

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6931442号  
(P6931442)

(45) 発行日 令和3年9月8日 (2021.9.8)

(24) 登録日 令和3年8月18日 (2021.8.18)

(51) Int.Cl. F I  
**G 0 6 F 16/00 (2019.01)** G O 6 F 16/00  
**G 0 6 F 16/901 (2019.01)** G O 6 F 16/901

請求項の数 9 (全 21 頁)

|           |                               |           |                                |
|-----------|-------------------------------|-----------|--------------------------------|
| (21) 出願番号 | 特願2017-97668 (P2017-97668)    | (73) 特許権者 | 000005223                      |
| (22) 出願日  | 平成29年5月16日 (2017.5.16)        |           | 富士通株式会社                        |
| (65) 公開番号 | 特開2018-195028 (P2018-195028A) |           | 神奈川県川崎市中原区上小田中4丁目1番1号          |
| (43) 公開日  | 平成30年12月6日 (2018.12.6)        | (74) 代理人  | 110002147                      |
| 審査請求日     | 令和2年2月13日 (2020.2.13)         |           | 特許業務法人酒井国際特許事務所                |
|           |                               | (72) 発明者  | 片岡 正弘                          |
|           |                               |           | 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 |
|           |                               | (72) 発明者  | 川久保 賢一                         |
|           |                               |           | 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 |
|           |                               | (72) 発明者  | 岩本 昭次                          |
|           |                               |           | 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 |

最終頁に続く

(54) 【発明の名称】 符号化プログラム、インデックス生成プログラム、検索プログラム、符号化装置、インデックス生成装置、検索装置、符号化方法、インデックス生成方法および検索方法

(57) 【特許請求の範囲】

【請求項1】

コンピュータに、

対象文章データに含まれる符号化対象データを符号化する際に、単語と前記単語を圧縮したときの符号である基本コードとを対応付けて記憶する静的辞書を参照し、前記符号化対象データが前記静的辞書に登録されている単語の変化形であるか否かを判定し、

前記符号化対象データが前記単語の変化形である場合、前記符号化対象データを圧縮する符号である動的符号を前記符号化対象データに割り与え、前記符号化対象データと前記動的符号と前記符号化対象データに対応する前記基本コードとを対応付けて動的辞書に登録する、

処理を実行させる符号化プログラム。

【請求項2】

コンピュータに、

対象文章データに含まれる符号化対象データを符号化して圧縮ファイルを生成する際に、単語と前記単語を圧縮したときの符号である基本コードとを対応付けて記憶する静的辞書を参照し、前記符号化対象データが単語の変化形であるか否かを判定し、

前記符号化対象データが前記静的辞書に登録されている前記単語の変化形である場合、前記符号化対象データを圧縮する符号である動的符号を前記符号化対象データに割り与え、前記符号化対象データと前記動的符号と前記符号化対象データに対応する前記基本コードとを対応付けて動的辞書に登録し、

前記単語の基本形および前記単語の変化形ごとに、前記単語の基本形および前記単語の変化形が含まれる圧縮ファイルを特定する符号ビット列を記憶するビットマップ型転置インデックスに、符号と出現位置を対応付けた、ビット列を登録する、

処理を実行させるインデックス生成プログラム。

【請求項 3】

コンピュータに、

対象文章データを符号化した符号化文章データに対する検索要求を受け付け、

単語と前記単語を圧縮したときの符号である基本コードとを対応付けて記憶する静的辞書を参照し、前記検索要求に含まれる検索キーが、単語の変化形であるか否かを判定し、

前記検索キーが前記単語の変化形である場合、前記単語の基本形および前記単語の変化形ごとに、前記単語の基本形および前記単語の変化形が含まれる圧縮ファイルを特定する符号ビット列を記憶するビットマップ型転置インデックスの符号と出現位置との対応付けを参照して前記検索キーを含む圧縮ファイルの検索を行う、

処理を実行させる検索プログラム。

【請求項 4】

対象文章データに含まれる符号化対象データを符号化する際に、単語と前記単語を圧縮したときの符号である基本コードとを対応付けて記憶する静的辞書を参照し、前記符号化対象データが前記静的辞書に登録されている単語の変化形であるか否かを判定する判定部と、

前記符号化対象データが前記単語の変化形である場合、前記符号化対象データを圧縮する符号である動的符号を前記符号化対象データに割り与え、前記符号化対象データと前記動的符号と前記符号化対象データに対応する前記基本コードとを対応付けて動的辞書に登録する登録部と、

を有する符号化装置。

【請求項 5】

対象文章データに含まれる符号化対象データを符号化して圧縮ファイルを生成する際に、単語と前記単語を圧縮したときの符号である基本コードとを対応付けて記憶する静的辞書を参照し、前記符号化対象データが単語の変化形であるか否かを判定する判定部と、

前記符号化対象データが前記静的辞書に登録されている前記単語の変化形である場合、前記符号化対象データを圧縮する符号である動的符号を前記符号化対象データに割り与え、前記符号化対象データと前記動的符号と前記符号化対象データに対応する前記基本コードとを対応付けて動的辞書に登録する登録部と、

前記単語の基本形および前記単語の変化形ごとに、前記単語の基本形および前記単語の変化形が含まれる圧縮ファイルを特定する符号ビット列を記憶するビットマップ型転置インデックスに、符号と出現位置を対応付けた、ビット列を登録するインデックス登録部と

、  
を有するインデックス生成装置。

【請求項 6】

対象文章データを符号化した符号化文章データに対する検索要求を受け付ける受付部と

単語と前記単語を圧縮したときの符号である基本コードとを対応付けて記憶する静的辞書を参照し、前記検索要求に含まれる検索キーが、単語の変化形であるか否かを判定する検索キー判定部と、

前記検索キーが前記単語の変化形である場合、前記単語の基本形および前記単語の変化形ごとに、前記単語の基本形および前記単語の変化形が含まれる圧縮ファイルを特定する符号ビット列を記憶するビットマップ型転置インデックスの符号と出現位置との対応付けを参照して前記検索キーを含む圧縮ファイルの検索を行う検索部と、

を有する検索装置。

【請求項 7】

コンピュータが、

対象文章データに含まれる符号化対象データを符号化する際に、単語と前記単語を圧縮したときの符号である基本コードとを対応付けて記憶する静的辞書を参照し、前記符号化対象データが前記静的辞書に登録されている単語の変化形であるか否かを判定し、

前記符号化対象データが前記単語の変化形である場合、前記符号化対象データを圧縮する符号である動的符号を前記符号化対象データに割り与え、前記符号化対象データと前記動的符号と前記符号化対象データに対応する前記基本コードとを対応付けて動的辞書に登録する、

処理を実行する符号化方法。

【請求項 8】

コンピュータが、

対象文章データに含まれる符号化対象データを符号化して圧縮ファイルを生成する際に、単語と前記単語を圧縮したときの符号である基本コードとを対応付けて記憶する静的辞書を参照し、前記符号化対象データが単語の変化形であるか否かを判定し、

前記符号化対象データが前記静的辞書に登録されている前記単語の変化形である場合、前記符号化対象データを圧縮する符号である動的符号を前記符号化対象データに割り与え、前記符号化対象データと前記動的符号と前記符号化対象データに対応する前記基本コードとを対応付けて動的辞書に登録し、

前記単語の基本形および前記単語の変化形ごとに、前記単語の基本形および前記単語の変化形が含まれる圧縮ファイルを特定する符号ビット列を記憶するビットマップ型転置インデックスに、符号と出現位置を対応付けた、ビット列を登録する、

処理を実行するインデックス生成方法。

【請求項 9】

コンピュータが、

対象文章データを符号化した符号化文章データに対する検索要求を受け付け、

単語と前記単語を圧縮したときの符号である基本コードとを対応付けて記憶する静的辞書を参照し、前記検索要求に含まれる検索キーが、単語の変化形であるか否かを判定し、

前記検索キーが前記単語の変化形である場合、前記単語の基本形および前記単語の変化形ごとに、前記単語の基本形および前記単語の変化形が含まれる圧縮ファイルを特定する符号ビット列を記憶するビットマップ型転置インデックスの符号と出現位置との対応付けを参照して前記検索キーを含む圧縮ファイルの検索を行う、

処理を実行する検索方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、符号化プログラム、インデックス生成プログラムおよび検索プログラム等に関する。

