

(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
2015 年 4 月 2 日(02.04.2015)



(10) 国際公開番号
WO 2015/045155 A1

- (51) 国際特許分類:
G06F 17/21 (2006.01) **G06F 17/30** (2006.01)
- (21) 国際出願番号: PCT/JP2013/076545
- (22) 国際出願日: 2013 年 9 月 30 日(30.09.2013)
- (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: 楽天株式会社(RAKUTEN, INC.) [JP/JP]; 〒1400002 東京都品川区東品川四丁目 1 2 番 3 号 Tokyo (JP).
- (72) 発明者: 新里 圭司(SHINZATO Keiji); 〒1400002 東京都品川区東品川四丁目 1 2 番 3 号 楽天株式会社内 Tokyo (JP).
- (74) 代理人: 長谷川 芳樹, 外(HASEGAWA Yoshiki et al.); 〒1000005 東京都千代田区丸の内二丁目 1 番 1 号丸の内 MY PLAZA (明治安田生命ビル) 9 階 創英国際特許法律事務所 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, 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, 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, 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, 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))

(54) Title: CORPUS GENERATION DEVICE, CORPUS GENERATION METHOD, AND CORPUS GENERATION PROGRAM

(54) 発明の名称: コーパス生成装置、コーパス生成方法、及びコーパス生成プログラム

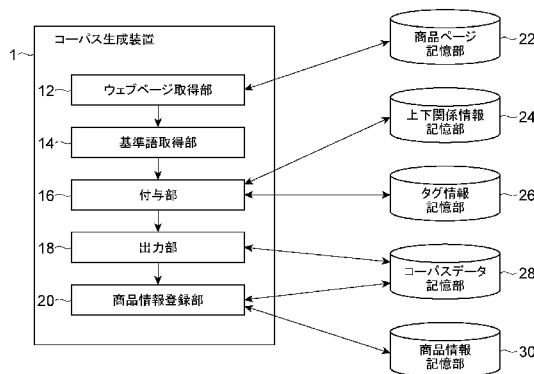


FIG. 1:
1 Corpus generation device
12 Web page acquisition unit
14 Reference term acquisition unit
16 Providing unit
18 Output unit
20 Merchandise information registration unit
22 Merchandise page storage unit
24 Upper-lower relationship information storage unit
26 Tag information storage unit
28 Corpus data storage unit
30 Merchandise information storage unit

(57) Abstract: A corpus generation device according to an embodiment is provided with: a Web page acquisition unit; a reference term acquisition unit; a providing unit; and an output unit. The Web page acquisition unit acquires a Web page including description data regarding a presentation object. The reference term acquisition unit acquires from the Web page a reference term which is an attribute value regarding the presentation object. The providing unit extracts, from a storage unit storing upper/lower relationship information representing an upper/lower relationship between attribute values, an upper term belonging to a higher level than the reference term acquired by the reference term acquisition unit, and provides an attribute tag corresponding to the reference term to the upper term included in the description data. The output unit outputs the description data to which the attribute tag has been provided by the providing unit as corpus data.

(57) 要約: 実施形態に係るコーパス生成装置は、ウェブページ取得部、基準語取得部、付与部及び出力部を備える。ウェブページ取得部は、提示対象に関する説明文データを含むウェブページを取得する。基準語取得部は、ウェブページから提示対象に関する属性値である基準語を取得する。付与部は、属性値間の上下関係を表す上下関係情報を記憶する記憶部から、基準語取得部により取得された基準語よりも上位に属する上位語を抽出し、説明文データに含まれる上位語に対して、基準語に対応する属性タグを付与する。出力部は、付与部により属性タグが付与された説明文データをコーパスデータとして出力する。

明 細 書

発明の名称：

コーパス生成装置、コーパス生成方法、及びコーパス生成プログラム

技術分野

[0001] 本発明の一形態は、コーパス生成装置、コーパス生成方法、及びコーパス生成プログラムに関する。

背景技術

[0002] 電子商取引サイトでは、提示対象である商品の特徴等を記載した商品の説明文を含む商品ページがユーザに対して提示される。従来から、商品の説明文に含まれる商品の特徴を示す語句にタグを自動的に付与してコーパスデータを生成することが試みられている。このようなコーパスデータによれば、タグを手がかりとして機械処理することにより商品に関する構造化されたデータを自動生成することが可能となる。

[0003] 商品の説明文等の自然文に対して自動的にタグを付与する技術として、例えば特許文献１に記載の技術が知られている。特許文献１には、辞書データを参照して入力文書から汎用的な固有表現及び意味役割語を抽出し、汎用的な固有表現及び意味役割語に基づいて入力文書の文章構造を推定すると共に、文章構造に対して予め定義されているルールに従い入力文書に意味タグの付与を行う文章処理装置が記載されている。

先行技術文献

特許文献

[0004] 特許文献１：特開２００７－９４８５５号公報

発明の概要

発明が解決しようとする課題

[0005] しかし、特許文献１に記載の技術では、辞書を参照して文章に含まれる固有表現及び意味役割語等の語句を抽出している。このため、特許文献１に記載の技術を用いて提示対象の説明文にタグを付与しようとした場合、提示対

象の説明文の内容によっては提示対象の特徴を示すものでない語句に対して誤って属性タグを付与してしまい、その結果、コーパスデータの品質を低下させてしまう場合がある。したがって、提示対象の特徴を示すものではない語句に対して属性タグを誤って付与することを抑制することができる技術が要請されている。

課題を解決するための手段

[0006] 本発明の一側面に係るコーパス生成装置は、提示対象に関する説明文データを含むウェブページを取得するウェブページ取得部と、ウェブページから提示対象に関する属性値である基準語を取得する基準語取得部と、属性値間の上下関係を表す上下関係情報を記憶する記憶部から、基準語取得部により取得された基準語よりも上位に属する上位語を抽出し、説明文データに上位語が含まれる場合には、説明文データに含まれる上位語に対して、基準語に対応する属性タグを付与する付与部と、付与部により属性タグが付与された説明文データをコーパスデータとして出力する出力部と、を備える。

[0007] 本発明の一側面に係るコーパス生成方法は、提示対象に関する説明文データを含むウェブページを取得するウェブページ取得ステップと、ウェブページから提示対象に関する属性値である基準語を取得する基準語取得ステップと、属性値間の上下関係を表す上下関係情報を記憶する記憶部から、基準語取得ステップにおいて取得された基準語よりも上位に属する上位語を抽出し、説明文データに上位語が含まれる場合には、説明文データに含まれる上位語に対して、基準語に対応する属性タグを付与する付与ステップと、付与ステップにおいて属性タグが付与された説明文データをコーパスデータとして出力する出力ステップと、を含む。

[0008] 本発明の一側面に係るコーパス生成プログラムは、コンピュータを、提示対象に関する説明文データを含むウェブページを取得するウェブページ取得部と、ウェブページから提示対象に関する属性値である基準語を取得する基準語取得部と、属性値間の上下関係を表す上下関係情報を記憶する記憶部から、基準語取得部により取得された基準語よりも上位に属する上位語を抽出

し、説明文データに上位語が含まれる場合には、説明文データに含まれる上位語に対して、基準語に対応する属性タグを付与する付与部と、付与部により属性タグが付与された説明文データをコーパスデータとして出力する出力部と、して機能させる。

[0009] 上記形態によれば、説明文データに提示対象に関する属性値である基準語の上位語が含まれる場合には、説明文データに含まれる上位語に対して基準語に対応する属性タグが付与される。基準語が提示対象に関する属性値である場合には、その上位語も提示対象に関する属性値であることが推認される。このため、説明文データ内の上位語に対して属性タグを付与することにより、提示対象の特徴を示すものではない語句に対して誤って属性タグを付与することを抑制することができる。そして、このように属性タグが付与された提示対象の説明文データをコーパスデータとして出力することにより、品質の高いコーパスデータを生成することができる。

[0010] 一形態では、付与部は、前記説明文データに前記基準語が含まれる場合には、前記説明文データに含まれる前記基準語に対して、該基準語に対応する属性タグを更に付与してもよい。このような形態によれば、説明文データ中のより多くの提示対象の属性値に正確な属性タグを付与することができる。

[0011] 一形態では、ウェブページは、提示対象に関する属性名と属性値とが対応付けられた属性リストを更に含み、基準語取得部は、属性リストにおける属性値を基準語として取得してもよい。属性リストにおける属性値は、提示対象に関する属性値である蓋然性が高い。したがって、本形態によれば適切な基準語を取得することができる。

[0012] 一形態では、基準語取得部は、構文解析器を用いて、説明文データから提示対象に関する属性値である蓋然性が高い語句を検索し、検索された語句を基準語として取得してもよい。このような形態によれば、属性リストを含まないウェブページからでも基準語を取得することができる。

[0013] 一形態では、上下関係情報は、属性値をノードとする木構造をなし、付与部は、上下関係情報から基準語をルートノードとし、且つ分岐を有しない部

分木を検索して、該部分木に含まれる基準語以外の 1 以上の属性値を抽出し、説明文データに 1 以上の属性値が含まれる場合には、説明文データに含まれる該 1 以上の属性値に対して、基準語に対応する属性タグを更に付与してもよい。基準語が提示対象に関する属性値である場合には、分岐を有しない部分木に含まれる属性値も提示対象に関する属性値であることが推認される。したがって、本形態によれば、説明文データ中のより多くの提示対象の属性値に正確な属性タグを付与することができる。

[0014] 一形態では、上下関係情報は、属性値間の上下関係を階層的に表しており、付与部は、上下関係情報において基準語の直下の階層に複数の属性値が存在する場合であって、該複数の属性値の内、一の属性値のみが説明文データに含まれる場合には、説明文データに含まれる該一の属性値に対して、基準語に対応する属性タグを更に付与してもよい。基準語の直下の階層に複数の属性値が存在する場合であっても、説明文に当該複数の属性値のうちの 1 つのみが含まれている場合には、当該属性値は提示対象に関する属性値である蓋然性が高い。したがって、本形態によれば、説明文データ中のより多くの提示対象の属性値に、正確な属性タグを付与することができる。

[0015] 一形態では、出力部から出力されたコーパスデータから取得される提示対象に関する属性名と属性値との組をウェブページに関連付けて記憶する提示対象情報登録部を更に備えてもよい。このような形態によれば、属性名と属性値との組を提示対象の検索の手がかりとして利用することができる。

[0016] 本発明の別の側面に係るコーパス生成装置は、属性値に対して属性タグが付与されたコーパスデータを取得するコーパス取得部と、コーパスデータから提示対象に関する属性値である基準語を取得する基準語取得部と、属性値間の上下関係を表す上下関係情報を記憶する記憶部から、基準語取得部により取得された基準語よりも上位に属する上位語を抽出し、上位語とは異なる属性値がコーパスデータに含まれている場合に、上位語とは異なる属性値に付与された属性タグをコーパスデータから除去する除去部と、除去部により属性タグが除去されたコーパスデータを出力する出力部と、を備える。

[0017] 本発明の別の側面に係るコーパス生成方法は、属性値に対して属性タグが付与されたコーパスデータを取得するコーパス取得ステップと、コーパスデータから提示対象に関する属性値である基準語を取得する基準語取得ステップと、属性値間の上下関係を表す上下関係情報を記憶する記憶部から、基準語取得ステップにおいて取得された基準語よりも上位に属する上位語を抽出し、上位語とは異なる属性値がコーパスデータに含まれている場合に、上位語とは異なる属性値に付与された属性タグをコーパスデータから除去する除去ステップと、除去ステップにおいて属性タグが除去されたコーパスデータを出力する出力ステップと、を含む。

[0018] 本発明の別の側面に係るコーパス生成プログラムは、コンピュータを、属性値に対して属性タグが付与されたコーパスデータを取得するコーパス取得部と、コーパスデータから提示対象に関する属性値である基準語を取得する基準語取得部と、属性値間の上下関係を表す上下関係情報を記憶する記憶部から、基準語取得部により取得された基準語よりも上位に属する上位語を抽出し、上位語とは異なる属性値がコーパスデータに含まれている場合に、上位語とは異なる属性値に付与された属性タグをコーパスデータから除去する除去部と、除去部により属性タグが除去されたコーパスデータを出力する出力部と、して機能させる。

[0019] 上記形態によれば、上位語とは異なる属性値がコーパスデータに含まれている場合には、上位語とは異なる属性値に付与された属性タグがコーパスデータから除去される。基準語が提示対象に関する属性値である場合には、その上位語も提示対象に関する属性値であることが推認される。反対に、上位語とは異なる属性値は提示対象に関する属性値ではないと推認される。このため、コーパスデータに含まれる上位語とは異なる属性値に付与された属性タグを除去することにより、正確に属性タグが付与されたコーパスデータを生成することができる。

[0020] 本発明の別の側面に係るコーパス生成装置は、提示対象に関する説明文データを含むウェブページを取得するウェブページ取得部と、属性名と属性値

とが関連付けられた辞書データを記憶する辞書データ記憶部を参照し、説明文データに含まれる提示対象に関する属性値の候補を抽出する候補抽出部と、ウェブページから提示対象に関する属性値である基準語を取得する基準語取得部と、属性値間の上下関係を表す上下関係情報を記憶する記憶部から、基準語取得部により取得された基準語よりも上位に属する上位語を抽出し、説明文データに含まれる属性値の候補の内、上位語と一致する属性値の候補に属性値の候補に対応する属性タグを付与する付与部と、付与部により属性タグが付与された説明文データをコーパスデータとして出力する出力部と、を備える。

