

(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
2011年5月5日(05.05.2011)

PCT

(10) 国際公開番号
WO 2011/052526 A1

- (51) 国際特許分類:
G06Q 50/00 (2006.01)
- (21) 国際出願番号: PCT/JP2010/068820
- (22) 国際出願日: 2010年10月25日(25.10.2010)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2009-250594 2009年10月30日(30.10.2009) JP
特願 2009-250646 2009年10月30日(30.10.2009) JP
- (71) 出願人 (米国を除く全ての指定国について): 楽天株式会社 (Rakuten, Inc.) [JP/JP]; 〒1400002 東京都品川区東品川四丁目1番3号 Tokyo (JP).
- (72) 発明者: および
- (75) 発明者/出願人 (米国についてのみ): 岩淵 志学 (IWABUCHI Shigaku) [JP/JP]; 〒1400002 東京都品川区東品川四丁目1番3号 楽天株式会社内 Tokyo (JP).
- (74) 代理人: 特許業務法人 インテクト国際特許事務所, 外 (INTECT INTERNATIONAL PATENT OFFICE et al.); 〒1020083 東京都千代田区麹町四丁目7番2号 サンライン第7ビル4階 Tokyo (JP).

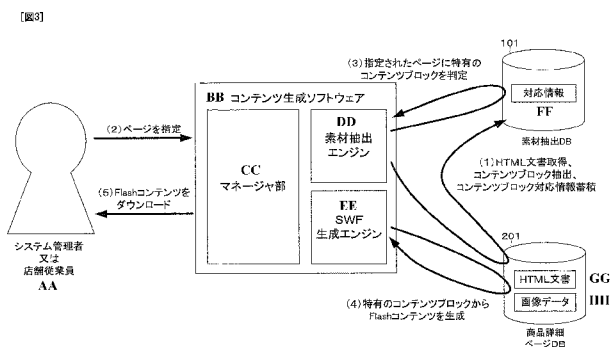
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, 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, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:

- 一 国際調査報告 (条約第21条(3))

(54) Title: CHARACTERISTIC CONTENT DETERMINATION PROGRAM, CHARACTERISTIC CONTENT DETERMINATION DEVICE, CHARACTERISTIC CONTENT DETERMINATION METHOD, RECORDING MEDIUM, CONTENT GENERATION DEVICE, AND RELATED CONTENT INSERTION DEVICE

(54) 発明の名称: 特有コンテンツ判定プログラム、特有コンテンツ判定装置、特有コンテンツ判定方法、記録媒体、コンテンツ生成装置及び関連コンテンツ挿入装置



(57) Abstract: Disclosed is a device that easily extracts content that is characteristic of a Web page from content constituting the Web page. The disclosed device causes a computer to function as an extraction means for extracting contents constituting a Web page, a calculation means for calculating the frequency of appearance of each content element constituting the designated Web page, and a determination means for determining content that is characteristic in the designated Web page from among the contents constituting the Web page, on the basis of the calculated frequency of appearance.

(57) 要約: Web ページを構成しているコンテンツの中からのそのWeb ページ特有のコンテンツを容易に抽出する。コンピュータを、指定されたWeb ページを構成しているコンテンツを抽出する抽出手段、指定されたWeb ページを構成している各コンテンツの出現頻度を計算する計算手段、及び、計算された出現頻度に基づいて、指定されたWeb ページを構成しているコンテンツのうち、当該Web ページに特有であるコンテンツを判定する判定手段、として機能させる。

- 101 SOURCE MATERIAL EXTRACTION DATABASE
- 201 PRODUCT DETAIL PAGE DATABASE
- (1) ACQUIRE HTML DOCUMENT, EXTRACT CONTENT BLOCK, ACCUMULATE CONTENT BLOCK CORRESPONDENCE INFORMATION
- (2) DESIGNATE PAGE
- (3) DETERMINE CONTENT BLOCK CHARACTERISTIC IN DESIGNATED PAGE
- (4) GENERATE FLASH CONTENT FROM CHARACTERISTIC CONTENT BLOCK
- (5) DOWNLOAD FLASH CONTENT
- AA SYSTEM ADMINISTRATOR OR SHOP EMPLOYEE
- BB CONTENT GENERATION SOFTWARE
- CC MANAGER UNIT
- DD SOURCE MATERIAL EXTRACTION ENGINE
- EE SWF GENERATION ENGINE
- FF CORRESPONDING INFORMATION
- GG HTML DOCUMENT
- HH IMAGE DATA

WO 2011/052526 A1

明 細 書

発明の名称：

特有コンテンツ判定プログラム、特有コンテンツ判定装置、特有コンテンツ判定方法、記録媒体、コンテンツ生成装置及び関連コンテンツ挿入装置

技術分野

[0001] 本発明は、Webページを構成するコンテンツを抽出する技術分野に関する。

背景技術

[0002] 従来、Webサイト上に公開されているWebページを構成している素材たるコンテンツを取得し、取得したコンテンツに基づいて新たなコンテンツを生成する技術が知られている。例えば、非特許文献1には、ユーザにより画像データのURLが指定されると、当該URLに対応する画像データをWeb上から取得し、取得した画像データに基づいてバナーを自動作成する技術が開示されている。

先行技術文献

非特許文献

[0003] 非特許文献1：“バナー自動作成”、[online]、[平成21年10月21日検索]、インターネット<URL：<http://hyperbannermaker.com/>>

発明の概要

発明が解決しようとする課題

[0004] Webサイトを構成する各Webページには、そのWebサイトの目的に沿った内容が掲載される。そのため、Webサイトを構成する各Webページの内容は、基本的には互いに関連性を有しているのであるが、夫々何らかの特徴を有してる場合がある。そして、そのWebページの内容を決める要因が、Webページを構成しているコンテンツ（例えば、テキストデータ、画像データ等）の内容である。従って、Webページを構成しているコンテ

ンツの中で、そのWebページを特徴付けるコンテンツ、すなわち、そのWebページ特有のコンテンツが存在することがある。

[0005] 非特許文献1に記載の技術は、Webページに特有のコンテンツを抽出するものであるが、自動的に抽出するものではなく、ユーザが手作業でコンテンツを指定しなければならず、該Webページに特有のコンテンツを容易に抽出することはできない。そのため、どのコンテンツがWebページ特有のコンテンツであるかをユーザが判断することができない場合、又はユーザの嗜好によって好みのコンテンツが偏ってしまう場合など、Webページに特有のコンテンツを的確に抽出することができない。また、対象とするWebページのページ数が多いと、ユーザの作業が甚大となってしまう問題があった。

[0006] また、例えば、HTML (HyperText Markup Language) 文書のタグの記述に基づいて、画像だけ又はテキストだけといったように、特定種類のコンテンツを全て抽出することは可能である。しかしながら、抽出されたコンテンツの中には、Webページ特有のものではないありふれたコンテンツも含まれるので、抽出結果としての信頼性が低く、そのため抽出結果からユーザが特有のコンテンツを探さなければならなかった。

[0007] 本発明は以上の点に鑑みてなされたものであり、Webページを構成しているコンテンツの中からのそのWebページ特有のコンテンツを容易に抽出することができる特有コンテンツ判定装置、特有コンテンツ判定方法、特有コンテンツ判定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

[0008] 上記課題を解決するために、請求項1に記載の発明は、コンピュータを、指定されたWebページを構成しているコンテンツを抽出する抽出手段、前記指定されたWebページを構成している各コンテンツの出現頻度を計算する計算手段、及び、前記計算された出現頻度に基づいて、前記指定されたWebページを構成しているコンテンツのうち、当該Webページに特有であるコンテンツを判定する判定手段、として機能させることを特徴とする。

- [0009] この発明によれば、指定されたWebページを構成している各コンテンツの出現頻度が夫々計算される。この出現頻度が小さいコンテンツであるほど、指定されたWebページ以外にはあまり出現しないコンテンツである。そのため、出現頻度に基づいて、指定されたWebページに特有のコンテンツを特定することができる。よって、Webページに特有のコンテンツを容易に抽出することができる。
- [0010] 請求項2に記載の発明は、請求項1に記載の特有コンテンツ判定プログラムにおいて、前記判定手段が、前記指定されたWebページを構成しているコンテンツのうち、出現頻度が最も小さいコンテンツを当該Webページに特有のコンテンツであると判定するように、前記コンピュータを機能させることを特徴とする。
- [0011] この発明によれば、最も出現頻度が小さいコンテンツを判断することで、指定されたWebページに特有のコンテンツが特定される。
- [0012] 請求項3に記載の発明は、請求項1に記載の特有コンテンツ判定プログラムにおいて、前記判定手段が、前記指定されたWebページを構成しているコンテンツのうち、出現頻度が所定値以下のコンテンツを当該Webページに特有のコンテンツであると判定するように、前記コンピュータを機能させることを特徴とする。
- [0013] この発明によれば、出現頻度が所定値以下であるかを判定することで、当該条件を満たす全てのコンテンツが、指定されたWebページに特有のコンテンツであると特定される。
- [0014] 請求項4に記載の発明は、請求項1乃至3の何れか1項に記載の特有コンテンツ判定プログラムにおいて、前記計算手段が、所定のサイトに含まれる複数のWebページ上における各コンテンツの出現頻度を計算するように、前記コンピュータを機能させることを特徴とする。
- [0015] この発明によれば、所定のサイトに含まれる或る複数のWebページ上において、指定されたWebページを構成している各コンテンツの出現頻度が計算されるので、サイト内で共通して用いられるコンテンツは、Webペー

ジに特有のコンテンツではないと判定することが可能となり、Web ページに特有のコンテンツを判断する精度を上げることができる。

[0016] 請求項 5 に記載の発明は、請求項 1 乃至 4 の何れか 1 項に記載の特有コンテンツ判定プログラムにおいて、前記抽出手段が、前記所定のサイトに含まれる予め定められた種類の各 Web ページについて Web ページを構成しているコンテンツを抽出し、抽出したコンテンツを示すコンテンツ情報を予め記憶手段に記憶しておき、前記計算手段が、前記記憶されたコンテンツ情報に基づいて、前記指定された Web ページを構成している各コンテンツの出現頻度を計算するように、前記コンピュータを機能させることを特徴とする。

[0017] この発明によれば、所定のサイトに含まれる予め定められた種類の各 Web ページについて予め抽出したておいた結果としてのコンテンツ情報に基づいて、指定された Web ページを構成している各コンテンツの出現頻度が計算されるので、出現頻度が正確に計算され、Web ページに特有のコンテンツを判断する精度を上げることができる。

[0018] 請求項 6 に記載の発明は、請求項 1 乃至 5 の何れか 1 項に記載の特有コンテンツ判定プログラムにおいて、前記抽出手段が、1 つ以上のコンテンツで構成されたコンテンツグループの単位で、Web ページを構成しているコンテンツを抽出し、前記計算手段が、前記指定された Web ページを構成しているコンテンツグループの出現頻度を計算し、前記判定手段が、前記指定された Web ページを構成しているコンテンツグループのうち、当該 Web ページに特有であるコンテンツグループを判定するように、前記コンピュータを機能させることを特徴とする。

[0019] この発明によれば、コンテンツグループの単位で Web ページに特有のコンテンツが判断されるので、例えば、Web ページ上において或るまとまりをもって表示されていたり、互いに関連性を有しているようなコンテンツをコンテンツグループとしたときに、Web ページに特有のコンテンツとなるものを抽出することができる。

- [0020] 請求項 7 に記載の発明は、請求項 6 に記載の特有コンテンツ判定プログラムにおいて、前記抽出手段が、所定のマークアップ言語で記述され、Web ページを構成するコンテンツを示すドキュメントデータに基づいて、コンテンツグループを抽出するように、前記コンピュータを機能させることを特徴とする。
- [0021] この発明によれば、Web ページを構成するコンテンツを示すドキュメントデータに基づいてコンテンツグループが抽出されるので、的確にコンテンツグループを抽出することができる。
- [0022] 請求項 8 に記載の発明は、請求項 7 に記載の特有コンテンツ判定プログラムにおいて、前記抽出手段が、前記コンテンツを示すドキュメントデータにおいて予め定められたタグに基づいてコンテンツグループを定めるように、前記コンピュータを機能させることを特徴とする。
- [0023] この発明によれば、予め定められたタグに基づいてコンテンツグループが抽出されるので、Web ページに特有のコンテンツと、特有ではないコンテンツとが夫々予め定められたタグでグループ化されている場合に、Web ページに特有のコンテンツを判断する精度を上げることができる。
- [0024] 請求項 9 に記載の発明は、請求項 1 乃至 8 の何れか 1 項に記載の特有コンテンツ判定プログラムにおいて、特有のコンテンツであると判定されたコンテンツに基づいて、新たなコンテンツを生成する生成手段として前記コンピュータを更に機能させることを特徴とする。
- [0025] この発明によれば、Web ページに特有のコンテンツに基づいて新たなコンテンツが生成されるので、例えば、当該Web ページに掲載されている内容の特徴を示すコンテンツを生成することができる。
- [0026] 請求項 10 に記載の発明は、請求項 9 に記載の特有コンテンツ判定プログラムにおいて、前記生成手段が、特有のコンテンツであると判定されたコンテンツの表示サイズを、予め設定された表示サイズに合うように調整し、表示サイズが調整されたコンテンツを含む新たなコンテンツを生成するように、前記コンピュータを機能させることを特徴とする。

- [0027] 請求項 11 に記載の発明は、請求項 9 又は請求項 10 に記載の特有コンテンツ判定プログラムにおいて、前記生成手段が、特有のコンテンツであると判定されたコンテンツにエフェクトが施されて当該コンテンツが再生される新たなコンテンツを生成するように、前記コンピュータを機能させることを特徴とする。
- [0028] 請求項 12 に記載の発明は、請求項 1 乃至 8 の何れか 1 項に記載の特有コンテンツ判定プログラムにおいて、特有のコンテンツであると判定されたコンテンツに関連する関連コンテンツを、前記指定された Web ページに挿入する挿入手段として前記コンピュータを更に機能させることを特徴とする。
- [0029] この発明によれば、特有のコンテンツであると判定されたコンテンツに関連するコンテンツが、指定された Web ページに挿入されるので、Web ページの特徴と関連する情報を当該 Web ページに追加することができる。
- [0030] 請求項 13 に記載の発明は、請求項 12 に記載の特有コンテンツ判定プログラムにおいて、前記判定手段が、前記指定された Web ページを構成しているコンテンツとして、ブログの記事のテキストデータが含まれている場合に、当該テキストデータを当該 Web ページに特有のコンテンツであると判定し、前記挿入手段が、前記特有コンテンツ判定装置により特有のコンテンツであると判定されたブログの記事のテキストデータから前記指定された Web ページの特徴語を抽出し、当該特徴語に関連する関連コンテンツを、当該 Web ページに挿入するように、前記コンピュータを機能させることを特徴とする。
- [0031] この発明によれば、各ブログの記事のテキストデータにその記事特有の内容が含まれているのであれば、特有コンテンツ判定装置により各ブログの記事のテキストデータを抽出することができる。これにより、Web ページに掲載されているブログの内容に関連する情報を当該 Web ページに追加することができる。
- [0032] 請求項 14 に記載の発明は、指定された Web ページを構成しているコンテンツを抽出する抽出手段と、前記指定された Web ページを構成している

各コンテンツの出現頻度を計算する計算手段と、前記計算された出現頻度に基づいて、前記指定されたWebページを構成しているコンテンツのうち、当該Webページに特有であるコンテンツを判定する判定手段と、を備えることを特徴とする。

[0033] 請求項15に記載の発明は、請求項14に記載の特有コンテンツ判定装置において、前記判定手段は、前記指定されたWebページを構成しているコンテンツのうち、出現頻度が最も小さいコンテンツを当該Webページに特有のコンテンツであると判定することを特徴とする。

[0034] 請求項16に記載の発明は、請求項14に記載の特有コンテンツ判定装置において、前記判定手段は、前記指定されたWebページを構成しているコンテンツのうち、出現頻度が所定値以下のコンテンツを当該Webページに特有のコンテンツであると判定することを特徴とする。

[0035] 請求項17に記載の発明は、請求項14乃至16の何れか1項に記載の特有コンテンツ判定装置において、前記計算手段は、所定のサイトに含まれる複数のWebページ上における各コンテンツの出現頻度を計算することを特徴とする。

[0036] 請求項18に記載の発明は、請求項14乃至17の何れか1項に記載の特有コンテンツ判定装置において、前記抽出手段は、前記所定のサイトに含まれる予め定められた種類の各WebページについてWebページを構成しているコンテンツを抽出し、抽出したコンテンツを示すコンテンツ情報を予め記憶手段に記憶しておき、前記計算手段は、前記記憶されたコンテンツ情報に基づいて、前記指定されたWebページを構成している各コンテンツの出現頻度を計算することを特徴とする。

[0037] 請求項19に記載の発明は、請求項14乃至18の何れか1項に記載の特有コンテンツ判定装置において、前記抽出手段は、1つ以上のコンテンツで構成されたコンテンツグループの単位で、Webページを構成しているコンテンツを抽出し、前記計算手段は、前記指定されたWebページを構成しているコンテンツグループの出現頻度を計算し、前記判定手段は、前記指定さ