【背景技術】

【0002】

圧縮対象のテキストファイルから単語を抽出し、抽出した単語をアルファベット単位で圧縮する技術が知られている。かかる技術では、a～zまでの各アルファベットに符号を割り当て、圧縮対象のテキストファイルから抽出した単語に含まれる各アルファベットを符号に変換する。さらに、変換された符号に、例えば全角または半角の指定、大文字または小文字の指定などに関する情報を表す符号を付加することで単語を符号に変換する。さらに近年では、出現頻度の高い単語に個別の符号を割り当て、圧縮対象のテキストファイルから抽出した出現頻度の高い単語を直接符号に変換する技術が提案されている。

【先行技術文献】

【特許文献】

【0003】

【特許文献 1】 国際公開第 2008/047432 号

【特許文献 2】 特開 2013-150041 号公報

10

20

30

40

50

【特許文献3】特開平05-174064号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかしながら、上記技術では、出現頻度の高い単語について、基本形（例えば現在形）のみならず過去形や過去分詞形、複数形などの変化形にもそれぞれ個別の符号を割り当て、かかる符号と単語の変化形とを対応付けて静的辞書に登録した場合、静的辞書に必要な容量が増加するため、圧縮率が低下する場合がある。

【0005】

一つの側面では、出現頻度の高い単語に個別の符号を割り当てて符号化する場合の圧縮率を向上させる符号化プログラム、インデックス生成プログラムおよび検索プログラム等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】

【0006】

第1の案では、符号化プログラムは、コンピュータに、対象文章データに含まれる符号化対象データを符号化する際に、前記符号化対象データが単語の変化形であるか否かを判定する処理を実行させる。符号化プログラムは、コンピュータに、前記符号化対象データが前記単語の変化形である場合、当該符号化対象データと符号とを対応付けて動的辞書に登録する処理を実行させる。

【発明の効果】

【0007】

一実施形態によれば、出現頻度の高い単語に個別の符号を割り当てて符号化する場合の圧縮率を向上できる。

【図面の簡単な説明】

【0008】

【図1】図1は、実施例1の単語の符号化処理を説明するための図である。

【図2】図2は、情報処理装置のシステム構成の例を説明するための図である。

【図3】図3は、実施例1の符号化処理に係るシステム構成の例を示す図である。

【図4】図4は、ビットフィルタの第1の例を示す図である。

【図5】図5は、動的辞書の一例を示す図である。

【図6】図6は、ビットフィルタの第2の例を示す図である。

【図7】図7は、実施例1の符号化処理の流れの例を示す図である。

【図8】図8は、実施例2のインデックス生成処理に係るシステム構成の例を示す図である。

【図9】図9は、ビットマップ型転置インデックスの一例を示す図である。

【図10】図10は、実施例2のインデックス生成処理の流れの例を示す図である。

【図11】図11は、実施例3の検索処理に係るシステム構成の例を示す図である。

【図12】図12は、圧縮ファイル選択の処理の流れの一例を示す図である。

【図13】図13は、圧縮ファイル選択の処理の流れの別の例を示す図である。

【図14】図14は、実施例3の検索処理の流れの例を示す図である。

【図15】図15は、実施例1～3の情報処理装置のハードウェア構成を示す図である。

【図16】図16は、コンピュータで動作するプログラムの構成例を示す図である。

【図17】図17は、実施形態のシステムにおける装置の構成例を示す図である。

【発明を実施するための形態】

【0009】

以下に、本願の開示する符号化プログラム、インデックス生成プログラムおよび検索プログラム等の実施例を図面に基づいて詳細に説明する。なお、この実施例によりこの権利範囲が限定されるものではない。各実施例は、処理内容を矛盾させない範囲で適宜組み合わせることが可能である。

【実施例1】

## 【0010】

(実施例1の符号化処理)

図1を用いて、実施例1の情報処理装置100(図2参照)による符号化処理について説明する。図1は、実施例1の単語の符号化処理を説明するための図である。図1の例のように、情報処理装置100は、符号化処理の対象である対象文章データ10に含まれる「She acted in a movie. . .」を「She」「acted」「in」「a」「movie」のように数値または単語ごとに区切り、符号化対象データである各数値または各単語を取得する。以下、1以上の数字を有する数値を数値文字列と呼ぶ。数値文字列は、数字以外にプラス・マイナス、カンマおよび小数点等の符号を含んでもよい。

## 【0011】

情報処理装置100は、符号化対象データとして、例えば「acted」を抽出してビットフィルタ121に出力し、抽出された符号化対象データが単語の変化形であるか否かを判定する。ビットフィルタ121は、各単語に対してかかる単語の基本形と、動的符号とを対応付ける辞書である。ビットフィルタ121は、それぞれの単語の基本形と、かかる単語の基本形に対応する基本コードとがあらかじめ登録されている。例えば、ビットフィルタ121には、単語の変化形「acted」「acting」「acts」…に対応する単語の基本形「act」と、かかる単語の基本形「act」に対応する基本コード「c(act)」とがあらかじめ登録されている。なお、図1において「c(単語)」の表記は、カッコ内の単語に対応する基本コードを表す。例えば、「c(act)」は、「act」の基本コードを表す。ビットフィルタ121のデータ構造に関する詳細は後述する。

## 【0012】

次いで、情報処理装置100は、抽出された符号化対象データ「acted」が単語の変化形である場合、ビットフィルタ121から単語「acted」の基本形「act」に対応する基本コード「c(act)」を取得する。次いで、情報処理装置100は、取得した基本コード「c(act)」と、抽出された単語「acted」の変化種別(ここでは「過去形」と、登録された順に動的に付される動的符号(ここでは「A100h」と)を対応づけて動的辞書122に登録する。なお、動的符号「A100h」の末尾に示す「h」は、16進数で表記されていることを示す符号である。

## 【0013】

次いで、情報処理装置100は、動的辞書122において動的に付された動的符号「A100h」を、単語の基本形「act」および基本コード「c(act)」と対応づけてビットフィルタ121に登録する。そして、情報処理装置100は、ビットフィルタ121および動的辞書122を基にして単語「acted」に対応する動的符号「A100h」を取得し、圧縮ファイル11に出力する。

## 【0014】

また、情報処理装置100は、次回、対象文章データ10に単語「acted」が出現した場合、既にビットフィルタ121に登録されている動的符号「A100h」を取得し、圧縮ファイル11に出力する。

## 【0015】

(実施例1と参考例との比較)

上述したように実施例1では、単語の変化形を静的辞書であるビットフィルタ121ではなく、動的辞書122に登録して符号化する。一方、参考例では、単語の変化形を、単語の基本形と同様に静的辞書であるビットフィルタ121に登録して符号化するものとする。参考例のように単語の変化形も静的辞書に登録すると、静的辞書の容量が増加してしまうことから、符号化した場合の圧縮率が低下する場合がある。これに対して、実施例1のように単語の変化形を動的辞書122に登録して符号化すれば、静的辞書の容量増加を抑制することができる。したがって、参考例と比べて符号化した際の圧縮率を向上させることができる。

## 【0016】

(実施例1の符号化処理に関する処理部の構成)

図2を用いて、情報処理装置100の圧縮部110と記憶部120との関係について説明する。図2は、情報処理装置100のシステム構成の例を説明するための図である。図2の例に示すように、情報処理装置100の記憶部120は、圧縮部110と処理部150とに接続される。記憶部120は例えば、RAM(Random Access Memory)、ROM(Read Only Memory)、フラッシュメモリなどの半導体メモリ素子、ハードディスクや光ディスクなどの記憶装置に対応する。

#### 【0017】

また、情報処理装置100は、圧縮部110と、処理部150とを有する。圧縮部110および処理部150の機能は例えば、CPU(Central Processing Unit)が所定のプログラムを実行することで実現することができる。また、圧縮部110および処理部150の機能は例えば、ASIC(Application Specific Integrated Circuit)やFPGA(Field Programmable Gate Array)などの集積回路により実現することができる。

10

#### 【0018】

圧縮部110は、入力されたデータから単語を抽出する。圧縮部110は、単語と符号とを対応付けて記憶する辞書を参照し、抽出された単語に対応する符号に変換する。

#### 【0019】

かかる辞書は、所定の複数の単語の基本形を予め記憶する第1の辞書と、第1の辞書に登録されている単語の変化形を記憶する第2の辞書とである。圧縮部110は、抽出された単語の基本形が第1の辞書に登録されており、かつ、かかる単語の変化形に対応する動的符号が第1の辞書に登録されていない場合、かかる単語に動的符号を付与して第1の辞書におけるかかる単語の基本形に対応付けてかかる動的符号を登録する。なお、第1の辞書は、例えば、ビットフィルタであり、第2辞書は、例えば、動的辞書である。

