した形態素を符号化して格納する文書情報格納装置も知られている（例えば、特許文献1を参照）。この文書情報格納装置は、形態素解析処理を施すことにより、入力された文書情報から文書情報構成要素としての形態素を抽出し、抽出された形態素を符号化し、符号化された形態素に圧縮処理を施し、圧縮された符号化形態素を記憶部に格納する。

先行技術文献

特許文献

[0004] 特許文献1：特開平11-85790号公報

発明の概要

発明が解決しようとする課題

[0005] 上述した従来の符号化処理では、圧縮率を向上させるために2種類の辞書を用いて2段階の符号化処理が行われる場合、処理時間が長くなることがある。

[0006] 1つの側面において、本発明は、符号化対象テキストに対する2段階の符号化処理を効率化することを目的とする。

課題を解決するための手段

[0007] 1つの案では、符号化プログラムは、以下の処理をコンピュータに実行させる。

(1) コンピュータは、形態素の区分に用いられ、登録文字列と符号とを対

応付けたレコードを複数有する文字列符号化情報に含まれる各レコードの登録文字列の先頭形態素を複数のレコード各々に含む形態素区分情報に基づいて、符号化対象テキストを解析する。

(2) コンピュータは、形態素区分情報に含まれる複数のレコードのうち、ある登録文字列に対応したレコードに含まれる、ある形態素の識別に応じて、符号化対象テキストに含まれるある形態素から始まる符号化対象文字列を文字列符号化情報において探索する。

(3) コンピュータは、符号化対象文字列がある登録文字列に対応する場合、符号化対象文字列を文字列符号化情報においてある登録文字列に対応付けられた符号に変換する。

(4) コンピュータは、変換後の符号を含む符号化テキストを出力する。

発明の効果

[0008] 実施形態によれば、符号化対象テキストに対する2段階の符号化処理を効率化することができる。

図面の簡単な説明

[0009] [図1]形態素解析辞書を示す図である。

[図2]郵便番号辞書を示す図である。

[図3]形態素解析辞書及び郵便番号辞書を用いた符号化処理を示す図である。

[図4]郵便番号辞書の符号を追加した形態素解析辞書を示す図である。

[図5]符号化装置の機能的構成図である。

[図6]符号化処理のフローチャートである。

[図7]符号化装置の具体例を示す機能的構成図である。

[図8]切り替え情報を含む形態素解析辞書を示す図である。

[図9]住所辞書を示す図である。

[図10]切り替え情報設定処理のフローチャートである。

[図11]符号化処理の具体例を示すフローチャートである。

[図12]登録されていない文字列の探索を省略する符号化装置の機能的構成図である。

[図13]都道府県を複数のブロックに分割した住所辞書を示す図である。

[図14]形態素位置情報を示す図である。

[図15]登録されていない文字列の探索を省略する符号化処理のフローチャートである。

[図16]ブロック絞り込み処理のフローチャートである。

[図17]作業領域に格納されるフィルタ情報を示す図である。

[図18]エラーメッセージ辞書を示す図である。

[図19]情報処理装置の構成図である。

発明を実施するための形態

[0010] 以下、図面を参照しながら、実施形態を詳細に説明する。

特許文献1の文書情報格納装置は、文書から形態素解析により抽出した1つ以上の形態素に対して、郵便番号辞書、同義語辞書等に登録された符号を割り当てることで、文書を符号化する。

[0011] 図1は、形態素解析で用いられる形態素解析辞書の例を示している。図1の形態素解析辞書は、形態素、品詞、及び符号を含む。品詞は、各形態素の品詞を表し、符号は、各形態素に対応付けられた符号を表す。

[0012] 図2は、郵便番号辞書の例を示している。図2の郵便番号辞書は、文字列及び符号を含む。文字列は、郵便番号が付与された住所を表す複数の形態素を含み、符号は、各文字列に対応付けられた郵便番号を表す。

[0013] 図3は、図1の形態素解析辞書及び図2の郵便番号辞書を用いた符号化処理の例を示している。「神奈川県川崎市中原区上小田中に住む」という符号化対象テキスト301が入力された場合、まず、第1パスの形態素解析により符号化対象テキスト301が複数の形態素に分割される。そして、それぞれの形態素が形態素解析辞書に登録された符号に変換され、符号化テキスト302が生成される。例えば、符号化対象テキスト301の先頭の形態素「神奈川県」は、16進数の符号0x0105に変換されている。

[0014] 次に、第2パスの符号化により符号化テキスト302に含まれる文字列が郵便番号辞書に登録された文字列と比較され、2つの文字列が一致する場合

、その文字列が郵便番号辞書に登録された符号に変換されて、符号化テキスト303が生成される。この例では、「神奈川県川崎市中原区上小田中」という文字列が符号211-0053に変換されている。このように、第2パスの符号化を行うことで、住所を表す複数の形態素が1つの符号に置き換えられるため、第1パスの符号化結果をさらに圧縮することができる。

[0015] 第1パスの形態素解析では、品詞又は意味を持つ最小の単位である形態素を高速に抽出するために形態素解析辞書が採用される。形態素解析辞書では、1つの形態素に対して1つの符号が割り当てられる。一方、第2パスの符号化では、住所を表す文字列を高速に抽出するために郵便番号辞書が採用される。郵便番号辞書では、複数の形態素に対して1つの符号が割り当てられる。

[0016] このように、2つの辞書の間で符号が割り当てられる形態素の個数が異なるため、第1パス及び第2パスの2パス（2段階）に分けて符号化処理が行われる。2パスの符号化処理は、1パスの符号化処理よりも時間がかかるとともに、記憶領域等の計算資源をより多く使用する。

[0017] 図4は、2パスの符号化処理と同等の処理を1パスで行うために、郵便番号辞書の符号を追加した形態素解析辞書の例を示している。図4の形態素解析辞書では、「神奈川県川崎市中原区上小田中」という文字列が形態素として登録されており、符号211-0053と対応付けられている。

[0018] しかし、テキスト中における住所を表す文字列の出現頻度は、その住所に含まれる個々の地名を表す形態素の出現頻度よりも低いため、形態素解析において統計的出現頻度の高い順に形態素を抽出する場合、住所を表す文字列が抽出される可能性は極めて低い。このため、図4の形態素解析辞書を用いたとしても、符号化対象テキスト301から符号化テキスト303が生成されることはなく、符号化テキスト302が生成される。結果的に、郵便番号辞書の符号が使用されず、圧縮率が向上しない。

[0019] また、第2パスの符号化において、住所を表す文字列に対して郵便番号辞書の符号を割り当てるために、形態素解析結果のすべての形態素に対して、

完全最長一致法により郵便番号辞書が探索される。この場合、形態素の個数と同じ回数だけ、郵便番号辞書を参照して完全最長一致検索が行われるため、膨大な時間がかかる。

[0020] なお、かかる問題は、第２パスの符号化において郵便番号辞書を用いる場合に限らず、住所以外の他の文字列を登録した辞書を用いる場合においても生ずるものである。

[0021] 図５は、実施形態の符号化装置の機能的構成例を示している。図５の符号化装置５０１は、記憶部５１１、符号化部５１２、及び出力部５１３を含む。

[0022] 記憶部５１１は、形態素区分情報５２１及び文字列符号化情報５２２を記憶する。文字列符号化情報５２２は、複数のレコードを有し、各レコードには、登録文字列とその登録文字列に対応付けられた符号とが登録される。形態素区分情報５２１は、複数の形態素それぞれを含む複数のレコードを有し、形態素の区分に用いられる。形態素区分情報５２１には、文字列符号化情報５２２に含まれる各レコードの登録文字列の先頭形態素を含むレコードが含まれる。

[0023] 符号化部５１２は、形態素区分情報５２１及び文字列符号化情報５２２に基づいて符号化対象テキストを符号化して、符号化テキストを生成し、出力部５１３は、符号化テキストを出力する。

[0024] 図６は、図５の符号化装置５０１が行う符号化処理の例を示すフローチャートである。まず、符号化部５１２は、形態素区分情報５２１に基づいて符号化対象テキストを解析する（ステップ６０１）。そして、符号化部５１２は、形態素区分情報５２１に含まれる複数のレコードのうち、ある登録文字列に対応したレコードに含まれる、ある形態素の識別に応じて、符号化対象テキストに含まれるその形態素から始まる符号化対象文字列を文字列符号化情報内で探索する（ステップ６０２）。

[0025] 符号化対象文字列がある登録文字列に対応する場合、符号化部５１２は、符号化対象文字列をその登録文字列に対応付けられた符号に変換する（ステ

ップ603)。そして、出力部513は、変換後の符号を含む符号化テキストを出力する(ステップ604)。

[0026] このような符号化装置501によれば、符号化対象テキストに対する2段階の符号化処理を効率化することができる。

[0027] 図7は、図5の符号化装置501の具体例を示している。図7の符号化部512は、解析部711、判定部712、変換部713、及び変換部714を含み、記憶部511は、形態素解析辞書721及び住所辞書722を記憶する。形態素解析辞書721及び住所辞書722は、図5の形態素区分情報521及び文字列符号化情報522にそれぞれ対応する。

[0028] 図8は、形態素解析辞書721の例を示している。図8の形態素解析辞書721は、形態素、品詞、符号、及びフラグを含む。品詞は、各形態素の品詞を表し、符号は、各形態素に対応付けられた符号を表す。フラグは、形態素解析辞書721から住所辞書722への切り替えを行うか否かを示す切り替え情報であり、フラグが論理“1”のとき、切り替えを行うことを示し、フラグが論理“0”のとき、切り替えを行わないことを示す。