れたWebページを構成しているコンテンツグループのうち、当該Webページに特有であるコンテンツグループを判定することを特徴とする。

[0038] 請求項20に記載の発明は、請求項19に記載の特有コンテンツ判定装置において、前記抽出手段は、所定のマークアップ言語で記述され、Webページを構成するコンテンツを示すドキュメントデータに基づいて、コンテンツグループを抽出することを特徴とする。

[0039] 請求項21に記載の発明は、請求項20に記載の特有コンテンツ判定装置において、前記抽出手段は、前記コンテンツを示すドキュメントデータにおいて予め定められたタグに基づいてコンテンツグループを定めることを特徴とする。

[0040] 請求項22に記載の発明は、指定されたWebページを構成しているコンテンツを抽出する抽出行程と、前記指定されたWebページを構成している各コンテンツの出現頻度を計算する計算行程と、前記計算された出現頻度に基づいて、前記指定されたWebページを構成しているコンテンツのうち、当該Webページに特有であるコンテンツを判定する判定行程と、を有することを特徴とする。

[0041] 請求項23に記載の発明は、コンピュータを、指定されたWebページを構成しているコンテンツを抽出する抽出手段、前記指定されたWebページを構成している各コンテンツの出現頻度を計算する計算手段、及び、前記計算された出現頻度に基づいて、前記指定されたWebページを構成しているコンテンツのうち、当該Webページに特有であるコンテンツを判定する判定手段、として機能させる特有コンテンツ判定プログラムがコンピュータ読み取り可能に記録されていることを特徴とする。

[0042] 請求項24に記載の発明は、請求項14乃至21の何れか1項に記載の特有コンテンツ判定装置と、前記特有コンテンツ判定装置により特有のコンテンツであると判定されたコンテンツに基づいて、新たなコンテンツを生成する生成手段と、を備えることを特徴とする。

[0043] 請求項25に記載の発明は、請求項24に記載のコンテンツ生成装置にお

いて、前記生成手段は、特有のコンテンツであると判定されたコンテンツの表示サイズを、予め設定された表示サイズに合うように調整し、表示サイズが調整されたコンテンツを含む新たなコンテンツを生成することを特徴とする。

[0044] 請求項 26 に記載の発明は、請求項 24 又は請求項 25 に記載のコンテンツ生成装置において、前記生成手段は、特有のコンテンツであると判定されたコンテンツにエフェクトが施されて当該コンテンツが再生される新たなコンテンツを生成することを特徴とする。

[0045] 請求項 27 に記載の発明は、請求項 14 乃至 21 の何れか 1 項に記載の特有コンテンツ判定装置と、前記特有コンテンツ判定装置により特有のコンテンツであると判定されたコンテンツに関連する関連コンテンツを、前記指定された Web ページに挿入する挿入手段と、を備えることを特徴とする。

[0046] 請求項 28 に記載の発明は、請求項 27 に記載の関連コンテンツ挿入装置において、前記特有コンテンツ判定装置は、前記指定された Web ページを構成しているコンテンツとして、ブログの記事のテキストデータが含まれている場合に、当該テキストデータを当該 Web ページに特有のコンテンツであると判定し、前記挿入手段は、前記特有コンテンツ判定装置により特有のコンテンツであると判定されたブログの記事のテキストデータから前記指定された Web ページの特徴語を抽出し、当該特徴語に関連する関連コンテンツを、当該 Web ページに挿入することを特徴とする。

発明の効果

[0047] 本発明によれば、出現頻度が小さいコンテンツであるほど、指定された Web ページ以外にはあまり出現しないコンテンツである。そのため、出現頻度に基づいて、指定された Web ページに特有のコンテンツを特定することができる。よって、Web ページに特有のコンテンツを容易に抽出することができる。

図面の簡単な説明

[0048] [図1]一実施形態に係るショッピングシステム S の概要構成の一例を示す図で

ある。

[図2]一実施形態に係るコンテンツ生成サーバ1の概要構成の一例を示すブロック図である。

[図3]Webページが指定されてからFlashコンテンツが生成されるまでの処理の概要を示す図である。

[図4]Webページの構成例を示す図である。

[図5]HTML文書から生成されたDOMツリーの一例を示す図である。

[図6]素材抽出DB101に登録される情報の内容の一例を示す図である。

[図7]一実施形態に係るコンテンツ生成サーバ1のシステム制御部20の素材抽出処理における処理例を示すフローチャートである。

[図8]一実施形態に係るコンテンツ生成サーバ1のシステム制御部20の1ページ対応抽出処理における処理例を示すフローチャートである。

[図9]一実施形態に係るコンテンツ生成サーバ1のシステム制御部20のツリー探索処理における処理例を示すフローチャートである。

[図10]一実施形態に係るコンテンツ生成サーバ1のシステム制御部20のコンテンツ生成処理における処理例を示すフローチャートである。

[図11]一実施形態に係るコンテンツ生成サーバ1のシステム制御部20の特有コンテンツブロック判定処理における処理例を示すフローチャートである。

[図12]一実施形態に係るコンテンツ生成サーバ1のシステム制御部20のFlashコンテンツ生成処理における処理例を示すフローチャートである。

[図13]一実施形態に係るブログシステムBSの概要構成の一例を示す図である。

[図14]一実施形態に係るブログサーバ6の概要構成の一例を示すブロック図である。

[図15]ブロガーが指定されてからブログページに広告コンテンツが挿入されるまでの処理の概要を示す図である。

[図16]Webページの構成例を示す図である。

[図17] HTML文書から生成されたDOMツリーの一例を示す図である。

[図18] 記憶部65に記憶されたコンテンツブロック対応情報の内容の一例を示す図である。

[図19] 一実施形態に係るブログサーバ6のシステム制御部70の広告コンテンツ挿入処理における処理例を示すフローチャートである。

[図20] 一実施形態に係るブログサーバ6のシステム制御部70の1ページ対応抽出処理における処理例を示すフローチャートである。

[図21] 一実施形態に係るブログサーバ6のシステム制御部70の特有コンテンツブロック判定処理における処理例を示すフローチャートである。

[図22] 一実施形態の変形例に係るブログサーバ6のシステム制御部70のブログ更新時処理における処理例を示すフローチャートである。

発明を実施するための形態

[0049] [1. 第1実施形態]

以下、図面を参照して本発明の実施形態について詳細に説明する。なお、以下に説明する実施の形態は、ネットワーク上の電子的なやりとりにより商品の売買が行われるショッピングシステムにおいて、ショッピングサイトのWebページから抽出されたWebページに特有のコンテンツに基づいて新たなコンテンツを生成するサーバ装置に対して本発明を適用した場合の実施形態である。

[0050] [1-1. ショッピングシステムの構成及び機能概要]

先ず、本実施形態に係るショッピングシステムSの構成及び概要機能について、図1を用いて説明する。

[0051] 図1は、本実施形態に係るショッピングシステムSの概要構成の一例を示す図である。

[0052] 図1に示すように、ショッピングシステムSは、特有コンテンツ判定装置及びコンテンツ生成装置の一例としてのコンテンツ生成サーバ1と、ショッピングサーバ2と、管理端末3と、複数の店舗端末4と、複数のユーザ端末5と、を含んで構成されている。そして、コンテンツ生成サーバ1と、ショ

ッピングサーバ2と、各店舗端末4と、各ユーザ端末5とは、ネットワークNWを介して、例えば、通信プロトコルにTCP/IP等を用いて相互にデータの送受信が可能になっている。なお、ネットワークNWは、例えば、インターネット、専用通信回線（例えば、CATV（Community Antenna Television）回線）、移動体通信網（基地局等を含む）、及びゲートウェイ等により構築されている。また、コンテンツ生成サーバ1と管理端末3とは、LAN（Local Area Network）等のネットワークを介して接続されている。なお、コンテンツ生成サーバ1とショッピングサーバ2とが、同じくLAN等のネットワークを介して接続されても良い。

- [0053] このような構成のショッピングシステムSにおいて、ショッピングサーバ2は、店舗端末4やユーザ端末5からのリクエストに応じて、ショッピングサイトを構成するWebページを送信するWebサーバである。また、ショッピングサーバ2は、店舗端末4からのリクエストに基づいて、ショッピングサイトで販売される商品を登録し、その商品の詳細な説明等が掲載される商品詳細ページを生成する。そして、ショッピングサーバ2は、商品詳細ページDB201を備え、生成した商品詳細ページ（商品詳細ページのHTML文書（ドキュメントデータの一例）、商品詳細ページの素材である画像データ等）を商品詳細ページDB201に登録する。また、ショッピングサーバ2は、商品詳細ページを閲覧したユーザのユーザ端末5からのリクエストに基づいて、商品の購入の処理を行う。
- [0054] 店舗端末4は、ショッピングサイトで商品を販売する店舗の従業員に使用される端末装置である。店舗端末4としては、例えば、パーソナルコンピュータ等が用いられる。
- [0055] ユーザ端末5は、ショッピングサイトで商品を購入するユーザに使用される端末装置である。ユーザ端末5としては、例えば、パーソナルコンピュータ、PDA、携帯電話機等が用いられる。
- [0056] コンテンツ生成サーバ1は、管理端末3や店舗端末4からのリクエストに基づいて、指定された商品詳細ページの特徴を示す（ひいては、商品の特徴

を示す) Flashコンテンツ (Adobe Systems社により規格化されたソフトウェア) を生成する。生成されるFlashコンテンツは、例えば、商品のバナー画像、商品を紹介するスライドショーコンテンツ、動画コンテンツ等である。そして、Flashコンテンツは、例えば、店舗が運営するWebサイトに掲載されたり、ショッピングサイトを構成するWebページの素材として用いられたりする。

[0057] このようなFlashコンテンツを生成するため、コンテンツ生成サーバ1は、素材抽出DB101を備え、商品詳細ページDB201に登録されている商品詳細ページを構成するWeb素材としてのコンテンツ (画像データ、HTML文書に記述されているテキストデータ等) を抽出して、その抽出結果を素材抽出DB101に登録しておく。そして、コンテンツ生成サーバ1は、指定された商品詳細ページから抽出されたコンテンツの中からその商品詳細ページ特有のコンテンツを特定し、特定されたコンテンツに基づいてFlashコンテンツを生成する。

[0058] 管理端末3は、ショッピングシステムSのシステム管理者により使用される端末装置である。管理端末3としては、例えば、パーソナルコンピュータ等が用いられる。

[0059] [1-2. コンテンツ生成サーバの構成及び機能]

次に、コンテンツ生成サーバ1の構成及び機能について、図2を用いて説明する。

[0060] 図2は、本実施形態に係るコンテンツ生成サーバ1の概要構成の一例を示すブロック図である。また、図3は、Webページが指定されてからFlashコンテンツが生成されるまでの処理の概要を示す図である。また、図4は、Webページの構成例を示す図である。また、図5は、HTML文書から生成されたDOMツリーの一例を示す図である。また、図6は、素材抽出DB101に登録される情報の内容の一例を示す図である。

[0061] 図2に示すように、コンテンツ生成サーバ1は、操作部11と、表示部12と、通信部13と、ドライブ部14と、記憶手段の一例としての記憶部1

5と、入出カインタフェース部16と、システム制御部20と、を備えている。そして、システム制御部20と入出カインタフェース部16とは、システムバス21を介して接続されている。

[0062] 操作部11は、例えば、キーボード、マウス等により構成されており、システム管理者等からの操作指示を受け付け、その指示内容を指示信号としてシステム制御部20に出力するようになっている。表示部12は、例えば、CRT (Cathode Ray Tube) ディスプレイ、液晶ディスプレイ等により構成されており、文字や画像等の情報を表示するようになっている。通信部13は、ネットワークNW等に接続して、ショッピングサーバ2、管理端末3、店舗端末4、ユーザ端末5等との通信状態を制御するようになっている。ドライブ部14は、例えば、フレキシブルディスク、CD (Compact Disc)、DVD (Digital Versatile Disc) 等のディスクDKからデータ等を読み出す一方、当該ディスクDKに対してデータ等を記録するようになっている。記憶部15は、例えば、ハードディスクドライブ等により構成されており、各種プログラム及びデータ等を記憶するようになっている。また、記憶部15には、素材抽出DB101が構築されている。入出カインタフェース部16は、操作部11～記憶部15とシステム制御部20との間のインタフェース処理を行うようになっている。システム制御部20は、CPU (Central Processing Unit) 17、ROM (Read Only Memory) 18、RAM (Random Access Memory) 19等により構成されている。

[0063] システム制御部20は、CPU17が、ROM18や記憶部15に記憶された各種プログラムを読み出し実行することによりコンテンツ生成サーバ1の各部を制御する。また、システム制御部20は、コンテンツ生成ソフトウェア (特有コンテンツ判定プログラムの一例) を実行することにより、抽出手段、計算手段、判定手段及び生成手段として機能する。なお、コンテンツ生成ソフトウェア等は、例えば、他のサーバ装置等からネットワークNWを介して取得されるようにしても良いし、CD-ROM等のディスクDKに記録されてドライブ部14を介して読み込まれるようにしても良い。

[0064] コンテンツ生成ソフトウェアは、商品詳細ページ特有のコンテンツに基づいてFlashコンテンツを生成するためのプログラムである。図3に示すように、コンテンツ生成ソフトウェアは、マネージャ部、素材抽出エンジン、SWF (ShockWave Flash Object) 生成エンジン等により構成されている。マネージャ部は、素材抽出エンジン及びSWFエンジンの実行を制御するとともに、コンテンツ生成ソフトウェアを利用するユーザ（店舗従業員やシステム管理者）に対して、Flashコンテンツを生成するためのGUI (Graphical User Interface) を提供するためのソフトウェアである。素材抽出エンジンは、商品詳細ページのHTML文書からWeb素材としてのコンテンツを抽出するとともに、商品詳細ページ特有のコンテンツを判定するためのソフトウェアである。コンテンツの抽出は、後述のコンテンツブロック（コンテンツグループの一例）の単位で行われる。SWFエンジンは、与えられた1又は複数のコンテンツ（Web素材）に基づいてFlashコンテンツを生成するためのソフトウェアである。なお、新たなコンテンツとしてFlashコンテンツ以外のリッチインターネットアプリケーションを生成する場合、SWF生成エンジンに代えて、例えば、マイクロソフト社のSilverlight（商標）の生成エンジンを適用しても良い。また、Ajax (Asynchronous JavaScript（登録商標）+ XML) 等の技術を利用して動的ページを実現するスクリプトを生成するソフトウェアを適用しても良い。

[0065] 以下に、Flashコンテンツの生成の概要について説明する。図3に示すように、システム制御部20は、ショッピングサーバ2から商品詳細ページDB201に登録されているHTML文書を取得して解析し、Web素材たるコンテンツをコンテンツブロック単位で抽出する。そして、その抽出結果として、抽出したコンテンツブロック毎にコンテンツブロック対応情報（コンテンツ情報の一例）を素材抽出DB101に登録する（1）。この処理は、Flashコンテンツの生成前に予め行われ、基本的には、商品詳細ページDB201に登録されている全HTML文書、すなわち、ショッピングサイトを構成する全ての商品詳細ページについて抽出が行われる。

[0066] その後、システム管理者又は店舗従業員により、Flashコンテンツの生成対象となる商品詳細ページのHTML文書のURLが指定される(2)。すると、システム制御部20は、指定されたURLに基づいてショッピングサーバ2からHTML文書を取得し、コンテンツブロックを抽出する。そして、制御部20は、素材抽出DB101を参照して、抽出した各コンテンツブロックの全商品詳細ページにおける出現頻度を計算する。計算される出現頻度としては、出現回数(度数)であっても良いし、全商品詳細ページの全コンテンツブロックに対する出現回数の割合(相対度数)であっても良い。そして、システム制御部20は、出現頻度に基づいて、指定されたURLに対応する商品詳細ページに特有であるコンテンツブロックを判定する。具体的に、システム制御部20は、出現頻度の最も小さいコンテンツブロックを、指定されたURLに対応する商品詳細ページ特有のコンテンツブロックであると判定する(3)。

[0067] システム制御部20は、特有のコンテンツブロックであると判定したコンテンツブロックに含まれるコンテンツをショッピングサーバ2を介して商品詳細ページDB201から取得する。システム制御部20は、取得したコンテンツに基づいてFlashコンテンツを生成する(4)。そして、システム管理者又は店舗従業員は、生成されたFlashコンテンツをダウンロードする(5)。なお、Flashコンテンツのダウンロードの前に、システム管理者や店舗従業員によって適宜Flashコンテンツを修正することができるようにしても良い。

[0068] 次に、コンテンツブロックの抽出方法について説明する。なお、本実施形態においては、テキストデータ及び画像データがWeb素材として抽出されるものとする。

[0069] 例えば、商品詳細ページの構成(レイアウト)が図4に示すようなものであるとする。Web素材としての各コンテンツは、商品詳細ページ上において、それぞれ或るまとまり(かたまり)毎に表示されている。その各まとまりがコンテンツブロックに相当する。各コンテンツは、HTML文書に記述されているDIVタグ及びTABLEタグ(予め定められたタグの一例)に

より夫々コンテンツブロックに分けられる。つまり、各コンテンツは、DIVタグ及びTABLEタグによりブロック化（グループ化）される。なお、DIVタグ及びTABLEタグを、「ブロック化タグ」という。