20

#### 【0020】

図3を用いて、実施例1の符号化処理に係るシステム構成について説明する。図3は、実施例1の符号化処理に係るシステム構成の例を示す図である。図3の例に示されるように、情報処理装置100は、圧縮部110と、記憶部120とを有する。圧縮部110は、ファイルリード部111、判定部112、動的符号付与部113およびファイルライト部114を有する。記憶部120は、ビットフィルタ121および動的辞書122を有する。なお、ビットフィルタ121は静的辞書の一例であり、動的符号付与部113は登録部113の一例である。以下、実施例1の圧縮部110および記憶部120の構成について詳細に説明する。

30

#### 【0021】

(記憶部の各構成)

まずは、ビットフィルタ121について説明する。ビットフィルタ121は、出現頻度の高い単語に基本コードおよび動的符号を対応付けた辞書である。図4は、ビットフィルタ121の第1の例を示す図である。図4の例に示すように、ビットフィルタ121は、2グラムと、ビットマップと、単語と、文字列長と、出現頻度と、符号長と、基本コードと、動的符号とを対応付ける。「2グラム」は、各単語に含まれる連数字である。例えば「act」は、「ac」「ct」に対応する2グラムを有する。

40

#### 【0022】

「ビットマップ」は、2グラムの文字列に対応するビットマップを表す。例えば、「act」は、各単語へのポイントによって、2グラム「ac」のビットマップ「1\_\_0\_\_0\_\_0\_\_0」と、2グラム「ct」のビットマップ「0\_\_1\_\_0\_\_0\_\_0」とに対応付けられる。例えば、情報処理装置100は、対象文章データ10から「act」を取得した場合に、2グラム「ac」のビットマップ「1\_\_0\_\_0\_\_0\_\_0」と、2グラム「ct」のビットマップ「0\_\_1\_\_0\_\_0\_\_0」および、2グラム「t△」のビットマップ「0\_\_0\_\_1\_\_0\_\_0」とを用いて単語「act」にアクセスする。「単語」は、ビットフィルタ121に登録された単語である。例えば、「単語」とは、辞典、テキスト群等からあらかじめ抽出された単語である。なお、「単語」の登録数は任意でよい。

50

## 【0023】

「文字列長」は、登録された単語における文字列の長さである。「出現頻度」は、頻度集計用のテキストファイル群において各単語が出現した回数である。ここで、頻度集計用のテキストファイル群とは、対象文章データ10とは別に用意された各単語の出現頻度を集計するための1以上のテキストファイルである。

## 【0024】

単語の出現頻度の集計について説明する。情報処理装置100は、頻度集計用のテキストファイルを読み込み、頻度集計用のテキストファイル群に存在する単語を適宜抽出してビットフィルタ121に登録する。さらに、情報処理装置100は、ビットフィルタ121に登録された各々の単語に関し、頻度集計用のテキストファイル群において出現した回数を出現頻度としてカウントする。例えば、図4の例において、ビットフィルタ121は、単語「a」が、頻度集計用のテキストファイル群において「9738」回出現したことを示す。

10

## 【0025】

「基本コード」は、単語の基本形にそれぞれ割り当てられた符号である。情報処理装置100は、出現頻度が高い単語に対してより短い符号長を有する符号を割り当てる。なお、図4の例では、各項目のデータがレコードとして関連づけられて記憶されている例を示したが、上記説明において互いに関連づけられた項目どうしの関係が保たれれば、データは他の記憶のされ方をしても構わない。これは、後述する図5、図6に示されるビットフィルタ121および動的辞書122に関しても同様である。

20

## 【0026】

次に、動的辞書122について説明する。動的辞書122は、単語の基本形に対応する基本コードと、かかる単語の変化種別と、付与された動的符号とを対応付ける辞書である。図5は、動的辞書122の一例を示す図である。図5の例に示すように、動的辞書122は、動的符号と、単語の基本コードおよび変化種別とを対応付ける。「基本コード」は、ビットフィルタ121に登録された基本コードである。「変化種別」は、単語の基本形がどのように変化したかを示す変化種別である。「動的符号」は、例えば動的辞書122に登録された順番に基本コードおよび変化種別に昇順に付与される固定長の動的符号である。

## 【0027】

例えば、情報処理装置100が、対象文章データ10から単語の変化形を順番に抽出した場合、各々の単語の変化形に対して、動的符号「A100h」「A101h」「A102h」「A103h」…を割り当てる。例えば、対象文章データ10から「acted」「eating」…の順番に単語を抽出した場合、抽出された単語が変化形であるか否かを判定し、抽出された単語が変化形である場合には抽出された単語の変化種別を判定する。そして、抽出された単語が変化形である場合、図5の例に示すように、単語の基本形「act」における基本コード「6041h」および変化種別「過去形」に対して、動的符号「A100h」を割り当て、次に単語の基本形「eat」における基本コード「60F5h」および変化種別「進行形」に対して、動的符号「A101h」を割り当てる。

30

## 【0028】

(圧縮部の各構成)

図3に戻り、圧縮部110の各構成について説明する。ファイルリード部111は、対象文章データ10を読み出して、対象文章データ10から符号化対象データを抽出する処理部である。ファイルリード部111は、対象文章データ10における文字列中の空白文字やカンマなどによって文字列を単語または数値表記文字列ごとに区切り、対象文章データ10から各符号化対象データを抽出する。そして、ファイルリード部111は、抽出した符号化対象データを判定部112に出力する。

40

## 【0029】

判定部112は、対象文章データ10から抽出された符号化対象データが単語の変化形であるか否かを判定し、抽出された符号化対象データが単語の変化形である場合、抽出さ

50

れた単語の基本形および変化種別を判定する処理部である。そして、抽出された符号化対象データが単語の変化形である場合、判定部 1 1 2 は、抽出された単語の変化形と、かかる単語の基本形および変化種別の情報とを動的符号付与部 1 1 3 に出力する。

#### 【0030】

動的符号付与部 1 1 3 は、対象文章データ 1 0 から抽出された単語の変化形に動的符号を付与する処理部である。動的符号付与部 1 1 3 は、判定部 1 1 2 から単語の変化形を受け付けると、ビットフィルタ 1 2 1 に登録されているかかる単語の基本形を参照して、抽出された単語の変化形にアクセスする。

#### 【0031】

動的符号付与部 1 1 3 は、アクセスした単語の変化形に対応する動的符号がビットフィルタ 1 2 1 に既に登録されていた場合、ビットフィルタ 1 2 1 から単語の変化形に対応する動的符号を取得してファイルライト部 1 1 4 に出力する。

#### 【0032】

一方、動的符号付与部 1 1 3 は、アクセスした単語の変化形に対応する動的符号がビットフィルタ 1 2 1 に登録されていない場合、ビットフィルタ 1 2 1 から単語の基本形に対応する基本コードを取得する。次いで、動的符号付与部 1 1 3 は、ビットフィルタ 1 2 1 から取得した基本コードを、動的辞書 1 2 2 への登録順に付与される動的符号と、かかる単語の変化種別とに対応付けて動的辞書 1 2 2 に登録する。

#### 【0033】

図 5 を用いて、動的辞書 1 2 2 への動的符号の登録の具体例について説明する。図 5 の例に示すように、動的符号付与部 1 1 3 は、1 番目に登録された単語「acted」について、かかる単語の基本形「act」に対応する基本コード「6041h」および変化種別「過去形」に、動的符号「A100h」を対応づけて動的辞書 1 2 2 に登録する。また、動的符号付与部 1 1 3 は、2 番目に登録された単語「eating」について、かかる単語の基本形「eat」に対応する基本コード「60F5h」および変化種別「進行形」に、動的符号「A101h」を対応づけて動的辞書 1 2 2 に登録する。このように、動的符号付与部 1 1 3 は、抽出された単語の変化形に対して動的辞書 1 2 2 へ登録された順番に対応する動的符号を付与し、単語の基本形に対応する基本コードおよび変化種別と、付与された動的符号とを対応付けて動的辞書 1 2 2 に登録する。

#### 【0034】

次いで、動的符号付与部 1 1 3 は、動的辞書 1 2 2 に登録した動的符号を、単語の基本形に対応付けてビットフィルタ 1 2 1 に登録する。

#### 【0035】

図 6 を用いて、登録番号および動的符号が登録された後のビットフィルタ 1 2 1 の具体例について説明する。図 6 は、ビットフィルタ 1 2 1 の第 2 の例を示す図である。図 6 の例に示すように、ビットフィルタ 1 2 1 に動的符号が登録される。ビットフィルタ 1 2 1 において、「動的符号」は、単語の変化形に対応する動的符号である。例えば、ビットフィルタ 1 2 1 には、単語の基本形「act」に対応する単語の変化形「acted」に付与された動的符号「A100h」が登録される。なお、動的符号には、かかる単語の変化種別についても登録される。