[0029] この例では、都道府県を表す文字列の先頭の形態素に対応するフラグが論理“1”に設定されている。例えば、「東京」の品詞は名詞であり、符号はC1であり、フラグは論理“1”である。一方、「川崎」の品詞は名詞であり、符号はC4であり、フラグは論理“0”である。

[0030] 図9は、住所辞書722の例を示している。図9の住所辞書722は、文字列及び符号を含む。文字列は、住所を表す複数の形態素を含み、符号は、各文字列に対応付けられた符号を表す。符号は、郵便番号であってもよい。住所辞書722に登録された複数の文字列は、都道府県に対応する複数のブロックに分割されている。セパレータ901は、「東京都」のブロックの開始位置を表し、セパレータ902は、「神奈川県」のブロックの開始位置を表す。例えば、「東京都杉並区」の符号はD1であり、「神奈川県川崎市中原区」の符号はD3である。

[0031] 解析部711は、形態素解析辞書721に切り替え情報を設定するととも

に、形態素解析辞書 7 2 1 に基づいて符号化対象テキストに対する形態素解析を行い、符号化対象テキストを複数の形態素に分割する。判定部 7 1 2 は、解析部 7 1 1 が生成した形態素毎に住所辞書 7 2 2 への切り替えを行うか否かを判定する。変換部 7 1 3 は、形態素を形態素解析辞書 7 2 1 の符号に変換し、変換部 7 1 4 は、住所を表す複数の形態素を住所辞書 7 2 2 の符号に変換する。

[0032] 図 1 0 は、解析部 7 1 1 が行う切り替え情報設定処理の例を示すフローチャートである。切り替え情報設定処理の開始時には、形態素解析辞書 7 2 1 内のすべての形態素のフラグが論理 “0” に設定されている。まず、解析部 7 1 1 は、形態素解析辞書 7 2 1 に基づいて、住所辞書 7 2 2 に登録されている各文字列に対する形態素解析を行い、各文字列を複数の形態素に分割する（ステップ 1 0 0 1）。そして、解析部 7 1 1 は、形態素解析辞書 7 2 1 内の形態素の中から各文字列の先頭の形態素を選択し、選択した形態素のフラグを論理 “1” に変更することで切り替え情報を設定する（ステップ 1 0 0 2）。

[0033] 解析部 7 1 1 は、フラグの代わりに住所辞書 7 2 2 のブロックの開始位置を指すポインタを、切り替え情報として設定してもよい。この場合、「東京」の切り替え情報としては、セパレータ 9 0 1 の位置を指すポインタが設定され、「神奈川」の切り替え情報としては、セパレータ 9 0 2 の位置を指すポインタが設定される。

[0034] 図 1 1 は、図 7 の符号化装置 5 0 1 が行う符号化処理の具体例を示すフローチャートである。まず、解析部 7 1 1 は、形態素解析辞書 7 2 1 に基づいて符号化対象テキストに対する形態素解析を行い（ステップ 1 1 0 1）、符号化対象テキストの先頭から順に形態素を抽出する（ステップ 1 1 0 2）。次に、判定部 7 1 2 は、形態素解析辞書 7 2 1 において、解析部 7 1 1 が抽出した形態素の切り替え情報が切り替えを示しているか否かを判定する（ステップ 1 1 0 3）。

[0035] 切り替え情報が切り替えを示している場合（ステップ 1 1 0 3, YES）

、判定部 7 1 2 は、符号化対象テキストに含まれる、抽出された形態素から始まる符号化対象文字列（形態素群）を、住所辞書 7 2 2 内で探索する（ステップ 1 1 0 4）。このとき、判定部 7 1 2 は、切り替え情報が示すブロック内で符号化対象文字列を探索し、符号化対象文字列がそのブロックに登録されているか否かをチェックする。判定部 7 1 2 は、例えば、完全最長一致検索により符号化対象文字列とブロック内の各登録文字列とを比較することで、符号化対象文字列がそのブロックに登録されているか否かをチェックすることができる。

[0036] 符号化対象文字列が住所辞書 7 2 2 に登録されている場合（ステップ 1 1 0 4, Y E S）、変換部 7 1 4 は、その符号化対象文字列を住所辞書 7 2 2 に登録された符号に変換する（ステップ 1 1 0 5）。そして、判定部 7 1 2 は、符号化対象テキストの最後の形態素が符号に変換されたか否かをチェックする（ステップ 1 1 0 6）。最後の形態素が符号に変換されていない場合（ステップ 1 1 0 6, N O）、判定部 7 1 2 は、残りのテキストに対して、ステップ 1 1 0 1 以降の処理を繰り返す。

[0037] 一方、切り替え情報が切り替えを示していない場合（ステップ 1 1 0 3, N O）、変換部 7 1 3 は、解析部 7 1 1 が抽出した形態素を、形態素解析辞書 7 2 1 に登録された符号に変換する（ステップ 1 1 0 8）。そして、符号化部 5 1 2 は、ステップ 1 1 0 6 以降の処理を行う。また、符号化対象文字列が住所辞書 7 2 2 に登録されていない場合（ステップ 1 1 0 4, N O）、符号化部 5 1 2 は、ステップ 1 1 0 8 以降の処理を行う。

[0038] そして、最後の形態素が符号に変換された場合（ステップ 1 1 0 6, Y E S）、出力部 5 1 3 は、符号列を含む符号化テキストを出力する（ステップ 1 1 0 7）。

[0039] 例えば、「神奈川県川崎市中原区に住む」という符号化対象テキストは、「神奈川／県／川崎／市／中原／区／に／住む」のように分割される。図 8 の形態素解析辞書 7 2 1 では、「神奈川」のフラグ“1”が切り替えを示しているため、「神奈川」から始まる文字列が図 9 の住所辞書 7 2 2 の「神奈

川」のブロック内で探索され、「神奈川県川崎市中原区」という登録文字列が検出される。そこで、「神奈川／県／川崎／市／中原／区」が対応する符号D3に変換され、後続する「に」及び「住む」は形態素解析辞書721の符号に変換される。

[0040] 形態素解析辞書721においてフラグの代わりにポインタが用いられる場合、形態素に対応するポインタが存在すれば、切り替え情報が切り替えを示していると判定され、ポインタが存在しなければ、切り替え情報が切り替えを示していないと判定される。

[0041] このような符号化処理によれば、住所辞書722の登録文字列が出現した時点で形態素解析辞書721から住所辞書722へ切り替えられるため、2パスの符号化処理を1パスで行うことが可能になる。これにより、2段階の符号化処理が高速化されるとともに、符号化処理に使用する記憶領域等の計算資源が削減される。

[0042] また、複数の登録文字列を先頭の形態素に対応する複数のブロックに分割して住所辞書722に登録することで、符号化対象文字列の先頭の形態素に対応するブロックのみを探索対象にすることが可能になる。これにより、探索対象ブロックが限定され、探索処理が効率化される。

[0043] しかし、図11の符号化処理では、切り替えを示す切り替え情報が検出される度に、符号化対象テキストに含まれる符号化対象文字列と、住所辞書722の対応するブロック内のすべての登録文字列とが、完全最長一致検索により比較される。このため、1ブロックに多数の登録文字列が含まれている場合、ステップ1104の探索処理に長い時間がかかる。

[0044] 例えば、「神奈川に行く」という符号化対象テキストは、「神奈川／に／行く」のように分割される。図8の形態素解析辞書721では、「神奈川」のフラグ“1”が切り替えを示しているため、「神奈川」から始まる文字列が住所辞書722内で探索されるが、「に」は住所辞書722内のいずれの登録文字列にも含まれていない。このため、「神奈川」は形態素解析辞書721の符号に変換される。

- [0045] このように、符号化対象テキスト内の形態素が住所辞書 7 2 2 の登録文字列の先頭の形態素と一致しているが、その形態素から始まる符号化対象文字列が住所辞書 7 2 2 内のいずれの登録文字列とも一致しない場合は、探索時間が無駄になってしまう。そこで、住所辞書 7 2 2 に登録されていない文字列の探索を省略して、符号化処理をさらに高速化することが望ましい。
- [0046] 図 1 2 は、住所辞書 7 2 2 に登録されていない文字列の探索を省略する符号化装置 5 0 1 の具体例を示している。図 1 2 の符号化装置 5 0 1 の構成は、図 7 の符号化装置 5 0 1 と同様である。記憶部 5 1 1 は、形態素解析辞書 7 2 1 及び住所辞書 7 2 2 に加えて形態素位置情報 1 2 1 1 を記憶し、作業領域 1 2 1 2 を含む。形態素位置情報 1 2 1 1 は、住所辞書 7 2 2 の各登録文字列に含まれる各形態素について、登録文字列内における形態素の位置を示す情報である。作業領域 1 2 1 2 は、形態素位置情報 1 2 1 1 を用いて文字列が住所辞書 7 2 2 に登録されているか否かを判定するために用いられ、スタックであってもよい。
- [0047] 図 1 3 は、各都道府県の住所を複数のブロックに分割した住所辞書 7 2 2 の例を示している。「東京 1」及び「東京 2」は、「東京都」の住所のブロック名を表し、「神奈川 1」～「神奈川 5」は、「神奈川県」の住所のブロック名を表す。
- [0048] 図 1 4 は、図 1 3 の住所辞書 7 2 2 に対する形態素位置情報 1 2 1 1 の例を示している。図 1 4 の形態素位置情報 1 2 1 1 は、形態素、オフセット、及び都道府県のブロックに対するフィルタ情報を含む。形態素は、住所辞書 7 2 2 の各登録文字列に含まれる各形態素を表し、オフセットは、登録文字列内における形態素の位置を表す。例えば、「東京」とオフセット“1”との組み合わせは、登録文字列の先頭の形態素が「東京」であることを表し、「多摩」とオフセット“3”との組み合わせは、登録文字列の 3 番目の形態素が「多摩」であることを表す。
- [0049] フィルタ情報は、ブロックの個数だけのビット値を含むビット列で表される。あるブロックのビット値が論理“1”のとき、そのブロックに含まれる