[0021] 本発明の別の側面に係るコーパス生成方法は、提示対象に関する説明文データを含むウェブページを取得するウェブページ取得ステップと、属性名と属性値とが関連付けられた辞書データを記憶する辞書データ記憶部を参照し、説明文データに含まれる提示対象に関する属性値の候補を抽出する候補抽出ステップと、ウェブページから提示対象に関する属性値である基準語を取得する基準語取得ステップと、属性値間の上下関係を表す上下関係情報を記憶する記憶部から、基準語取得ステップにおいて取得された基準語よりも上位に属する上位語を抽出し、説明文データに含まれる属性値の候補の内、上位語と一致する属性値の候補に属性値の候補に対応する属性タグを付与する付与部と、付与ステップにおいて属性タグが付与された説明文データをコーパスデータとして出力する出力ステップと、を含む。

[0022] 本発明の別の側面に係るコーパス生成プログラムは、コンピュータを、提示対象に関する説明文データを含むウェブページを取得するウェブページ取得部と、属性名と属性値とが関連付けられた辞書データを記憶する辞書データ記憶部を参照し、説明文データに含まれる提示対象に関する属性値の候補を抽出する候補抽出部と、ウェブページから提示対象に関する属性値である基準語を取得する基準語取得部と、属性値間の上下関係を表す上下関係情報を記憶する記憶部から、基準語取得部により取得された基準語よりも上位に属する上位語を抽出し、説明文データに含まれる属性値の候補の内、上位語

と一致する属性値の候補に属性値の候補に対応する属性タグを付与する付与部と、付与部により属性タグが付与された説明文データをコーパスデータとして出力する出力部と、して機能させる。

- [0023] 上記の形態によれば、説明文データに含まれる属性値の候補の内、上位語と一致する属性値に基準語に対応する属性タグが付与される。基準語が提示対象に関する属性値である場合には、その上位語も提示対象に関する属性値であることが推認される。反対に、基準語及びその上位語とは異なる属性値は提示対象に関する属性値ではないと推認される。本形態では、説明文データに含まれる属性値の候補の内、上位語と一致する属性値の候補に対して、当該属性値の候補に対応する属性タグが付与されるので、提示対象に関する属性値ではない語句に対して誤って属性タグが付与されることを防止することができる。

発明の効果

- [0024] 本発明の一側面によれば、提示対象の特徴を示すものではない語句に対して属性タグを誤って付与することを抑制することができる。

図面の簡単な説明

- [0025] [図1]第1実施形態に係るコーパス生成装置の機能構成を示す図である。
[図2]図1に示すコーパス生成装置のハードウェア構成を示す図である。
[図3]商品ページの一例を示す図である。
[図4]上下関係情報の一例を示す図である。
[図5]タグ情報の一例を示す図である。
[図6] (a) は商品ページに含まれる説明文データの一例であり、(b) は属性リストの一例であり、(c) は出力部により出力されたコーパスデータの一例を示す図である。
[図7]商品情報の一例を示す図である。
[図8] (a) は機械学習の対象となるコーパスデータの一例であり、(b) は解析の対象となる説明文データの一例であり、(c) は属性タグが付与された説明文データの一例を示す図である。

[図9]上下関係情報の一例を示す図である。

[図10] (a) は商品ページに含まれる説明文データの一例であり、(b) は属性タグが付与された説明文データの一例である。

[図11] (a) は商品ページに含まれる説明文データの一例であり、(b) は属性タグが付与された説明文データの一例である。

[図12]第1実施形態に係るコーパス生成方法の処理を示すフロー図である。

[図13]第1実施形態に係るコーパス生成プログラムの構成を示す図である。

[図14]第2実施形態に係るコーパス生成装置の機能構成を示す図である。

[図15] (a) はコーパスデータの一例であり、(b) は更新されたコーパスデータの一例である。

[図16]第2実施形態に係るコーパス生成方法の処理を示すフロー図である。

[図17]第3実施形態に係るコーパス生成装置の機能構成を示す図である。

[図18] (a) は商品ページに含まれる説明文データの一例であり、(b) は属性タグが付与された説明文データの一例である。

[図19]第3実施形態に係るコーパス生成方法の処理を示すフロー図である。

発明を実施するための形態

[0026] 以下、添付図面を参照しながら本発明の実施形態を詳細に説明する。なお、図面の説明において同一又は同等の要素には同一の符号を付し、重複する説明を省略する。

[0027] (第1実施形態)

図1は、第1実施形態に係るコーパス生成装置1の機能的構成を示す図である。図1に示すように、コーパス生成装置1はウェブページ取得部12、基準語取得部14、付与部16、出力部18及び商品情報登録部(提示対象情報登録部)20を備えている。コーパス生成装置1は、提示対象に関するウェブページに含まれる説明文データに対して属性タグを付与し、属性タグが付与された説明文をコーパスデータとして出力する装置である。以下では、コーパスデータの生成に用いるウェブページにおける提示対象が商品である場合について説明するが、提示対象は商品に限定されない。

[0028] このコーパス生成装置 1 は、図 2 に示すように、オペレーティングシステムやアプリケーション・プログラムなどを実行する CPU 101 と、ROM 及び RAM で構成される主記憶部 102 と、ハードディスクなどで構成される補助記憶部 103 と、ネットワークカードなどで構成される通信制御部 104 と、キーボードやマウスなどの入力部 105 と、モニタなどの出力部 106 とで構成される。図 1 に示すこのコーパス生成装置 1 の各機能は、CPU 101 や主記憶部 102 の上に所定のソフトウェアを読み込ませ、CPU 101 の制御の下で通信制御部 104 や入力部 105、出力部 106 などを動作させ、主記憶部 102 や補助記憶部 103 におけるデータの読み出し及び書き込みを行うことで実現される。処理に必要なデータやデータベースは主記憶部 102 や補助記憶部 103 内に格納される。

[0029] ウェブページ取得部 12 は、商品に関する説明文データを含む商品ページ（ウェブページ）を取得する機能要素である。ウェブページ取得部 12 は、商品ページ記憶部 22 にアクセス可能とされており、商品ページ記憶部 22 に記憶されている商品ページを取得する。ウェブページ取得部 12 は、取得した商品ページを基準語取得部 14 に出力する。

[0030] 商品ページ記憶部 22 は、電子商取引サイトにおいて販売される商品に関する商品ページを記憶している機能要素である。図 3 は、商品ページ記憶部 22 に記憶されている商品ページ M1 の一例である。商品ページ M1 は、商品画像 B1、説明文データ D1 及び属性リスト L1 を含む。説明文データ D1 は、商品ページ M1 において販売される商品の特徴等を説明した自然文からなる文字列データである。属性リスト L1 は、商品に関する属性名 A1 と属性値 V1 とを対応付けたものである。ここで、属性名とは属性の項目名であり、属性値とは具体的な属性の値である。属性リスト L1 は、例えば商品を販売する店舗の運営者により作成される。

[0031] 基準語取得部 14 は、ウェブページ取得部 12 から基準語を取得する機能要素である。基準語とは、商品の特徴を表す属性値である蓋然性が高い語句であり、説明文データ D1 に含まれる属性値を抽出するための基準として用

いられる。一実施形態では、基準語取得部 14 は、ウェブページ取得部 12 から出力された商品ページに含まれる属性リスト L1 から属性値 V1 を基準語として取得し得る。属性リスト L1 に含まれる属性値 V1 は、商品を販売する店舗の運営者によって登録された情報であるので、商品ページ M1 で販売される商品の特徴を表す属性値である蓋然性が極めて高い語句である。

[0032] なお、基準語取得部 14 は、商品ページ M1 に含まれる属性リスト L1 からではなく、商品ページ M1 内の説明文データ D1 から基準語を取得してもよい。この場合には、基準語取得部 14 は、構文解析器を用いて、説明文データ D1 に含まれる属性値のうち、商品の特性を表す属性値である蓋然性が高い属性値を基準語として抽出する。例えば、基準語取得部 14 は、属性名及び属性値が連続して現れる文や語句を説明文データ D1 から検索し、その属性値を基準語として取得する。一例としては、説明文データ D1 に「産地はブルゴーニュです。」という文を検出した場合には、「ブルゴーニュ」を基準語として検出する。また、基準語取得部 14 は、例えば「フランス産」の「産」や「シャトーマルゴー」の「シャトー」というように、属性を表し、属性値に付加される接頭辞、接尾辞を用いて説明文データから基準語を取得してもよい。基準語取得部 14 は、取得した基準語を示す情報を付与部 16 に出力する。

[0033] 付与部 16 は、上下関係情報記憶部 24 に記憶される上下関係情報から、基準語取得部 14 により取得された基準語よりも上位階層に属する属性値である上位語を抽出し、説明文データ D1 に含まれる上位語に対して、基準語に対応する属性タグを付与する機能要素である。付与部 16 は、上下関係情報記憶部 24 及びタグ情報記憶部 26 にアクセス可能とされている。

[0034] 上下関係情報記憶部 24 は、属性値間の概念的な上下関係を階層的に表す上下関係情報を記憶している。ここでいう上下関係とは、上位下位関係及び部分全体関係を含む概念である。上限関係情報は複数の階層を有し、属性値をノードとする木構造を有している。上下関係情報には階層が深くなるにつれて概念的に小さい区分を表す属性値が格納されている。図 4 は、上下関係

情報のデータ構造を概念的に示す一例である。図4に示す上限関係データでは、最上位の階層に属性値「フランス」が格納されている。その下位階層には「フランス」の下位区分である属性値「ブルゴーニュ」、「ボルドー」、「シャンパーニュ」等が格納されている。更に、属性値「ブルゴーニュ」の下位階層には属性値「ブルゴーニュ」の下位区分である属性値「コート・ド・ニュイ」が格納されており、その下位階層には属性値「マルサネ」が格納されている。また、属性値「ボルドー」の下位階層には「ボルドー」の下位区分である属性値「サン・テステフ」、「ポイヤック」が格納されている。また、図4に示す上限関係データでは、属性値「フランス」に関する系列とは独立する系列として、最上位の階層に属性値「イタリア」が格納されている。その下位階層には属性値「イタリア」の下位区分を示す属性値「ピエモンテ」、「トスカーナ」等が格納されている。

[0035] タグ情報記憶部26は、属性名と属性値が関連付けられたタグ情報を記憶している。図5は、タグ情報の一例を示す図である。図5に示すように、タグ情報は、上下関係情報とは異なり階層構造を有しておらず、一の属性名に複数の属性値が関連付けられている。図5に示すタグ情報では、属性名「産地」に属性値として「フランス、イタリア、ブルゴーニュ、コート・ド・ニュイ」等が関連付けられている。このタグ情報は、例えば「フランス」、「イタリア」、「ブルゴーニュ」、「コート・ド・ニュイ」という語句が「産地」を表す属性値であることを示している。また、図5に示すタグ情報では、属性名「タイプ」、「品種」に対しても対応する属性値が関連付けられている。

[0036] 以下、図4～6を参照して、付与部16が説明文データに対して属性タグを付与する例を具体的に説明する。

[0037] 図6(a)は、商品ページに含まれる説明文データD2の一例を示す図である。説明文データD2は、産地に関する語句として「フランス」、「ブルゴーニュ」及び「ピエモンテ」を含んでいる。図6(b)は、商品ページに含まれる属性リストL2の一例を示す図である。属性リストL2に含まれる

属性値は、上述した基準語取得部 14 により取得され付与部 16 に出力される。以下では、基準語取得部 14 が属性リスト L2 に含まれる属性値 V2 である「ブルゴーニュ」を基準語として付与部 16 に出力したものとして説明する。

[0038] 付与部 16 は、上下関係情報から基準語である「ブルゴーニュ」を検索し、「ブルゴーニュ」よりも上位階層に属する上位語を抽出する。図 4 に示す上下関係情報では、最上位の階層から 1 つ下の階層に属性値「ブルゴーニュ」が含まれているので、付与部 16 は、「ブルゴーニュ」の上位階層に属する「フランス」を上位語として抽出する。次に、付与部 16 は、基準語「ブルゴーニュ」に対応する属性名をタグ情報記憶部 26 に記憶されるタグ情報を参照して取得する。図 5 に示すタグ情報では、基準語「ブルゴーニュ」には属性名「産地」が対応付けられている。付与部 16 は、このように上位語及び基準語に対応する属性名を取得すると、説明文データ D2 に含まれる上位語に対して、基準語に対応する属性名を示す属性タグを付与する。

[0039] また、一実施形態では、付与部 16 は説明文データ D2 に含まれる基準語に対しても当該基準語に対応する属性タグを更に追加してもよい。この場合、付与部 16 は、タグ情報記憶部 26 を参照し、基準語「ブルゴーニュ」に対応する属性名「産地」を説明文データ D2 に付与する。