#### 【0036】

そして、動的符号付与部 1 1 3 は、ビットフィルタ 1 2 1 に登録した動的符号をファイルライト部 1 1 4 に出力する。

#### 【0037】

ファイルライト部 1 1 4 は、動的符号付与部 1 1 3 から出力された動的符号を基に圧縮ファイル 1 1 を生成する処理部である。ファイルライト部 1 1 4 は、例えば動的符号付与部 1 1 3 から出力された単語の変化形に対応する各動的符号を、それぞれバッファに格納して圧縮データを生成する。ファイルライト部 1 1 4 は、バッファに生成された圧縮データを基に圧縮ファイル 1 1 を生成する。

#### 【0038】



(実施例１の符号化処理の流れ)

次に、実施例１の符号化処理の流れについて説明する。図７は、実施例１の符号化処理の流れの例を示す図である。図７の例に示すように、情報処理装置１００は、最初に、前処理をおこなう（ステップＳ１０）。例えば、情報処理装置１００は、前処理においてビットフィルタ１２１を保持する領域や、動的辞書１２２を作成する作業領域を確保する。次いで、ファイルリード部１１１は、対象文章データ１０を読み出す（ステップＳ１１）。次いで、判定部１１２は、読み出された対象文章データ１０から抽出された符号化対象データが、単語の変化形であるか否かを判定する（ステップＳ１２）。そして、抽出された符号化対象データが単語の変化形でない場合（ステップＳ１２，Ｎｏ）、ステップＳ２１の処理に移行する。

10

【００３９】

一方、抽出された符号化対象データが単語の変化形である場合（ステップＳ１２，Ｙｅｓ）、判定部１１２は、抽出された単語の基本形および変化種別を判定する（ステップＳ１３）。そして、動的符号付与部１１３は、対象文章データ１０から抽出された単語の変化形に対応する動的符号がビットフィルタ１２１に登録されているか否かを判定する（ステップＳ１４）。そして、動的符号付与部１１３は、ビットフィルタ１２１に動的符号が登録されている場合（ステップＳ１４，Ｙｅｓ）、ステップＳ１９の処理に移行する。

【００４０】

一方、動的符号付与部１１３は、ビットフィルタ１２１に動的符号が登録されていない場合（ステップＳ１４，Ｎｏ）、ビットフィルタ１２１から抽出された単語の基本形に対応する基本コードを取得する（ステップＳ１５）。次いで、動的符号付与部１１３は、ビットフィルタ１２１から取得した基本コードと、単語の変化形における変化種別と、動的辞書１２２に単語の変化形に登録する順に付与される動的符号とを対応付けて動的辞書１２２に登録する（ステップＳ１６）。例えば、動的符号付与部１１３は、単語の変化形が動的辞書１２２に登録される順番に、単語の変化形（すなわち、単語の基本コードと変化種別との組み合わせ）に動的符号「A100h」「A101h」「A102h」「A103h」「A104h」「A105h」…を付与する。動的符号付与部１１３は、動的辞書１２２から登録された動的符号を取得する（ステップＳ１７）。次いで、動的符号付与部１１３は、動的辞書１２２から取得された動的符号を、単語の基本形に対応づけてビットフィルタ１２１に登録する（ステップＳ１８）。

20

30

【００４１】

次いで、動的符号付与部１１３は、ビットフィルタ１２１から単語の変化形に対応する動的符号を取得する（ステップＳ１９）。そして、ファイルライト部１１４は、ビットフィルタ１２１から取得された動的符号を圧縮ファイル１１に書き込む（ステップＳ２０）。

【００４２】

次いで、ファイルリード部１１１は、ファイルの読み出し位置がファイルの終端であるか否かを判定する（ステップＳ２１）。ファイルリード部１１１は、読み出し位置がファイルの終端である場合（ステップＳ２１，Ｙｅｓ）、処理を終了させる。一方、ファイルリード部１１１は、読み出し位置がファイルの途中である場合（ステップＳ２１，Ｎｏ）、ステップＳ１１の処理に戻る。

40

【００４３】

以上のように実施例１の情報処理装置１００は、対象文章データ１０に含まれる各単語の変化形に動的符号を割り当てるので、符号化処理時に単語の変化形に割り当てる符号長を短くできる。

【００４４】

(効果)

圧縮部１１０は、対象文章データ１０に含まれる符号化対象データを符号化する際に、符号化対象データが単語の変化形であるか否かを判定し、符号化対象データが単語の変化形である場合、かかる符号化対象データと符号とを対応付けて動的辞書１２２に登録する

50

。これにより、出現頻度の高い単語に個別の符号を割り当てて符号化する場合の圧縮率を向上することができる。

#### 【0045】

圧縮部110における判定する処理は、符号化対象データが静的辞書（ビットフィルタ121）に登録されている単語の変化形であるか否かを判定する。これにより、出現頻度の高い単語に個別の符号を割り当てて、高い圧縮率で符号化することができる。

#### 【実施例2】

#### 【0046】

図8を用いて、実施例2のインデックス生成処理に係るシステム構成について説明する。図8は、実施例2のインデックス生成処理に係るシステム構成の例を示す図である。図8の例に示されるように、情報処理装置200は、圧縮部210と、記憶部220とを有する。圧縮部210は、ファイルリード部211、判定部212、動的符号付与部213およびファイルライト部214を有する。記憶部220は、ビットフィルタ221、動的辞書222およびビットマップ型転置インデックス223を記憶する。なお、実施例1と同じ構成に関しては、番号の下2桁を同一にして適宜説明を省略する。

10

#### 【0047】

圧縮部210は、抽出処理で単語の基本形および変化形が抽出された際に、複数の圧縮ファイルのうち単語の基本形および変化形が含まれる圧縮ファイルを示す情報ビット列を単語の基本形および変化形ごとに対応付けて保持するビットマップ型転置インデックス223を更新する。なお、ファイルライト部214は、インデックス登録部の一例であり、ビットマップ型転置インデックス223はインデックスの一例である。以下、実施例2の圧縮部210および記憶部220の構成について詳細に説明する。

20

#### 【0048】

実施例2の情報処理装置200は、記憶部220がビットマップ型転置インデックス223を記憶する点で、実施例1の情報処理装置100と異なる。図9を用いてビットマップ型転置インデックス223のデータ構造について説明する。図9は、ビットマップ型転置インデックス223の一例を示す図である。図9の例に示すように、ビットマップ型転置インデックス223は、単語の基本形にかかる基本コードや単語の変化形に係る動的付与部ごとにビットマップを対応付ける。ビットマップとは、単語の基本形および変化形がいずれの圧縮ファイルに含まれるかを表す符号ビット列である。ビットマップの各ビットが、各圧縮ファイルに単語の基本形または変化形が含まれているか否かを表す。

30

#### 【0049】

ビットマップ型転置インデックス223は、例えば、出現頻度の高い単語の基本形、およびかかる単語の変化形ごとにビットマップを対応付ける。出現頻度の高い単語とは、出現頻度集計用のテキストファイル群において各単語の出現頻度を集計した場合に、出現頻度の高い単語を表す。

#### 【0050】

例えば、ビットマップ型転置インデックス223の有効行1行目は、基本コード「25h」に対応する単語の基本形「a」のビットマップが「…0110111101」となっている。ビットマップ型転置インデックス223の有効行1行目のビットマップは、単語の基本形「a」の基本コードが含まれるファイルを表す。ビットマップ「…0110111101」は、レジスタの0番に「1」が格納されているのでファイル1に単語の基本形「a」が含まれ、レジスタの1番に「0」が格納されているのでファイル2に単語の基本形「a」が含まれず、レジスタの2番に「1」が格納されているのでファイル3に単語の基本形「a」が含まれることを表す。また、ビットマップ「…0110111101」は、レジスタの3番に「1」が格納されているのでファイル4に単語の基本形「a」が含まれ、レジスタの4番に「1」が格納されているのでファイル5に単語の基本形「a」が含まれていることを表す。なお、ビットマップ「…0110111101」は、ファイル6以降の他の各ファイルに単語の基本形「a」が含まれるか否かについても表す。

40

#### 【0051】

50

次に、ビットマップ型転置インデックス 2 2 3 の更新について説明する。ファイルライト部 2 1 4 は、動的符号付与部 2 1 3 から受け付けた基本コードや動的符号がビットマップ型転置インデックス 2 2 3 に登録されているか否かを判定する。ファイルライト部 2 1 4 は、ビットマップ型転置インデックス 2 2 3 に、受け付けた基本コードや動的符号に対応するビットマップが登録されている場合、受け付けた基本コードや動的符号に対応するビットマップを参照する。ファイルライト部 2 1 4 は、参照したビットマップのうち、対象ファイルに対応するビットが「0」の場合、ビットを「1」に更新する。なお、ファイルライト部 2 1 4 は、対象ファイルに対応するビットが「1」の場合、ビットマップを更新しない。