いずれかの登録文字列内において、オフセットが示す位置に形態素が存在することを示す。一方、あるブロックのビット値が論理“0”のとき、そのブロックに含まれるすべての登録文字列内において、オフセットが示す位置には形態素が存在しないことを表す。

[0050] 例えば、「東京」とオフセット“1”との組み合わせに対応する「東京1」のビット値“1”は、「東京1」のブロックには、先頭の形態素が「東京」である登録文字列が存在することを示す。また、「東京」とオフセット“1”との組み合わせに対応する「神奈川1」のビット値“0”は、「神奈川1」のブロックには、先頭の形態素が「東京」である登録文字列が存在しないことを示す。

[0051] 「多摩」とオフセット“3”との組み合わせに対応する「東京1」のビット値“1”は、「東京1」のブロックには、3番目の形態素が「多摩」である登録文字列が存在することを示す。また、「多摩」とオフセット“5”との組み合わせに対応する「神奈川3」のビット値“1”は、「神奈川3」のブロックには、5番目の形態素が「多摩」である登録文字列が存在することを示す。

[0052] このような形態素位置情報1211を用いることで、符号化対象文字列と登録文字列とを比較しなくても、その符号化対象文字列が住所辞書722に登録されているか否かを判定することが可能になる。さらに、符号化対象文字列が住所辞書722に登録されている場合、同じ先頭の形態素に対応する複数のブロックの中から、その符号化対象文字列を含むブロックを絞り込むことも可能になる。

[0053] 図15は、図12の符号化装置501が行う符号化処理の具体例を示すフローチャートである。ステップ1501～ステップ1503及びステップ1506～ステップ1509の処理は、図11のステップ1101～ステップ1103及びステップ1105～ステップ1108の処理と同様である。

[0054] 切り替え情報が切り替えを示している場合（ステップ1503，YES）、判定部712は、形態素位置情報1211を用いてブロックを絞り込むこ

とで、住所辞書 7 2 2 内の探索対象ブロックを特定する（ステップ 1 5 0 4）。そして、判定部 7 1 2 は、符号化対象文字列を探索対象ブロック内で探索する（ステップ 1 5 0 5）。探索対象ブロックが存在する場合は、符号化対象文字列が探索対象ブロックに登録されているため（ステップ 1 5 0 5, YES）、符号化部 5 1 2 は、ステップ 1 5 0 6 以降の処理を行う。

[0055] 一方、探索対象ブロックが存在しない場合は、符号化対象文字列が住所辞書 7 2 2 に登録されていないため（ステップ 1 5 0 5, NO）、符号化部 5 1 2 は、ステップ 1 5 0 9 以降の処理を行う。この場合、住所辞書 7 2 2 から符号化対象文字列を探索する処理は行われない。

[0056] 図 1 6 は、図 1 5 のステップ 1 5 0 4 で行われるブロック絞り込み処理の例を示すフローチャートである。まず、判定部 7 1 2 は、オフセットを表す変数 X に 1 を設定し（ステップ 1 6 0 1）、形態素を表す変数 W にステップ 1 5 0 2 で抽出された形態素を設定する（ステップ 1 6 0 2）。

[0057] 次に、判定部 7 1 2 は、形態素位置情報 1 2 1 1 から X 及び W に対応するフィルタ情報のビット列を検索し（ステップ 1 6 0 3）、対応するフィルタ情報が存在するか否かをチェックする（ステップ 1 6 0 4）。対応するフィルタ情報が存在する場合（ステップ 1 6 0 4, YES）、判定部 7 1 2 は、そのフィルタ情報を作業領域 1 2 1 2 に追加し（ステップ 1 6 0 7）、X を 1 だけインクリメントする（ステップ 1 6 0 8）。そして、判定部 7 1 2 は、符号化対象テキスト内の次の形態素を W に設定し（ステップ 1 6 0 2）、ステップ 1 6 0 3 以降の処理を繰り返す。

[0058] 一方、対応するフィルタ情報が存在しない場合（ステップ 1 6 0 4, NO）、判定部 7 1 2 は、作業領域 1 2 1 2 に格納されたフィルタ情報を参照して、すべての形態素に対してビット値 “1” を含むブロックを特定する（ステップ 1 6 0 5）。そして、判定部 7 1 2 は、特定したブロックの個数をチェックする（ステップ 1 6 0 6）。

[0059] 特定したブロックの個数が 1 個である場合（ステップ 1 6 0 6, YES）、判定部 7 1 2 は、そのブロックを探索対象ブロックに決定する（ステップ

1607)。一方、特定したブロックの個数が2個以上である場合（ステップ1606，NO）、判定部712は、符号化対象文字列が住所辞書722に登録されていないと判定して、処理を終了する。

[0060] 図17は、作業領域1212に格納されるフィルタ情報の例を示している。例えば、「神奈川県川崎市多摩区登戸に住む」という符号化対象テキストは、「神奈川／県／川崎／市／多摩／区／登戸／に／住む」のように分割される。

[0061] 図8の形態素解析辞書721では、先頭の「神奈川」のフラグ“1”が切り替えを示しているため、図14の形態素位置情報1211から、「神奈川」とオフセット“1”との組み合わせに対応するフィルタ情報が検索される。そして、対応するフィルタ情報のビット列は、「神奈川1」～「神奈川5」のブロックにビット値“1”を含んでいるため、このビット列が作業領域1212に追加される。

[0062] 同様にして、以下のような形態素及びオフセットの組み合わせに対応するフィルタ情報が形態素位置情報1211から検索されて、作業領域1212に追加される。

「県」及び“2”

「川崎」及び“3”

「市」及び“4”

「多摩」及び“5”

「区」及び“6”

「登戸」及び“7”

[0063] しかし、次の形態素「に」とオフセット“8”との組み合わせに対応するフィルタ情報は形態素位置情報1211に存在しないため、図17の7個のフィルタ情報を参照して、すべての形態素に対してビット値“1”を含むブロックが特定される。このとき、判定部712は、図17の各ブロックに対応する7個のビット値を含む縦方向のビット列と、7個のビット値“1”を含むビット列との論理積を求めることで、すべての形態素に対してビット値

“1”を含むブロックを特定してもよい。

[0064] この場合、すべての形態素に対してビット値“1”を含むブロックは、「神奈川3」のブロックのみであるため、そのブロックが探索対象ブロックに決定される。次に、図13の住所辞書722における「神奈川3」のブロック内で、「神奈川／県／川崎／市／多摩／区／登戸」という符号化対象文字列が探索される。そして、「神奈川／県／川崎／市／多摩／区／登戸」が対応する符号に変換され、後続する「に」及び「住む」は形態素解析辞書721の符号に変換される。

[0065] また、「東京都多摩市一ノ宮に住む」という符号化対象テキストは、「東京／都／多摩／市／一ノ宮／に／住む」のように分割される。図8の形態素解析辞書721では、先頭の「東京」のフラグ“1”が切り替えを示しているため、図14の形態素位置情報1211から、「東京」とオフセット“1”との組み合わせに対応するフィルタ情報が検索される。そして、対応するフィルタ情報のビット列は、「東京1」及び「東京2」のブロックにビット値“1”を含んでいるため、このビット列が作業領域1212に追加される。

[0066] 同様にして、以下のような形態素及びオフセットの組み合わせに対応するフィルタ情報が形態素位置情報1211から検索されて、作業領域1212に追加される。

「都」及び“2”

「多摩」及び“3”

「市」及び“4”

「一ノ宮」及び“5”

[0067] しかし、次の形態素「に」とオフセット“6”との組み合わせに対応するフィルタ情報は形態素位置情報1211に存在しないため、5個のフィルタ情報を参照して、すべての形態素に対してビット値“1”を含むブロックが特定される。そして、該当するブロックが「東京1」のブロックのみである場合、そのブロックが探索対象ブロックに決定される。