[0040] 図 6 (c) は、付与部 16 により属性タグが付与された説明文データ C2 の一例を示す図である。図 6 (c) に示すように、説明文データ C2 には属性名「産地」を示す属性タグ「<産地>・・・</産地>」が説明文データ C2 内の上位語「フランス」及び基準語「ブルゴーニュ」に付されている。なお、付与部 16 は、説明文データ D2 に含まれる語句「ピエモンテ」には、産地に関する属性タグを付与しない。「ピエモンテ」は上位語又は基準語ではないので、産地に関する語句であるものの、商品の特徴を示す属性値ではないと推認されるためである。

[0041] なお、属性名を示す属性タグは、必ずしもタグ情報記憶部 26 から取得されなくてもよい。例えば、付与部 16 は、属性リスト L2 において基準語と

なる属性値V 2に対応する属性名A 2を属性名として取得し、その属性名を示す属性タグを基準語及び上位語に付与してもよい。図6（b）に示す例では、属性リストL 2から基準語として「ブルゴーニュ」を取得した場合には、付与部1 6は属性リストL 2において属性値「ブルゴーニュ」に対応する属性名「産地」を取得し、属性名「産地」に対応する属性タグを説明文データD 2内の上位語及び基準語に付与する。

[0042] 図1に戻り、付与部1 6は、上述のように属性タグが付与された説明文データC 2を出力部1 8に出力する。出力部1 8は、付与部1 6から出力された属性タグ付きの説明文データC 2をコーパスデータとしてコーパスデータ記憶部2 8に出力する。コーパスデータ記憶部2 8は、出力部1 8から出力されたコーパスデータを記憶する機能要素である。

[0043] 商品情報登録部2 0は、出力部1 8から出力されたコーパスデータから取得される属性名と属性値との組からなる商品タグを商品ページに関連付けて記憶する機能要素である。商品情報登録部2 0は、コーパスデータ記憶部2 8及び商品情報記憶部3 0にアクセス可能とされている。

[0044] 商品情報記憶部3 0は、仮想店舗が販売する商品に関する商品情報を記憶する機能要素である。図7は、商品情報の一例を示す図である。図7に示す商品情報の各レコードは、商品を提供する仮想店舗の店舗ID、商品を一意に特定する商品ID、商品名、カテゴリ、価格、商品ページのURL、在庫数及び商品タグを含んでいる。ただし、商品情報に含まれる情報はこれらに限定されない。

[0045] 商品情報登録部2 0は、コーパスデータ記憶部2 8からコーパスデータを取得し、当該コーパスデータに付与された属性タグを用いて属性名と属性値との組を取得する。そして、コーパスデータが取得された商品ページに対応する商品情報のレコードを抽出し、そのレコードの商品タグに属性名と属性値との組を登録する。この商品タグは、電子商取引サイトからユーザが目的とする商品ページを絞り込むためのキーとして利用される。

[0046] また、コーパスデータ記憶部2 8に記憶されたコーパスデータは、例えば

、属性リストを自動生成する解析装置を機械学習を用いて構築することができる。このような解析装置を用いることにより商品ページに含まれる説明文データから属性リストを自動生成することができる。例えば、例えば、図8（a）に示されるようなコーパスデータC Xを用いて機械学習を適用すると、解析装置は、文脈の中で「産地」の属性値が現れる箇所を認識できるようになる。

[0047] このようなコーパスデータC Xを用いて機械学習を適用した後に、解析装置が図8（b）に示されるような説明文データD 3の解析を実施すると、解析装置は、その文の構造及び文脈等から「チリ」が「産地」を表す属性値であることを検出し、図8（c）に示すように、「チリ」の部分にタグ「＜産地＞・・・＜／産地＞」が付された出力データC 3を生成できる。そして、解析装置は、出力データC 3に基づき、属性名「産地」と属性値「チリ」を対応付けた属性リストを生成できる。これにより、説明文データD 3が含まれていた商品ページが属性リストを有していなくとも、属性名「産地」と属性値「チリ」を対応付けた属性リストを当該商品ページに含ませることが可能となる。

[0048] また、付与部1 6による属性タグの付与処理は、種々の態様に変形可能である。以下では、付与部1 6の第1の変形例について説明する。

[0049] （第1の変形例）

第1の変形例に係る付与部1 6は、上下関係情報から基準語をルートノードとし、且つ分岐を有しない部分木を検索して、部分木に含まれる基準語以外の1以上の属性値を抽出し、説明文データに含まれる1以上の属性値に対して、基準語に対応する属性タグを更に付与する。以下、図9を参照して、第1の変形例に係る付与部1 6の処理を具体的に説明する。図9は、上下関係情報の別の一例を示す図である。

[0050] 例えば基準語が「ブルゴーニュ」である場合、付与部1 6は、「ブルゴーニュ」をルートノードとし、且つ分岐を有しない部分木として「ブルゴーニュ」、「コート・ド・ニュイ」及び「マルサネ」をノードとして含む部分木

P Tを抽出する。次に、部分木P Tに含まれる属性値のうち、基準語である「ブルゴーニュ」を除いた「コート・ド・ニュイ」及び「マルサネ」を抽出する。次に、付与部16は基準語「ブルゴーニュ」に対応する属性名「産地」をタグ情報記憶部26に記憶されるタグ情報を参照して取得する。そして、付与部16は、説明文データに含まれる基準語「ブルゴーニュ」及び上位語「フランス」に加えて、「コート・ド・ニュイ」及び「マルサネ」にも属性タグ「産地」を付与する。

[0051] なお、この第1の変形例において、基準語が「ボルドー」である場合には、「ボルドー」をルートノードとし、且つ分岐を有しない部分木は存在しない。即ち、「ボルドー」の直下の階層には複数の属性値「サン・テステフ」及び「ポイヤック」が存在しており、分岐を有さない部分木は存在しない。したがって、この場合には、付与部16は、説明文データに含まれる基準語「ボルドー」及びその上位語「フランス」のみに対して属性タグ「産地」を付与する。

[0052] 本変形例の作用効果について説明する。基準語が商品に関する属性値である場合には、分岐を有しない部分木P Tに含まれる属性値も商品に関する属性値であることが推認される。よって、第1の変形例に係る付与部16によれば、説明文データ中のより多くの商品の属性値に正確な属性タグを付与することができる。

[0053] 次に、付与部16の第2の変形例について説明する。

[0054] (第2の変形例)

第2の変形例に係る付与部16は、上下関係情報において基準語の直下の階層に複数の属性値が存在する場合であって、当該複数の属性値の内、一の属性値のみが説明文データに含まれる場合には、説明文データに含まれる当該一の属性値に対して、基準語に対応する属性タグを更に付与する。以下、図9～11を参照して、第2の変形例に係る付与部16に係る付与部16の処理を具体的に説明する。

[0055] 図10(a)は、商品ページに含まれる説明文データD4の一例である。

説明文データD4は、産地に関する語句として「フランス」、「ボルドー」及び「サン・テステフ」を含んでいる。ここで、例えば基準語が「ボルドー」である場合、付与部16は基準語「ボルドー」の直下の階層に属する属性値を検索する。図9に示す上下関係情報では、「ボルドー」の下位階層には「サン・テステフ」及び「ポイヤック」の2つの属性値が存在する。説明文データD4は、これら2つの属性値のうち、一の属性地である「サン・テステフ」のみを含んでいる。この場合、付与部16は「サン・テステフ」も商品に関する属性値であるとして基準語に対応する属性タグを付与する。図10(b)は、本変形例の付与部16により説明文データD4に属性タグが付与された説明文データC4を示している。図10(b)に示すように、説明文データC4には属性名「産地」を示す属性タグ「<産地>・・・</産地>」が説明文データC4内の上位語「フランス」及び基準語「ボルドー」に加えて、属性値「サン・テステフ」にも付されている。

- [0056] 図11(a)は、商品ページに含まれる説明文データの別の一例である。この説明文データD5は、産地に関する語句として「フランス」、「ボルドー」、「サン・テステフ」及び「ポイヤック」を含んでいる。この説明文データD5は、基準語「ボルドー」の直下の階層に属する複数の属性値「サン・テステフ」及び「ポイヤック」の双方を含んでいる。この場合には、「サン・テステフ」及び「ポイヤック」のうち、どちらが商品の関する属性値であるか判断することが困難であるので、付与部16は、「サン・テステフ」及び「ポイヤック」には属性タグを付与せず、基準語「ボルドー」及びその上位語「フランス」のみに対して属性タグ「産地」を付与する。図11(b)は、本変形例の付与部16により説明文データD5に属性タグが付与された説明文データC5を示している。図11(b)に示すように、説明文データC5では属性名「産地」を示す属性タグ「<産地>・・・</産地>」が説明文データC4内の上位語「フランス」及び基準語「ボルドー」にのみ付されている。その一方で、基準語「ボルドー」の下位語である「サン・テステフ」および「ポイヤック」を含む文は最終的なコーパスに含めないように

してもよい。これはコーパスの品質を向上させるためである。

[0057] 本変形例の作用効果について説明する。基準語の直下の階層に複数の属性値が存在する場合であっても、説明文に当該複数の属性値のうちの1つのみが含まれている場合には、当該属性値は商品に関する属性値である蓋然性が高い。したがって、本形態によれば、説明文データ中のより多くの商品の属性値に、正確な属性タグを付与することができる。

[0058] 次に、図12を用いて、コーパス生成装置1の動作を説明するとともに第1実施形態に係るコーパス生成方法について説明する。

[0059] まず、コーパス生成装置1のウェブページ取得部12は、商品ページ記憶部22から商品に関する説明文データを含むウェブページを取得する（ステップS1、ウェブページ取得ステップ）。次に、基準語取得部14は、ステップS1において取得したウェブページから基準語を取得する（ステップS2、基準語取得ステップ）。この際、基準語はウェブページ内に含まれる属性リストから取得され得る。続いて、付与部16は、上下関係情報記憶部24に記憶される上下関係情報から、ステップS2において取得された基準語よりも上位階層に属する属性値である上位語を抽出する（ステップS3）。

[0060] その後、付与部16は、商品ページの説明文データに含まれる上位語に対して、基準語に対応する属性タグを付与する（ステップS4、付与ステップ）。なおこの際、商品ページの説明文データに含まれる基準語に対しても基準語に対応する属性タグを付与してもよい。更に付与部16は、上下関係情報から基準語をルートノードとし、且つ分岐を有しない部分木を検索して、該部分木に含まれる基準語以外の1以上の属性値を抽出し、説明文データに含まれる該1以上の属性値に対して、基準語に対応する属性タグを更に付与してもよい。あるいは、付与部16は、上下関係情報において基準語の直下の階層に複数の属性値が存在する場合であって、当該複数の属性値の内、一の属性値のみが説明文データに含まれる場合には、説明文データに含まれる一の属性値に対して、基準語に対応する属性タグを更に付与してもよい。

[0061] 続いて、出力部18は、ステップS4において属性タグが付与された説明

文データをコーパスデータとしてコーパスデータ記憶部 28 に格納する（ステップ S 5、出力ステップ）。以上のようにして、コーパス生成装置 1 はコーパスデータを生成する。

[0062] 次に、図 13 を用いて、上述した一連の処理をコンピュータに実行させるためのコーパス生成プログラム P 1 を説明する。図 13 は、実施形態に係るコーパス生成プログラム P 1 の構成を示す図である。なお、このプログラムをコンピュータに適用することによりコンピュータを上記機能を有するコーパス生成装置 1 として動作させることができる。

[0063] コーパス生成プログラム P 1 は、メインモジュール P 2、ウェブページ取得モジュール P 3、基準語取得モジュール P 4、付与モジュール P 5、出力モジュール P 6、及び商品情報登録モジュール P 7 を備えている。

[0064] メインモジュール P 2 は、処理を統括的に制御する部分である。ウェブページ取得モジュール P 3、基準語取得モジュール P 4、付与モジュール P 5、出力モジュール P 6、及び商品情報登録モジュール P 7 を実行することにより実現される機能はそれぞれ、上記のウェブページ取得部 12、基準語取得部 14、付与部 16、出力部 18 及び商品情報登録部 20 の機能と同様である。

[0065] コーパス生成プログラム P 1 は、例えば、CD-ROM や DVD-ROM、半導体メモリ等の有形の記録媒体に記録された上で提供される。また、コーパス生成プログラム P 1 は、搬送波に重畳されたデータ信号としてネットワーク N を介して提供されてもよい。

[0066] 以上説明したように、コーパス生成装置 1 によれば、説明文データに提示対象に関する属性値である基準語の上位語が含まれる場合には、説明文データに含まれる上位語に対して基準語に対応する属性タグが付与される。基準語が提示対象に関する属性値である場合には、その上位語も提示対象に関する属性値であることが推認される。このため、説明文データ内の上位語に対して属性タグを付与することにより、提示対象の特徴を示すものではない語句に対して誤って属性タグを付与することを抑制することができる。そして

、このように属性タグが付与された提示対象の説明文データをコーパスデータとして出力することにより、品質の高いコーパスデータを生成することができる。また、付与部 16 は、説明文データに含まれる基準語に対して、該基準語に対応する属性タグを更に付与しているので、説明文データ中のより多くの商品の属性値に正確な属性タグを付与することができる。