[0070] 図4には、コンテンツブロック301～306が表示されている。コンテンツブロック301は、例えば、ページのヘッダ部分のコンテンツブロックであり、テキストA及び画像aにより構成されている。また、コンテンツブロック302は、例えば、各種カテゴリーの商品に関するWebページに移動するためのナビゲーション部分のコンテンツブロックであり、例えば、他のWebページへのリンクを示すテキストB、テキストC及びテキストDにより構成されている。また、コンテンツブロック303は、例えば、商品に関する情報が表示されるコンテンツブロックであり、商品名等の見出しを示すテキストE、コンテンツブロック304及びコンテンツブロック305により構成されている。このように、コンテンツブロックが入れ子状、つまり、階層構造になっている場合もある。この場合、コンテンツブロック303に含まれるコンテンツはテキストEのみとされ、コンテンツブロック304及びコンテンツブロック305は、コンテンツブロック303から独立しているものとされる。コンテンツブロック304は、例えば、商品の詳細を示すコンテンツブロックであり、詳細な説明を示すテキストF、商品等の画像としての画像b及び画像cにより構成されている。コンテンツブロック305は、例えば、商品購入の際の一般的な注意事項等を示すコンテンツブロックであり、テキストG及びテキストHにより構成されている。コンテンツブロック306は、例えば、コピーライト表示を示すコンテンツブロックであり、テキストIにより構成されている。

[0071] これらのコンテンツブロックのうち、コンテンツブロック301、302、305及び306は、図4に示す商品詳細ページ以外の商品詳細ページ上でも比較的頻繁に現れる。一方、コンテンツブロック303（テキストE）及びコンテンツブロック304は、当該商品詳細ページにだけに用いられるか、又は、他の商品詳細ページで用いられるとしても、その頻度はコンテン

ツブロック 301、302、305 及び 306 に比べて小さい。よって、例えば、コンテンツブロック 303 又はコンテンツブロック 304 が、当該商品詳細ページ特有のコンテンツブロックであると判断されることとなる。

[0072] 図 4 に示す商品詳細ページの HTML 文書を DOM (Document Object Model) ツリー、すなわち、木構造で表したものが図 5 である。なお、図 5 に示す DOM ツリーにおいて、本実施形態の説明に必要なないタグのノードの図示は省略している。

[0073] DOM ツリーにおいては、DIV タグを示す DIV ノード及び TABLE タグを示す TABLE ノードが、各コンテンツをコンテンツブロックにブロック化するノードとされる（以下、「ブロック化ノード」という）。システム制御部 20 は、例えば、深さ優先探索により DOM ツリーを探索し、コンテンツブロックを確定していく。具体的に、システム制御部 20 は、ブロック化ノードを発見すると、当該ノードを頂点とする部分木の各ノードに規定されているコンテンツを一まとめにしてコンテンツブロックとする。ただし、コンテンツブロックが階層的に規定された結果、或るブロック化ノード（以下、「上位のブロック化ノード」という）を発見した後、当該ブロック化ノードの子孫のノードの中から更にブロック化ノード（以下、「下位のブロック化ノード」という）を発見すると、コンテンツブロックが分割される。例えば、ルートノードからの距離が近いノードほど階層的により上位のノードであるとする、上位のブロック化ノードを頂点とする部分木（以下、「上位部分木」という）に相当するコンテンツブロックを、下位のブロック化ノードを頂点とする部分木（以下、「下位部分木」という）に相当するコンテンツブロックと、上位部分木のうち下位部分木を除く部分に相当するコンテンツブロックとに分ける（例えば、コンテンツブロック 304 とコンテンツブロック 303）。この場合においては、前者のコンテンツブロックを、後者のコンテンツブロックよりも階層的に下位にあるものとする。例えば、コンテンツブロック 301、302、303 及び 306 の階層は 1 であり、コンテンツブロック 304 及び 305 の階層は 2 である。つまり、階層の値

が小さいほど階層的により上位となる。

[0074] これを、HTML文書のタグの記述で説明すると、基本的にはブロック化タグが記述されていると、ブロック化タグに挟まれた範囲内に規定が記述されているコンテンツが、まとめて当該ブロック化タグに対応するコンテンツブロックとされる。ただし、ブロック化タグが階層的に記述されている場合、或るブロック化タグに挟まれた範囲内に規定が記述されているコンテンツのうち、当該ブロック化タグよりも下位に記述されているブロック化タグに挟まれた範囲に規定が記述されているコンテンツを除いたものが、まとめて当該ブロック化タグに対応するコンテンツブロックとされる。

[0075] このようにしてコンテンツブロックを抽出すると、システム制御部20は、抽出結果を示すコンテンツブロック対応情報を素材抽出DB101に登録する。図6に示すように、コンテンツブロック対応情報(符号401)は、コンテンツブロック毎に登録される。コンテンツブロック対応情報には、抽出元のHTML文書のURL設定部分(符号402)とブロック構成情報(符号403)とにより構成されている。ブロック構成情報には、抽出された各コンテンツが設定される。ここで、テキストデータについては、DOMツリーにおけるテキストノードの内容が設定される。一方、画像データについては、画像データそのものの代わりとして、DOMツリーにおいてIMGタグを示すIMGノードのsrc属性としての画像データのURLが設定される。

[0076] [1-3. ショッピングシステムの動作]

次に、ショッピングシステムSの動作について、図7乃至図12を用いて説明する。

[0077] [1-3-1 全商品詳細ページからのコンテンツブロック抽出時]

図7は、本実施形態に係るコンテンツ生成サーバ1のシステム制御部20の素材抽出処理における処理例を示すフローチャートである。

[0078] 素材抽出処理は、例えば、定期的に行なわれたり、システム管理者の操作に基づいて管理端末3から素材抽出処理の実行のリクエストが送信されてき

たときに開始される。素材抽出処理において、システム制御部20は、商品詳細ページDB201に登録されている全てのHTML文書を解析する。そのために、システム制御部20は、例えば、商品詳細ページDB201に登録されているHTML文書の一覧の情報を予めショッピングサーバ2から取得しておき、この一覧の情報に基づいてHTML文書を取得しても良いし、ショッピングサイトのトップページのHTML文書から次々とリンクを辿って、商品詳細ページのHTML文書を順次取得しても良い。

[0079] 図7に示すように、システム制御部20は、素材抽出DB101を初期化する（ステップS1）。具体的に、システム制御部20は、素材抽出DB101にコンテンツブロック対応情報が登録されている場合には、全てのコンテンツブロック対応情報を素材抽出DB101から削除する。

[0080] 次いで、システム制御部20は、全商品詳細ページのうち最初に取得すべき商品詳細ページのHTML文書のURLを特定し（ステップS2）、特定したURLを設定したリクエストをショッピングサーバ2に送信することにより、当該ショッピングサーバ2からHTML文書を取得する（ステップS3）。次いで、システム制御部20は、取得したHTML文書を指定して、後述する1ページ対応抽出処理を実行する（ステップS4）。この1ページ対応抽出処理では、取得したHTML文書からコンテンツブロックが抽出され、コンテンツブロック対応情報が登録される。

[0081] 次いで、システム制御部20は、全ての商品詳細ページのコンテンツブロックを抽出したか否かを判定する（ステップS5）。このとき、システム制御部20は、コンテンツブロックを抽出していない商品詳細ページが存在する場合には（ステップS5：NO）、次の商品詳細ページのHTML文書のURLを特定して（ステップS6）、ステップS3に移行する。そして、システム制御部20は、ステップS3～S6の処理を繰り返して全ての商品詳細ページのコンテンツブロックを抽出すると（ステップS5：YES）、素材抽出処理を終了させる。

[0082] なお、システム制御部20は、素材抽出DB101を初期化してコンテン

ツブロック対応情報を登録し直さなくても良い。例えば、システム制御部20は、素材抽出DB101を初期化を行わず、前回素材抽出処理を実行した後に新規生成された商品詳細ページについてコンテンツブロック対応情報を生成して素材抽出DB101に追加登録し、また、前回素材抽出処理を実行した後に更新された商品詳細ページについてコンテンツブロック対応情報を生成して素材抽出DB101に更新登録しても良い。

- [0083] 図8は、本実施形態に係るコンテンツ生成サーバ1のシステム制御部20の1ページ対応抽出処理における処理例を示すフローチャートである。
- [0084] 図8に示すように、システム制御部20は、先ず、取得したHTML文書のDOMツリーをRAM19上に生成する（ステップS21）。
- [0085] 次いで、システム制御部20は、ブロック数NUMに0を設定するとともに、階層LVに0を設定する（ステップS22）。ブロック数NUMは、現時点で発見済みのコンテンツブロックの個数である。また、階層LVは、DOMツリーにおいて現在探索中のノードが属するコンテンツブロックの階層である。NUM及びLVは何れもグローバル変数であり、1ページ対応抽出処理及び後述するツリー探索処理からアクセスが可能である。
- [0086] 次いで、システム制御部20は、DOMツリーのルートノードを指定して（ステップS23）、ツリー探索処理を実行する（ステップS24）。ツリー探索処理は再帰呼び出しが可能であり、このツリー探索処理により、Webページから全てのコンテンツブロックが抽出され、コンテンツブロック対応情報が生成される。
- [0087] 次いで、システム制御部20は、ツリー探索処理により生成された各コンテンツブロック対応情報を素材抽出DB101に登録する（ステップS25）。システム制御部20は、この処理を終えると、1ページ対応抽出処理を終了させる。
- [0088] 図9は、本実施形態に係るコンテンツ生成サーバ1のシステム制御部20のツリー探索処理における処理例を示すフローチャートである。
- [0089] 図9に示すように、システム制御部20は、先ず、指定されたノードの種

類を判定する（ステップS31）。このとき、システム制御部20は、指定されたノードの種類が、DIVノード又はTABLEノード（ブロック化ノード）である場合には、すなわち、コンテンツブロックが発見された場合には（ステップS31：DIV又はTABLE）、ステップS32に移行する。

[0090] ステップS32において、システム制御部20は、ブロック数NUMに1を加算するとともに、階層LVに1を加算する。次いで、システム制御部20は、ブロック番号BN[LV]にNUMを設定する（ステップS33）。ブロック番号BN[LV]は、現在探索中のノードが属する階層LVで示されるコンテンツブロックのブロック番号である。このブロック番号は、コンテンツブロックの発見順に付与される。また、BN[LV]は、グローバル変数である。

[0091] 次いで、システム制御部20は、ブロック番号BN[LV]のコンテンツブロックに対応するコンテンツブロック対応情報を初期化する（ステップS34）。具体的に、システム制御部20は、コンテンツブロック対応情報を格納する領域をRAM19上に設定し、取得したHTML文書のURLを、当該領域に設定する。

[0092] 次いで、システム制御部20は、指定されたノードの子ノードのうち、未だ探索されていない子ノードが存在するか否かを判定する（ステップS35）。このとき、システム制御部20は、未だ探索されていない子ノードが存在する場合には（ステップS35：YES）、ステップS36に移行する。

[0093] ステップS36において、システム制御部20は、探索されていない子ノードのうちの1つの子ノードを指定して、ツリー探索処理を実行する（ステップS37）。システム制御部20は、ツリー探索処理を終えると、ステップS35に移行する。

[0094] そして、システム制御部20は、ステップS35～S37の処理を繰り返して全ての子ノードのツリー探索処理を終えると（ステップS35：NO）、ステップS38に移行する。なお、システム制御部20は、指定されたノ

ードの子ノードが1つも存在しない場合にも、ステップS 3 8に移行する。ステップS 3 8において、システム制御部2 0は、階層LVから1を減算して、ツリー探索処理を終了させる。

[0095] ステップS 3 1において、システム制御部2 0は、指定されたノードの種類がテキストノードである場合には（ステップS 3 1：テキスト）、指定されたノードの内容（テキストデータ）を、ブロック番号BN [LV] のコンテンツブロックに対応するコンテンツブロック対応情報中のブロック構成情報に追加設定する（ステップS 3 9）。システム制御部2 0は、この処理を終えたと、ツリー探索処理を終了させる。

[0096] ステップS 3 1において、システム制御部2 0は、指定されたノードの種類がIMGノードである場合には（ステップS 3 1：IMG）、指定されたノードのsrc属性として設定されている画像データのURLを取得し、取得したURLを、ブロック番号BN [LV] のコンテンツブロックに対応するコンテンツブロック対応情報中のブロック構成情報に追加設定する（ステップS 4 0）。システム制御部2 0は、この処理を終えたと、ツリー探索処理を終了させる。

[0097] ステップS 3 1において、システム制御部2 0は、指定されたノードの種類が、DIVノード、TABLEノード、テキストノード、及びIMGノードの何れでもない場合には（ステップS 3 1：その他）、指定されたノードの子ノードのうち、未だ探索されていない子ノードが存在するか否かを判定する（ステップS 4 1）。このとき、システム制御部2 0は、未だ探索されていない子ノードが存在する場合には（ステップS 4 1：YES）、探索されていない子ノードのうちの1つの子ノードを指定して（ステップS 4 2）、ツリー探索処理を実行する（ステップS 4 3）。システム制御部2 0は、ツリー探索処理を終えたと、ステップS 4 1に移行する。

[0098] 一方、システム制御部2 0は、指定されたノードの全ての子ノードのツリー探索処理を終えた場合、又は、指定されたノードの子ノードが1つも存在しない場合には（ステップS 4 1：NO）、ツリー探索処理を終了させる。

[0099] [1-3-2 Flashコンテンツ生成時]

図10は、本実施形態に係るコンテンツ生成サーバ1のシステム制御部20のコンテンツ生成処理における処理例を示すフローチャートである。

[0100] コンテンツ生成処理は、システム管理者の操作に基づいて管理端末3からコンテンツ生成処理の実行のリクエストが送信されてきたとき、又は、店舗従業員の操作に基づいて店舗端末4からコンテンツ生成処理の実行のリクエストが送信されてきたときに開始される。

[0101] そして、システム管理者又は店舗従業員がFlashコンテンツの生成対象とするHTML文書のURLを指定すると、図10に示すように、システム制御部20は、指定されたURLを管理端末3又は店舗端末4から受信する（ステップS51）。次いで、システム制御部20は、受信したURLを設定したリクエストをショッピングサーバ2に送信することにより、当該ショッピングサーバ2からHTML文書を取得する（ステップS52）。

[0102] 次いで、システム制御部20は、取得したHTML文書を指定して、後述する特有コンテンツブロック判定処理を実行する（ステップS53）。この特有コンテンツブロック判定処理では、取得したHTML文書からコンテンツブロックが抽出され、当該HTML文書特有のコンテンツブロックが判定される。

[0103] 次いで、システム制御部20は、特有と判定されたコンテンツブロックを構成する各コンテンツを取得する（ステップS54）。このとき、システム制御部20は、テキストデータを取得する場合には、特有と判定されたコンテンツブロックに対応するコンテンツブロック対応情報からテキストデータを取得する。一方、システム制御部20は、画像データを取得する場合には、特有と判定されたコンテンツブロックに対応するコンテンツブロック対応情報から画像データのURLを取得し、取得されたURLを設定したリクエストをショッピングサーバ2に送信することにより、商品詳細ページDB201に登録されている画像データをショッピングサーバ2から取得する。

[0104] 次いで、システム制御部20は、取得した全てのコンテンツを指定して、

後述するFlashコンテンツ生成処理を実行する（ステップS55）。

[0105] 次いで、システム制御部20は、Flashコンテンツ生成処理において生成されたFlashコンテンツを、生成要求元の管理端末3又は店舗端末4に送信する（ステップS56）。システム制御部20は、この処理を終えると、コンテンツ生成処理を終了させる。

[0106] 図11は、本実施形態に係るコンテンツ生成サーバ1のシステム制御部20の特有コンテンツブロック判定処理における処理例を示すフローチャートである。

[0107] 図11に示すように、システム制御部20は、先ず、1ページ対応抽出処理と同様に、取得したHTML文書のDOMツリー生成（ステップS61）、ブロック数NUM及び階層LVに対して0の設定を行い（ステップS62）、DOMツリーのルートノードを指定して（ステップS63）、ツリー探索処理を実行する（ステップS64）。

[0108] 次いで、システム制御部20は、ブロック番号iに1を設定する（ステップS65）。次いで、システム制御部20は、ブロック番号iのコンテンツブロックの出現頻度を計算する（ステップS66）。

[0109] 具体的に、システム制御部20は、ツリー探索処理において生成されたコンテンツブロック対応情報i（ブロック番号iのコンテンツブロックに対応するコンテンツブロック対応情報）のブロック構成情報と、素材抽出DB101に登録されている各コンテンツブロック対応情報のブロック構成情報とを比較する。このとき、システム制御部20は、ブロック構成情報の内容が一致する場合には、出現回数1回としてカウントする。このとき、システム制御部20は、ブロック構成情報中におけるコンテンツの規定順は無視してかまわない。また、システム制御部20は、素材抽出DB101に登録されているコンテンツブロック対応情報のブロック構成情報に規定されている一部のコンテンツがコンテンツブロック対応情報iのブロック構成情報に規定されている全部のコンテンツに一致する場合も、出現回数1回としてカウントしても良い。更に、システム制御部20は、コンテンツブロック対応情報