- [0068] 次に、図 13 の住所辞書 722 における「東京 1」のブロック内で、「東京／都／多摩／市／一ノ宮」という符号化対象文字列が探索される。そして、「東京／都／多摩／市／一ノ宮」が対応する符号に変換され、後続する「に」及び「住む」は形態素解析辞書 721 の符号に変換される。
- [0069] 一方、「神奈川に行く」という符号化対象テキストは、「神奈川／に／行く」のように分割される。図 8 の形態素解析辞書 721 では、先頭の「神奈川」のフラグ“1”が切り替えを示しているため、図 14 の形態素位置情報 1211 から、「神奈川」とオフセット“1”との組み合わせに対応するフィルタ情報が検索される。そして、対応するフィルタ情報のビット列は、「神奈川 1」～「神奈川 5」のブロックにビット値“1”を含んでいるため、このビット列が作業領域 1212 に追加される。
- [0070] しかし、次の形態素「に」とオフセット“2”との組み合わせに対応するフィルタ情報は形態素位置情報 1211 に存在しないため、「神奈川」のフィルタ情報のみを参照してビット値“1”を含むブロックが特定される。この場合、「神奈川 1」～「神奈川 5」の 5 個のブロックが該当するため、「神奈川」から始まる文字列が住所辞書 722 に登録されていないと判定される。したがって、住所辞書 722 内で「神奈川」が探索されることはなく、「神奈川」は形態素解析辞書 721 の符号に変換される。
- [0071] このように、図 12 の符号化装置 501 によれば、符号化対象テキスト内の形態素が住所辞書 722 の登録文字列の先頭の形態素と一致しているが、その形態素から始まる符号化対象文字列がいずれの登録文字列とも一致しない場合は、住所辞書 722 の探索が省略される。これにより、不要な探索処理が削減されるため、符号化処理がさらに高速化される。
- [0072] ところで、第 2 パスの符号化において、住所辞書 722 の代わりに、住所以外の他の文字列を登録した辞書を用いることも可能である。例えば、固有名詞の文字列を登録した固有名詞辞書、成句の文字列を登録した成句辞書、慣用句の文字列を登録した慣用句辞書、エラーメッセージの文字列を登録したエラーメッセージ辞書等を住所辞書 722 の代わりに用いてもよい。

[0073] 図18は、エラーメッセージ辞書の例を示している。図18のエラーメッセージ辞書は、エラーメッセージ及び符号を含む。エラーメッセージは、エラーメッセージを表す複数の形態素を含み、符号は、各エラーメッセージに対応付けられた符号を表す。例えば、“File Deliver Start”の符号は0xe001であり、“Memory Allocate Error”の符号は0xe002である。図9の住所辞書722と同様に、複数のエラーメッセージを先頭の形態素に対応する複数のブロックに分割して、エラーメッセージ辞書に登録してもよい。

[0074] 図5、図7、及び図12の符号化装置501の構成は一例に過ぎず、符号化装置501の用途や条件に応じて一部の構成要素を省略又は変更してもよい。

[0075] 図6、図10、図11、図15、及び図16のフローチャートは一例に過ぎず、符号化装置501の構成や条件に応じて一部の処理を省略又は変更してもよい。例えば、形態素解析辞書721にあらかじめ切り替え情報が設定されている場合は、図10の切り替え情報設定処理を省略することができる。

[0076] 図8の形態素解析辞書、図9及び図13の住所辞書、図14の形態素位置情報、及び図18のエラーメッセージ辞書は一例に過ぎず、符号化装置501の構成や条件に応じて別の形態素解析辞書、住所辞書、形態素位置情報、又はエラーメッセージ辞書を用いてもよい。例えば、図12の符号化装置501において、図13の住所辞書の代わりに図9の住所辞書を用いてもよい。

[0077] 図5、図7、及び図12の符号化装置501は、例えば、図19に示すような情報処理装置（コンピュータ）を用いて実現可能である。図19の情報処理装置は、Central Processing Unit（CPU）1901、メモリ1902、入力装置1903、出力装置1904、補助記憶装置1905、媒体駆動装置1906、及びネットワーク接続装置1907を含む。これらの構成要素はバス1908により互いに接続されている。

[0078] メモリ 1902 は、例えば、Read Only Memory (ROM)、Random Access Memory (RAM)、フラッシュメモリ等の半導体メモリである。メモリ 1902 は、符号化処理のためのプログラム及びデータを格納する。メモリ 1902 は、図 5、図 7、及び図 12 の記憶部 511 として用いることができる。

[0079] CPU 1901 (プロセッサ) は、例えば、メモリ 1902 を利用してプログラムを実行することにより、図 5、図 7、及び図 12 の符号化部 512、解析部 711、判定部 712、変換部 713、及び変換部 714 として動作し、符号化処理を行う。

[0080] 入力装置 1903 は、例えば、キーボード、ポインティングデバイス等であり、ユーザ又はオペレータからの指示や情報の入力に用いられる。出力装置 1904 は、例えば、表示装置、プリンタ、スピーカ等であり、ユーザ又はオペレータへの問い合わせや処理結果の出力に用いられる。

[0081] 補助記憶装置 1905 は、例えば、磁気ディスク装置、光ディスク装置、光磁気ディスク装置、テープ装置等である。補助記憶装置 1905 は、ハードディスクドライブ又はフラッシュメモリであってもよい。情報処理装置は、補助記憶装置 1905 にプログラム及びデータを格納しておき、それらをメモリ 1902 にロードして使用することができる。補助記憶装置 1905 は、図 5、図 7、及び図 12 の記憶部 511 として用いることができる。

[0082] 媒体駆動装置 1906 は、可搬型記録媒体 1909 を駆動し、その記録内容にアクセスする。可搬型記録媒体 1909 は、メモリデバイス、フレキシブルディスク、光ディスク、光磁気ディスク等である。可搬型記録媒体 1909 は、Compact Disk Read Only Memory (CD-ROM)、Digital Versatile Disk (DVD)、Universal Serial Bus (USB) メモリ等であってもよい。ユーザ又はオペレータは、この可搬型記録媒体 1909 にプログラム及びデータを格納しておき、それらをメモリ 1902 にロードして使用することができる。

[0083] このように、プログラム及びデータを格納するコンピュータ読み取り可能

な記録媒体は、メモリ 1902、補助記憶装置 1905、及び可搬型記録媒体 1909のような、物理的な（非一時的な）記録媒体である。

[0084] ネットワーク接続装置 1907は、Local Area Network（LAN）、インターネット等の通信ネットワークに接続され、通信に伴うデータ変換を行う通信インタフェースである。情報処理装置は、ネットワーク接続装置 1907を介して外部の装置からプログラム及びデータを受信し、それらをメモリ 1902にロードして使用することができる。ネットワーク接続装置 1907は、図5、図7、及び図12の出力部513として用いることができる。

[0085] なお、情報処理装置が図19のすべての構成要素を含む必要はなく、用途や条件に応じて一部の構成要素を省略することも可能である。例えば、ユーザ又はオペレータからの指示や情報の入力を行わない場合は、入力装置 1903を省略してもよく、ユーザ又はオペレータへの問い合わせや処理結果の出力を行わない場合は、出力装置 1904を省略してもよい。情報処理装置が可搬型記録媒体 1909又は通信ネットワークにアクセスしない場合は、媒体駆動装置 1906又はネットワーク接続装置 1907を省略してもよい。

[0086] 開示の実施形態とその利点について詳しく説明したが、当業者は、特許請求の範囲に明確に記載した本発明の範囲から逸脱することなく、様々な変更、追加、省略をすることができるであろう。

請求の範囲

- [請求項1] 形態素の区分に用いられ、登録文字列と符号とを対応付けたレコードを複数有する文字列符号化情報に含まれる各レコードの登録文字列の先頭形態素を複数のレコード各々に含む形態素区分情報に基づいて、符号化対象テキストを解析し、
- 前記形態素区分情報に含まれる前記複数のレコードのうち、ある登録文字列に対応したレコードに含まれる、ある形態素の識別に応じて、前記符号化対象テキストに含まれる前記ある形態素から始まる符号化対象文字列を前記文字列符号化情報において探索し、
- 前記符号化対象文字列が前記ある登録文字列に対応する場合、前記符号化対象文字列を前記文字列符号化情報において前記ある登録文字列に対応付けられた符号に変換し、
- 前記符号を含む符号化テキストを出力する、
- 処理をコンピュータに実行させる符号化プログラム。
- [請求項2] 前記文字列符号化情報は、前記ある形態素を含む複数の先頭形態素それぞれに対応する複数のブロックを含み、前記複数のブロックの各々は、前記複数の先頭形態素の各々から始まる複数の登録文字列と前記複数の登録文字列それぞれに対応付けられた複数の符号とを含み、前記コンピュータは、前記複数のブロックのうち前記ある形態素に対応するブロックに含まれる前記複数の登録文字列の中から、前記符号化対象文字列を探索することを特徴とする請求項1記載の符号化プログラム。
- [請求項3] 前記コンピュータは、各ブロックに含まれる各登録文字列内の複数の形態素それぞれの各登録文字列内における位置を示す形態素位置情報に基づいて、前記ある登録文字列を含むブロックを特定し、特定したブロックに含まれる前記複数の登録文字列の中から前記符号化対象文字列を探索することを特徴とする請求項2記載の符号化プログラム。

[請求項4] 前記形態素位置情報は、形態素と登録文字列内における位置との組み合わせ毎に、前記複数のブロックそれぞれが前記組み合わせに対応する登録文字列を含むか否かを示す情報を含み、前記コンピュータは、前記符号化対象文字列に含まれる各形態素と前記符号化対象文字列内における各形態素の位置とに基づいて、前記形態素位置情報を参照することで、前記ある登録文字列を含む前記ブロックを特定することを特徴とする請求項3記載の符号化プログラム。

[請求項5] 登録文字列と符号とを対応付けたレコードを複数有する文字列符号化情報と、形態素の区分に用いられ、前記文字列符号化情報に含まれる各レコードの登録文字列の先頭形態素を複数のレコード各々に含む形態素区分情報とを記憶する記憶部と、

前記形態素区分情報に基づいて符号化対象テキストを解析し、前記形態素区分情報に含まれる前記複数のレコードのうち、ある登録文字列に対応したレコードに含まれる、ある形態素の識別に応じて、前記符号化対象テキストに含まれる前記ある形態素から始まる符号化対象文字列を前記文字列符号化情報において探索し、前記符号化対象文字列が前記ある登録文字列に対応する場合、前記符号化対象文字列を前記文字列符号化情報において前記ある登録文字列に対応付けられた符号に変換する符号化部と、