[0067] また、コーパス生成装置 1 では、基準語取得部 14 は、属性リストにおける属性値を基準語として取得している。属性リストにおける商品の属性値は商品に関する属性値である蓋然性が高いので、コーパス生成装置 1 は適切な基準語を取得することができる。

[0068] (第 2 実施形態)

次に、第 2 実施形態に係るコーパス生成装置 1 A について説明する。コーパス生成装置 1 A は、既に属性タグが付与されたコーパスデータから上位語とは異なる属性値に対して付与されている属性タグを除去する装置である。以下では、主に第 1 実施形態と相違する事項について説明し、第 1 実施形態と同一又は同等の事項については説明を省略する。

[0069] 図 14 は、コーパス生成装置 1 A の機能的構成を示す図である。図 14 に示すように、コーパス生成装置 1 A はコーパス取得部 32、基準語取得部 14、除去部 36、出力部 18、及び商品情報登録部 20 を備えている。

[0070] コーパス取得部 32 は、コーパスデータ記憶部 28 から属性値に対して属性タグが付与されたコーパスデータを取得する機能要素である。コーパス取得部 32 により取得されるコーパスデータは、説明文データに含まれる属性値に属性タグが付与されたテキストデータであってもよいし、説明文データ及び属性リストを含み、その説明文データに含まれる属性値に属性タグが付与されたウェブページデータであってもよい。

[0071] 基準語取得部 14 は、コーパス取得部 32 によって取得されたコーパスデータから商品に関する属性値である基準語を取得する機能要素である。基準語取得部 14 は、例えばコーパス取得部 32 によって取得されたコーパスデータが説明文データ及び属性リストを含むウェブページデータである場合に

は、コーパスデータに含まれる属性リストから属性値を基準語として取得し得る。また、基準語取得部 14 は、例えば例えばコーパス取得部 32 によって取得されたコーパスデータが説明文データに含まれる属性値に属性タグが付与されたテキストデータである場合には、構文解析器を用いてコーパスデータに含まれる属性値のうち、商品の特性を表す属性値である蓋然性が高い属性値を基準語として抽出し得る。

[0072] 除去部 36 は、属性値間の上下関係を表す上下関係情報を記憶する上下関係情報記憶部 24 から、基準語取得部 14 により取得された基準語よりも上位に属する上位語を抽出し、上位語とは異なる属性値がコーパスデータに含まれている場合に、当該上位語とは異なる属性値に付与された属性タグをコーパスデータから除去する機能要素である。除去部 36 は、基準語がコーパスデータに含まれている場合には、基準語に付与された属性タグをコーパスデータから除去せずに残すこととしてもよい。

[0073] 以下、図 4、5、15 を参照して、除去部 36 がコーパスデータから属性タグを除去する例を具体的に説明する。ここで除去部 36 は、上位語及び基準語とは異なる属性値がコーパスデータに含まれている場合に、上位語及び基準語とは異なる属性値に付与された属性タグをコーパスデータから除去するものとする。

[0074] 図 15 (a) は、コーパス取得部 32 によって取得されたコーパスデータ C6a の一例を示す図である。コーパスデータ C6a は、属性名「産地」を示す属性タグ「<産地>・・・</産地>」がコーパスデータ C6a 内の語句「フランス」、「ブルゴーニュ」及び「ピエモンテ」に付されている。ここでは、基準語取得部 14 がコーパスデータ C6a から「ブルゴーニュ」を基準語として付与部 16 に出力したものとして説明する。

[0075] 除去部 36 は、上下関係情報から基準語である「ブルゴーニュ」を検索し、「ブルゴーニュ」よりも上位階層に属する上位語を抽出する。図 4 に示す上下関係情報では、最上位の階層から 1 つ下の階層に基準語「ブルゴーニュ」が含まれているので、付与部 16 は、「ブルゴーニュ」の上位階層に属す

る「フランス」を上位語として抽出する。次に、除去部 36 は、コーパスデータ C 6 a に属性タグが付与された語句を取得し、上位語とは異なる語句を検索する。ここでは、コーパスデータ C 6 a に属性タグが付与された語句「フランス」、「ブルゴーニュ」及び「ピエモンテ」の内、「フランス」は上位語と同一であり、「ブルゴーニュ」は基準語と同一であるが、「ピエモンテ」は上位語及び基準語の何れとも異なっている。そこで、除去部 36 は、コーパスデータ C 6 a に含まれる語句「ピエモンテ」に付与された属性タグを除去する。

[0076] 図 15 (b) は、除去部 36 により属性タグが除去されたコーパスデータ C 6 b の一例を示す図である。図 15 (b) に示すように、コーパスデータ C 6 b では、語句「フランス」、「ブルゴーニュ」に付与された属性タグは維持されているが、語句「ピエモンテ」に付与された属性タグは除去されている。「ピエモンテ」は上位語又は基準語とは異なっているので、商品の特徴を示す属性値ではないと推認されるためである。

[0077] 除去部 36 は、属性タグの一部が除去されたコーパスデータを出力部 18 に出力する。出力部 18 は、除去部 36 から出力されたコーパスデータをコーパスデータ記憶部 28 に出力し、コーパスデータ記憶部 28 内のコーパスデータを更新する。商品情報登録部 20 の機能は第 1 実施形態で説明した通りであるので説明を省略する。

[0078] なお、一実施形態では、除去部 36 は、上下関係情報から基準語をルートノードとし、且つ分岐を有しない部分木を検索して、部分木に含まれる基準語以外の 1 以上の属性値を抽出し、当該 1 以上の属性値に付与された属性タグをコーパスデータから除去せずに残してもよい。更に、除去部 36 は、上下関係情報において基準語の直下の階層に複数の属性値が存在する場合であって、当該複数の属性値の内、一の属性値のみがコーパスデータに含まれる場合には、当該一の属性値に付与された属性タグをコーパスデータから除去せずに残してもよい。

[0079] 次に、図 16 を用いて、コーパス生成装置 1 A の動作を説明するとともに

第2実施形態に係るコーパス生成方法について説明する。

[0080] まず、コーパス生成装置1Aのコーパス取得部32は、コーパスデータ記憶部28から属性値に対して属性タグが付与されたコーパスデータを取得する（ステップS11、コーパス取得ステップ）。次に、基準語取得部14は、ステップS1において取得したコーパスデータから基準語を取得する（ステップS12）。続いて、除去部36は、上下関係情報記憶部24に記憶される上下関係情報から、ステップS12において取得された基準語よりも上位階層に属する属性値である上位語を抽出する（ステップS13）。

[0081] その後、除去部36は、上位語とは異なる属性値がコーパスデータに含まれている場合に、上位語とは異なる属性値に付与された属性タグをコーパスデータから除去する（ステップS14、除去ステップ）。なおこの際、除去部36は、基準語とは異なる属性値がコーパスデータに含まれている場合に、基準語に付与された属性タグをコーパスデータから除去せずに残してもよい。また、除去部36は、上下関係情報から基準語をルートノードとし、且つ分岐を有しない部分木を検索して、部分木に含まれる基準語以外の1以上の属性値を抽出し、当該1以上の属性値に付与された属性タグをコーパスデータから除去せずに残してもよい。あるいは、除去部36は、上下関係情報において基準語の直下の階層に複数の属性値が存在する場合であって、当該複数の属性値の内、一の属性値のみがコーパスデータに含まれる場合には、当該一の属性値に付与された属性タグをコーパスデータから除去せずに残してもよい。

[0082] 続いて、出力部18は、ステップS14において属性タグが除去されたコーパスデータをコーパスデータ記憶部28に格納する（ステップS15）。以上のようにして、コーパス生成装置1Aは更新されたコーパスデータを生成する。

[0083] 次に、第2実施形態に係るコーパス生成プログラムを説明する。

[0084] コーパス生成プログラムは、メインモジュール、コーパス取得モジュール、基準語取得モジュール、除去モジュール、出力モジュール、及び商品情報

登録モジュールを備えている。

[0085] メインモジュールは、コーパス生成プログラムを統括的に制御する部分である。コーパス取得モジュール、基準語取得モジュール、除去モジュール、出力モジュール、及び商品情報登録モジュールを実行することにより実現される機能はそれぞれ、上記のコーパス取得部 32、基準語取得部 14、除去部 36、出力部 18、及び商品情報登録部 20の機能と同様である。

[0086] 本実施形態に係るコーパス生成プログラムも、第 1 実施形態のコーパス生成プログラム P1 と同様に様々な方法で配布しうる。

[0087] 以上説明したように、コーパス生成装置 1A によれば、上位語とは異なる属性値がコーパスデータに含まれている場合には、上位語とは異なる属性値に付与された属性タグがコーパスデータから除去される。基準語が商品に関する属性値である場合には、その上位語も商品に関する属性値であることが推認される。反対に、基準語及びその上位語とは異なる属性値は商品に関する属性値ではないと推認される。このため、コーパスデータに含まれる上位語とは異なる属性値に付与された属性タグを除去することにより、正確に属性タグが付与されたコーパスデータを生成することができる。

[0088] (第 3 実施形態)

次に、第 3 実施形態に係るコーパス生成装置 1B について説明する。コーパス生成装置 1B は、説明データを含むウェブページから辞書データを用いて属性値の候補を抽出し、説明文データに含まれる属性値の候補の内、上位語と一致する属性値の候補に属性タグを付与する装置である。以下では、主に第 1 実施形態と相違する事項について説明し、第 1 実施形態と同一又は同等の事項については説明を省略する。

[0089] 図 17 は、コーパス生成装置 1B の機能的構成を示す図である。図 16 に示すように、ウェブページ取得部 12、基準語取得部 14、付与部 16、出力部 18、商品情報登録部 20 に加えて候補抽出部 42 を備えている。

[0090] 候補抽出部 42 は、属性名と属性値とが関連付けられた辞書データを記憶するタグ情報記憶部（辞書データ記憶部）26 を参照し、説明文データに含

まれる 1 以上の属性値の候補を抽出する機能要素である。候補抽出部 42 は、タグ情報記憶部 26 にアクセス可能とされている。第 1 実施形態で説明したように、タグ情報記憶部 26 は、属性名と属性値が関連付けられたタグ情報を記憶している。本実施形態では、このタグ情報は辞書データとしても機能する。候補抽出部 42 は、説明文データから辞書データの属性値と一致する語句を抽出し、抽出した語句を商品に関する属性値の候補とする。また、候補抽出部 42 は、抽出した属性値の候補に対応する属性名を辞書データから取得する。候補抽出部 42 は、この属性値の候補及びその属性名を付与部 16 に出力する。

[0091] 本実施形態では、付与部 16 は、上下関係情報記憶部 24 から、基準語取得部 14 により取得された基準語よりも上位に属する上位語を抽出し、説明文データに含まれる属性値の候補の内、上位語と一致する属性値の候補に、属性値の候補に対応する属性タグを付与する。一方、説明文データに含まれる属性値の候補の内、上位語と一致しない属性値の候補には属性タグを付与しない。また、付与部 16 は、説明文データに含まれる属性値の候補の内、基準語と一致する属性値の候補に、属性値の候補に対応する属性タグを更に付与してもよい。

[0092] 以下、図 4、5、18 を参照して、付与部 16 が説明文データに属性タグを付与する例を具体的に説明する。ここで付与部 16 は、説明文データに含まれる属性値の候補の内、上位語又は基準語と一致する属性値の候補に、属性値の候補に対応する属性タグを付与するものとする。

[0093] 図 18 (a) は、商品ページに含まれる説明文データ D7 の一例を示す図である。説明文データ D7 は、産地に関する語句として「フランス」、「ブルゴーニュ」及び「ピエモンテ」を含んでいる。ここでは、基準語取得部 14 が商品ページから「ブルゴーニュ」を基準語として取得し、この基準語を付与部 16 に出力したものとして説明する。

[0094] 候補抽出部 42 は、タグ情報記憶部 26 に記憶された辞書データを参照し、説明文データに含まれる 1 以上の属性値の候補を抽出する。ここでは、候

補抽出部４２は、説明文データＤ７に含まれる語句「フランス」、「ブルゴーニュ」及び「ピエモンテ」を属性値の候補として抽出する。また、候補抽出部４２は、辞書データを参照し、これらの語句に対応する属性名「産地」を取得する。

[0095] 付与部１６は、上下関係情報から基準語である「ブルゴーニュ」を検索し、「ブルゴーニュ」よりも上位階層に属する上位語を抽出する。図４に示す上下関係情報では、最上位の階層から１つ下の階層に基準語「ブルゴーニュ」が含まれているので、付与部１６は、「ブルゴーニュ」の上位階層に属する「フランス」を上位語として抽出する。次に、付与部１６は、属性値の候補の内、上位語と一致する属性値の候補を選択する。ここでは、属性値の候補「フランス」、「ブルゴーニュ」及び「ピエモンテ」の内、「フランス」は上位語と一致し、「ブルゴーニュ」は基準語と一致するが、「ピエモンテ」は上位語及び基準語の何れとも一致しない。そこで、付与部１６は、上位語又は基準語と一致する「フランス」及び「ブルゴーニュ」に対して属性名「産地」を示す属性タグを付与する。