のブロック構成情報中に規定されているテキストデータ同士を比較する場合には、テキストデータが示す文章等そのものが一致するか否かを判定するのではなく、その文章等により表現されている実質的な内容を比較しても良い。例えば、システム制御部20は、夫々のテキストデータの形態素解析等を行うことによりテキストデータから単語を抽出し、抽出した単語同士を比較しても良い。そして、システム制御部20は、全ての単語が一致した場合にテキストデータ同士が一致したと判断しても良いし、所定の割合以上で単語が一致した場合にテキストデータ同士が一致したと判断しても良い。システム制御部20は、このようにしてコンテンツブロック対応情報iのブロック構成情報と、素材抽出DB101に登録されている全てのコンテンツブロック対応情報のブロック構成情報とを比較して、出現頻度を計算する。

[0110] システム制御部20は、出現頻度を計算すると、ブロック番号iに1を加算して（ステップS67）、ブロック番号iがブロック数NUMの値より大きいか否かを判定する（ステップS68）。このとき、システム制御部20は、ブロック番号iがブロック数NUMの値以下である場合には（ステップS68:NO）、ステップS66に移行する。そして、システム制御部20は、ツリー探索処理において抽出された全てのコンテンツブロックの出現頻度を計算すると（ステップS68:YES）、ステップS69に移行する。

[0111] ステップS69において、システム制御部20は、コンテンツブロック1からブロック数NUMが示すブロック番号までの全てのコンテンツブロックの出現頻度同士を比較して、出現頻度が最も少ないコンテンツブロックを、特有のコンテンツブロックであると判定する（ステップS69）。システム制御部20は、この処理を終えると、特有コンテンツブロック判定処理を終了させる。

[0112] 図12は、本実施形態に係るコンテンツ生成サーバ1のシステム制御部20のFlashコンテンツ生成処理における処理例を示すフローチャートである。なお、本実施形態においては、商品詳細ページに掲載されている商品を紹介するコンテンツとして、特有のコンテンツブロックであると判定されたコン

テンツブロックに含まれる各コンテンツをスライドショー表示するFlashコンテンツを生成するものとして以下説明する。

- [0113] 図12に示すように、システム制御部20は、先ず、指定された各コンテンツの表示サイズを調整する（ステップS71）。例えば、システム制御部20は、Flashコンテンツ再生時の実際の表示サイズに合うように、画像データの縦横の画素数を調整したり、テキストデータのフォントサイズ等を調整する。また、システム制御部20は、コンテンツの表示サイズがFlashコンテンツ再生時の実際の表示サイズに比べて大きすぎる場合には、当該コンテンツを複数に分割する。また、システム制御部20は、コンテンツの表示サイズがFlashコンテンツ再生時の実際の表示サイズに比べて小さすぎる場合には、複数のコンテンツを1つに結合する。
- [0114] 次いで、システム制御部20は、各コンテンツの表示順序を決定する（ステップS72）。各コンテンツの表示順序は、基本的にはツリー探索処理におけるコンテンツブロック対応情報に対するコンテンツの設定順と同じ順序である。つまり、HTML文書において当該文書の先頭の近くに規定されているコンテンツほど、表示順序が早くなる。
- [0115] 次いで、システム制御部20は、各コンテンツの遷移方法を決定する（ステップS73）。つまり、システム制御部20は、スライドショー表示において、表示するコンテンツを切り換える際に施されるエフェクト（表示効果）を決定する。エフェクトとしては、例えば、フェードイン／フェードアウト、スライド、ランダムブロック、ワイプ、エフェクトなし等がある。
- [0116] 次いで、システム制御部20は、ステップS71において調整が行われた各コンテンツに基づき、ステップS72及びS73において決定された条件に基づいてFlashコンテンツを生成する（ステップS74）。システム制御部20は、この処理を終えると、Flashコンテンツ生成処理を終了させる。
- [0117] 以上説明したように、本実施形態によれば、コンテンツ生成サーバ1のシステム制御部20が、指定されたURLに対応する商品詳細ページを構成しているコンテンツを抽出し、指定されたURLに対応する商品詳細ページを

構成している各コンテンツの出現頻度を計算し、指定されたURLに対応する商品詳細ページを構成するコンテンツのうち、出現頻度が最も小さいコンテンツを当該商品詳細ページに特有のコンテンツであると判断する。

[0118] 従って、出現頻度が小さいコンテンツであるほど、指定された商品詳細ページ以外にはあまり出現しないコンテンツであるので、最も出現頻度が小さいコンテンツを判断することで、指定された商品詳細ページに特有のコンテンツを特定される。よって、商品詳細ページに特有のコンテンツを容易に抽出することができる。

[0119] また、コンテンツ生成サーバ1のシステム制御部20が、商品詳細ページに特有のコンテンツであると判断されたコンテンツに基づいてFlashコンテンツを生成する。

[0120] 従って、当該商品詳細ページに掲載されている商品の内容等を端的に示し又は特徴を示すコンテンツを生成することができる。

[0121] また、コンテンツ生成サーバ1のシステム制御部20が、ショッピングサイトに含まれる複数の商品詳細ページ上における各コンテンツの出現頻度を計算する。

[0122] 従って、ショッピングサイトに含まれる複数のWebページ上において、指定された商品詳細ページを構成している各コンテンツの出現頻度が計算されるので、ショッピングサイト内で共通して用いられるコンテンツは、特有のコンテンツではないと判定することが可能となり、判断精度を上げることができる。

[0123] また、コンテンツ生成サーバ1のシステム制御部20が、ショッピングサイトを構成する全ての商品詳細ページについて、商品詳細ページを構成しているコンテンツを抽出し、抽出したコンテンツを示すコンテンツブロック対応情報を予め素材抽出DB101に登録しておき、素材抽出DB101に登録された各コンテンツブロック対応情報に基づいて、指定されたURLに対応する商品詳細ページを構成している各コンテンツの出現頻度を計算する。

[0124] 従って、全ての商品詳細ページに基づいて出現頻度計算されるので、特有

のコンテンツを判断する精度を上げることができる。

[0125] また、コンテンツ生成サーバ1のシステム制御部20が、1つ以上のコンテンツで構成されるコンテンツブロックの単位で、商品詳細ページを構成しているコンテンツを抽出し、指定されたURLに対応する商品詳細ページを構成している各コンテンツブロック出現頻度を計算し、指定されたURLに対応する商品詳細ページを構成するコンテンツブロックのうち、出現頻度が最も小さいコンテンツブロックを当該商品詳細ページに特有のコンテンツブロックであると判断する。

[0126] 従って、商品詳細ページ上において、例えば、ヘッダ部分、ナビゲーション部分、商品の詳細を示す部分、商品購入の際の一般的な注意事項等を示す部分、コピーライト表示の部分等のように、1つ以上のコンテンツがまとまりをもってコンテンツブロックとして表示されている場合に、商品詳細ページに特有のコンテンツブロックを抽出することができる。

[0127] また、コンテンツ生成サーバ1のシステム制御部20が、商品詳細ページを構成しているコンテンツを当該商品詳細ページのHTML文書に基づいて抽出し、HTML文書においてDIVタグ又はTABLEタグに基づいてコンテンツブロックを定める。

[0128] 従って、DIVタグにより、HTML文書の作成の際に明示的にブロック化された1つ以上のコンテンツを特定することができ、また、TABLEタグにより、表形式でブロック化されて表示される1つ以上のコンテンツを特定することができるので、例えば、これらのタグにより、商品詳細ページに特有のコンテンツと、特有ではないコンテンツとがブロック化されている場合に、Webページに特有のコンテンツを判断する精度を上げることができる。

[0129] なお、上記第1実施形態においては、指定された商品詳細ページを構成する各コンテンツブロックに対応するコンテンツブロック対応情報を、素材抽出DB101に登録されている全てのコンテンツブロック対応情報と比較することによって各出現頻度が計算されていた。つまり、指定された商品詳細

ページを構成する各コンテンツブロックの出現頻度を計算する場合に、ショッピングサイトに含まれる全ての商品詳細ページを対象とした範囲に出現する頻度を計算するようになっていたが、全ての商品詳細ページを対象としなくても良い。例えば、対象となる店舗の指定を可能とし、指定された店舗に対応する全ての商品詳細ページを対象とした範囲に出現する頻度を計算するようにしても良い。また、例えば、予め定められたページ数分の商品詳細ページを対象としても良い。

[0130] また、予め各商品詳細ページからコンテンツブロックの抽出しておかなくても良い。例えば、Flashコンテンツ生成時に、出現頻度を計算するのに必要な各商品詳細ページについてコンテンツブロックの抽出を行えば良い。

[0131] また、上記第1実施形態においては、システム管理者又は店舗従業員から商品詳細ページのHTML文書のURLを指定されたときに、Flashコンテンツを生成するようになっていたが、例えば、新規の商品詳細ページが作成されたときや、商品詳細ページが更新されたときに、新規作成され又は更新された商品詳細ページについてのFlashコンテンツを生成しても良い。

[0132] また、上記第1実施形態においては、出現頻度が最も小さいコンテンツブロックのみをWebページに特有のコンテンツとしていたが、例えば、出現頻度が1番目に小さいコンテンツブロックからN番目（Nは2以上の自然数）に小さいコンテンツまでのN個のコンテンツブロックをWebページに特有のコンテンツブロックとしても良い。これは、例えば、所望の処理に必要なコンテンツブロックの個数が2個以上で予め定められている場合に適用することができる。また、例えば、所望の処理に必要なコンテンツ（コンテンツブロックではない）の個数が2個以上で予め定められている場合に、出現頻度が1番目に小さいコンテンツブロックに含まれるコンテンツのみでは足りないときには、出現頻度が2番目に小さいコンテンツブロックをWebページに特有のコンテンツであると追加認定し、出現頻度が1番目に小さいコンテンツブロックと2番目に小さいコンテンツブロックに含まれるコンテンツのみで足りないときには、出現頻度が3番目に小さいコンテンツブロック

をWebページに特有のコンテンツであると追加認定する、といった処理を行っても良い。

[0133] また、上記第1実施形態においては、Webページに特有のコンテンツを用いてFlashコンテンツを生成していたが、Flashコンテンツ以外のコンテンツ（例えば、動画データ、静止画像データ、電子文書等）を生成しても良い。

[0134] また、Webページに特有のコンテンツの用途としては新たなコンテンツの生成のみに限られるものではない。例えば、Webページに特有の画像データを判定し、特有の画像データと判定された画像データを、当該Webページを代表する画像データとして、検索結果等に表示しても良い。

[0135] [2. 第2実施形態]

以下、図面を参照して本発明の実施形態について詳細に説明する。なお、以下に説明する実施の形態は、ブログサービスを提供するブログシステムにおいて、ブログページ送信するサーバ装置に対して本発明を適用した場合の実施形態である。

[0136] [2-1. ブログシステムの構成及び機能概要]

まず、本実施形態に係るブログシステムBSの構成及び概要機能について、図13を用いて説明する。

[0137] 図13は、本実施形態に係るブログシステムBSの概要構成の一例を示す図である。図13において、図1と同様の要素については同様の符号を付してある。

[0138] 図13に示すように、ブログシステムBSは、特有コンテンツ判定装置及び関連コンテンツ挿入装置の一例としてのブログサーバ6と、管理端末3と、複数のユーザ端末5と、を含んで構成されている。そして、ブログサーバ6と、各ユーザ端末5とは、ネットワークNWを介して、例えば、通信プロトコルにTCP/IP等を用いて相互にデータの送受信が可能になっている。また、ブログサーバ6と管理端末3とは、LAN等のネットワークを介して接続されている。

[0139] このような構成のブログシステムBSにおいて、ブログサーバ6は、ユーザ端末5からのリクエストに応じて、ブログサービスサイトを構成するWebページを送信するWebサーバである。ユーザ端末5を利用するユーザがブログサービスサイトのユーザ登録をすると、そのユーザは、当該ブログサービスサイトにおいてユーザ自身のブログを運営することができるようになっている。そして、ユーザ登録されたユーザ（ブロガー）は、ブログサービスサイトにアクセスして、自己のブログを更新（ブログ記事（ブログ1件毎の記録）を追加）することができるようになっている。そのため、ブログサーバ6は、ブログの更新に応じて、ブログのWebページとして、1又は複数のブログ記事が掲載されるブログページを生成又は更新する。そして、ブログサーバ6は、ブログページDB601を備え、ブログページを当該ブログページDB601に登録する。

[0140] また、ブログサーバ6は、システム管理者により指定されたブロガーのブログページに対して、広告コンテンツ（関連コンテンツの一例）を挿入する。広告コンテンツとしては、例えば、広告文章のテキストデータ、バナーの画像データ、動画データ、Adobe Flash（商標）やSilverlight（商標）等により生成されたリッチインターネットアプリケーション（RIA）等がある。各ブログページに挿入される広告コンテンツは、対象のブログページに掲載されているブログ記事に関連する商品やサービスに関する広告を示すコンテンツである。そのため、ブログサーバ6は、複数の広告コンテンツが登録されている広告DB602を備える。そして、ブログサーバ6は、ブログページからブログ記事を抽出し、更にブログ記事から特徴語を抽出して、抽出した特徴語に関連する広告コンテンツを選択する。

[0141] ユーザ端末5は、ブロガーとしてのユーザや、ブログを閲覧するユーザにより利用される端末装置である。ユーザ端末5としては、例えば、パーソナルコンピュータ、PDA、携帯電話機等が用いられる。

[0142] 管理端末3は、ブログシステムBSのシステム管理者により使用される端末装置である。管理端末3としては、例えば、パーソナルコンピュータ等が

用いられる。

[0143] [2-2. ブログサーバの構成及び機能]

次に、ブログサーバ6の構成及び機能について、図14を用いて説明する。

[0144] 図14は、本実施形態に係るブログサーバ6の概要構成の一例を示すブロック図である。また、図15は、ブロガーが指定されてからブログページに広告コンテンツが挿入されるまでの処理の概要を示す図である。また、図16は、Webページの構成例を示す図である。また、図17は、HTML文書から生成されたDOMツリーの一例を示す図である。また、図18は、記憶部65に記憶されたコンテンツブロック対応情報の内容の一例を示す図である。

[0145] 図14に示すように、ブログサーバ6は、操作部61と、表示部62と、通信部63と、ドライブ部64と、記憶手段の一例としての記憶部65と、入出インタフェース部66と、システム制御部70と、を備えている。そして、システム制御部70と入出インタフェース部66とは、システムバス71を介して接続されている。

[0146] 操作部61は、例えば、キーボード、マウス等により構成されており、システム管理者等からの操作指示を受け付け、その指示内容を指示信号としてシステム制御部70に出力するようになっている。表示部62は、例えば、CRTディスプレイ、液晶ディスプレイ等により構成されており、文字や画像等の情報を表示するようになっている。通信部63は、ネットワークNW等に接続して、管理端末3、ユーザ端末5等との通信状態を制御するようになっている。ドライブ部64は、例えば、フレキシブルディスク、CD、DVD等のディスクDKからデータ等を読み出す一方、当該ディスクDKに対してデータ等を記録するようになっている。

[0147] 記憶部65は、例えば、ハードディスクドライブ等により構成されており、各種プログラム及びデータ等を記憶するようになっている。また、記憶部65には、ブログページDB601及び広告DB602が構築されている。

ブログページDB601には、ブログサービスサイトを構成する各ブログページ（ブログページのHTML文書（ドキュメントデータの一例）、ブログページの素材である画像データ等）が、例えば、そのページのURL及びブロガーの識別情報であるユーザIDに対応付けて登録されている。広告DB602には、複数の広告コンテンツが、その広告コンテンツによる広告対象の商品やサービスに関連するキーワードに対応付けて登録されている。ここで、広告コンテンツにテキストデータ以外のコンテンツが含まれている場合には、当該コンテンツのURLも対応付けて登録されている。また、ブログページ上に表示された広告コンテンツをユーザが選択したときに広告対象の商品やサービスに関するWebページに移動するため、当該WebページのURLも対応付けて登録されている。

[0148] 入出カインタフェース部66は、操作部61～記憶部65とシステム制御部70との間のインタフェース処理を行うようになっている。システム制御部70は、CPU67、ROM68、RAM69等により構成されている。

[0149] システム制御部70は、CPU67が、ROM68や記憶部65に記憶された各種プログラムを読み出し実行することによりブログサーバ6の各部を制御する。また、システム制御部70は、広告コンテンツ挿入ソフトウェア（特有コンテンツ判定プログラムの一例）を実行することにより、抽出手段、計算手段、判定手段及び挿入手段として機能する。なお、広告コンテンツ挿入ソフトウェア等は、例えば、他のサーバ装置等からネットワークNWを介して取得されるようにしても良いし、CD-ROM等のディスクDKに記録されてドライブ部64を介して読み込まれるようにしても良い。

[0150] 広告コンテンツ挿入ソフトウェアは、ブログページに広告コンテンツを挿入するためのプログラムである。図15に示すように、広告コンテンツ挿入ソフトウェアは、マネージャ部、素材抽出エンジン、文章解析エンジン、広告選択部等により構成されている。マネージャ部は、素材抽出エンジン、文章解析エンジン及び広告選択部の実行を制御する。素材抽出エンジンは、ブログページのHTML文書からWeb素材としてのコンテンツを抽出すると