前記符号を含む符号化テキストを出力する出力部と、
を備えることを特徴とする符号化装置。

[請求項6] 前記文字列符号化情報は、前記ある形態素を含む複数の先頭形態素それぞれに対応する複数のブロックを含み、前記複数のブロックの各々は、前記複数の先頭形態素の各々から始まる複数の登録文字列と前記複数の登録文字列それぞれに対応付けられた複数の符号とを含み、前記符号化部は、前記複数のブロックのうち前記ある形態素に対応するブロックに含まれる前記複数の登録文字列の中から、前記符号化対象文字列を探索することを特徴とする請求項5記載の符号化装置。

[請求項7] 前記記憶部は、各ブロックに含まれる各登録文字列内の複数の形態素それぞれの各登録文字列内における位置を示す形態素位置情報を記憶し、前記符号化部は、前記形態素位置情報に基づいて、前記ある登録文字列を含むブロックを特定し、特定したブロックに含まれる前記複数の登録文字列の中から前記符号化対象文字列を探索することを特徴とする請求項6記載の符号化装置。

[請求項8] 前記形態素位置情報は、形態素と登録文字列内における位置との組み合わせ毎に、前記複数のブロックそれぞれが前記組み合わせに対応する登録文字列を含むか否かを示す情報を含み、前記符号化部は、前記符号化対象文字列に含まれる各形態素と前記符号化対象文字列内における各形態素の位置とに基づいて、前記形態素位置情報を参照することで、前記ある登録文字列を含む前記ブロックを特定することを特徴とする請求項7記載の符号化装置。

[請求項9] コンピュータが、
形態素の区分に用いられ、登録文字列と符号とを対応付けたレコードを複数有する文字列符号化情報に含まれる各レコードの登録文字列の先頭形態素を複数のレコード各々に含む形態素区分情報に基づいて、符号化対象テキストを解析し、
前記形態素区分情報に含まれる前記複数のレコードのうち、ある登録文字列に対応したレコードに含まれる、ある形態素の識別に応じて、前記符号化対象テキストに含まれる前記ある形態素から始まる符号化対象文字列を前記文字列符号化情報において探索し、
前記符号化対象文字列が前記ある登録文字列に対応する場合、前記符号化対象文字列を前記文字列符号化情報において前記ある登録文字列に対応付けられた符号に変換し、
前記符号を含む符号化テキストを出力する、
ことを特徴とする符号化方法。

[請求項10] 前記文字列符号化情報は、前記ある形態素を含む複数の先頭形態素

それぞれに対応する複数のブロックを含み、前記複数のブロックの各々は、前記複数の先頭形態素の各々から始まる複数の登録文字列と前記複数の登録文字列それぞれに対応付けられた複数の符号とを含み、前記コンピュータは、前記複数のブロックのうち前記ある形態素に対応するブロックに含まれる前記複数の登録文字列の中から、前記符号化対象文字列を探索することを特徴とする請求項 9 記載の符号化方法。

[請求項11] 前記コンピュータは、各ブロックに含まれる各登録文字列内の複数の形態素それぞれの各登録文字列内における位置を示す形態素位置情報に基づいて、前記ある登録文字列を含むブロックを特定し、特定したブロックに含まれる前記複数の登録文字列の中から前記符号化対象文字列を探索することを特徴とする請求項 10 記載の符号化方法。

[請求項12] 前記形態素位置情報は、形態素と登録文字列内における位置との組み合わせ毎に、前記複数のブロックそれぞれが前記組み合わせに対応する登録文字列を含むか否かを示す情報を含み、前記コンピュータは、前記符号化対象文字列に含まれる各形態素と前記符号化対象文字列内における各形態素の位置とに基づいて、前記形態素位置情報を参照することで、前記ある登録文字列を含む前記ブロックを特定することを特徴とする請求項 11 記載の符号化方法。

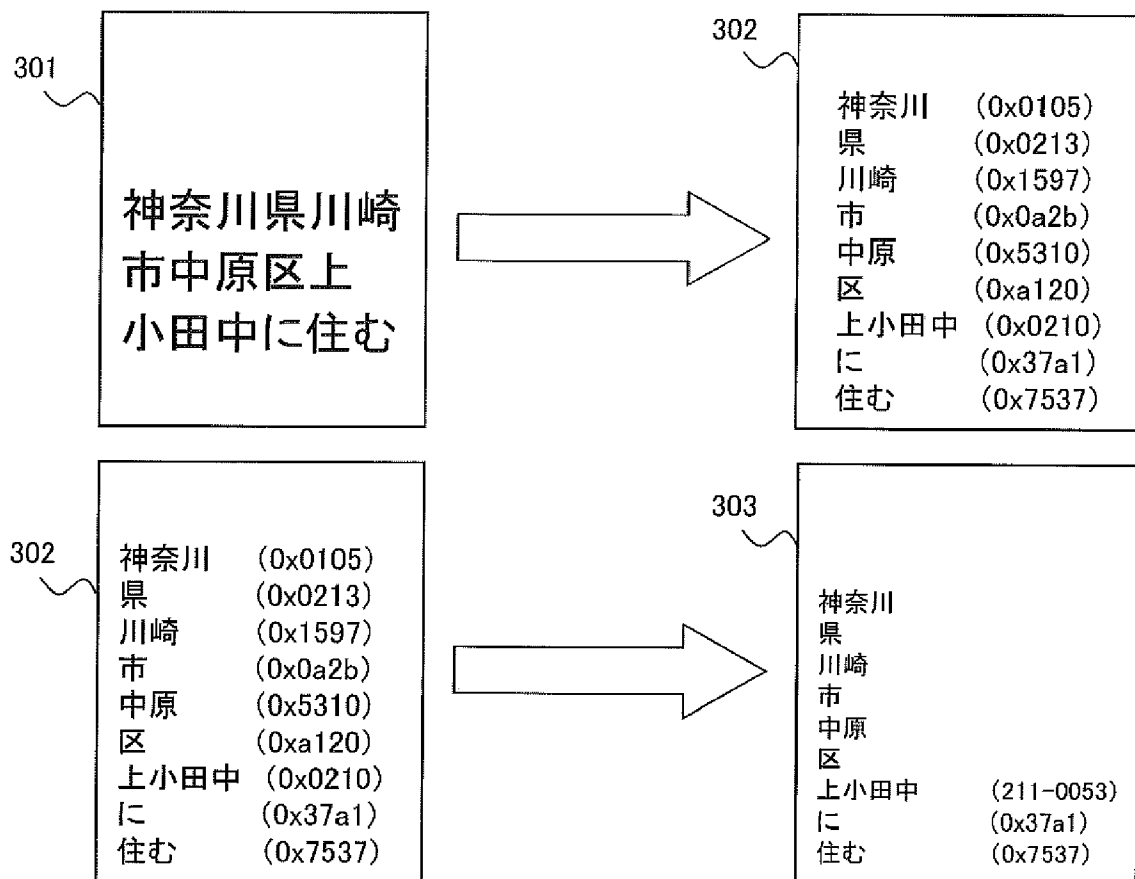
[図1]

形態素	品詞	符号
神奈川	名詞	0x0105
県	名詞	0x0213
	⋮	
住む	動詞	0x7537

[図2]

文字列	符号
神奈川県川崎市中原区等々力	211-0052
神奈川県川崎市中原区上小田中	211-0053
神奈川県川崎市中原区小杉	213-0061
⋮	

[図3]

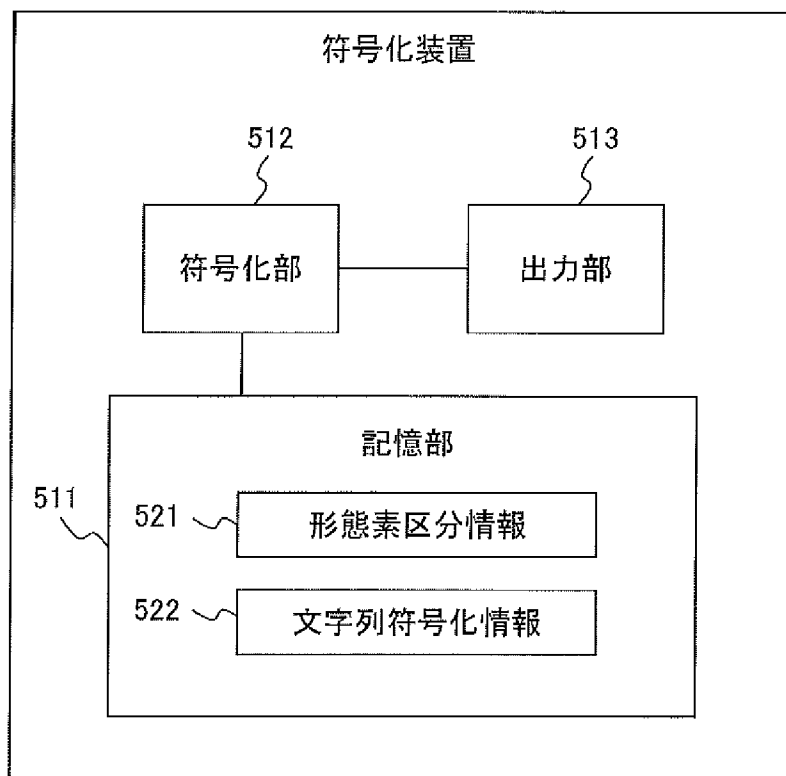


[図4]