**【0052】**

一方、ファイルライト部 2 1 4 は、受け付けた基本コードや動的符号に対応するビットマップがビットマップ型転置インデックス 2 2 3 に登録されていない場合、ビットマップ型転置インデックス 2 2 3 に新しくビットマップを登録する。

**【0053】**

具体的には、ファイルライト部 2 1 4 は、動的符号付与部 2 1 3 が対象ファイル中の単語の変化形を動的辞書 2 2 2 に登録した場合に、単語の変化形に付与された動的符号を取得する。かかる場合において、ファイルライト部 2 1 4 は、取得した動的符号に係るビットマップをビットマップ型転置インデックス 2 2 3 に登録する。このビットマップには、対象ファイルの数分のビット「0」が含まれる。さらに、ファイルライト部 2 1 4 は、登録したビットマップのビットのうち、対象ファイルに対応するビットを「1」に更新する。すなわち、ファイルライト部 2 1 4 は、対象ファイルにおいて初出である単語の変化形を動的辞書 2 2 2 に登録した際に、ビットマップ型転置インデックス 2 2 3 に登録された単語の変化形に対応するビットマップを登録する。このようにしてビットマップ型転置インデックス 2 2 3 を生成する。

**【0054】**

（実施例 2 のインデックス生成処理の流れ）

次に、実施例 2 のインデックス生成処理の流れについて説明する。図 1 0 は、実施例 2 のインデックス生成処理の流れの例を示す図である。図 1 0 の例に示すように、情報処理装置 2 0 0 は、最初に、前処理をおこなう（ステップ S 3 0）。例えば、情報処理装置 2 0 0 は、前処理においてビットフィルタ 2 2 1 を保持する領域や、動的辞書 2 2 2 を作成する作業領域を確保する。次いで、ファイルリード部 2 1 1 は、対象文章データ 1 0 を読み出す（ステップ S 3 1）。次いで、判定部 2 1 2 は、読み出された対象文章データ 1 0 から抽出された符号化対象データが、単語の変化形であるか否かを判定する（ステップ S 3 2）。そして、抽出された符号化対象データが単語の変化形でない場合（ステップ S 3 2, No）、ステップ S 4 2 の処理に移行する。

**【0055】**

一方、抽出された符号化対象データが単語の変化形である場合（ステップ S 3 2, Yes）、判定部 2 1 2 は、抽出された単語の基本形および変化種別を判定する（ステップ S 3 3）。そして、動的符号付与部 2 1 3 は、対象文章データ 1 0 から抽出された単語の変化形に対応する動的符号がビットフィルタ 2 2 1 に登録されているか否かを判定する（ステップ S 3 4）。そして、動的符号付与部 2 1 3 は、ビットフィルタ 2 2 1 に動的符号が登録されている場合（ステップ S 3 4, Yes）、ステップ S 3 9 の処理に移行する。

**【0056】**

一方、動的符号付与部 2 1 3 は、ビットフィルタ 2 2 1 に動的符号が登録されていない場合（ステップ S 3 4, No）、ビットフィルタ 2 2 1 から抽出された単語の基本形に対応する基本コードを取得する（ステップ S 3 5）。次いで、動的符号付与部 2 1 3 は、ビットフィルタ 2 2 1 から取得した基本コードと、単語の変化形における変化種別と、動的辞書 2 2 2 に単語の変化形を登録する順に付与される動的符号とを対応付けて動的辞書 2 2 2 に登録する（ステップ S 3 6）。例えば、動的符号付与部 2 1 3 は、単語の変化形が動的辞書 2 2 2 に登録される順番に、単語の変化形（すなわち、単語の基本コードと変化

種別との組み合わせ)に動的符号「A100h」「A101h」「A102h」「A103h」「A104h」「A105h」…を付与する。動的符号付与部213は、動的辞書222から登録された動的符号を取得する(ステップS37)。次いで、動的符号付与部213は、動的辞書222から取得された動的符号を、単語の基本形に対応づけてビットフィルタ221に登録する(ステップS38)。次いで、動的符号付与部213は、ビットフィルタ221から単語の変化形に対応する動的符号を取得する(ステップS39)。

#### 【0057】

次に、ファイルライト部214は、動的符号付与部213が取得した動的符号、およびかかる動的符号に対応付けられた基本コードを基にして、ビットマップ型転置インデックス223を更新する(ステップS40)。例えば、ファイルライト部214は、ビットマップ型転置インデックス223に動的符号に対応するビットマップが登録されている場合、ビットマップに含まれるビットのうち、対象ファイルに対応するビットを「1」に更新する。一方、ファイルライト部214は、ビットマップ型転置インデックス223に動的符号に対応するビットマップが登録されていない場合、ビットマップ型転置インデックス223に新しく動的符号に対応するビットマップを登録する。

#### 【0058】

そして、ファイルライト部214は、ビットフィルタ221から取得された動的符号を圧縮ファイル11に書き込む(ステップS41)。

#### 【0059】

次いで、ファイルリード部211は、ファイルの読み出し位置がファイルの終端であるか否かを判定する(ステップS42)。ファイルリード部211は、読み出し位置がファイルの終端である場合(ステップS42, Yes)、処理を終了させる。一方、ファイルリード部211は、読み出し位置がファイルの途中である場合(ステップS42, No)、ステップS31の処理に戻る。

#### 【0060】

以上のように実施例2の情報処理装置200は、ファイル圧縮を行う際にビットマップ型転置インデックス223を生成する。これにより、情報処理装置200は、複数の圧縮ファイルを基に単語の検索をおこなう際に、検索対象のうち、単語の基本形または変化形を有するファイルを特定でき、オープンする圧縮ファイルを絞り込むことができるので単語検索を高速化することができる。

#### 【0061】

(効果)

圧縮部210は、対象文章データ10に含まれる符号化対象データを符号化する際に、符号化対象データが単語の変化形であるか否かを判定し、符号化対象データが単語の変化形である場合、かかる符号化対象データと符号とを対応付けて動的辞書221に登録し、単語の基本形および変化形と、符号化対象データとを対応付けてインデックス(ビットマップ型転置インデックス223)に登録する。これにより、複数の圧縮ファイルを基に単語検索をおこなう際に、検索対象のうち単語の基本形または変化形を有するファイルを特定でき、オープンする圧縮ファイルを絞り込むことができるので単語検索を高速化することができる。

#### 【実施例3】

#### 【0062】

図11を用いて、実施例3の検索処理に係るシステム構成について説明する。図11は、実施例3の検索処理に係るシステム構成の例を示す図である。図11の例に示されるように、情報処理装置300は、圧縮部310と、記憶部320と、検索処理部330とを有する。圧縮部310は、ファイルリード部311、判定部312、動的符号付与部313およびファイルライト部314を有する。記憶部320は、ビットフィルタ321、動的辞書322およびビットマップ型転置インデックス323を記憶する。検索処理部330は、受付部331、検索キー判定部332および検索部333を有する。なお、ビットマップ型転置インデックス323はインデックスの一例である。また、実施例1、2と同

10

20

30

40

50

じ構成に関しては、番号の下 2 桁を同一にして適宜説明を省略する。

#### 【0063】

検索処理部 330 は、ビットマップ型転置インデックス 323 が作成された際に、かかるビットマップ型転置インデックス 323 を参照して、複数の圧縮ファイルの中から検索キーに対応する単語を検索する。以下、実施例 3 の検索処理部 330 の構成について詳細に説明する。

#### 【0064】

実施例 3 の情報処理装置 300 は、検索処理部 330 を有する点で、実施例 2 の情報処理装置 200 と異なる。受付部 331 は、実施例 2 に記載の方法で符号化された対象文章データ 10 に対する検索要求を受け付ける。検索キー判定部 332 は、かかる検索要求に含まれる検索キーが、単語の変化形であるか否かを判定する。検索部 333 は、かかる検索キーが単語の変化形である場合に、ビットマップ型転置インデックス 323 を参照して検索要求に対応した検索を行う。

#### 【0065】

次に、実施例 3 に係る単語の検索処理の流れを説明する。図 12 は、圧縮ファイル選択の処理の流れの一例を示す図である。図 12 の例に示すように、例えば、検索キーが「act」であった場合、検索キー判定部 332 は、かかる検索キーが単語の変化形であるか否かを判定する。ここで、検索キー「act」は単語の変化形ではなく基本形であることから、検索部 333 は、ビットフィルタ 321 からかかる検索キーに対応する基本コード「6041h」を取得し、取得された基本コードのビットマップを抽出する。次いで、検索部 333 は、検索範囲内で抽出されたビットマップの AND 演算を行い、選択結果マップ「…0110000111」を生成する。かかる生成された選択結果マップ「…0110000111」は、単語「act」を含む圧縮ファイルを示す。