[0096] 図１８（ｂ）は、付与部１６により属性タグが付与された説明文データＣ７の一例を示す図である。図１８（ｂ）に示すように、説明文データＣ７では、語句「フランス」、「ブルゴーニュ」に属性タグが付与され、語句「ピエモンテ」には属性タグが付与されていない。「ピエモンテ」は上位語又は基準語と一致していないので、商品の特徴を示す属性値ではないと推認されるためである。

[0097] なお、一実施形態では、付与部１６は上下関係情報から基準語をルートノードとし、且つ分岐を有しない部分木を検索して、部分木に含まれる基準語以外の１以上の属性値を抽出し、説明文データに含まれる属性値の候補の内、当該１以上の属性値と一致する属性値に、属性値の候補に対応する属性タグを更に付与してもよい。あるいは、付与部１６は、上下関係情報において基準語の直下の階層に複数の属性値が存在する場合であって、当該複数の属性値の内、一の属性値のみがコーパスデータに含まれる場合には、説明文デ

ータに含まれる属性値の候補の内、当該一の属性値と一致する属性値に、属性値の候補に対応する属性タグを更に付与してもよい。

[0098] 付与部 16 は、属性タグが付与された説明文データを出力部 18 に出力する。出力部 18 は、付与部 16 から出力された説明文データをコーパスデータとしてコーパスデータ記憶部 28 に出力する。

[0099] 次に、図 19 を用いて、コーパス生成装置 1B の動作を説明するとともに第 3 実施形態に係るコーパス生成方法について説明する。

[0100] まず、コーパス生成装置 1B のウェブページ取得部 12 は、商品ページ記憶部 22 から商品に関する説明文データを含むウェブページを取得する（ステップ S21）。次に、候補抽出部 42 は、辞書データを記憶するタグ情報記憶部 26 を参照し、説明文データに含まれる 1 以上の属性値の候補を抽出する（ステップ S22、候補抽出ステップ）。続いて、基準語取得部 14 は、ステップ S21 において取得したウェブページから基準語を取得する（ステップ S23）。続いて、付与部 16 は、上下関係情報記憶部 24 に記憶される上下関係情報から、ステップ S23 において取得された基準語よりも上位階層に属する属性値である上位語を抽出する（ステップ S24）。

[0101] その後、付与部 16 は、説明文データに含まれる属性値の候補の内、上位語と一致する属性値に、属性値の候補に対応する属性タグを付与する（ステップ S25）。この際、付与部 16 は、基準語と一致する属性値が説明文データに含まれている場合に、説明文データに含まれる属性値の候補の内、基準語と一致する属性値に、属性値の候補に対応する属性タグを更に付与してもよい。更に、付与部 16 は、上下関係情報から基準語をルートノードとし、且つ分岐を有しない部分木を検索して、該部分木に含まれる基準語以外の 1 以上の属性値を抽出し、説明文データに含まれる属性値の候補の内、当該 1 以上の属性値と一致する属性値の候補に、属性値の候補に対応する属性タグを更に付与してもよい。あるいは、付与部 16 は、上下関係情報において基準語の直下の階層に複数の属性値が存在する場合であって、当該複数の属性値の内、一の属性値のみが説明文データに含まれる場合には、説明文デー

タに含まれる属性値の候補の内、当該一の属性値と一致する属性値の候補に、当該属性値の候補に対応する属性タグを更に付与してもよい。

[0102] 続いて、出力部 18 は、ステップ S 25 において属性タグが付与された説明文データをコーパスデータ記憶部 28 に格納する（ステップ S 26）。以上のようにして、コーパス生成装置 1B はコーパスデータを生成する。

[0103] 次に、第 3 実施形態に係るコーパス生成プログラムを説明する。

[0104] コーパス生成プログラムは、メインモジュール、ウェブページ取得モジュール、基準語取得モジュール、付与モジュール、出力モジュール、商品情報登録モジュール、及び候補抽出モジュールを備えている。

[0105] メインモジュールは、コーパス生成プログラムを統括的に制御する部分である。ウェブページ取得モジュール、基準語取得モジュール、付与モジュール、出力モジュール、商品情報登録モジュール、及び候補抽出モジュールを実行することにより実現される機能はそれぞれ、上記のウェブページ取得部 12、基準語取得部 14、付与部 16、出力部 18、商品情報登録部 20、及び候補抽出部 42 の機能と同様である。

[0106] 本実施形態に係るコーパス生成プログラムも、第 1 実施形態のコーパス生成プログラム P1 と同様に様々な方法で配布しうる。

[0107] 以上説明したように、コーパス生成装置 1B によれば、説明文データに含まれる属性値の候補の内、上位語と一致する属性値に基準語に対応する属性タグが付与される。基準語が商品に関する属性値である場合には、その上位語も商品に関する属性値であることが推認される。反対に、基準語及びその上位語とは異なる属性値は商品に関する属性値ではないと推認される。コーパス生成装置 1B では、説明文データに含まれる属性値の候補の内、上位語と一致する属性値の候補に、当該属性値の候補に対応する属性タグが付与されるので、商品に関する属性値ではない語句に対して誤って属性タグが付与されることを防止することができる。

[0108] 以上、本発明をその実施形態に基づいて詳細に説明した。しかし、本発明は上記実施形態に限定されるものではない。本発明は、その要旨を逸脱しな

い範囲で様々な変形が可能である。

[0109] 例えば、上記実施形態では、ワインの商品ページに含まれる商品の説明文データに対して属性タグを付与する例について説明したが、ワインとは異なる商品に関する商品ページの説明文データに属性タグを付与してもよい。また、コーパス生成装置 1 は、例えば「品種」、「タイプ」、「容量」など、産地以外の属性タグを説明文データに付与してもよい。

[0110] 上記実施形態では、説明文データ中の上位語、基準語等に同一の属性タグを付与しているが、それぞれに異なる属性タグを付与してもよい。例えば、上位語が「フランス」であり、基準語が「ブルゴーニュ」である場合に、上位語には属性名「国」に関する属性タグを付与し、基準語には属性名「地方」に関する属性タグを付与してもよい。

符号の説明

[0111] 1…コーパス生成装置、12…ウェブページ取得部、14…基準語取得部、16…付与部、18…出力部、20…商品情報登録部、22…商品ページ記憶部、24…上下関係情報記憶部、26…タグ情報記憶部、28…コーパスデータ記憶部、30…商品情報記憶部、32…コーパス取得部、36…除去部、42…候補抽出部、D1～D5, D7…説明文データ、C2～C7…説明文データ、L1, L2…属性リスト、M1…商品ページ、N…ネットワーク、P1…コーパス生成プログラム、P2…メインモジュール、P3…ウェブページ取得モジュール、P4…基準語取得モジュール、P5…付与モジュール、P6…出力モジュール、P7…商品情報登録モジュール、PT…部分木。

請求の範囲

- [請求項1] 提示対象に関する説明文データを含むウェブページを取得するウェブページ取得部と、
- 前記ウェブページから前記提示対象に関する属性値である基準語を取得する基準語取得部と、
- 属性値間の上下関係を表す上下関係情報を記憶する記憶部から、前記基準語取得部により取得された前記基準語よりも上位に属する上位語を抽出し、前記説明文データに前記上位語が含まれる場合には、前記説明文データに含まれる前記上位語に対して、前記基準語に対応する属性タグを付与する付与部と、
- 前記付与部により属性タグが付与された説明文データをコーパスデータとして出力する出力部と、
- を備えるコーパス生成装置。
- [請求項2] 前記付与部は、前記説明文データに前記基準語が含まれる場合には、前記説明文データに含まれる前記基準語に対して、該基準語に対応する属性タグを更に付与する、請求項1に記載のコーパス生成装置。
- [請求項3] 前記ウェブページは、前記提示対象に関する属性名と属性値とが対応付けられた属性リストを更に含み、
- 前記基準語取得部は、前記属性リストにおける属性値を前記基準語として取得する、請求項1又は2に記載のコーパス生成装置。
- [請求項4] 前記基準語取得部は、構文解析器を用いて、前記説明文データから前記提示対象に関する属性値である蓋然性が高い語句を検索し、検索された語句を前記基準語として取得する、請求項1又は2に記載のコーパス生成装置。
- [請求項5] 前記上下関係情報は、属性値をノードとする木構造をなし、
- 前記付与部は、前記上下関係情報から前記基準語をルートノードとし、且つ分岐を有しない部分木を検索して、該部分木に含まれる前記基準語以外の1以上の属性値を抽出し、前記説明文データに前記1以

上の属性値が含まれる場合には、前記説明文データに含まれる該 1 以上の属性値に対して、前記基準語に対応する属性タグを更に付与する、請求項 1 ～ 4 の何れか一項に記載のコーパス生成装置。

[請求項6] 前記上下関係情報は、属性値間の上下関係を階層的に表しており、前記付与部は、前記上下関係情報において前記基準語の直下の階層に複数の属性値が存在する場合であって、該複数の属性値の内、一の属性値のみが前記説明文データに含まれる場合には、前記説明文データに含まれる該一の属性値に対して、前記基準語に対応する属性タグを更に付与する、請求項 1 ～ 4 の何れか一項に記載のコーパス生成装置。

[請求項7] 前記出力部から出力されたコーパスデータから取得される前記提示対象に関する属性名と属性値との組を前記ウェブページに関連付けて記憶する提示対象情報登録部を更に備える、請求項 1 ～ 6 の何れか一項に記載のコーパス生成装置。

[請求項8] 提示対象に関する説明文データを含むウェブページを取得するウェブページ取得ステップと、

前記ウェブページから前記提示対象に関する属性値である基準語を取得する基準語取得ステップと、

属性値間の上下関係を表す上下関係情報を記憶する記憶部から、前記基準語取得ステップにおいて取得された前記基準語よりも上位に属する上位語を抽出し、前記説明文データに前記上位語が含まれる場合には、前記説明文データに含まれる前記上位語に対して、前記基準語に対応する属性タグを付与する付与ステップと、

前記付与ステップにおいて属性タグが付与された説明文データをコーパスデータとして出力する出力ステップと、を含むコーパス生成方法。

[請求項9] コンピュータを、提示対象に関する説明文データを含むウェブページを取得するウェブ

ブページ取得部と、

前記ウェブページから前記提示対象に関する属性値である基準語を取得する基準語取得部と、

属性値間の上下関係を表す上下関係情報を記憶する記憶部から、前記基準語取得部により取得された前記基準語よりも上位に属する上位語を抽出し、前記説明文データに前記上位語が含まれる場合には、前記説明文データに含まれる前記上位語に対して、前記基準語に対応する属性タグを付与する付与部と、

前記付与部により属性タグが付与された説明文データをコーパスデータとして出力する出力部と、

して機能させる、コーパス生成プログラム。

[請求項10]

属性値に対して属性タグが付与されたコーパスデータを取得するコーパス取得部と、

前記コーパスデータから提示対象に関する属性値である基準語を取得する基準語取得部と、

属性値間の上下関係を表す上下関係情報を記憶する記憶部から、前記基準語取得部により取得された前記基準語よりも上位に属する上位語を抽出し、前記上位語とは異なる属性値が前記コーパスデータに含まれている場合に、前記上位語とは異なる属性値に付与された属性タグを前記コーパスデータから除去する除去部と、

前記除去部により属性タグが除去された前記コーパスデータを出力する出力部と、

を備えるコーパス生成装置。

[請求項11]

属性値に対して属性タグが付与されたコーパスデータを取得するコーパス取得ステップと、

前記コーパスデータから提示対象に関する属性値である基準語を取得する基準語取得ステップと、

属性値間の上下関係を表す上下関係情報を記憶する記憶部から、前

記基準語取得ステップにおいて取得された前記基準語よりも上位に属する上位語を抽出し、前記上位語とは異なる属性値が前記コーパスデータに含まれている場合に、前記上位語とは異なる属性値に付与された属性タグを前記コーパスデータから除去する除去ステップと、

前記除去ステップにおいて属性タグが除去された前記コーパスデータを出力する出力ステップと、
を含むコーパス生成方法。

[請求項12]

コンピュータを、

属性値に対して属性タグが付与されたコーパスデータを取得するコーパス取得部と、

前記コーパスデータから提示対象に関する属性値である基準語を取得する基準語取得部と、

属性値間の上下関係を表す上下関係情報を記憶する記憶部から、前記基準語取得部により取得された前記基準語よりも上位に属する上位語を抽出し、前記上位語とは異なる属性値が前記コーパスデータに含まれている場合に、前記上位語とは異なる属性値に付与された属性タグを前記コーパスデータから除去する除去部と、

前記除去部により属性タグが除去された前記コーパスデータを出力する出力部と、

して機能させる、コーパス生成プログラム。

[請求項13]

提示対象に関する説明文データを含むウェブページを取得するウェブページ取得部と、

属性名と属性値とが関連付けられた辞書データを記憶する辞書データ記憶部を参照し、前記説明文データに含まれる前記提示対象に関する属性値の候補を抽出する候補抽出部と、

前記ウェブページから前記提示対象に関する属性値である基準語を取得する基準語取得部と、

属性値間の上下関係を表す上下関係情報を記憶する記憶部から、前

記基準語取得部により取得された前記基準語よりも上位に属する上位語を抽出し、前記説明文データに含まれる前記属性値の候補の内、前記上位語と一致する属性値の候補に前記属性値の候補に対応する属性タグを付与する付与部と、