ともに、ブログページ特有のコンテンツを判定するためのソフトウェアである。コンテンツの抽出は、コンテンツブロック（コンテンツグループの一例）の単位で行われる。本実施形態においては、例えば、その記事特有の内容を含むブログ記事が、ブログページに特有のコンテンツブロックに相当する。

[0151] 文章解析エンジンは、ブログページ特有のコンテンツとして抽出されたブログ記事から当該ブログページの特徴語を抽出するためのソフトウェアである。広告選択部は、抽出された特徴語をキーワードとして、ブログページに関連する広告コンテンツを選択するためのソフトウェアである。

[0152] 以下に、広告コンテンツの挿入の概要について説明する。図15に示すように、システム管理者により対象のブロガーのユーザIDが指定される（1）。すると、システム制御部70は、指定されたユーザIDに対応する全ブログページのHTML文書をブログページDB601から取得して解析し、Web素材たるコンテンツをコンテンツブロック単位で抽出する。そして、その抽出結果として、抽出したコンテンツブロック毎にコンテンツブロック対応情報（コンテンツ情報の一例）を生成する（2）。次いで、システム制御部70は、抽出した各コンテンツブロックの、指定されたユーザIDに対応する全ブログページにおける出現頻度を計算する。本実施形態において計算される出現頻度は、例えば、出現回数（度数）である。そして、システム制御部70は、出現頻度に基づいて、各ブログページにおいて特有であるコンテンツブロックを判定する。具体的に、システム制御部70は、各ブログページにおいて、出現頻度が所定の閾値以下であるコンテンツブロックを、そのブログページ特有のコンテンツブロックであると判定する（3）。

[0153] システム制御部70は、特有のコンテンツブロックであると判定したコンテンツブロック、すなわち、ブログ記事に対して形態素解析等の分析を行って、ブロックページ毎の特徴語を抽出する（4）。特徴語の抽出方法としては、種々の方法があり、また公知であるので、詳細な説明は省略する。一例としては、出現頻度の最も高い単語を特徴語とする。

- [0154] 次いで、システム制御部70は、広告DB602を参照し、抽出した特徴語に関連する広告コンテンツを選択する(5)。そして、システム制御部70は、選択した広告コンテンツをブログページに挿入して表示させるための規定(タグやデータそのものの記述等)を、ブログページのHTML文書に挿入する(6)。
- [0155] 次に、コンテンツブロックの抽出方法について説明する。なお、本実施形態においては、テキストデータ及び画像データがWeb素材として抽出されるものとする。
- [0156] 例えば、ブログページの構成(レイアウト)が図16に示すようなものであるとする。Web素材としての各コンテンツは、ブログページ上において、それぞれ或るまとまり(かたまり)毎に表示されている。その各まとまりがコンテンツブロックに相当する。各コンテンツは、HTML文書に記述されているDIVタグ及びTABLEタグ(予め定められたタグの一例)により夫々コンテンツブロックに分けられる。つまり、各コンテンツは、DIVタグ及びTABLEタグによりブロック化(グループ化)される。
- [0157] 図16には、コンテンツブロック701~706が表示されている。コンテンツブロック701は、例えば、ページのヘッダ部分のコンテンツブロックであり、テキストA及び画像aにより構成されている。また、コンテンツブロック702は、例えば、他のWebページに移動するためのナビゲーション部分のコンテンツブロックであり、例えば、他のWebページへのリンクを示すテキストB、テキストC及びテキストDにより構成されている。また、コンテンツブロック703は、例えば、ブログの表示領域に対応するコンテンツブロックであり、ブログ等の見出しを示すテキストE、コンテンツブロック704及びコンテンツブロック705により構成されている。このように、コンテンツブロックが入れ子状、つまり、階層構造になっている場合もある。この場合、コンテンツブロック703に含まれるコンテンツはテキストEのみとされ、コンテンツブロック704及びコンテンツブロック705は、コンテンツブロック703から独立しているものとされる。コン

テンツブロック704及び705は、夫々1件のブログ記事である。コンテンツブロック704には、ブログ記事の表題や本文を示すテキストF及びGにより構成されている。コンテンツブロック705には、ブログ記事の表題や本文を示すテキストH、I及びJと、ブログ記事に関連してブロガーにより登録された画像b及びcとにより構成されている。コンテンツブロック706は、例えば、コピーライト表示を示すコンテンツブロックであり、テキストIにより構成されている。

[0158] これらのコンテンツブロックのうち、コンテンツブロック701、702、703及び706は、図16に示すブログページ以外のブログページ上でも比較的頻繁に現れる。一方、コンテンツブロック704及びコンテンツブロック705は、基本的に当該ブログページにだけに用いられる。よって、コンテンツブロック704又はコンテンツブロック705が、当該ブログページ特有のコンテンツブロックであると判断されることとなる。

[0159] 本実施形態においては、その記事特有の内容を含むブログ記事に相当するコンテンツブロックを特有のコンテンツブロックと判定されるようにする必要がある。こうした特有の内容を含むブログ記事は、1ページ内に複数含まれている場合がある。そのため、出現頻度が所定の閾値以下であるコンテンツブロックは、全て特有のコンテンツブロックとされる。例えば、閾値の値を1回に設定する。そうすると、特有の内容を含むブログ記事は、特有のコンテンツブロックと判定され、他のブログ記事と同じような内容のみを含むブログ記事は、特有のコンテンツブロックとは判定されない。また、ヘッダ部分、ナビゲーション部分、コピーライト表示部分等の各ブログページに共通するようなコンテンツブロックの出現頻度は夫々2回以上となるので、これらも特有のコンテンツブロックとは判定されない。なお、閾値は記憶部65に予め記憶される。

[0160] 図16に示すブログページのHTML文書をDOMツリー、すなわち、木構造で表したものが図17である。なお、図17に示すDOMツリーにおいて、本実施形態の説明に必要なないタグのノードの図示は省略している。

[0161] 第1実施形態の場合と同様にコンテンツブロックを抽出すると、システム制御部70は、抽出結果を示すコンテンツブロック対応情報を一時的に記憶部65に保存する。図18に示すように、コンテンツブロック対応情報（符号401）は、コンテンツブロック毎に保存される。なお、本実施形態においては、ブログページ特有のコンテンツブロックであると判定されたコンテンツブロック、すなわち、ブログ記事から特徴語を抽出するので、テキストデータが抽出されれば良く、画像データについては抽出しなくても良い。

[0162] [2-3. ブログシステムの動作]

次に、ブログシステムBSの動作について、図19乃至図21を用いて説明する。

[0163] 図19は、本実施形態に係るブログサーバ6のシステム制御部70の広告コンテンツ挿入処理における処理例を示すフローチャートである。

[0164] 広告コンテンツ挿入処理は、例えば、システム管理者の操作に基づいて管理端末3から広告コンテンツ挿入処理の実行のリクエストが送信されてきたときに開始される。

[0165] そして、システム管理者が広告コンテンツの挿入対象とするブログを運営するブロガーのユーザIDを指定すると、図19に示すように、システム制御部70は、指定されたユーザIDを管理端末3から受信する（ステップS101）。

[0166] 次いで、システム制御部70は、ブロック数NUMに0を設定する（ステップS102）。ブロック数NUMは、現時点で発見済みのコンテンツブロックの個数である。また、NUMはグローバル変数であり、後述する1ページ対応抽出処理及びびツリー探索処理からアクセスが可能である。

[0167] 次いで、システム制御部70は、受信したユーザIDに対応する最初のブログページのHTML文書をブログページDB601から取得する（ステップS103）。次いで、システム制御部70は、取得したHTML文書を指定して、後述する1ページ対応抽出処理を実行する（ステップS104）。この1ページ対応抽出処理では、取得したHTML文書からコンテンツプロ

ックが抽出され、コンテンツブロック対応情報が保存される。

[0168] 次いで、システム制御部70は、受信したユーザIDに対応する全てのブログページのコンテンツブロックを抽出したか否かを判定する（ステップS105）。このとき、システム制御部70は、コンテンツブロックを抽出していないブログページが存在する場合には（ステップS105：NO）、次のブログページのHTML文書をブログページDB601から取得して（ステップS106）、ステップS104に移行する。そして、システム制御部70は、ステップS104～S106の処理を繰り返して全てのブログページのコンテンツブロックを抽出すると（ステップS105：YES）、ステップS107に移行する。

[0169] ステップS107において、システム制御部70は、受信したユーザIDに対応する最初のブログページのHTML文書を特定する。

[0170] 次いで、システム制御部70は、取得したHTML文書を指定して、後述する特有コンテンツブロック判定処理を実行する（ステップS108）。この特有コンテンツブロック判定処理では、特定したHTML文書からコンテンツブロックが抽出され、ブログページ特有のコンテンツブロックが判定される。

[0171] 次いで、システム制御部70は、特有と判定されたコンテンツブロックを構成する各テキストデータからブログページの特徴語を抽出する（ステップS109）。次いで、システム制御部70は、抽出した特徴語に基づいて、ブログページに関連する広告ページを当該ブログページに挿入する（ステップS110）。具体的に、システム制御部70は、抽出した特徴語をキーワードとし、広告DB602を参照して当該キーワードに対応する広告コンテンツを選択する。次いで、システム制御部70は、特定したHTML文書上の所定の位置に、選択した広告コンテンツの規定を挿入する。例えば、システム制御部70は、広告コンテンツにテキストデータが含まれている場合には、当該テキストデータの内容をHTML文書に追加する。また、例えば、システム制御部70は、広告コンテンツに画像データが含まれている場合に

は、当該画像データを表示するためのIMGタグをHTML文書に追加する。また、例えば、システム制御部70は、広告対象の商品やサービスに関するWebページへのリンク情報をHTML文書に追加する。

[0172] システム制御部70は、特定したHTML文書に広告コンテンツの規定を挿入すると、当該HTML文書で、ブログページDB601に登録されているHTML文書を更新する(ステップS111)。

[0173] 次いで、システム制御部70は、受信したユーザIDに対応する全てのブログページに広告コンテンツを挿入したか否かを判定する(ステップS112)。このとき、システム制御部70は、広告コンテンツを挿入していないブログページが存在する場合には(ステップS112:NO)、次のブログページのHTML文書を特定して(ステップS113)、ステップS108に移行する。そして、システム制御部70は、ステップS108~S113の処理を繰り返して全てのブログページに広告コンテンツを挿入すると(ステップS112:YES)、記憶部65に保存させておいた全てのコンテンツブロック対応情報を、記憶部65から削除する(ステップS114)。システム制御部70は、この処理を終えると、広告コンテンツ挿入処理を終了させる。

[0174] 図20は、本実施形態に係るブログサーバ6のシステム制御部70の1ページ対応抽出処理における処理例を示すフローチャートである。

[0175] 図20に示すように、システム制御部70は、先ず、取得したHTML文書のDOMツリーをRAM69上に生成する(ステップS121)。

[0176] 次いで、システム制御部70は、階層LVに0を設定する(ステップS122)。階層LVは、DOMツリーにおいて現在探索中のノードが属するコンテンツブロックの階層である。LVはグローバル変数であり、1ページ対応抽出処理及びツリー探索処理からアクセスが可能である。

[0177] 次いで、システム制御部70は、DOMツリーのルートノードを指定して(ステップS123)、ツリー探索処理を実行する(ステップS124)。ツリー探索処理の処理内容は、第1実施形態の場合と同様であるので、詳細

な説明は省略する。

- [0178] 次いで、システム制御部70は、ツリー探索処理により生成された各コンテンツブロック対応情報を記憶部65に保存する（ステップS125）。システム制御部70は、この処理を終えると、1ページ対応抽出処理を終了させる。
- [0179] 図21は、本実施形態に係るコンテンツ生成サーバ1のシステム制御部70の特有コンテンツブロック判定処理における処理例を示すフローチャートである。
- [0180] 図21に示すように、システム制御部70は、まず、1ページ対応抽出処理と同様に、指定されたHTML文書のDOMツリー生成（ステップS161）、ブロック数NUM及び階層LVに対して0の設定を行い（ステップS162）、DOMツリーのルートノードを指定して（ステップS163）、ツリー探索処理を実行する（ステップS164）。
- [0181] 次いで、システム制御部70は、ブロック番号iに1を設定する（ステップS165）。次いで、システム制御部70は、ブロック番号iのコンテンツブロックの出現頻度を計算する（ステップS166）。
- [0182] 具体的に、システム制御部70は、ステップS164のツリー探索処理において生成されたコンテンツブロック対応情報iのブロック構成情報と、記憶部65に保存されている各コンテンツブロック対応情報のブロック構成情報とを比較して、出現頻度を計算する。出現頻度の計算方法は、第1実施形態の場合と同様である。
- [0183] システム制御部70は、出現頻度を計算すると、計算した出現頻度が、記憶部65に記憶されている閾値以下であるか否かを判定する（ステップS167）。このとき、システム制御部70は、出現頻度が閾値以下である場合には（ステップS167：YES）、ブロック番号iのコンテンツブロックを、特有のコンテンツブロックの1つであると判定する（ステップS168）。つまり、システム制御部70は、ブロック番号iのコンテンツブロックを、指定されたHTML文書が対応するブログページに特有のコンテンツ

ロックに加える。

[0184] システム制御部70は、出現頻度が閾値よりも大きい場合（ステップS167：NO）、又は、ステップS168の処理を終えた場合には、ブロック番号iに1を加算して（ステップS169）、ブロック番号iがブロック数NUMの値より大きいか否かを判定する（ステップS170）。このとき、システム制御部70は、ブロック番号iがブロック数NUMの値以下である場合には（ステップS170：NO）、ステップS166に移行する。そして、システム制御部70は、ツリー探索処理において抽出された全てのコンテンツブロックの出現頻度を計算すると（ステップS170：YES）、特有コンテンツブロック判定処理を終了させる。

[0185] なお、システム制御部70は、ステップS164のツリー探索処理によりコンテンツブロックを抽出していたが、広告コンテンツ挿入処理から実行された1ページ対応抽出処理（図19ステップS104）において、受信したブロガーのユーザIDに対応する全てのブログページについてコンテンツブロックが抽出され、その結果としてコンテンツブロック対応情報が記憶部65に記憶されているので、再度コンテンツブロックを抽出しなくても良い。その場合には、指定されたHTML文書のURLに基づいて、当該HTML文書が対応するブログページを構成する各コンテンツブロックのコンテンツブロック対応情報を記憶部65から取得することができる。

[0186] [2-4. 変形例1]

次に、本実施形態の変形例について、図22を用いて説明する。

[0187] これまでの説明においては、システム管理者によりブロガーが指定されたときに、指定されたブロガーのブログページに広告コンテンツを挿入していたが、ブログが更新されたタイミングで広告コンテンツを挿入しても良い。

[0188] 図22は、本実施形態の変形例に係るブログサーバ6のシステム制御部70のブログ更新時処理における処理例を示すフローチャートである。なお、図22において、図19と同様の処理については同様のステップ番号を付してある。

- [0189] 先ず、ブログの更新に先立ち、ブロガーは、ユーザ端末5を操作してブログサービスサイトにアクセスし、自身のユーザIDとパスワードとを入力することによりブログサービスサイトにログインする。このログインにより、ブログサーバ6はユーザ端末5に対してセッションIDを発行し、セッションIDとユーザIDとを対応付けて管理する。ユーザ端末5からブログサーバ6へのリクエストにはセッションIDが含まれているので、ブログサーバ6は、どのブロガーからのリクエストであるかを特定することができる。
- [0190] そして、ブロガーが新しいブログ記事の登録操作を行うと、ユーザ端末5は、ブログ記事のデータ（表題や本文等のテキストデータ、画像データ等）をブログサーバ6に送信し、図22に示すように、ブログサーバ6のシステム制御部70は、ブログ記事のデータを受信する（ステップS171）。次いで、システム制御部70は、ブロガーのユーザIDに対応するブログページの中から、更新すべきブログページのHTML文書をブログページDB601から取得する（ステップS172）。次いで、システム制御部70は、受信したブログ記事のデータに基づいて、取得したHTML文書を更新する（ステップS173）。例えば、システム制御部70は、取得したHTML文書に、ブログ記事用のTABLEタグ又はDIVタグを追加し、当該タグに挟まれた形で、受信したブログ記事の表題や本文のテキストデータ等を追加する。次いで、システム制御部70は、ブログ記事のデータを追加したHTML文書で、ブログページDB601に登録されているHTML文書を更新する（ステップS174）。
- [0191] 次いで、システム制御部70は、ブロガーのユーザIDに対応する全てのブログページからコンテンツブロックを抽出する（ステップS103～S106）。
- [0192] 次いで、システム制御部70は、ステップS173において更新したHTML文書を指定して、特有コンテンツブロック判定処理を実行し（ステップS108）、特有と判定されたコンテンツブロックを構成する各テキストデータからブログページの特徴語を抽出する（ステップS109）。

[0193] 次いで、システム制御部70は、指定されたHTML文書から、既存の広告コンテンツの規定を削除し（ステップS775）、抽出した特徴語をキーワードとして、関連する広告コンテンツの規定を挿入する（ステップS110）。つまり、システム制御部70は、ブログページ上に表示される広告コンテンツを変更する。