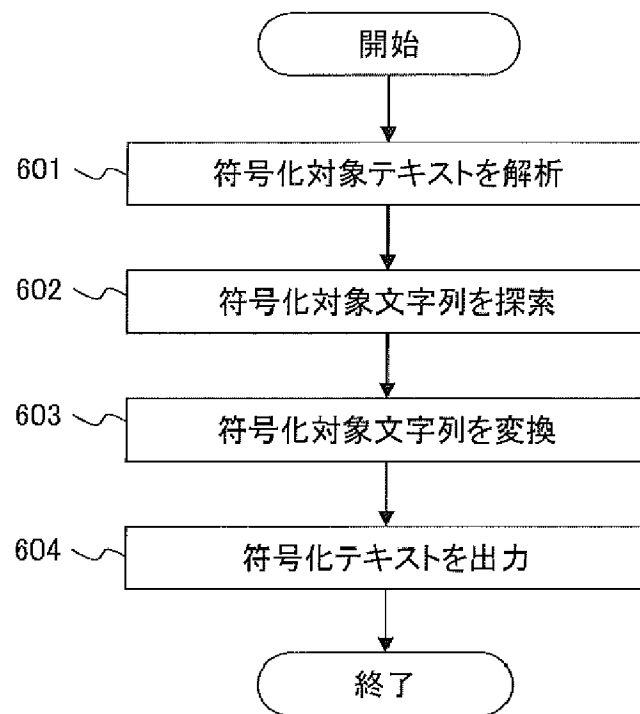
形態素	符号
神奈川	0x0105
県	0x0213
⋮	
神奈川県川崎市上小田中	211-0053

[図5]

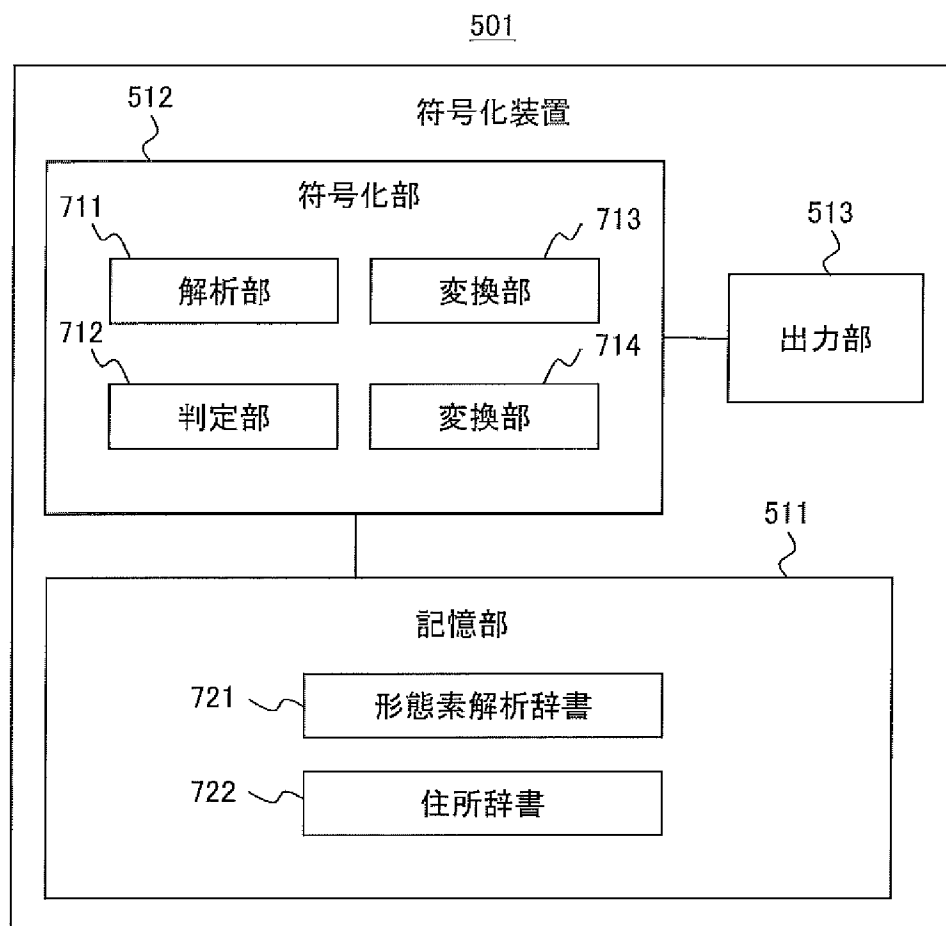
501



[図6]



[図7]



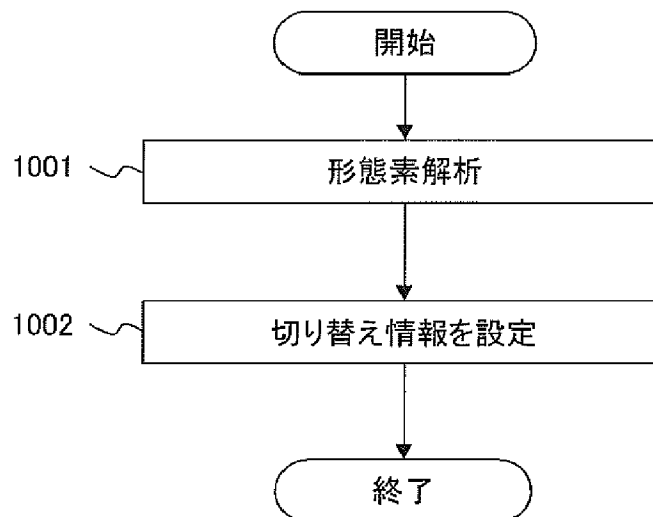
[図8]

形態素	品詞	符号	フラグ
東京	名詞	C1	1
都	名詞	C2	0
神奈川	名詞	C3	1
川崎	名詞	C4	0
市	名詞	C5	0
中原	名詞	C6	0
区	名詞	C7	0
⋮	⋮	⋮	⋮

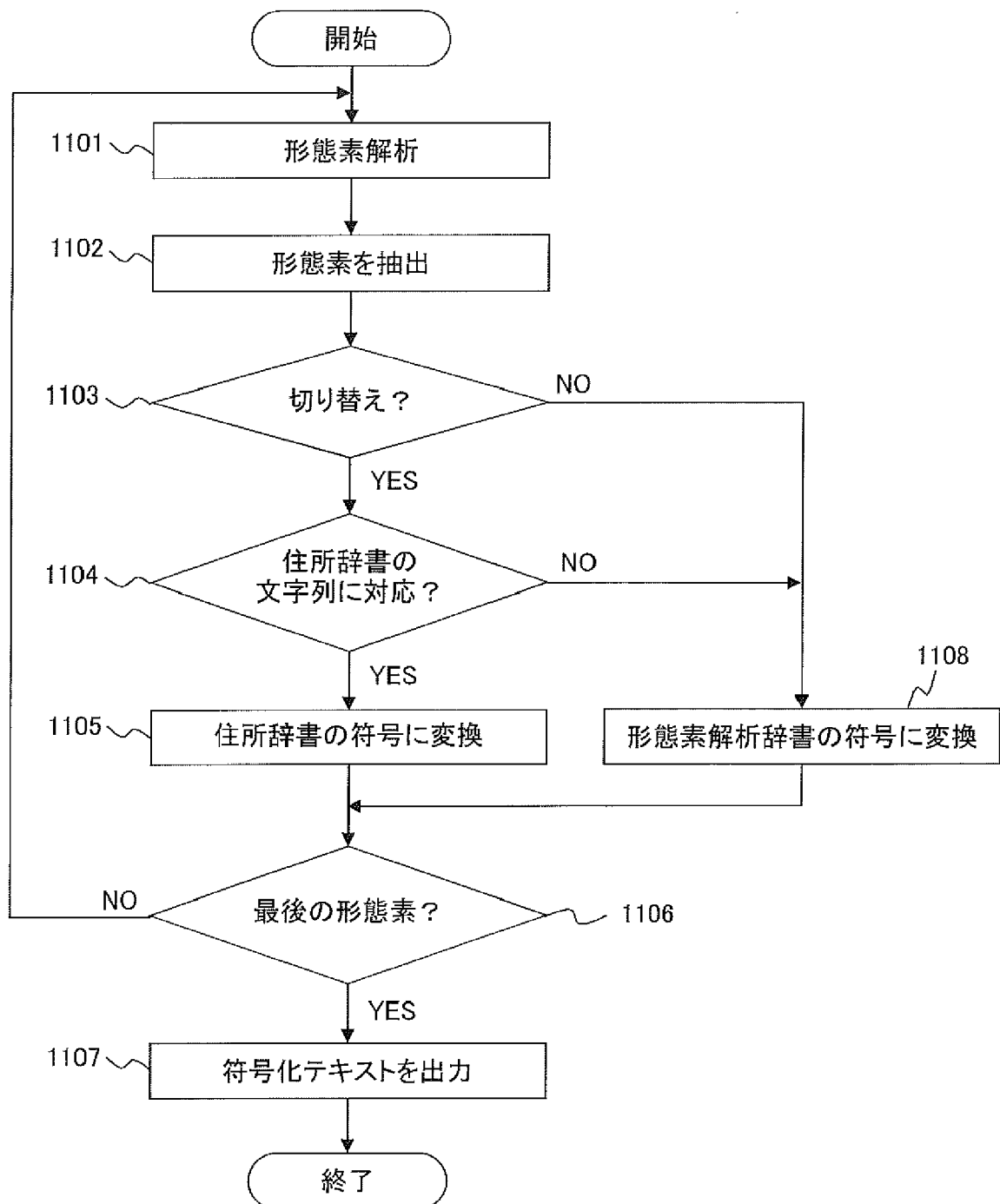
[図9]