前記付与部により属性タグが付与された説明文データをコーパスデータとして出力する出力部と、
を備えるコーパス生成装置。

[請求項14] 提示対象に関する説明文データを含むウェブページを取得するウェブページ取得ステップと、

属性名と属性値とが関連付けられた辞書データを記憶する辞書データ記憶部を参照し、前記説明文データに含まれる前記提示対象に関する属性値の候補を抽出する候補抽出ステップと、

前記ウェブページから前記提示対象に関する属性値である基準語を取得する基準語取得ステップと、

属性値間の上下関係を表す上下関係情報を記憶する記憶部から、前記基準語取得ステップにおいて取得された前記基準語よりも上位に属する上位語を抽出し、前記説明文データに含まれる前記属性値の候補の内、前記上位語と一致する属性値の候補に前記属性値の候補に対応する属性タグを付与する付与ステップと、

前記付与ステップにおいて属性タグが付与された説明文データをコーパスデータとして出力する出力ステップと、
を含むコーパス生成方法。

[請求項15] コンピュータを、

提示対象に関する説明文データを含むウェブページを取得するウェブページ取得部と、

属性名と属性値とが関連付けられた辞書データを記憶する辞書データ記憶部を参照し、前記説明文データに含まれる前記提示対象に関する属性値の候補を抽出する候補抽出部と、

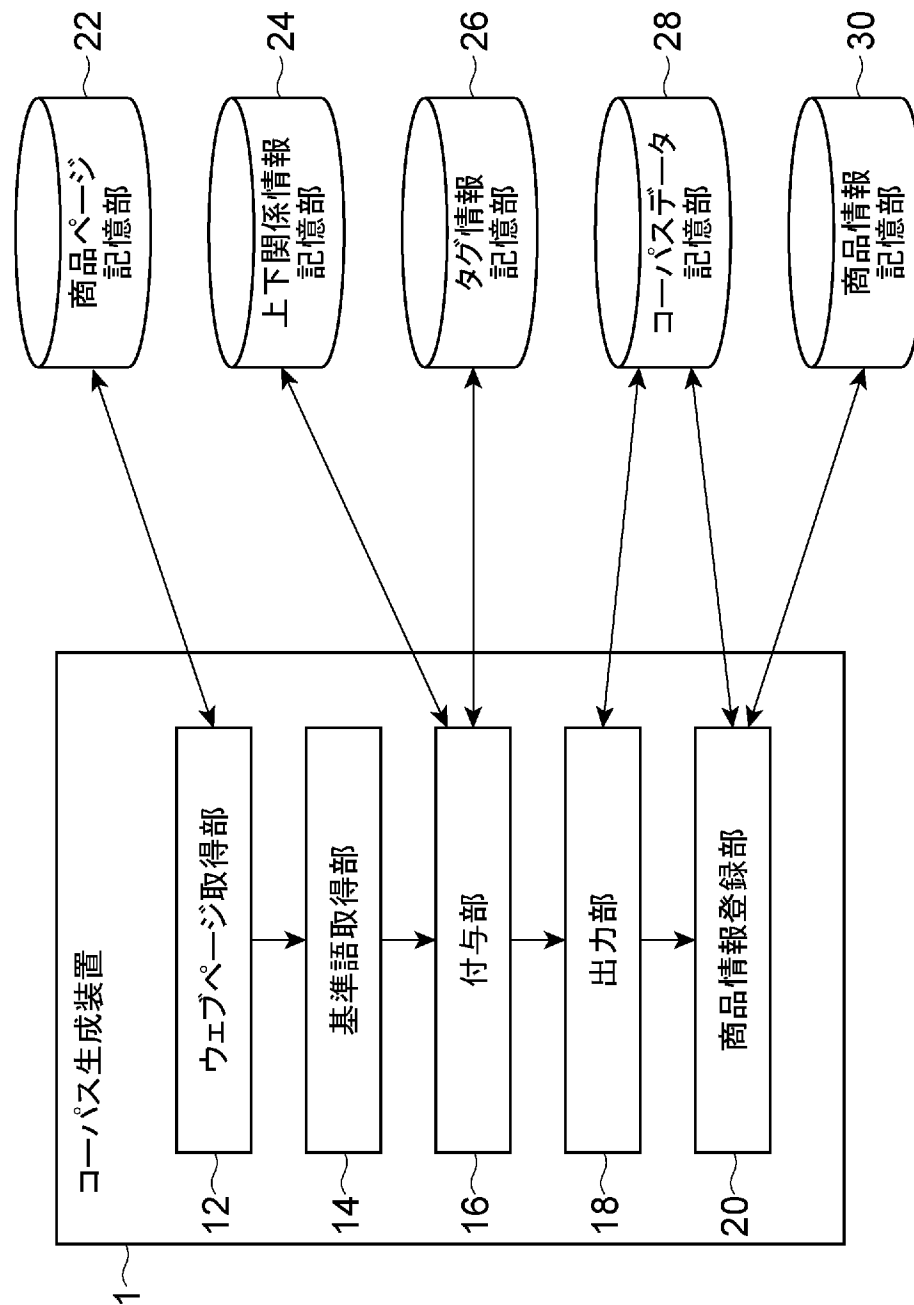
前記ウェブページから前記提示対象に関する属性値である基準語を取得する基準語取得部と、

属性値間の上下関係を表す上下関係情報を記憶する記憶部から、前記基準語取得部により取得された前記基準語よりも上位に属する上位語を抽出し、前記説明文データに含まれる前記属性値の候補の内、前記上位語と一致する属性値の候補に前記属性値の候補に対応する属性タグを付与する付与部と、

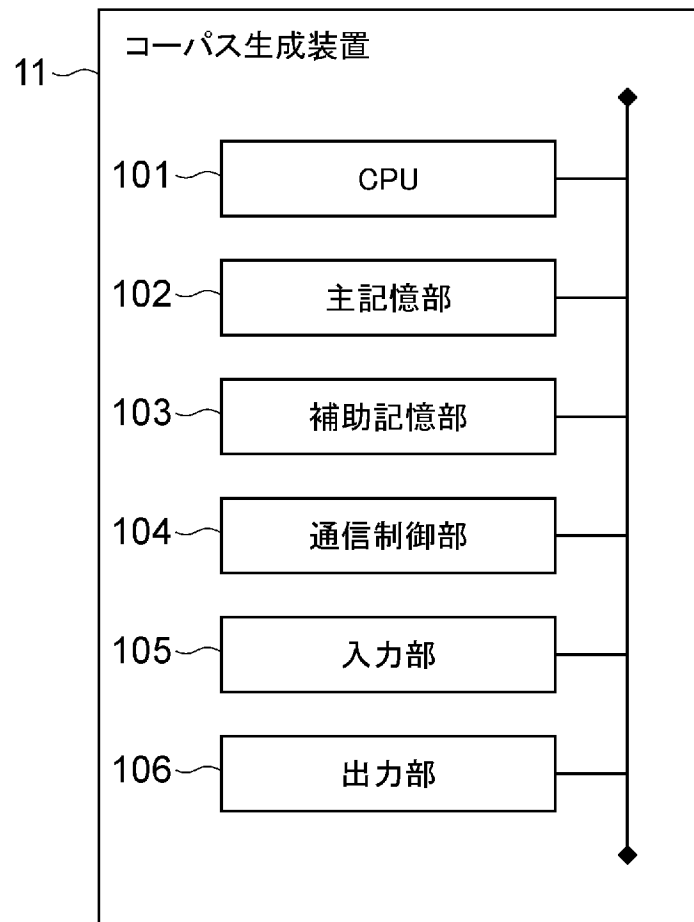
前記付与部により属性タグが付与された説明文データをコーパスデータとして出力する出力部と、

して機能させる、コーパス生成プログラム。

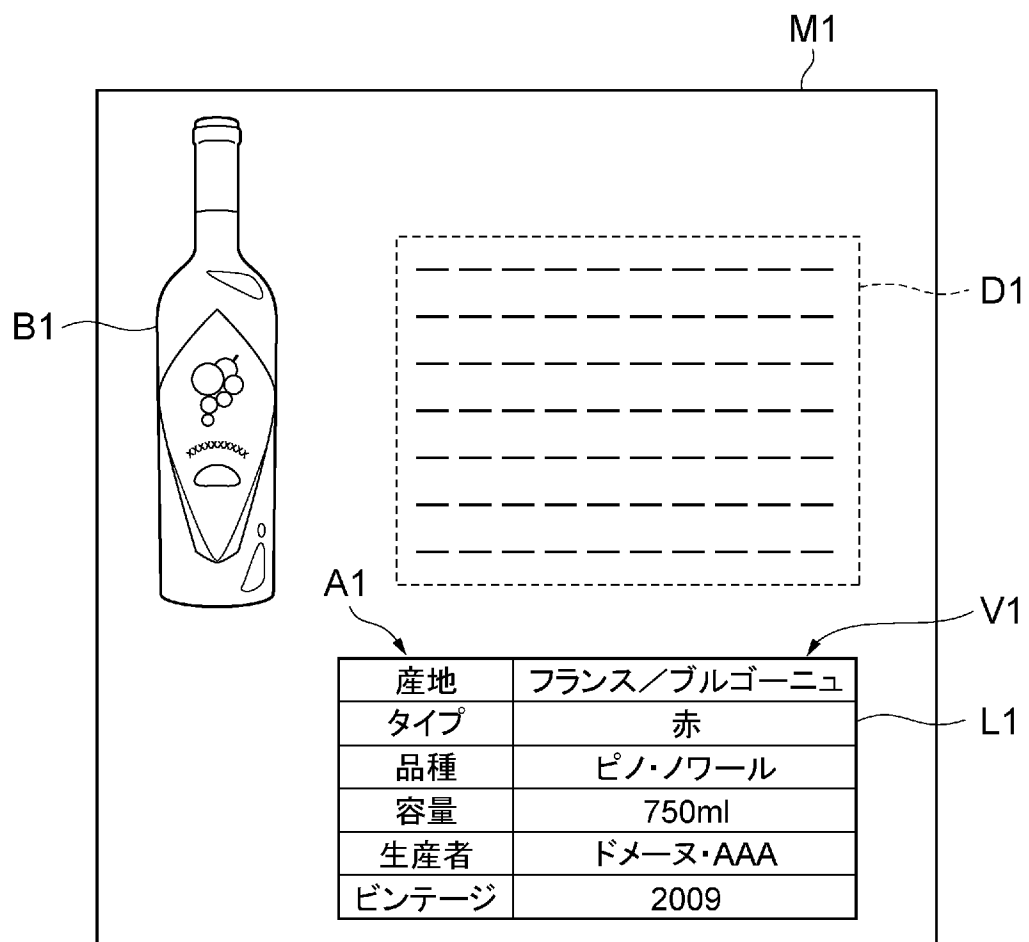
[図1]



[図2]

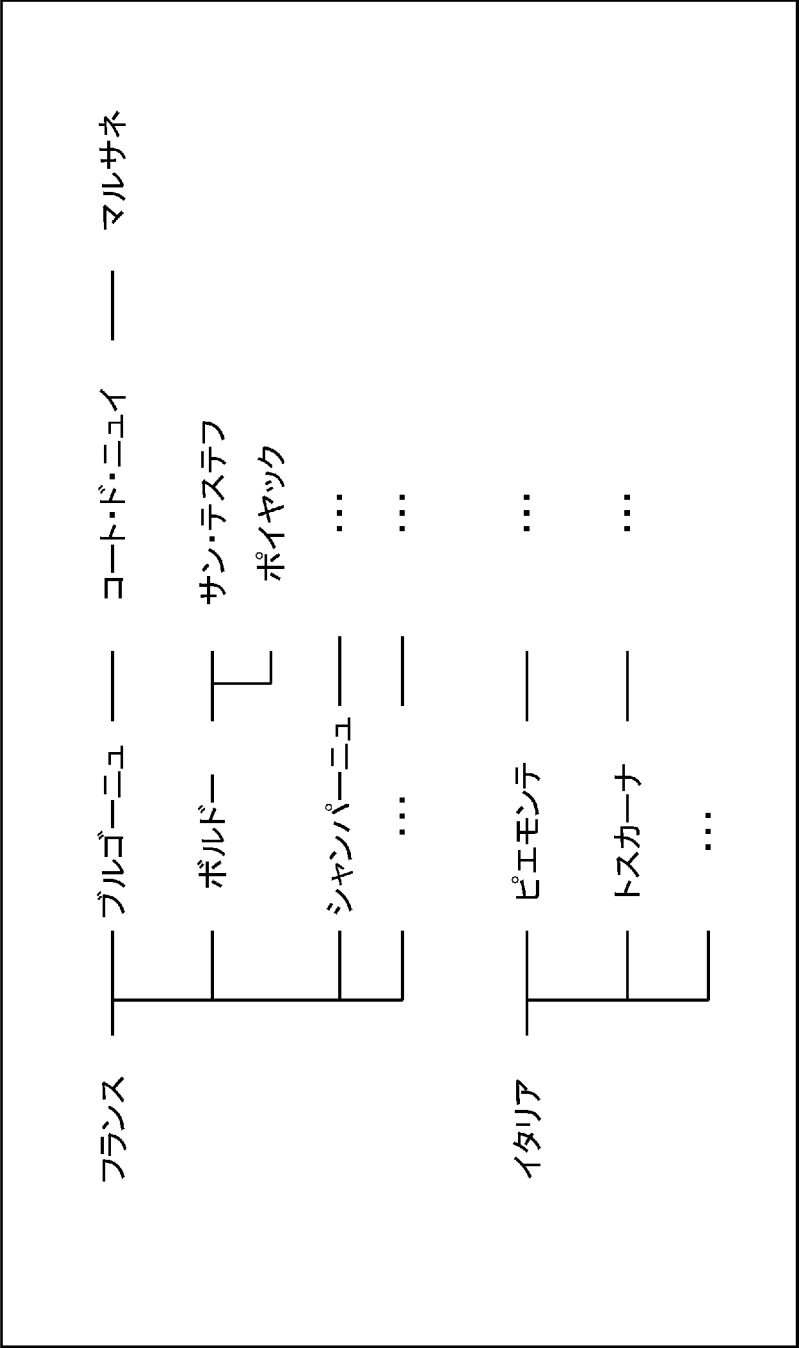


[図3]



[図4]

上下関係情報



[図5]

タグ情報

産地	フランス、イタリア、ブルゴーニュ、コート・ド・ニューイ、...
タイプ	赤、白、ロゼ、...
品種	シャルドネ、メルロー、ピノ・ノワール、...
...	...