#### 【0066】

ここで、図 12 の例に示すように、フォルダ 51 には、複数の圧縮ファイル 52a、52b、52c…が含まれる。検索部 333 は、選択結果マップが「…0110000111」となっているので、フォルダ 51 に含まれるファイルのうち、1 番目の圧縮ファイル 52a と、2 番目の圧縮ファイル 52b と、3 番目の圧縮ファイル 52c とをオープンする。なお、検索部 333 は、他にも 8 番目、9 番目のファイルをオープンする。検索部 333 によってオープンされた圧縮ファイル 52a には、例えば検索キー「act」に属する単語「acted」が格納されている。

#### 【0067】

図 13 は、圧縮ファイル選択の処理の流れの別の例を示す図である。図 13 の例に示すように、例えば、検索キーが「acting」であった場合、検索キー判定部 332 は、かかる検索キーが単語の変化形であるか否かを判定する。ここで、検索キー「acting」は単語の変化形であることから、検索キー判定部 332 は、かかる検索キー「acting」の基本形「act」および変化種別「進行形」を判定する。つづいて、検索部 333 は、かかる検索キーに対応する基本コード「6041h」および動的符号「A104h」を取得し、取得された基本コードおよび動的符号のビットマップを抽出する。次いで、検索部 333 は、検索範囲内で抽出されたビットマップの AND 演算を行い、選択結果マップ「…0000000010」を生成する。かかる生成された選択結果マップ「…0000000010」は、単語「acting」を含む圧縮ファイルを示す。

#### 【0068】

ここで、検索部 333 は、選択結果マップが「…0000000010」となっているので、フォルダ 51 に含まれるファイルのうち、2 番目の圧縮ファイル 52b をオープンする。検索部 333 によってオープンされた圧縮ファイル 52b には、検索キー「acting」に属する単語「acting」が格納されている。

#### 【0069】

(実施例 3 の検索処理の流れ)

次に、実施例 3 の検索処理の流れについて説明する。図 14 は、実施例 3 の検索処理の

流れの例を示す図である。図14の例に示すように、情報処理装置300は、前処理をおこなう（ステップS50）。例えば、情報処理装置300は、前処理として選択結果マップを生成するための作業領域を確保する。

#### 【0070】

次いで、受付部331は、検索要求を検索キーとして受け付ける（ステップS51）。そして、検索キー判定部332は、受け付けられた検索キーが単語の変化形であるか否かを判定する（ステップS52）。ここで、受け付けられた検索キーが単語の変化形でない場合（ステップS52, No）、検索キー判定部332は、かかる検索キーとして受け付けられた単語の基本形に対応する基本コードを取得する（ステップS53）。一方、受け付けられた検索キーが単語の変化形である場合（ステップS52, Yes）、検索キー判定部332は、かかる検索キーとして受け付けられた単語の変化形に対応する動的符号と、かかる単語の基本形に対応する基本コードとを取得する（ステップS54）。 10

#### 【0071】

次いで、検索部333は、取得された基本コードや動的符号に基づいて、検索範囲内に属するビットマップをビットマップ型転置インデックス323から抽出する（ステップS55）。つづいて、検索部333は、ビットマップ型転置インデックス323から抽出した複数のビットマップ間で検索クエリの複数の単語に対応したAND演算をおこなうことで選択結果マップを生成する（ステップS56）。そして、検索部333は、生成された選択結果マップに基づいてオープンする圧縮ファイルを選択し、選択された圧縮ファイルに関する検索結果を表示して（ステップS57）、処理を終了させる。 20

#### 【0072】

（効果）

情報処理装置300は、対象文章データ10を符号化した符号化文章データに対する検索要求を受け付け、かかる検索要求に含まれる検索キーが、単語の変化形であるか否かを判定し、検索キーが単語の変化形である場合、単語の基本形および変化形と、対象文章データ10に含まれる符号化対象データとを対応付けたインデックス（ビットマップ型転置インデックス323）を参照して検索要求に対応した検索を行う。これにより、単語の基本形および変化形といった類義語における検索の高速化を図ることができる。

#### 【0073】

（実施例1、2に関連する他の態様） 30

以下、上述の実施形態における変形例の一部を説明する。下記の変形例のみでなく、本発明の本旨を逸脱しない範囲の設計変更は適宜行われうる。

#### 【0074】

例えば、上述の実施例1～3では、ビットフィルタ121、221、321に登録される単語の基本形として、かかる単語の現在形を用いた例について示したが、単語の基本形はかかる単語の現在形に限られず、単語の変化形のうちいずれか1種類を用いてもよい。しかしながら、単語の基本形にかかる単語の現在形を用いることにより、単語の各種変化形といった類義語を容易に検索することができる。

#### 【0075】

また、符号化処理の対象は、ファイル内のデータ以外にも、システムから出力される監視メッセージなどでもよい。例えば、バッファに順次格納される監視メッセージを上述の符号化処理により圧縮し、ログファイルとして格納するなどの処理が行なわれる。また、例えば、データベース内のページ単位に圧縮が行なわれてもよいし、複数のページをまとめた単位で圧縮が行なわれてもよい。 40

#### 【0076】

また、実施例1に示した処理手順、制御手順、具体的名称、各種のデータやパラメータを含む情報については、特記する場合を除いて任意に変更することができる。

#### 【0077】

（情報処理装置のハードウェア構成）

図15は、実施例1～3の情報処理装置100、200、300のハードウェア構成を 50

示す図である。図15の例に示すように、コンピュータ400は、各種演算処理を実行するCPU401と、ユーザからのデータ入力を受け付ける入力装置402と、モニタ403とを有する。また、コンピュータ400は、記憶媒体からプログラム等を読み取る媒体読取装置404と、他の装置と接続するためのインターフェース装置405と、他の装置と無線により接続するための無線通信装置406とを有する。また、コンピュータ400は、各種情報を一時記憶するRAM407と、ハードディスク装置408とを有する。また、各装置401～408は、バス409に接続される。

#### 【0078】

ハードディスク装置408には、例えば図3に示したファイルリード部111、判定部112、動的符号付与部113およびファイルライト部114の各処理部と同様の機能を有する情報処理プログラムが記憶される。また、ハードディスク装置408には、情報処理プログラムを実現するための各種データが記憶される。

10

#### 【0079】

CPU401は、ハードディスク装置408に記憶された各プログラムを読み出して、RAM407に展開して実行することで各種の処理を行う。これらのプログラムは、コンピュータ400を、例えば図3に示したファイルリード部111、判定部112、動的符号付与部113およびファイルライト部114として機能させることができる。

#### 【0080】

なお、上記の情報処理プログラムは、必ずしもハードディスク装置408に記憶されている必要はない。例えば、コンピュータ400が読み取り可能な記憶媒体に記憶されたプログラムを、コンピュータ400が読み出して実行するようにしてもよい。コンピュータ400が読み取り可能な記憶媒体は、例えば、CD-ROMやDVDディスク、USB (Universal Serial Bus) メモリ等の可搬型記録媒体、フラッシュメモリ等の半導体メモリ、ハードディスクドライブ等が対応する。また、公衆回線、インターネット、LAN (Local Area Network) 等に接続された装置にこのプログラムを記憶させておき、コンピュータ400がこれらからプログラムを読み出して実行するようにしてもよい。

20

#### 【0081】

図16は、コンピュータ400で動作するプログラムの構成例を示す図である。コンピュータ400において、図15に示すハードウェア群26 (401～409) の制御を行なうOS (オペレーティング・システム) 27が動作する。OS 27に従った手順でCPU401が動作して、ハードウェア群26の制御・管理が行なわれることにより、アプリケーションプログラム29やミドルウェア28に従った処理がハードウェア群26で実行される。さらに、コンピュータ400において、ミドルウェア28またはアプリケーションプログラム29が、RAM407に読み出されてCPU401により実行される。

30

#### 【0082】

CPU401により圧縮機能が呼び出された場合、ミドルウェア28またはアプリケーションプログラム29の少なくとも一部に基づく処理を行なうことで、(それらの処理をOS 27に基づいてハードウェア群26を制御して) 圧縮部110の機能が実現される。圧縮機能は、それぞれアプリケーションプログラム29自体に含まれてもよいし、アプリケーションプログラム29に従って呼び出されることで実行されるミドルウェア28の一部であってもよい。