[0194] そして、システム制御部70は、広告コンテンツの規定が挿入されたHTML文書で、ブログページDB601に登録されているHTML文書を更新し（ステップS111）、記憶部65から全てのコンテンツブロック対応情報を削除する（ステップS114）。

[0195] なお、ブログの更新に伴ってブログページを新規に生成しなければならない場合の処理も、基本的に上述した処理と同様で良い。ただし、新規に生成されたブログページには、広告コンテンツは未だ挿入されていないので、ステップS175における広告コンテンツの規定の削除は行われない。

[0196] [2-5. 変形例2]

これまでの説明においては、ブログページに特有のコンテンツの判定に用いられる閾値として1回を設定していたが、2回以上の値を閾値として設定しても良い。

[0197] 例えば、閾値を1回とした場合には、出現頻度が1回であるコンテンツブロック（ブログ記事）がブログページに特有のコンテンツとして抽出され、抽出されたブログ記事のテキストデータから特徴語が抽出される。このとき、抽出された各ブログ記事のテキストデータのデータ量が少ないと、そこから抽出される単語の数は少なくなる。そして、十分な数の単語を抽出することができないと、どの単語が特徴語であるかを全く判断することができない場合や、的確に判断することができない場合がある。そこで、閾値の値を上げて、ブログページに特有のコンテンツと判定される条件をゆるめることで、特徴語を抽出する対象となるブログ記事を増やしていく。これにより、特徴語を抽出することが可能となる。

[0198] 具体的には、ブログサーバ6のシステム制御部70が、最初は閾値を1回

に設定して、ブログページに特有のコンテンツブロックを判定することにより、出現回数が1回のブログ記事を抽出して特徴語を抽出する。このとき、システム制御部70は、特徴語を抽出することができないと判定した場合には、閾値を2回に変更して、ブログ記事の抽出及び特徴語の抽出を行う。システム制御部70は、それでも特徴語を抽出することができないと判定した場合には、閾値を3回に変更して、ブログ記事の抽出及び特徴語の抽出を行う。システム制御部70は、こうした処理を、特徴語が抽出することができるまで継続する。つまり、特有のコンテンツブロックの抽出結果に基づく処理を正常に行うことができなかった場合に、閾値を上げるのである。

[0199] ただし、閾値を無制限に上げていくと、ブログ記事ではないものも抽出されてしまうので、閾値がある程度まで上がると処理を中断するものとする。例えば、閾値が、指定されたブロガーに対応するブログページのページ数の値にまで上がると、各ブログページで共通して用いられるコンテンツブロックを抽出してしまうので、閾値がブログページのページ数の値になったら処理を中断しても良い。

[0200] また例えば、ブログページの所定ページ数あたり1回のみ出現するコンテンツブロックをブログページ特有のコンテンツブロックであると、システム管理者側で予め定めても良い。この場合、指定されたブロガーに対応するブログページのページ数に比例して、閾値としての出現回数を変えても良い。

[0201] [2-6. 変形例3]

これまでの説明においては、ブログページに特有のコンテンツの判定に用いられる出現頻度として、出現回数（度数）を用いていたが、指定されたブロガーに対応するブログページの全コンテンツブロックに対する出現回数の割合（相対度数）を用いても良い。

[0202] 例えば、ブロガーが登録したブログ記事に対して、他のユーザからコメントを登録することができ、ブログ記事とともにコメントが閲覧可能になっているとする。このコメントのテキストデータもブログページを構成するコンテンツの1つとなる。ブログサーバ6のシステム制御部70は、コメントの

テキストデータをブログページに追加する場合、当該ブログページのHTML文書に、ブロック化タグの記述を追加した上で当該テキストデータを追加することにより、コメントのテキストデータを、ブログ記事や他のコメントのテキストデータとは独立したコンテンツブロックとする。そして、システム制御部70は、コンテンツブロックとしてコメントのテキストデータを抽出し、抽出したコメントのテキストデータが特有の内容を有している場合には、そのコメントに関連する広告コンテンツをブログページに挿入するようにする。

[0203] ところで、或るブログ記事に対して複数のコメントが登録された場合において、複数のコメントの各内容が、例えば、多数派の意見と少数派の意見といったように、頻繁に出現する内容と、あまり頻繁には出現しない内容とに分かれる場合がある。このとき、多数派の意見は、一般的な意見であり、あまり特徴的な内容ではないと考えることができる。一方、少数派の意見は、特異な意見であり、ブログページに特有の内容と考えることができる。そうした場合に、少数派の意見を示すコメントをブログページ特有のコンテンツとして抽出したい。

[0204] しかしながら、多数派の意見の数と少数派の意見の数は、相対的なものであり、コメントの総数によって変化する。こうした場合において、出現頻度として度数を用い、閾値を例えば1回とすると、頻繁に現れない内容（少数派の意見）を適切に抽出することができない場合がある。そこで、出現頻度として相対度数を用い、閾値を所定の割合に設定するのである。このときの閾値は任意に設定することができる。例えば、抽出されたコンテンツブロックの内容がN個のパターン（Nは2以上の整数）に分けられる場合、少数派の意見を区別するために、閾値には $1 \div N$ 未満の範囲で閾値を設定しても良い。このように、システム制御部70が、その時々状況に応じて閾値を変更しても良い。

[0205] なお、ブログ等のような記事に対してコメント等を登録することができるシステムとして、例えば、或るユーザが登録したつぶやきに対して、これを

フォローするつぶやきを他のユーザが登録することができるTwitter（商標）や、電子掲示板等がある。

[0206] 以上説明したように、本実施形態によれば、ブログサーバ6のシステム制御部70が、HTML文書が指定されることによって順次指定されたブログページを構成しているコンテンツを抽出し、指定されたブログページを構成している各コンテンツの出現頻度を計算し、指定されたブログページを構成するコンテンツのうち、出現頻度が所定の閾値以下のコンテンツを当該ブログページに特有のコンテンツであると判断する。

[0207] 従って、出現頻度が小さいコンテンツであるほど、指定されたブログページ以外にはあまり出現しないコンテンツであるので、出現頻度が閾値以下であるかを判定することで、当該条件を満たす全てのコンテンツが、指定されたブログページに特有のコンテンツであると特定される。よって、ブログページに特有のコンテンツを容易に抽出することができる。

[0208] また、ブログサーバ6のシステム制御部70が、指定されたブログページに特有のコンテンツに関連する広告コンテンツを当該ブログページに挿入する。

[0209] 従って、ブログページの特徴と関連する情報をWebページに追加することができる。

[0210] また、ブログサーバ6のシステム制御部70が、指定されたブログページを構成しているコンテンツとして、ブログの記事のテキストデータが含まれている場合に、当該テキストデータを、当該ブログページに特有のコンテンツであると判定し、ブログの記事のテキストデータから当該ブログページの特徴語を抽出し、当該特徴語をキーワードとして、予め関連付けられている広告コンテンツを、当該ブログページに挿入する。

[0211] 従って、ブログページに掲載されているブログの内容に関連する広告を当該ブログページに追加することができる。

[0212] また、ブログサーバ6のシステム制御部70が、ブログサービスサイトに含まれる複数のブログページ上における各コンテンツの出現頻度を計算する

- 。
- [0213] 従って、ブログサービスサイトに含まれる複数のWebページ（例えば、指定されたブロガーのユーザIDに対応する複数のブログページ）上において、指定されたブログページを構成している各コンテンツの出現頻度が計算されるので、ブログサービスサイト内で共通して用いられるコンテンツは、特有のコンテンツではないと判定することが可能となり、判断精度を上げることができる。
- [0214] また、ブログサーバ6のシステム制御部70が、1つ以上のコンテンツで構成されるコンテンツブロックの単位で、ブログページを構成しているコンテンツを抽出し、指定されたブログページを構成している各コンテンツブロック出現頻度を計算し、指定されたブログページを構成するコンテンツブロックのうち、出現頻度が閾値以下のコンテンツブロックを当該ブログページに特有のコンテンツブロックであると判断する。
- [0215] 従って、ブログページ上において、例えば、ヘッダ部分、ナビゲーション部分、ブログが表示される部分、コピーライト表示の部分等のように、1つ以上のコンテンツがまとまりをもってコンテンツブロックとして表示されている場合に、ブログページに特有のコンテンツブロックを抽出することができる。
- [0216] また、ブログサーバ6のシステム制御部70が、ブログページを構成しているコンテンツを当該ブログページのHTML文書に基づいて抽出し、HTML文書においてDIVタグ又はTABLEタグに基づいて、コンテンツブロックを定める。
- [0217] 従って、DIVタグにより、HTML文書の作成の際に明示的にブロック化された1つ以上のコンテンツを特定することができ、また、TABLEタグにより、表形式でブロック化されて表示される1つ以上のコンテンツを特定することができるので、例えば、これらのタグにより、ブログページに特有のコンテンツと、特有ではないコンテンツとがブロック化されている場合に、Webページに特有のコンテンツを判断する精度を上げることができる。

- 。
- [0218] なお、上記第2実施形態においては、指定されたブログページを構成する各コンテンツブロックに対応するコンテンツブロック対応情報を、指定されたブロガーのユーザIDに対応する全てのブログページを構成する各コンテンツブロック対応情報と比較することによって各出現頻度が計算されていた。つまり、指定されたブログページを構成する各コンテンツブロックの出現頻度を計算する場合に、指定されたブロガーに対応する全てのブログページを対象とした範囲に出現する頻度を計算するようになっていたが、対象とする範囲はこれだけに限られるものではない。例えば、予め定められたページ数分のブログページを対象としても良いし、ブログサービスサイトを構成する全てのブログページを対象としても良い。
- [0219] また、上記第2実施形態として、Webページに特有のコンテンツに関連するコンテンツとして、商品やサービスに関する広告を示す広告コンテンツを、当該Webページに挿入していたが、関連するコンテンツであれば広告コンテンツに限られるものではない。例えば、特有のコンテンツと判定されたブログ記事等のコンテンツに関連する画像データ（静止画や動画像）を、背景画像や挿入画像（挿絵等）として挿入しても良い。具体的には、例えば、画像データ用のデータベースを構築し、当該データベースに、画像データとキーワードとを対応付けて登録しておく。画像データに対応付けられるキーワードは、その画像データによって表される画像を示す単語や当該画像に関連する単語である。そして、特有のコンテンツと判定されたコンテンツから特徴語を抽出し、抽出した特徴語をキーワードとして、関連する画像データをデータベースから選択する。そして、対象のHTML文書のBODYタグに、選択した画像データのURLをbackground属性として挿入したり、対象のHTML文書の所定位置に、選択した画像データを表示するIMGタグを挿入したりする。これにより、特有のコンテンツと判定されたブログ記事等のコンテンツの内容に適した画像をWebページに挿入することができる。

- [0220] また、Webページに特有のコンテンツの用途としては、関連するコンテンツをWebページに挿入することのみに限られるものではない。例えば、Webページに特有のコンテンツに基づいて、新たなコンテンツを生成しても良い。
- [0221] また、上記各実施形態においては、Webページを構成しているコンテンツとして、テキストデータ及び画像データを抽出していたが、抽出対象のコンテンツはこれらに限られるものではない。例えば、Webページ上に表示されるコンテンツ、又は、Webページが表示されている際に再生されるコンテンツ（例えば、動画データ、音声データ、電子文書等）であれば良い。また、所定の種類のコンテンツのみを抽出しても良い。
- [0222] また、上記各実施形態においては、DIVタグに挟まれているコンテンツ、及び、TABLEタグに挟まれているコンテンツを、コンテンツブロックとしてグループ化して抽出していたが、コンテンツをグループ化するタグとしては、これらのみに限られるものではない。
- [0223] また、上記各実施形態においては、Webページに特有のコンテンツをコンテンツブロックの単位で抽出していたが、各コンテンツをそのまま一つずつ抽出しても良い。
- [0224] また、上記各実施形態においては、サーバ装置に対して本発明の特有コンテンツ判定装置を適用していたが、記憶手段やネットワーク上からHTML文書を取得することができれば、端末装置等に対して特有コンテンツ判定装置を適用しても良い。
- [0225] また、上記各実施形態においては、HTML文書に対して本発明のドキュメントデータを適用していたが、マークアップ言語で記述され、Webページを構成するコンテンツを示すデータ（例えば、XHTML（Extensible HyperText Markup Language）文書等）に対してドキュメントデータを適用しても良い。
- [0226] また、上記実施形態においては、ショッピングサイトにおける商品詳細ページを構成するコンテンツや、ブログサービスサイトにおけるブログページ

を構成するコンテンツを抽出していたが、対象とするサイト及びページの種類はこれらのみに限られるものではない。

符号の説明

- [0227] 1 コンテンツ生成サーバ
- 2 ショッピングサーバ
- 3 管理端末
- 4 店舗端末
- 5 ユーザ端末
- 1 1 操作部
- 1 2 表示部
- 1 3 通信部
- 1 4 ドライブ部
- 1 5 記憶部
- 1 6 入出力インタフェース部
- 1 7 CPU
- 1 8 ROM
- 1 9 RAM
- 2 0 システム制御部
- 2 1 システムバス
- 1 0 1 素材抽出DB
- 2 0 1 商品詳細ページDB
- NW ネットワーク
- S ショッピングシステム
- 6 ブログサーバ6
- 6 1 操作部
- 6 2 表示部
- 6 3 通信部
- 6 4 ドライブ部

- 6 5 記憶部
- 6 6 入出カインタフェース部
- 6 7 CPU
- 6 8 ROM
- 6 9 RAM
- 6 0 システム制御部
- 6 1 システムバス
- 6 0 1 ブログページDB
- 6 0 2 広告DB
- BS ブログシステム

請求の範囲

- [請求項1] コンピュータを、
指定されたWebページを構成しているコンテンツを抽出する抽出手段、
前記指定されたWebページを構成している各コンテンツの出現頻度を計算する計算手段、及び、
前記計算された出現頻度に基づいて、前記指定されたWebページを構成しているコンテンツのうち、当該Webページに特有であるコンテンツを判定する判定手段、
として機能させることを特徴とする特有コンテンツ判定プログラム。
- [請求項2] 請求項1に記載の特有コンテンツ判定プログラムにおいて、
前記判定手段が、前記指定されたWebページを構成しているコンテンツのうち、出現頻度が最も小さいコンテンツを当該Webページに特有のコンテンツであると判定するように、前記コンピュータを機能させることを特徴とする特有コンテンツ判定プログラム。
- [請求項3] 請求項1に記載の特有コンテンツ判定プログラムにおいて、
前記判定手段が、前記指定されたWebページを構成しているコンテンツのうち、出現頻度が所定値以下のコンテンツを当該Webページに特有のコンテンツであると判定するように、前記コンピュータを機能させることを特徴とする特有コンテンツ判定プログラム。
- [請求項4] 請求項1乃至3の何れか1項に記載の特有コンテンツ判定プログラムにおいて、
前記計算手段が、所定のサイトに含まれる複数のWebページ上における各コンテンツの出現頻度を計算するように、前記コンピュータを機能させることを特徴とする特有コンテンツ判定プログラム。
- [請求項5] 請求項1乃至4の何れか1項に記載の特有コンテンツ判定プログラムにおいて、

前記抽出手段が、前記所定のサイトに含まれる予め定められた種類の各WebページについてWebページを構成しているコンテンツを抽出し、抽出したコンテンツを示すコンテンツ情報を予め記憶手段に記憶しておき、

前記計算手段が、前記記憶されたコンテンツ情報に基づいて、前記指定されたWebページを構成している各コンテンツの出現頻度を計算するように、前記コンピュータを機能させることを特徴とする特有コンテンツ判定プログラム。

[請求項6] 請求項1乃至5の何れか1項に記載の特有コンテンツ判定プログラムにおいて、

前記抽出手段が、1つ以上のコンテンツで構成されたコンテンツグループの単位で、Webページを構成しているコンテンツを抽出し、

前記計算手段が、前記指定されたWebページを構成しているコンテンツグループの出現頻度を計算し、

前記判定手段が、前記指定されたWebページを構成しているコンテンツグループのうち、当該Webページに特有であるコンテンツグループを判定するように、前記コンピュータを機能させることを特徴とする特有コンテンツ判定プログラム。

[請求項7] 請求項6に記載の特有コンテンツ判定プログラムにおいて、

前記抽出手段が、所定のマークアップ言語で記述され、Webページを構成するコンテンツを示すドキュメントデータに基づいて、コンテンツグループを抽出するように、前記コンピュータを機能させることを特徴とする特有コンテンツ判定プログラム。

[請求項8] 請求項7に記載の特有コンテンツ判定プログラムにおいて、

前記抽出手段が、前記コンテンツを示すドキュメントデータにおいて予め定められたタグに基づいてコンテンツグループを定めるように、前記コンピュータを機能させることを特徴とする特有コンテンツ判定プログラム。