[図6]

(a)

フランスのブルゴーニュで生産されたワインです。このワインは、ピエモンテ産のワインにも引けをとらない・・・

D2

(b)

A2	V2	L2
産地	ブルゴーニュ	
タイプ	赤	
品種	ピノ・ノワール	
⋮	⋮	

(c)

〈産地〉フランス〈／産地〉の〈産地〉ブルゴーニュ〈／産地〉で生産されたワインです。このワインは、ピエモンテ産のワインにも引けをとらない・・・

C2

[図7]

商品情報

店舗ID	商品ID	商品名	カテゴリ	価格	URL	在庫数	商品タグ
S001	M001	商品A	...	...	...	...	産地: フランス、ブルゴーニュ タイプ: 赤
S002	M001	商品B	...	...	...	...	産地: フイタリア、ピエモンテ タイプ: 赤
...	...	...	...	...	...	...	...

[図8]

(a)

＜産地＞ナパ＜／産地＞には、ボルドースタイルのワイン
がたくさんあります。

Cx

(b)

チリには、ボルドースタイルのワインがたくさん
あります。

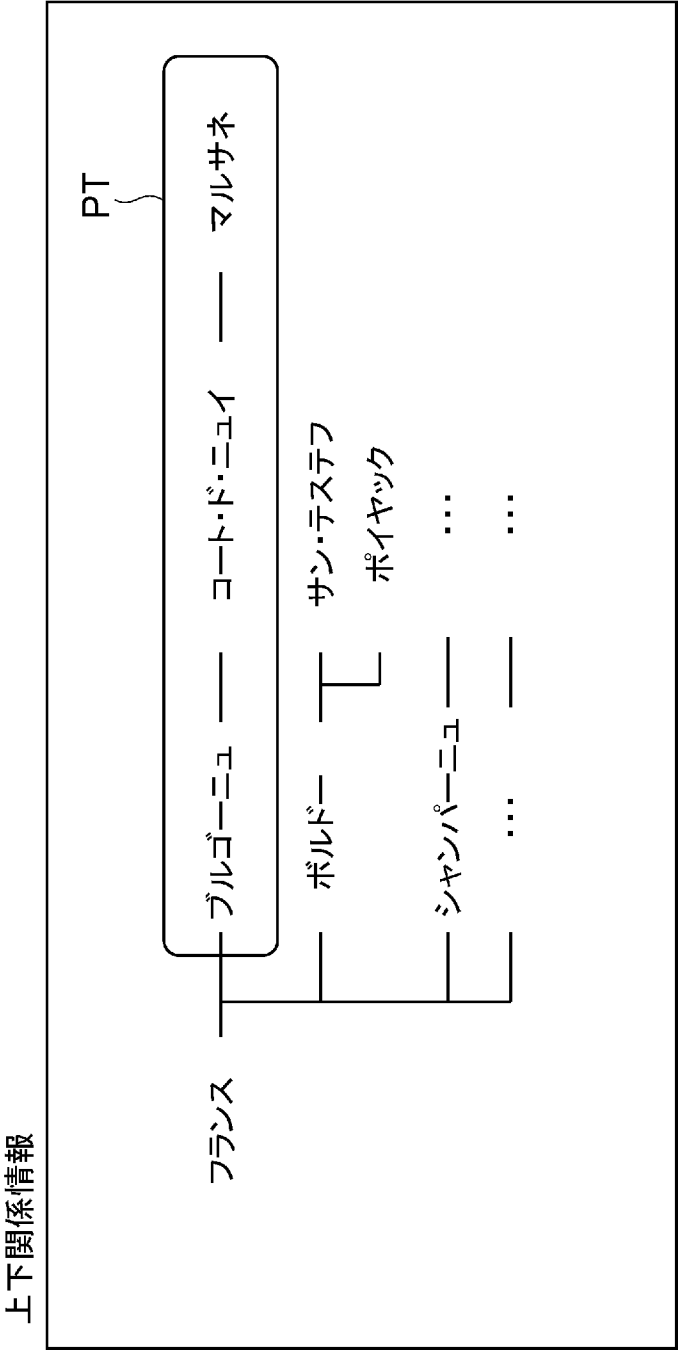
D3

(c)

＜産地＞チリ＜／産地＞には、ボルドースタイルのワイン
がたくさんあります。

C3

[図9]



[図10]

(a)

フランス、ボルドーのサン・テステフ地区で生産されたワインです。

D4

(b)

〈産地〉フランス〈/産地〉、〈産地〉ボルドー〈/産地〉の、〈産地〉サン・テステフ〈/産地〉地区で生産されたワインです。

C4

[図11]

(a)

フランス、ボルドー産のワイン。ワイナリーはポイヤック地区
の北に位置するサン・テステフ地区の名門XXX。

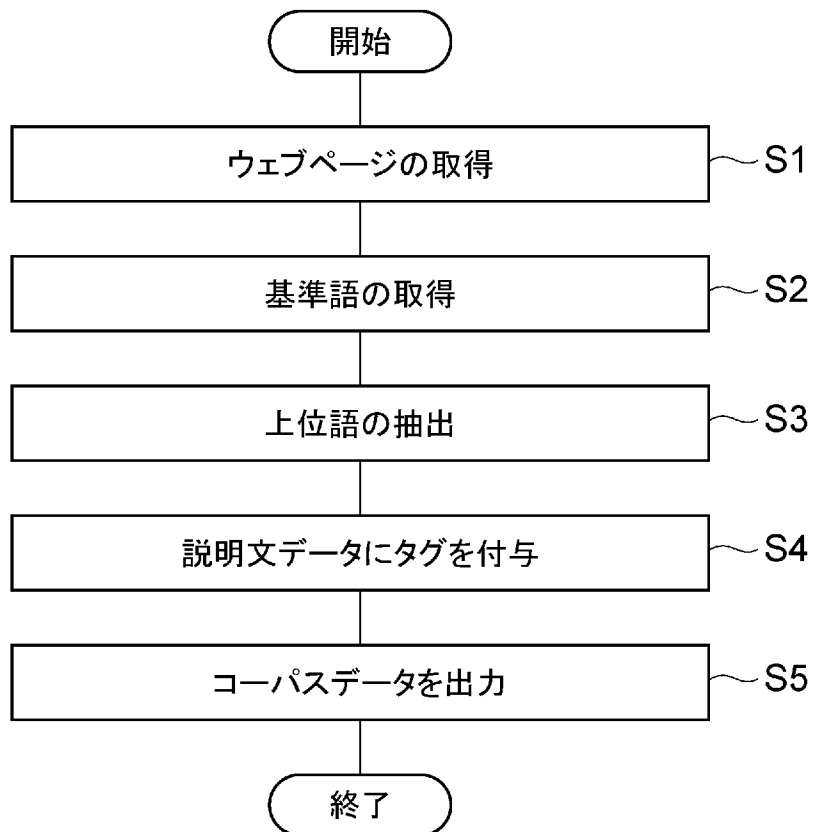
D5

(b)

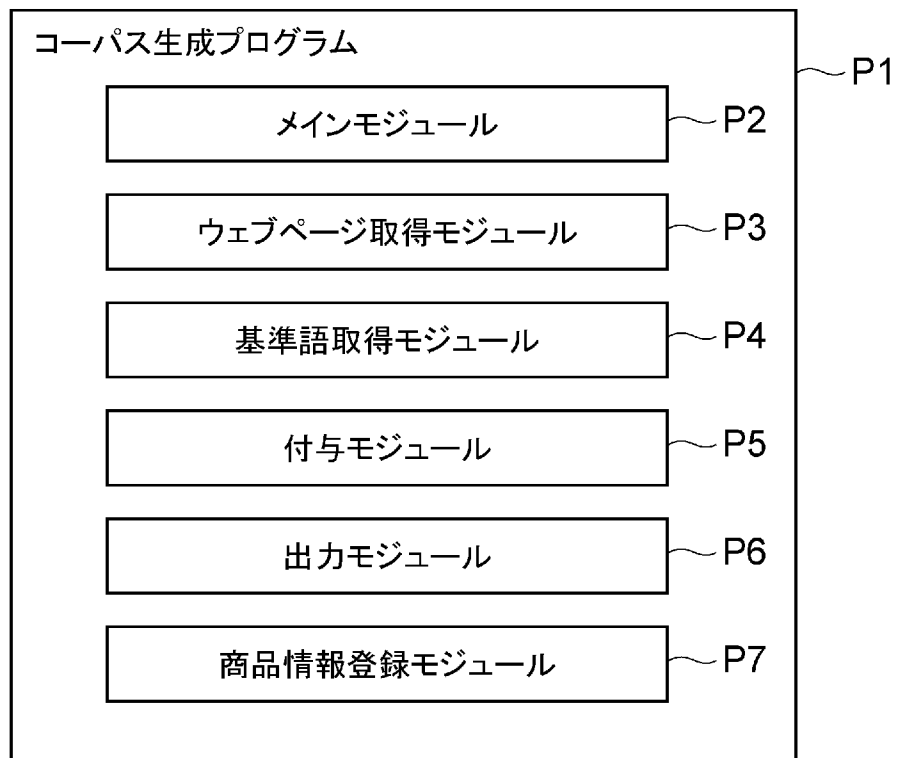
〈産地〉フランス〈／産地〉、〈産地〉ボルドー〈／産地〉産のワイ
ン。ワイナリーはポイヤック地区の北に位置するサン・テス
テフ地区の名門XXX。

C5

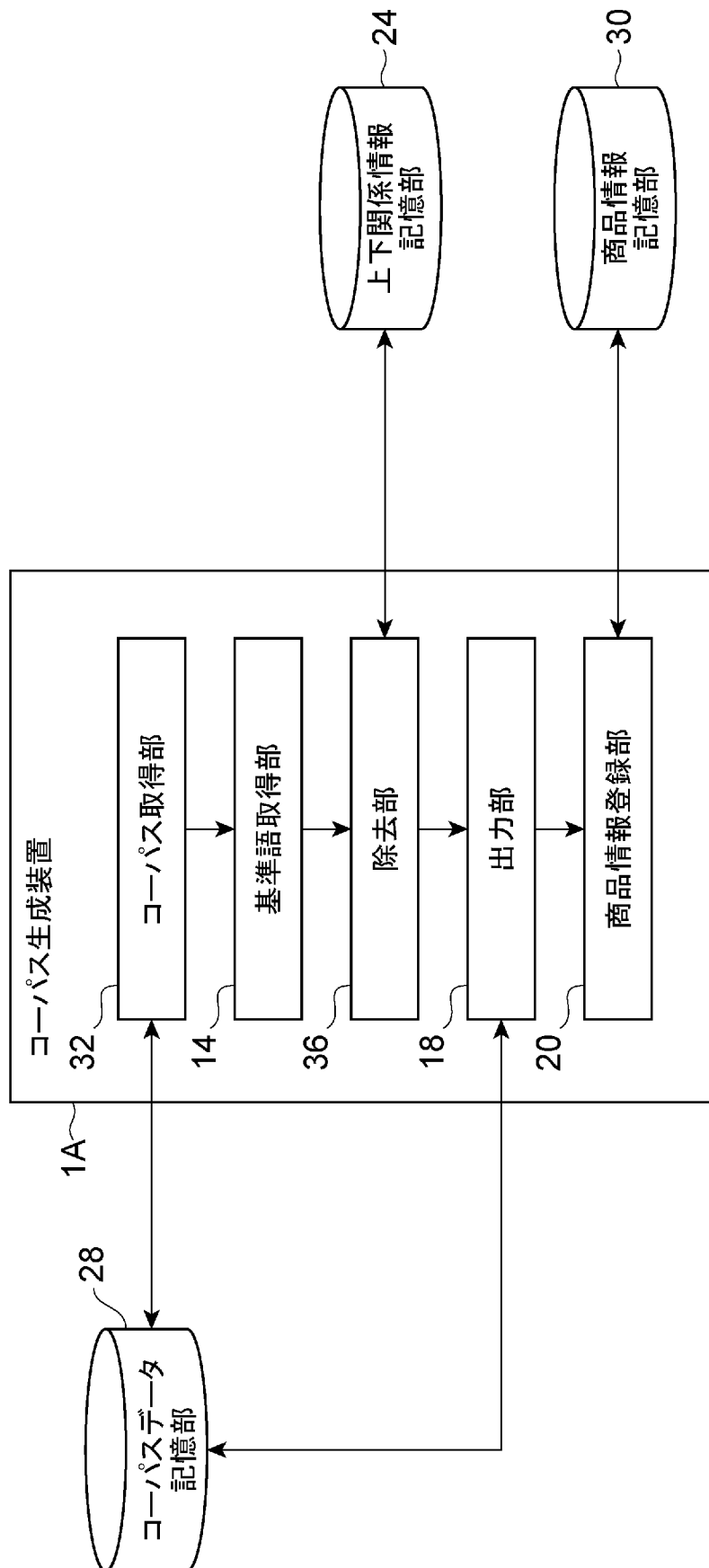
[図12]