40

#### 【0083】

アプリケーションプログラム29 (またはミドルウェア28) の圧縮機能により得られる圧縮ファイル11は、部分的に伸張することも可能である。圧縮ファイル11の途中を伸張する場合には、伸張対象の部分までの圧縮データの伸張処理が抑制されるため、CPU401の負荷が抑制される。また、伸張対象の圧縮データを部分的にRAM407上に展開するので、ワークエリアも削減される。

#### 【0084】

図17は、実施形態のシステムにおける装置の構成例を示す図である。図17のシステムは、コンピュータ400a、コンピュータ400b、基地局30およびネットワーク4

50

0を含む。コンピュータ400aは、無線または有線の少なくとも一方により、コンピュータ400bと接続されたネットワーク40に接続している。

【符号の説明】

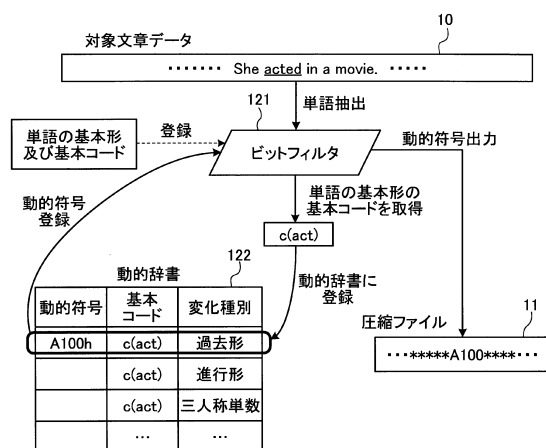
【0085】

|     |                 |
|-----|-----------------|
| 100 | 情報処理装置          |
| 110 | 圧縮部             |
| 111 | ファイルリード部        |
| 113 | 動的符号付与部         |
| 120 | 記憶部             |
| 121 | ビットフィルタ         |
| 122 | 動的辞書            |
| 223 | ビットマップ型転置インデックス |
| 330 | 検索処理部           |
| 331 | 受付部             |
| 332 | 検索キー判定部         |
| 333 | 検索部             |

10

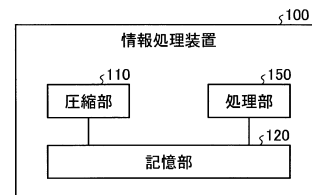
【図1】

実施例1の単語の符号化処理を説明するための図



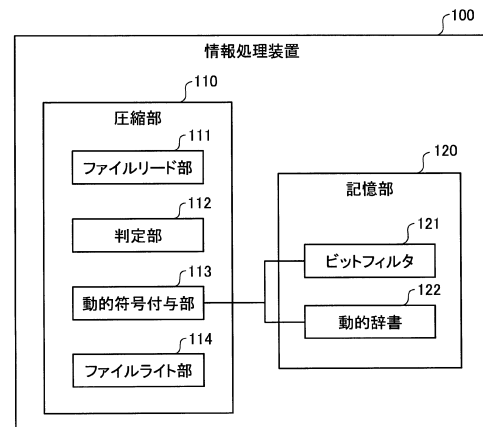
【図2】

情報処理装置のシステム構成の例を説明するための図



【図3】

実施例1の符号化処理に係るシステム構成の例を示す図





【図 4】

ビットフィルタの第1の例を示す図

| ビットフィルタ   |           |        |      |      |      |       |      |  |  |
|-----------|-----------|--------|------|------|------|-------|------|--|--|
| 静的辞書      |           |        |      |      | 動的辞書 |       |      |  |  |
| 各単語へのポイント |           |        |      |      | 動的辞書 |       |      |  |  |
| 2グラム      | ビットマップ    | 単語     | 文字列長 | 出現頻度 | 符号長  | 基本コード | 動的符号 |  |  |
| aa        | 0,0,0,0,0 | a      | 0    | 9738 | 8    | 25h   | 動的符号 |  |  |
| ab        | 1,0,0,0,0 | able   | 4    | 785  | 16   | 6005h | 動的符号 |  |  |
| ac        | 1,0,0,0,0 | about  | 6    | 17   | 16   | 6008h | 動的符号 |  |  |
| ..        | 0,1,0,0,0 | above  | 5    | 539  | 16   | 6028h | 動的符号 |  |  |
| aΔ        | 0,1,0,0,0 | acted  | 3    | 4832 | 16   | 6041h | 動的符号 |  |  |
| bl        | 0,1,0,0,0 | acting | 6    | 4832 | 16   | 6041h | 動的符号 |  |  |
| bo        | 0,1,0,0,0 | as     | 2    | 4832 | 16   | 32h   | 動的符号 |  |  |
| ct        | 0,1,0,0,0 | eat    | 3    | 4832 | 16   | 60F5h | 動的符号 |  |  |
| eΔ        | 0,0,0,1,1 | eating | 6    | 4832 | 16   | 60F5h | 動的符号 |  |  |
| le        | 0,0,1,0,0 |        |      |      |      |       |      |  |  |
| tΔ        | 0,0,1,0,0 |        |      |      |      |       |      |  |  |

【図 5】

動的辞書の一例を示す図

| 動的符号  | 基本コード | 変化種別  | (単語)     |
|-------|-------|-------|----------|
| A100h | 6041h | 過去形   | (acted)  |
|       | 6041h | 進行形   | (acting) |
|       | 6041h | 三人称単数 | (acts)   |
|       | ...   | ...   |          |
|       | 60F5h | 過去形   | (ate)    |
|       | 60F5h | 過去分詞  | (eaten)  |
| A101h | 60F5h | 進行形   | (eating) |
|       | 60F5h | 三人称単数 | (eats)   |
|       | ...   | ...   |          |
|       | ...   | ...   |          |

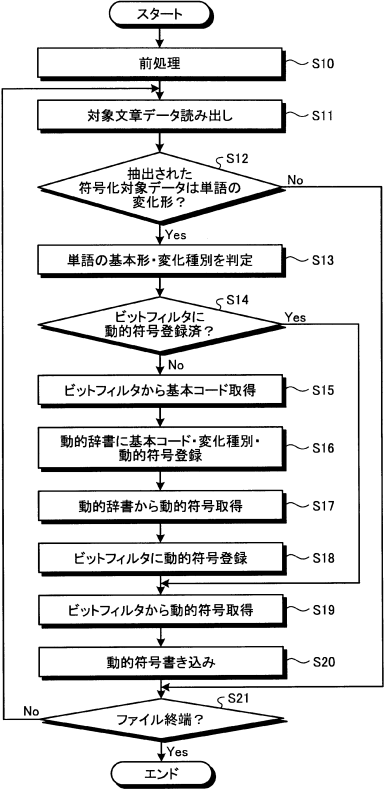
【図 6】

ビットフィルタの第2の例を示す図

| ビットフィルタ   |           |        |      |      |      |       |      |  |  |
|-----------|-----------|--------|------|------|------|-------|------|--|--|
| 静的辞書      |           |        |      |      | 動的辞書 |       |      |  |  |
| 各単語へのポイント |           |        |      |      | 動的辞書 |       |      |  |  |
| 2グラム      | ビットマップ    | 単語     | 文字列長 | 出現頻度 | 符号長  | 基本コード | 動的符号 |  |  |
| aa        | 0,0,0,0,0 | a      | 0    | 9738 | 8    | 25h   | 動的符号 |  |  |
| ab        | 1,0,0,0,0 | able   | 4    | 785  | 16   | 6005h | 動的符号 |  |  |
| ac        | 1,0,0,0,0 | about  | 6    | 17   | 16   | 6008h | 動的符号 |  |  |
| ..        | 0,1,0,0,0 | above  | 5    | 539  | 16   | 6028h | 動的符号 |  |  |
| aΔ        | 0,1,0,0,0 | acted  | 3    | 4832 | 16   | 6041h | 動的符号 |  |  |
| bl        | 0,1,0,0,0 | acting | 6    | 4832 | 16   | 6041h | 動的符号 |  |  |
| bo        | 0,1,0,0,0 | as     | 2    | 4832 | 16   | 32h   | 動的符号 |  |  |
| ct        | 0,1,0,0,0 | eat    | 3    | 4832 | 16   | 60F5h | 動的符号 |  |  |
| eΔ        | 0,0,0,1,1 | eating | 6    | 4832 | 16   | 60F5h | 動的符号 |  |  |
| le        | 0,0,1,0,0 |        |      |      |      |       |      |  |  |
| tΔ        | 0,0,1,0,0 |        |      |      |      |       |      |  |  |