- [請求項9] 請求項 1 乃至 8 の何れか 1 項に記載の特有コンテンツ判定プログラムにおいて、
- 特有のコンテンツであると判定されたコンテンツに基づいて、新たなコンテンツを生成する生成手段として前記コンピュータを更に機能させることを特徴とする特有コンテンツ判定プログラム。
- [請求項10] 請求項 9 に記載の特有コンテンツ判定プログラムにおいて、
- 前記生成手段が、特有のコンテンツであると判定されたコンテンツの表示サイズを、予め設定された表示サイズに合うように調整し、表示サイズが調整されたコンテンツを含む新たなコンテンツを生成するように、前記コンピュータを機能させることを特徴とする特有コンテンツ判定プログラム。
- [請求項11] 請求項 9 又は請求項 10 に記載の特有コンテンツ判定プログラムにおいて、
- 前記生成手段が、特有のコンテンツであると判定されたコンテンツにエフェクトが施されて当該コンテンツが再生される新たなコンテンツを生成するように、前記コンピュータを機能させることを特徴とする特有コンテンツ判定プログラム。
- [請求項12] 請求項 1 乃至 8 の何れか 1 項に記載の特有コンテンツ判定プログラムにおいて、
- 特有のコンテンツであると判定されたコンテンツに関連する関連コンテンツを、前記指定された Web ページに挿入する挿入手段として前記コンピュータを更に機能させることを特徴とする特有コンテンツ判定プログラム。
- [請求項13] 請求項 12 に記載の特有コンテンツ判定プログラムにおいて、
- 前記判定手段が、前記指定された Web ページを構成しているコンテンツとして、ブログの記事のテキストデータが含まれている場合に、当該テキストデータを当該 Web ページに特有のコンテンツであると判定し、

前記挿入手段が、前記特有コンテンツ判定装置により特有のコンテンツであると判定されたブログの記事のテキストデータから前記指定されたWebページの特徴語を抽出し、当該特徴語に関連する関連コンテンツを、当該Webページに挿入するように、前記コンピュータを機能させることを特徴とする特有コンテンツ判定プログラム。

[請求項14]

指定されたWebページを構成しているコンテンツを抽出する抽出手段と、

前記指定されたWebページを構成している各コンテンツの出現頻度を計算する計算手段と、

前記計算された出現頻度に基づいて、前記指定されたWebページを構成しているコンテンツのうち、当該Webページに特有であるコンテンツを判定する判定手段と、

を備えることを特徴とする特有コンテンツ判定装置。

[請求項15]

請求項14に記載の特有コンテンツ判定装置において、

前記判定手段は、前記指定されたWebページを構成しているコンテンツのうち、出現頻度が最も小さいコンテンツを当該Webページに特有のコンテンツであると判定することを特徴とする特有コンテンツ判定装置。

[請求項16]

請求項14に記載の特有コンテンツ判定装置において、

前記判定手段は、前記指定されたWebページを構成しているコンテンツのうち、出現頻度が所定値以下のコンテンツを当該Webページに特有のコンテンツであると判定することを特徴とする特有コンテンツ判定装置。

[請求項17]

請求項14乃至16の何れか1項に記載の特有コンテンツ判定装置において、

前記計算手段は、所定のサイトに含まれる複数のWebページ上における各コンテンツの出現頻度を計算することを特徴とする特有コンテンツ判定装置。

[請求項18] 請求項14乃至17の何れか1項に記載の特有コンテンツ判定装置において、

前記抽出手段は、前記所定のサイトに含まれる予め定められた種類の各WebページについてWebページを構成しているコンテンツを抽出し、抽出したコンテンツを示すコンテンツ情報を予め記憶手段に記憶しておき、

前記計算手段は、前記記憶されたコンテンツ情報に基づいて、前記指定されたWebページを構成している各コンテンツの出現頻度を計算することを特徴とする特有コンテンツ判定装置。

[請求項19] 請求項14乃至18の何れか1項に記載の特有コンテンツ判定装置において、

前記抽出手段は、1つ以上のコンテンツで構成されたコンテンツグループの単位で、Webページを構成しているコンテンツを抽出し、

前記計算手段は、前記指定されたWebページを構成しているコンテンツグループの出現頻度を計算し、

前記判定手段は、前記指定されたWebページを構成しているコンテンツグループのうち、当該Webページに特有であるコンテンツグループを判定することを特徴とする特有コンテンツ判定装置。

[請求項20] 請求項19に記載の特有コンテンツ判定装置において、

前記抽出手段は、所定のマークアップ言語で記述され、Webページを構成するコンテンツを示すドキュメントデータに基づいて、コンテンツグループを抽出することを特徴とする特有コンテンツ判定装置。

[請求項21] 請求項20に記載の特有コンテンツ判定装置において、

前記抽出手段は、前記コンテンツを示すドキュメントデータにおいて予め定められたタグに基づいてコンテンツグループを定めることを特徴とする特有コンテンツ判定装置。

[請求項22] 指定されたWebページを構成しているコンテンツを抽出する抽出

行程と、

前記指定されたW e b ページを構成している各コンテンツの出現頻度を計算する計算行程と、

前記計算された出現頻度に基づいて、前記指定されたW e b ページを構成しているコンテンツのうち、当該W e b ページに特有であるコンテンツを判定する判定行程と、

を有することを特徴とする特有コンテンツ判定方法。

[請求項23]

コンピュータを、

指定されたW e b ページを構成しているコンテンツを抽出する抽出手段、

前記指定されたW e b ページを構成している各コンテンツの出現頻度を計算する計算手段、及び、

前記計算された出現頻度に基づいて、前記指定されたW e b ページを構成しているコンテンツのうち、当該W e b ページに特有であるコンテンツを判定する判定手段、

として機能させる特有コンテンツ判定プログラムがコンピュータ読み取り可能に記録されていることを特徴とする記録媒体。

[請求項24]

請求項 1 4 乃至 2 1 の何れか 1 項に記載の特有コンテンツ判定装置と、

前記特有コンテンツ判定装置により特有のコンテンツであると判定されたコンテンツに基づいて、新たなコンテンツを生成する生成手段と、

を備えることを特徴とするコンテンツ生成装置。

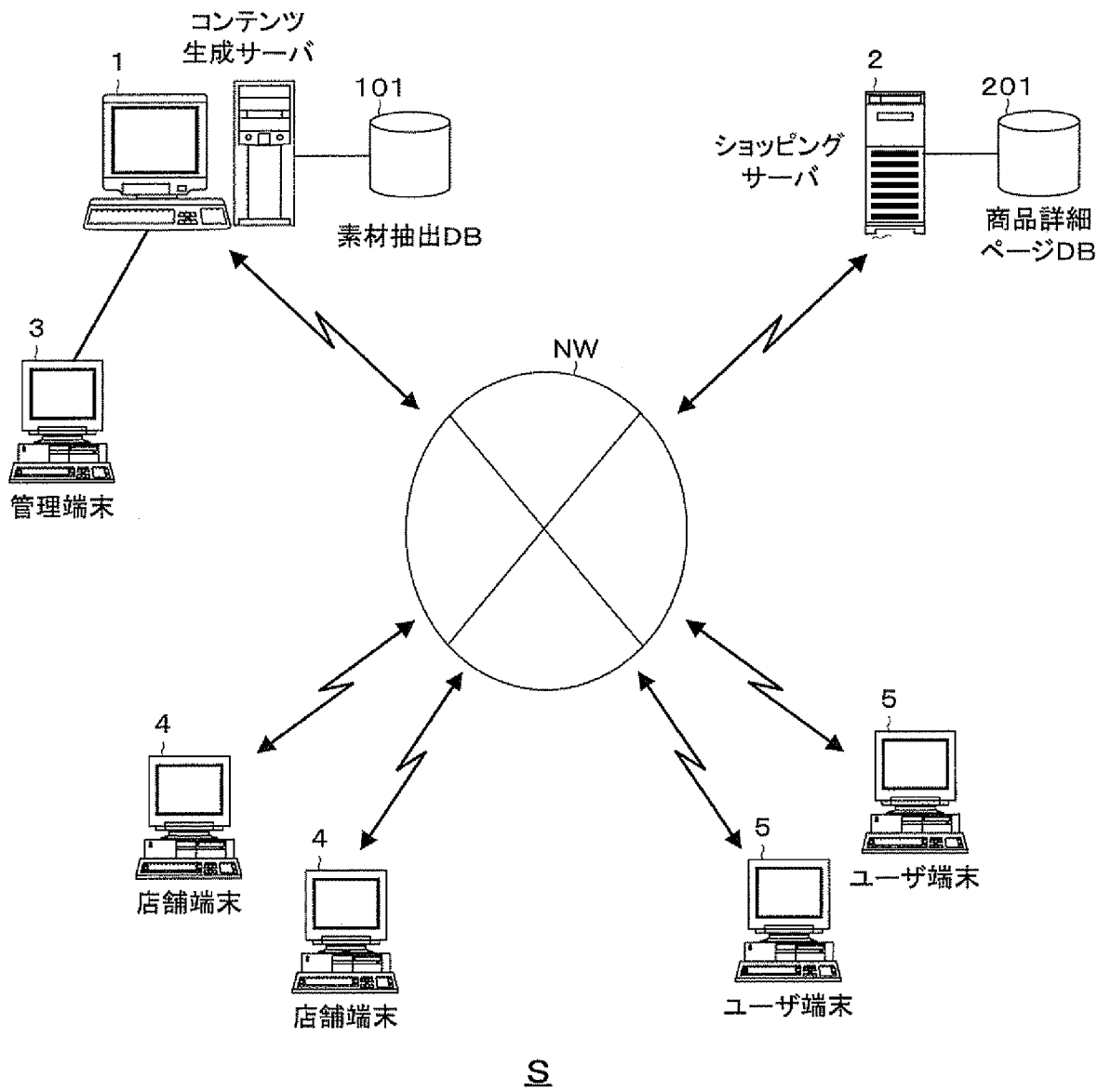
[請求項25]

請求項 2 4 に記載のコンテンツ生成装置において、

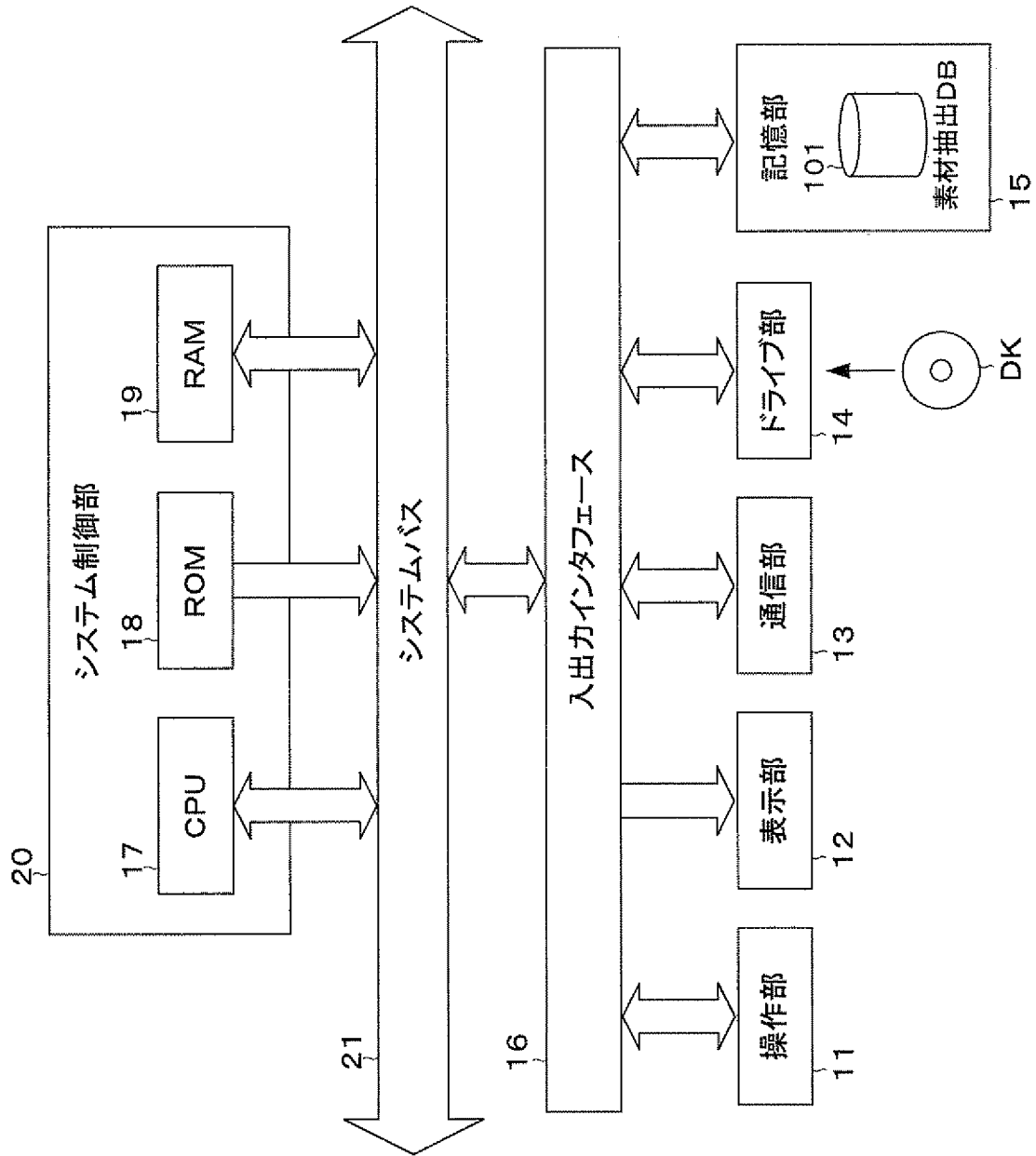
前記生成手段は、特有のコンテンツであると判定されたコンテンツの表示サイズを、予め設定された表示サイズに合うように調整し、表示サイズが調整されたコンテンツを含む新たなコンテンツを生成することを特徴とするコンテンツ生成装置。

- [請求項26] 請求項24又は請求項25に記載のコンテンツ生成装置において、前記生成手段は、特有のコンテンツであると判定されたコンテンツにエフェクトが施されて当該コンテンツが再生される新たなコンテンツを生成することを特徴とするコンテンツ生成装置。
- [請求項27] 請求項14乃至21の何れか1項に記載の特有コンテンツ判定装置と、
前記特有コンテンツ判定装置により特有のコンテンツであると判定されたコンテンツに関連する関連コンテンツを、前記指定されたWebページに挿入する挿入手段と、
を備えることを特徴とする関連コンテンツ挿入装置。
- [請求項28] 請求項27に記載の関連コンテンツ挿入装置において、
前記特有コンテンツ判定装置は、前記指定されたWebページを構成しているコンテンツとして、ブログの記事のテキストデータが含まれている場合に、当該テキストデータを当該Webページに特有のコンテンツであると判定し、
前記挿入手段は、前記特有コンテンツ判定装置により特有のコンテンツであると判定されたブログの記事のテキストデータから前記指定されたWebページの特徴語を抽出し、当該特徴語に関連する関連コンテンツを、当該Webページに挿入することを特徴とする関連コンテンツ挿入装置。

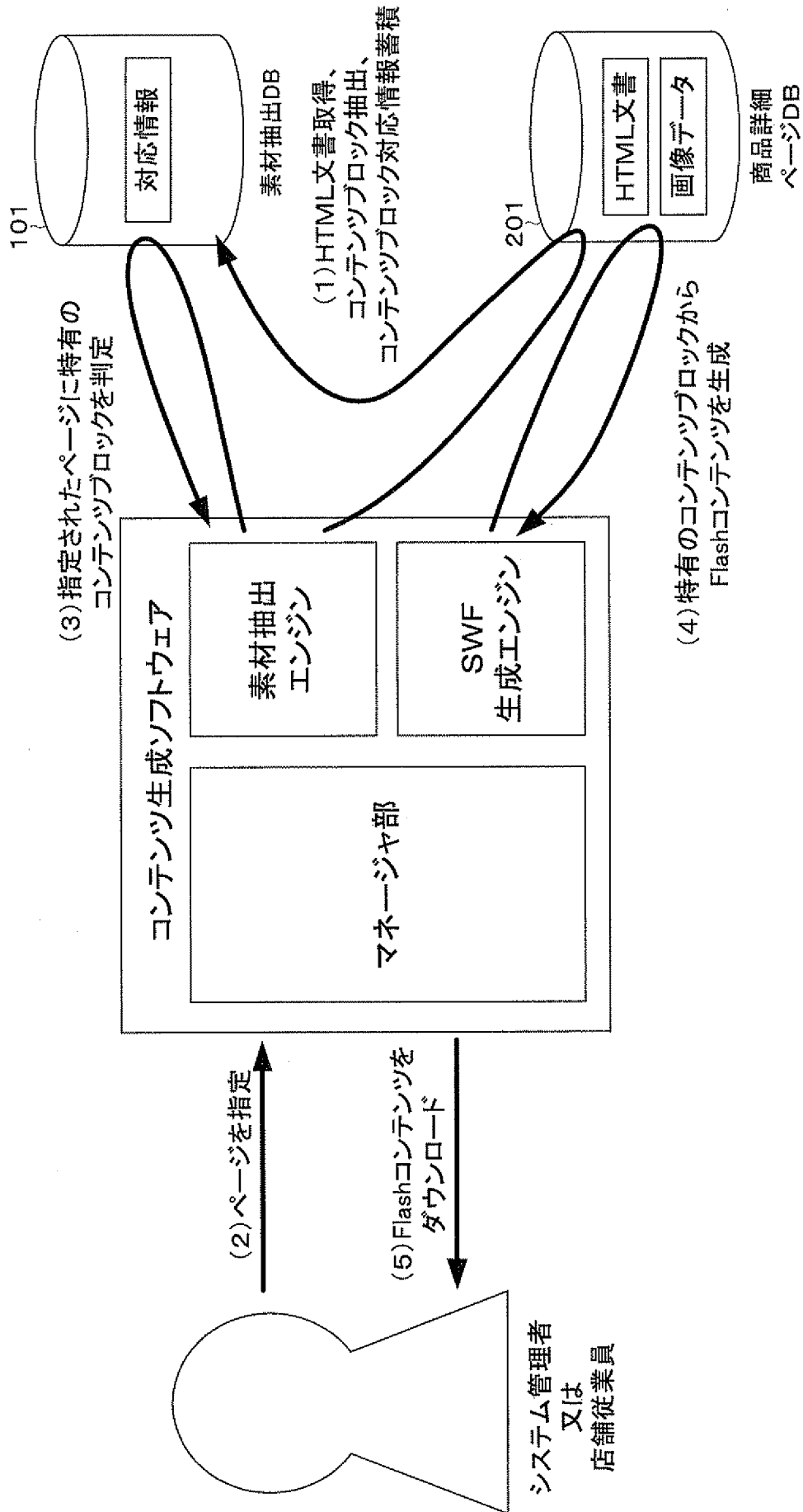
[図1]