	文字列	符号
901	東京都	
	東京都杉並区	D1
	⋮	⋮
902	神奈川県	
	神奈川県川崎市幸区	D2
	神奈川県川崎市中原区	D3
	⋮	⋮

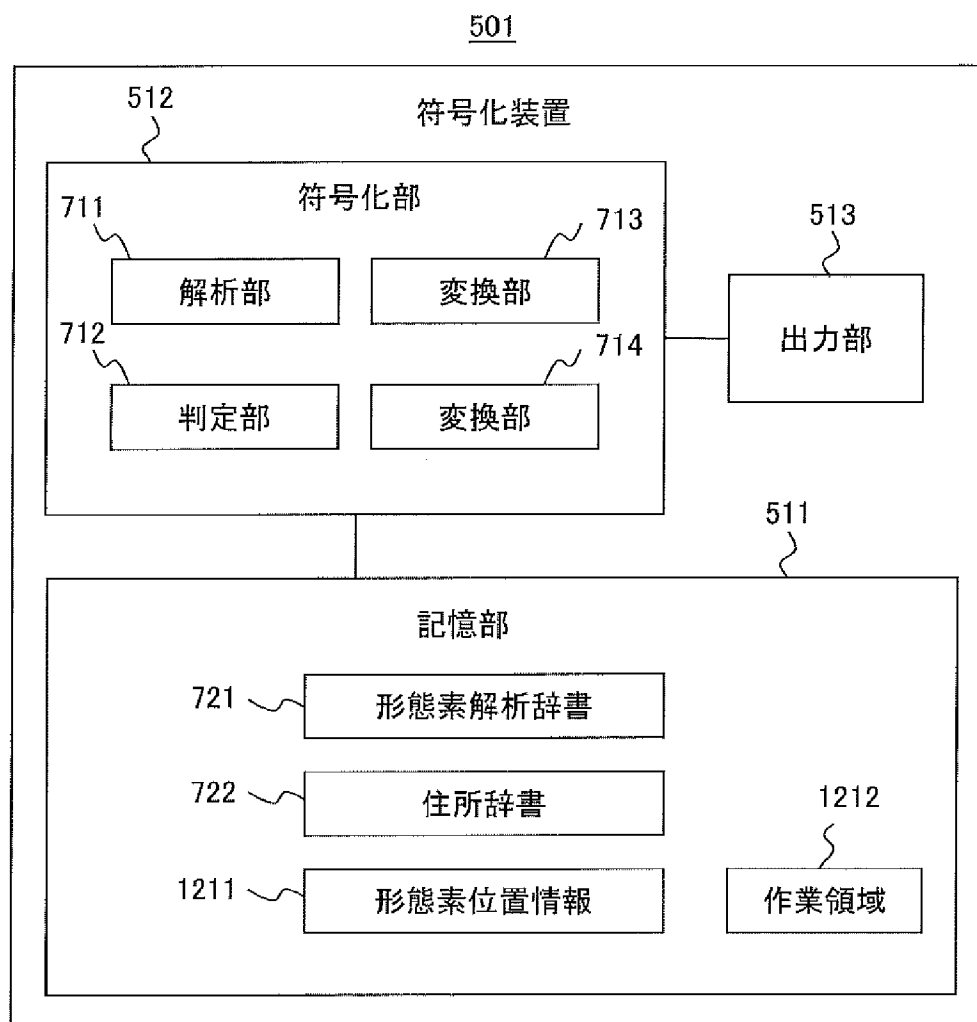
[図10]



[図11]



[図12]



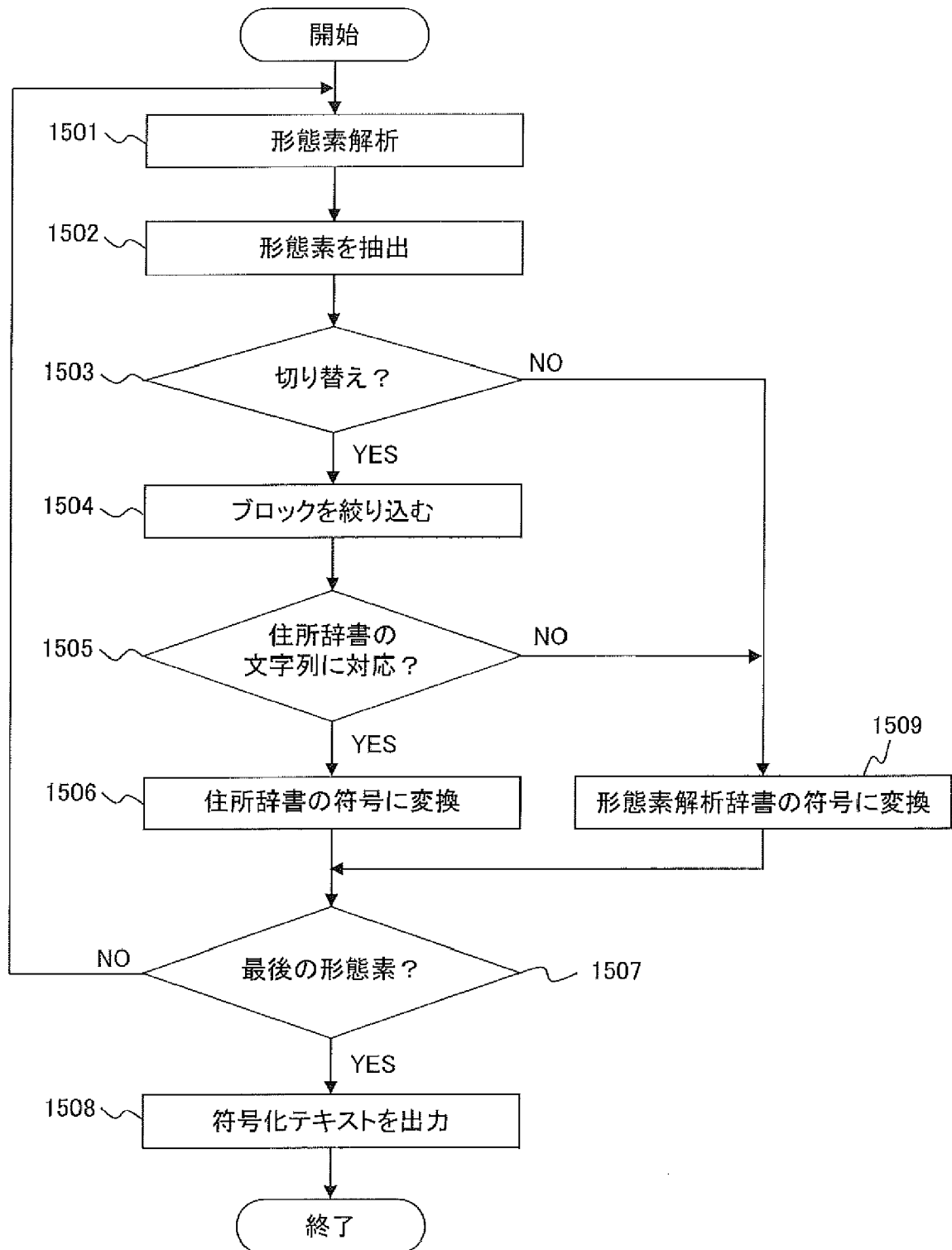
[図13]