[図13]



[図14]



[図15]

(a)

〈産地〉フランス〈／産地〉の〈産地〉ブルゴーニュ〈／産地〉で
生産されたワインです。このワインは、〈産地〉ピエモンテ〈／
産地〉産のワインにも引けをとらない・・・

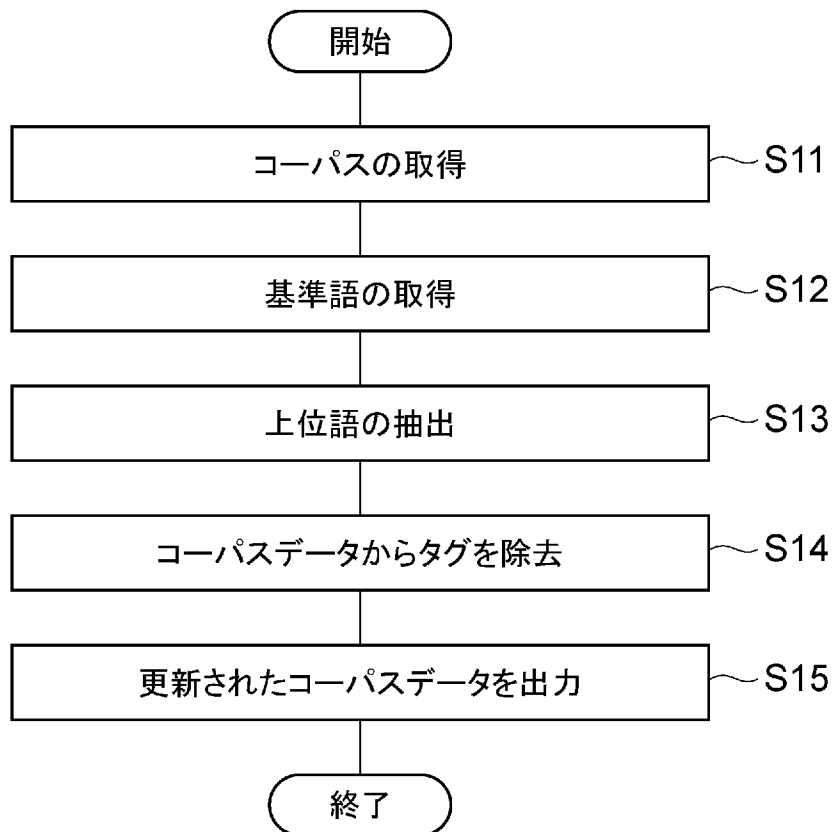
C6a

(b)

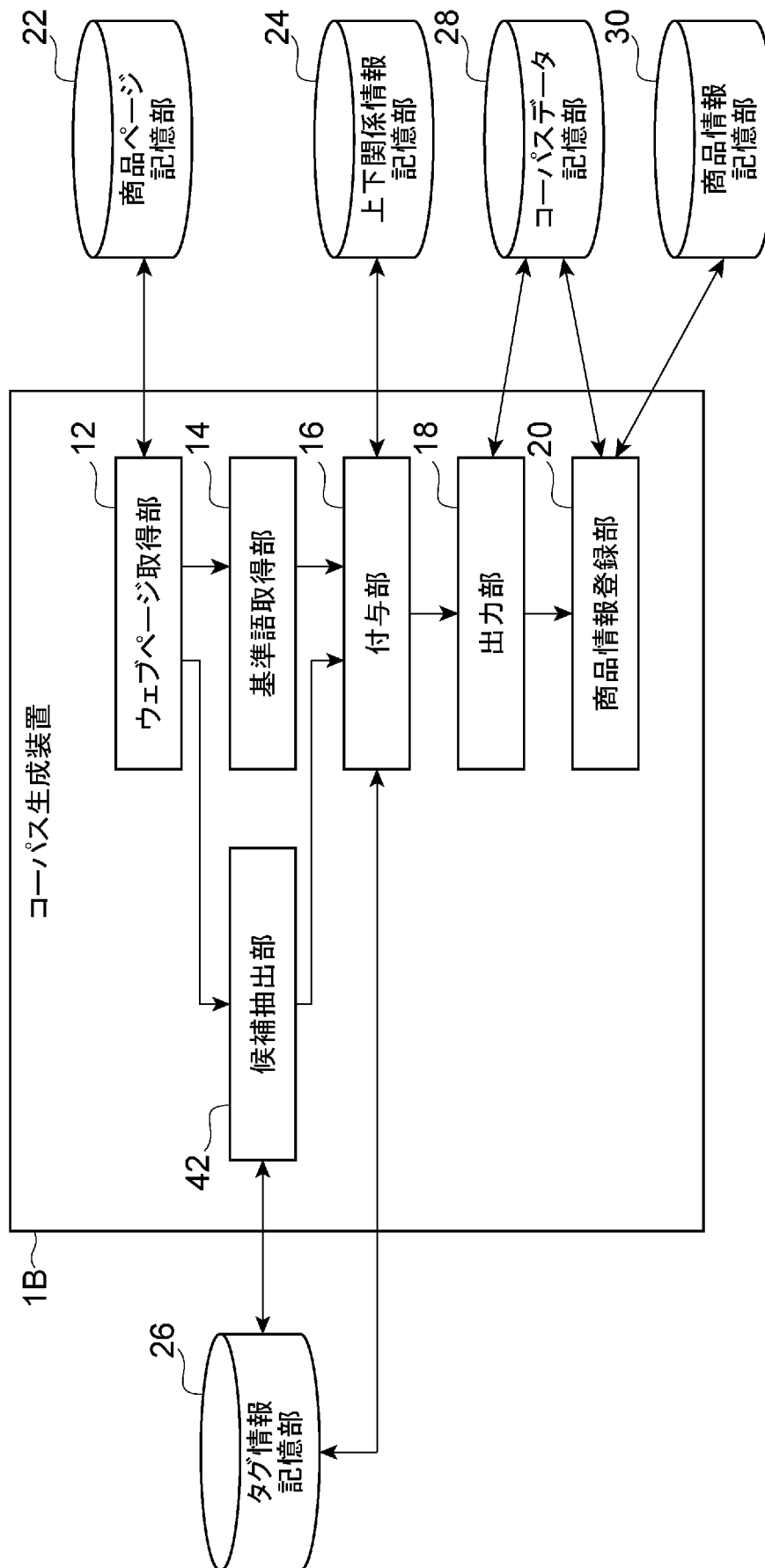
〈産地〉フランス〈／産地〉の〈産地〉ブルゴーニュ〈／産地〉で
生産されたワインです。このワインは、ピエモンテ産のワイ
ンにも引けをとらない・・・

C6b

[図16]



[図17]



[図18]

(a)

フランスのブルゴーニュで生産されたワインです。このワインは、ピエモンテ産のワインにも引けをとらない・・・

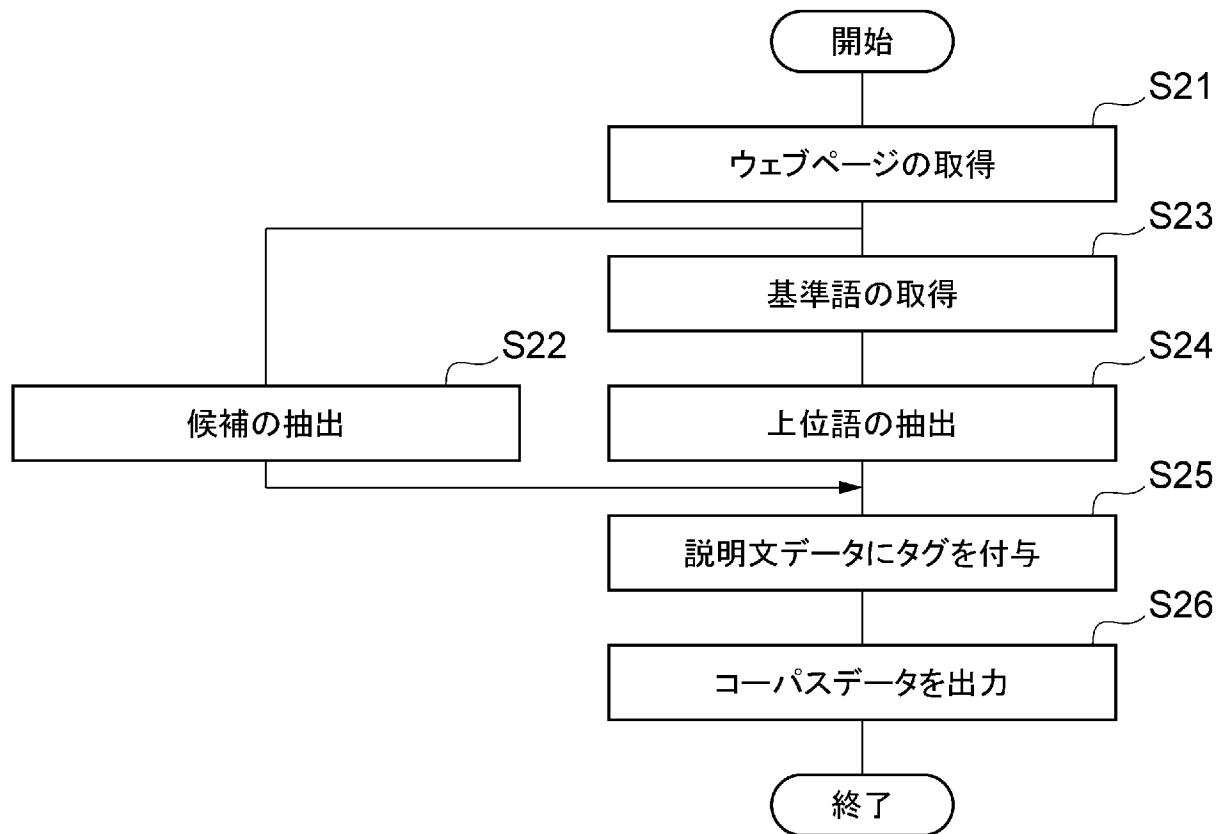
D7

(b)

〈産地〉フランス〈／産地〉の〈産地〉ブルゴーニュ〈／産地〉で生産されたワインです。このワインは、ピエモンテ産のワインにも引けをとらない・・・

C7

[図19]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2013/076545

A. CLASSIFICATION OF SUBJECT MATTER

G06F17/21(2006.01)i, G06F17/30(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)

G06F17/20-17/30

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 2009-181408 A (Nippon Telegraph and Telephone Corp.), 13 August 2009 (13.08.2009), entire text; all drawings (Family: none)	1-15
A	JP 2009-26195 A (Yokohama National University), 05 February 2009 (05.02.2009), entire text; all drawings (Family: none)	1-15
A	JP 7-325827 A (Mitsubishi Electric Corp.), 12 December 1995 (12.12.1995), paragraphs [0058] to [0070]; fig. 13, 15, 17 (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
22 November, 2013 (22.11.13)

Date of mailing of the international search report
03 December, 2013 (03.12.13)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

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. G06F17/21(2006.01)i, G06F17/30(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. G06F17/20-17/30

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 3 年
日本国実用新案登録公報	1 9 9 6 - 2 0 1 3 年
日本国登録実用新案公報	1 9 9 4 - 2 0 1 3 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 2009-181408 A（日本電信電話株式会社）2009.08.13, 全文・全図（ファミリーなし）	1-15
A	JP 2009-26195 A（国立大学法人横浜国立大学）2009.02.05, 全文・全図（ファミリーなし）	1-15
A	JP 7-325827 A（三菱電機株式会社）1995.12.12, 【0058】 - 【0070】段落, 【図 13】, 【図 15】, 【図 17】, （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

2 2 . 1 1 . 2 0 1 3

国際調査報告の発送日

0 3 . 1 2 . 2 0 1 3

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）

成瀬 博之

5 M

9 1 9 2

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