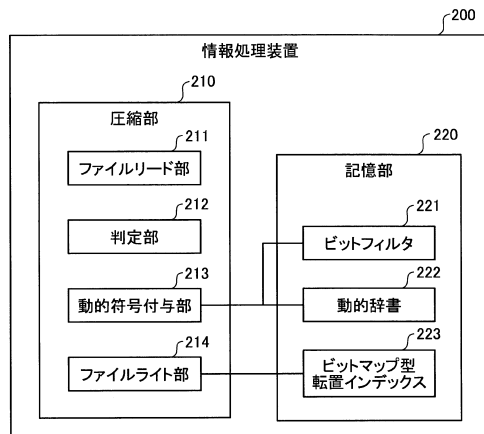
【図 7】

実施例1の符号化処理の流れの例を示す図



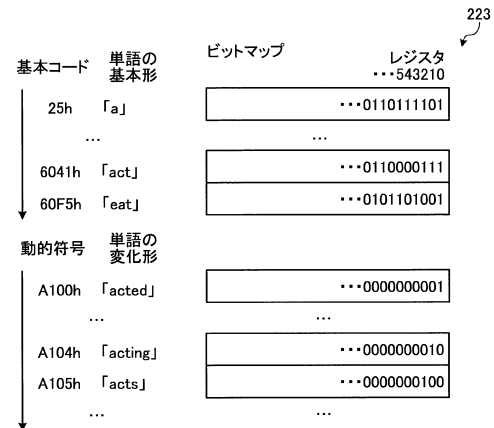
【図 8】

実施例2のインデックス生成処理に係るシステム構成の例を示す図



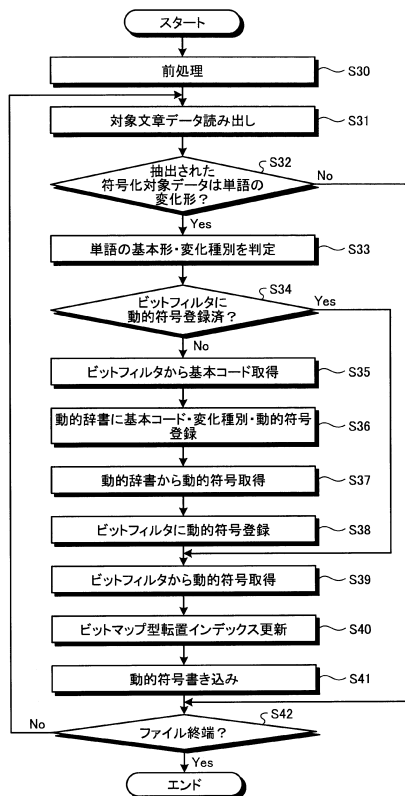
【図 9】

ビットマップ型転置インデックスの一例を示す図



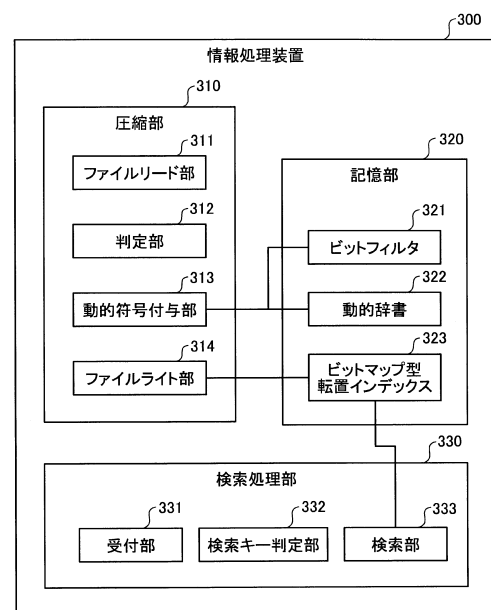
【図 10】

実施例2のインデックス生成処理の流れの例を示す図

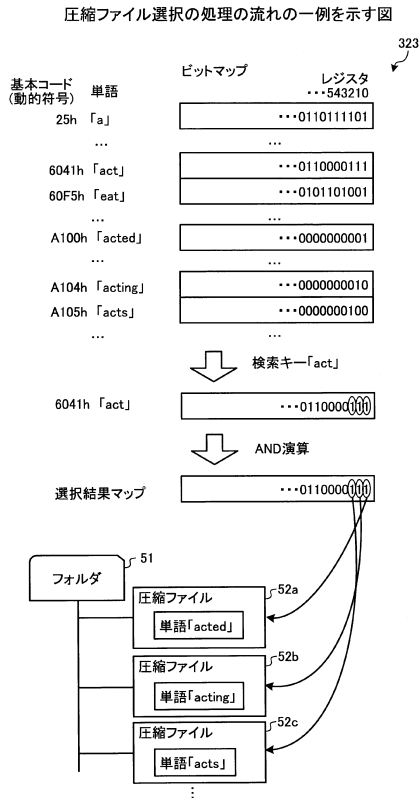


【図 11】

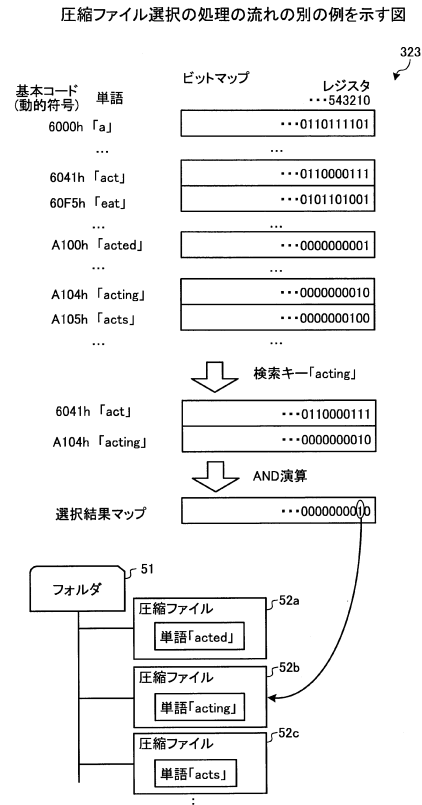
実施例3の検索処理に係るシステム構成の例を示す図



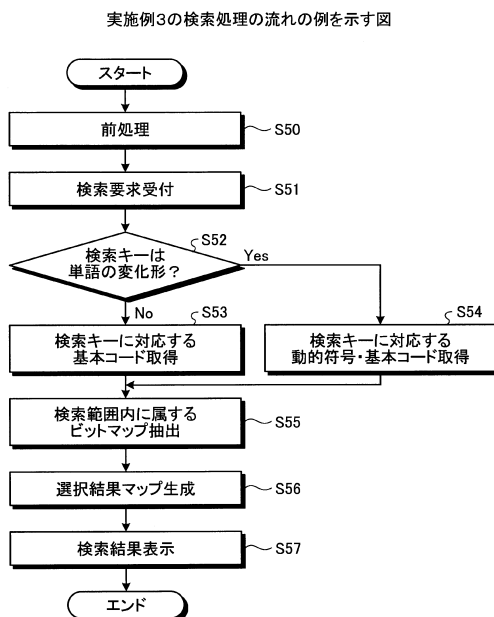
【図 1 2】



【図 1 3】

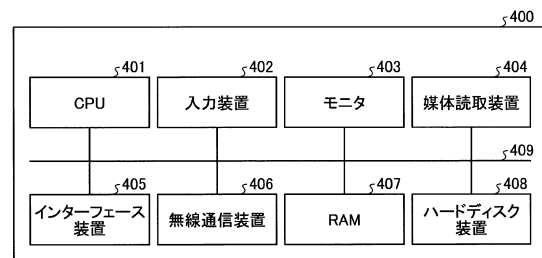


【図 1 4】



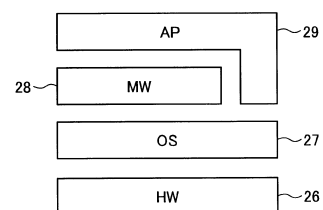
【図 1 5】

実施例1～3の情報処理装置のハードウェア構成を示す図



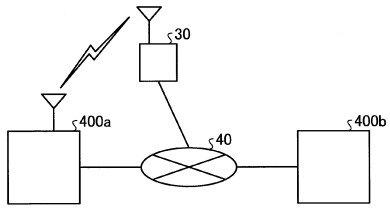
【図 1 6】

コンピュータで動作するプログラムの構成例を示す図



【図 17】

実施形態のシステムにおける装置の構成例を示す図



---

フロントページの続き

審査官 三橋 竜太郎

(56)参考文献 国際公開第2014/147672 (WO, A1)  
国際公開第2017/009958 (WO, A1)

(58)調査した分野(Int.Cl., DB名)  
G06F 16/00-16/958