文字列	符号
東京1	
⋮	⋮
東京2	
⋮	⋮
⋮	⋮
神奈川1	
⋮	⋮
神奈川2	
⋮	⋮
神奈川3	
⋮	⋮
神奈川4	
⋮	⋮
神奈川5	
⋮	⋮

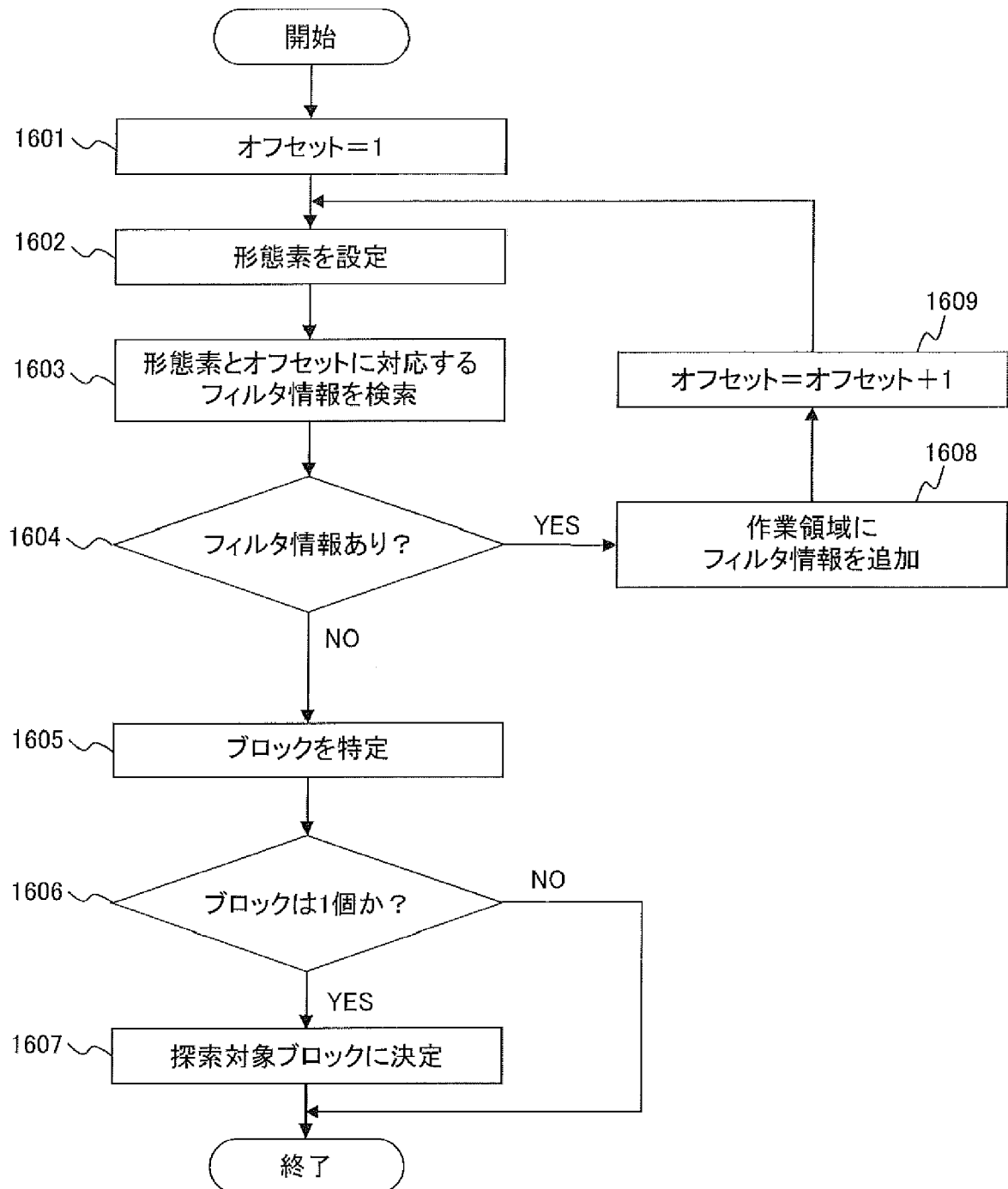
[図14]

形態素	オフセット	東京1	東京2	...	神奈川1	神奈川2	神奈川3	神奈川4	神奈川5	...
東京	1	1	1	...	0	0	0	0	0	...
：	：	：	：		：	：	：	：	：	
神奈川	1	0	0	...	1	1	1	1	1	...
神奈川	2	0	0	...	0	0	0	0	0	...
：	：	：	：		：	：	：	：	：	
多摩	1	0	0	...	0	0	0	0	0	...
多摩	2	0	0	...	0	0	0	0	0	...
多摩	3	1	0	...	0	0	0	0	0	...
多摩	4	0	0	...	0	0	0	0	0	...
多摩	5	0	1	...	0	0	1	0	0	...
：	：	：	：		：	：	：	：	：	

[図15]



[図16]



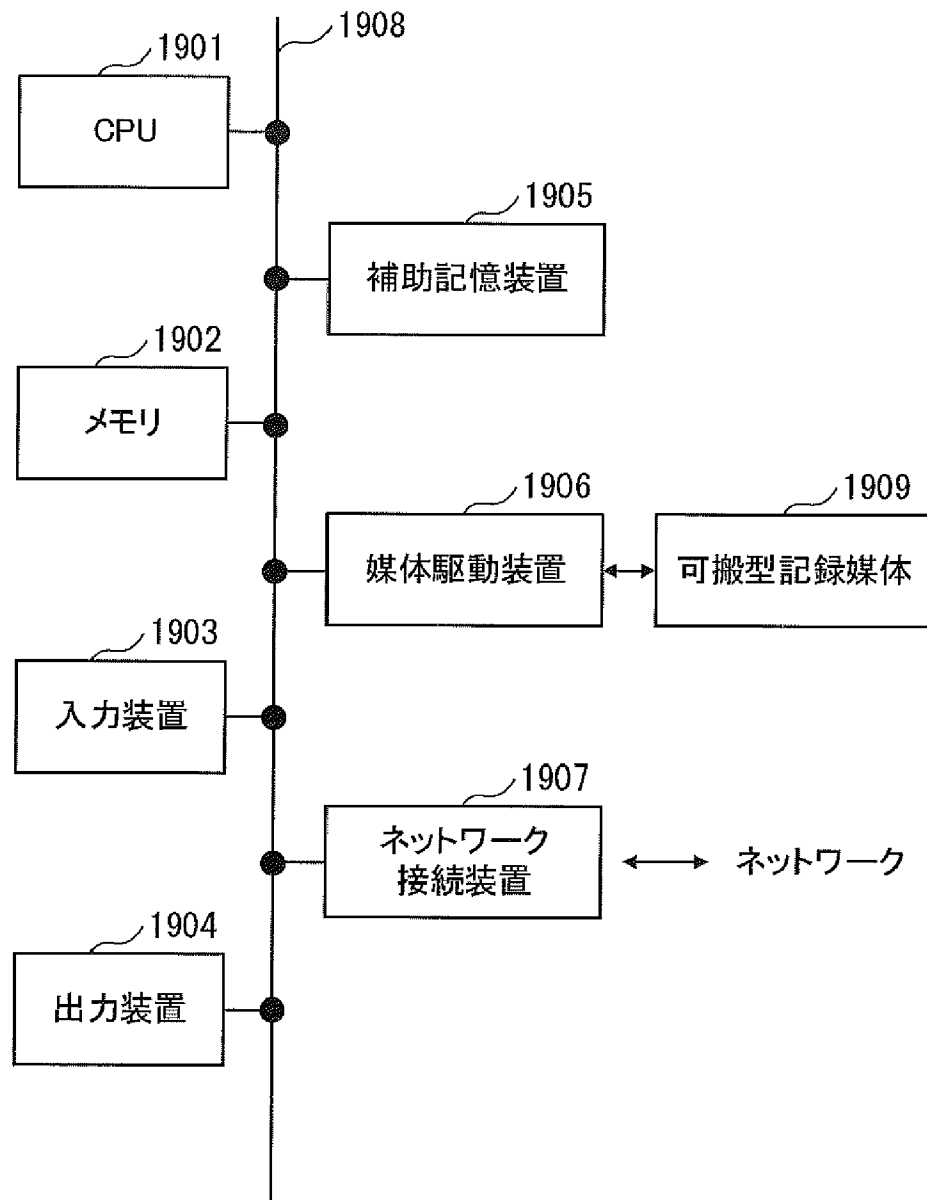
[図17]

形態素	オフセット	東京1	東京2	...	神奈川1	神奈川2	神奈川3	神奈川4	神奈川5	...
神奈川	1	0	0	...	1	1	1	1	1	...
県	2	0	0	...	1	1	1	1	1	...
川崎	3	0	0	...	1	0	1	0	0	...
市	4	0	0	...	1	1	1	0	0	...
多摩	5	0	1	...	0	0	1	0	0	...
区	6	1	0	...	0	1	1	1	0	...
登戸	7	0	0	...	0	0	1	0	0	...

[図18]

エラーメッセージ	符号
File Deliver Start	0xe001
Memory Allocate Error	0xe002
Multiple Start Error	0xe003
...	...

[図19]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2015/071171

A. CLASSIFICATION OF SUBJECT MATTER

G06F17/22(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)

G06F17/22

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	1996-2015
Kokai Jitsuyo Shinan Koho	1971-2015	Toroku Jitsuyo Shinan Koho	1994-2015

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 5-324427 A (Hitachi, Ltd.), 07 December 1993 (07.12.1993), paragraphs [0009], [0010] & US 5590317 A	1-12
A	JP 11-85790 A (Fujitsu Ltd.), 30 March 1999 (30.03.1999), paragraph [0007] & CN 1211013 A	1-12

☐ 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
12 August 2015 (12.08.15)

Date of mailing of the international search report
25 August 2015 (25.08.15)

Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

A. 発明の属する分野の分類（国際特許分類（I P C））

Int.Cl. G06F17/22 (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（I P C））

Int.Cl. G06F17/22

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	1 9 2 2 - 1 9 9 6 年
日本国公開実用新案公報	1 9 7 1 - 2 0 1 5 年
日本国実用新案登録公報	1 9 9 6 - 2 0 1 5 年
日本国登録実用新案公報	1 9 9 4 - 2 0 1 5 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 5-324427 A（株式会社日立製作所）1993. 12. 07, 段落 [0 0 0 9]、[0 0 1 0] & US 5590317 A	1 - 1 2
A	JP 11-85790 A（富士通株式会社）1999. 03. 30, 段落 [0 0 0 7] & CN 1211013 A	1 - 1 2

☐ 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

1 2 . 0 8 . 2 0 1 5

国際調査報告の発送日

2 5 . 0 8 . 2 0 1 5

国際調査機関の名称及びあて先

日本国特許庁（I S A / J P）
郵便番号 1 0 0 - 8 9 1 5
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）

川▲崎▼ 博章

5 M

5 2 8 4

電話番号 0 3 - 3 5 8 1 - 1 1 0 1 内線 3 5 9 9