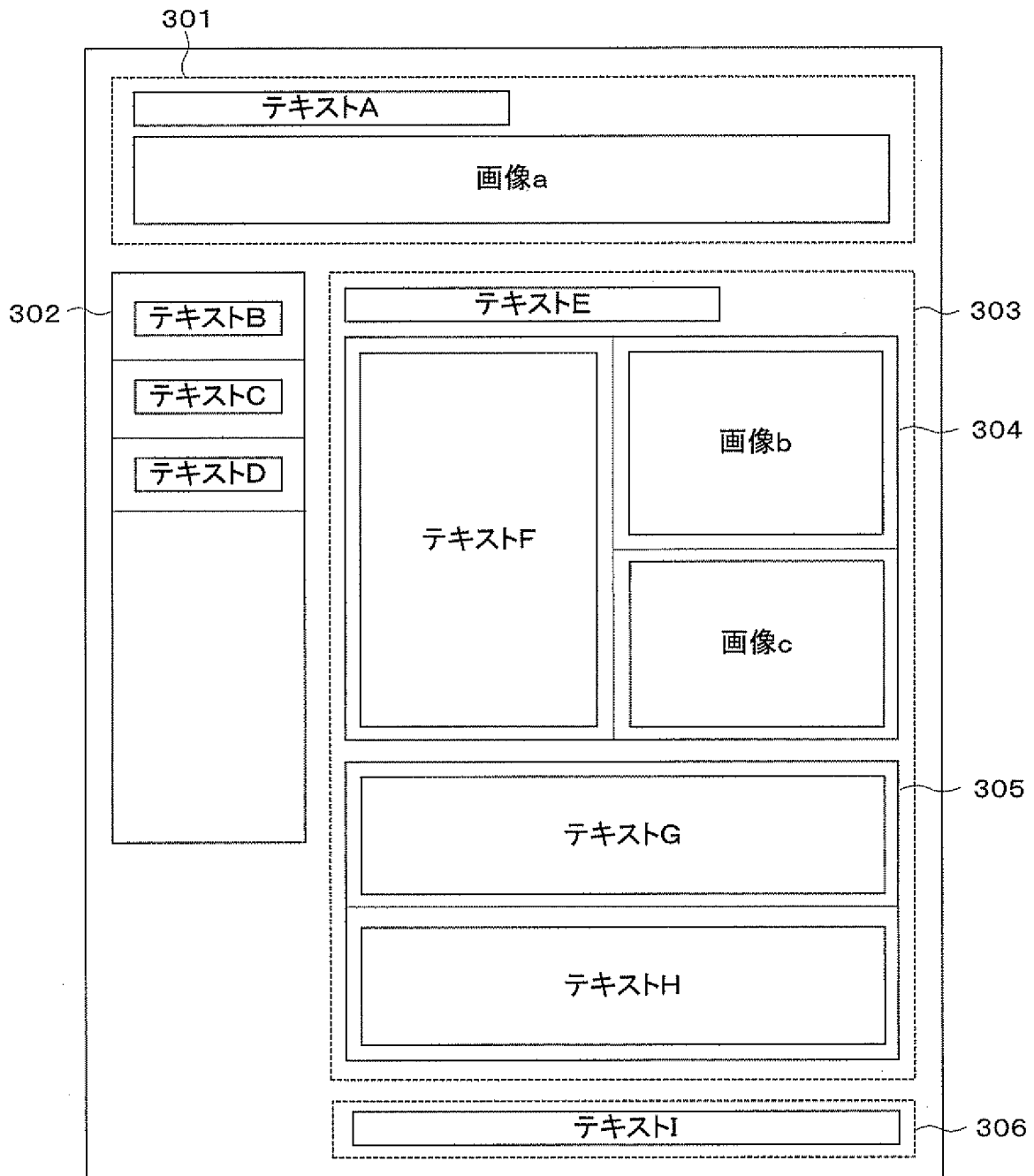
[図2]



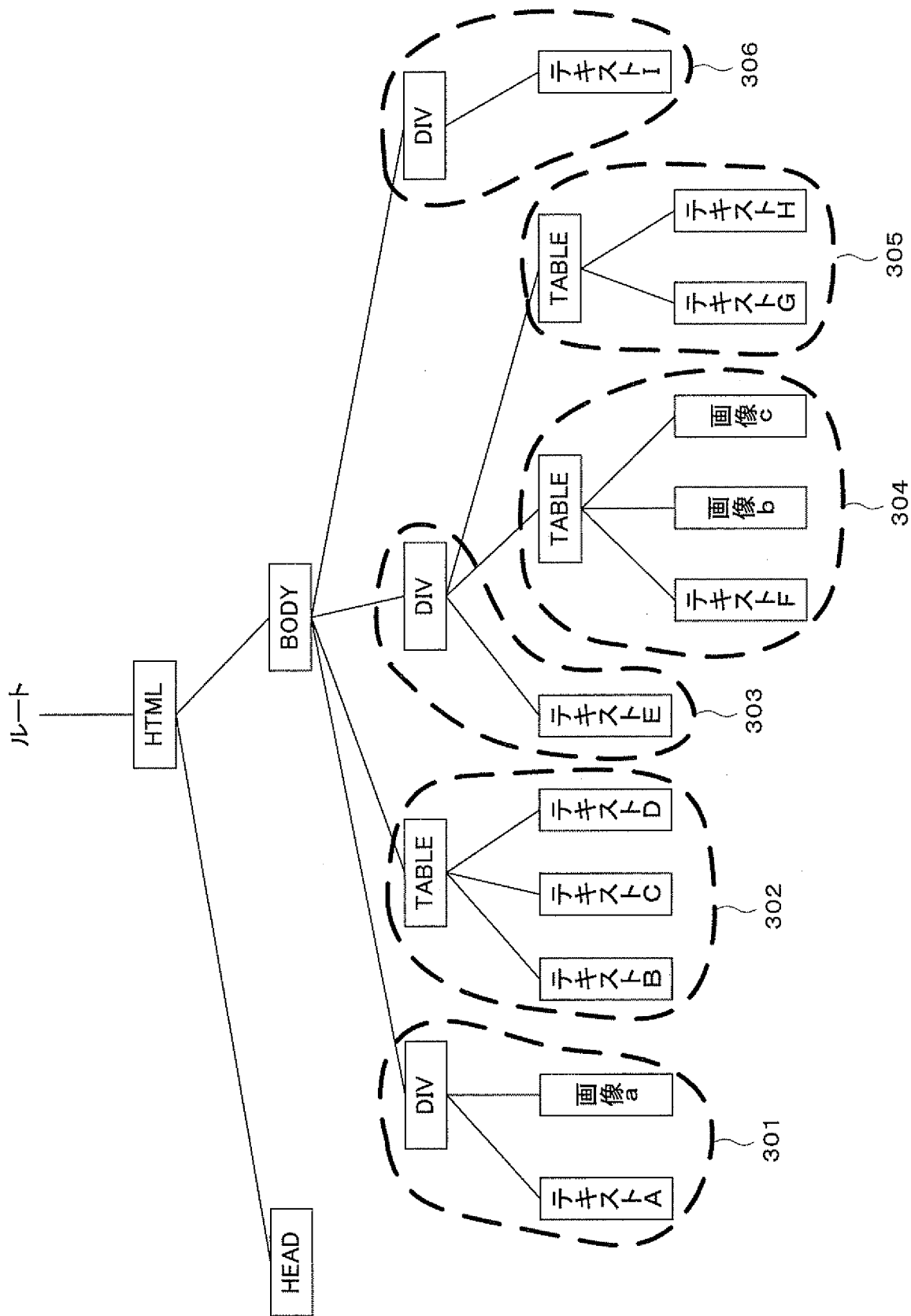
[図3]



[図4]



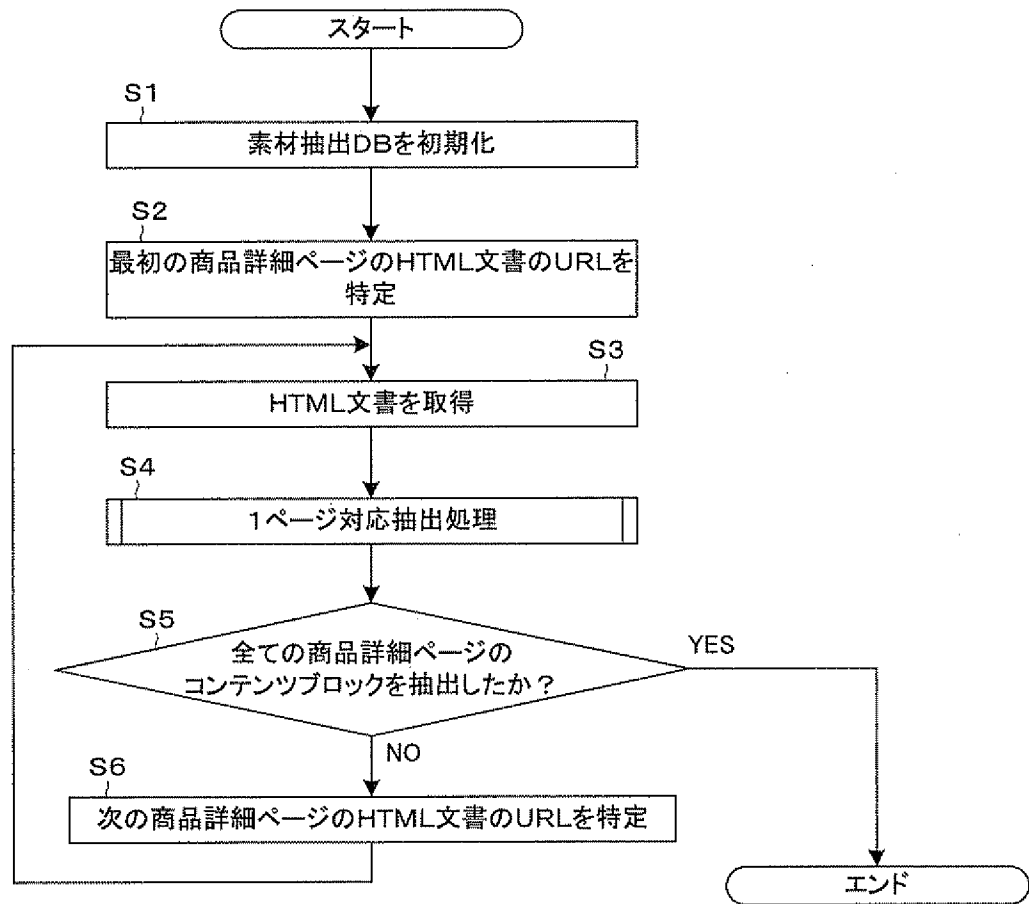
[図5]



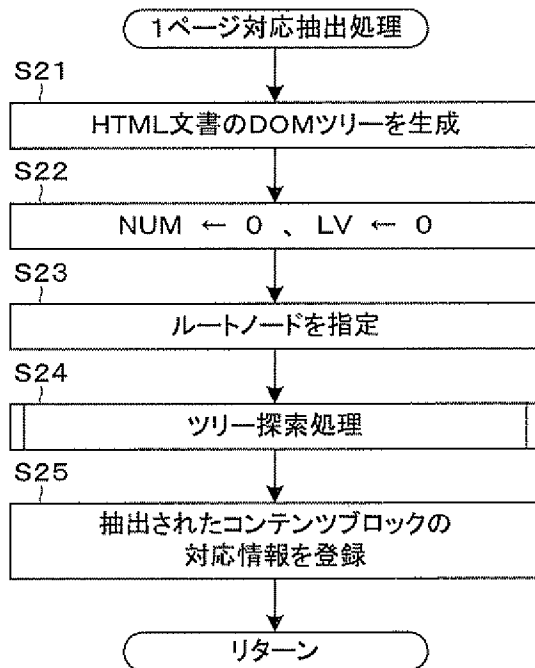
[図6]



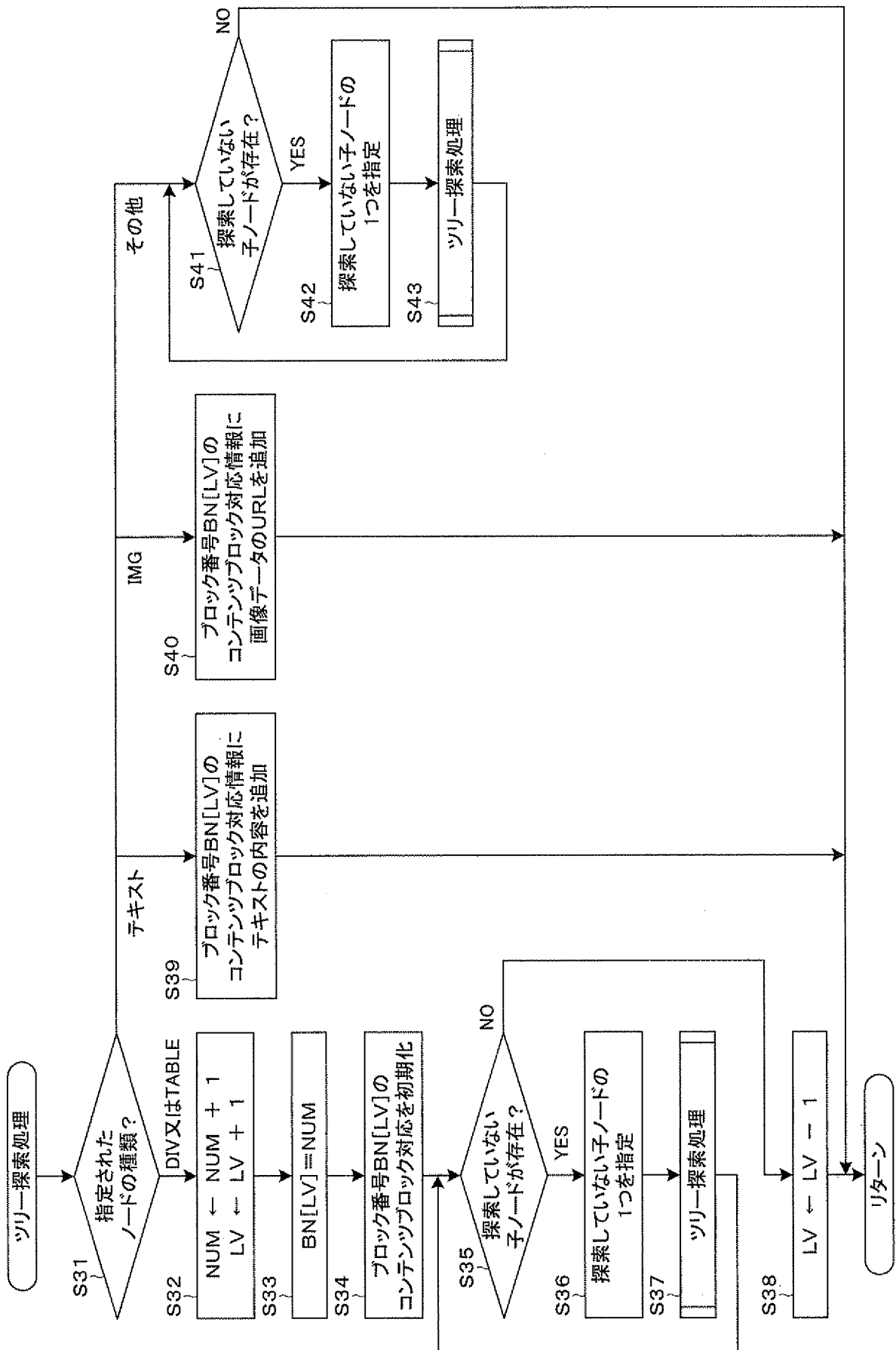
[図7]



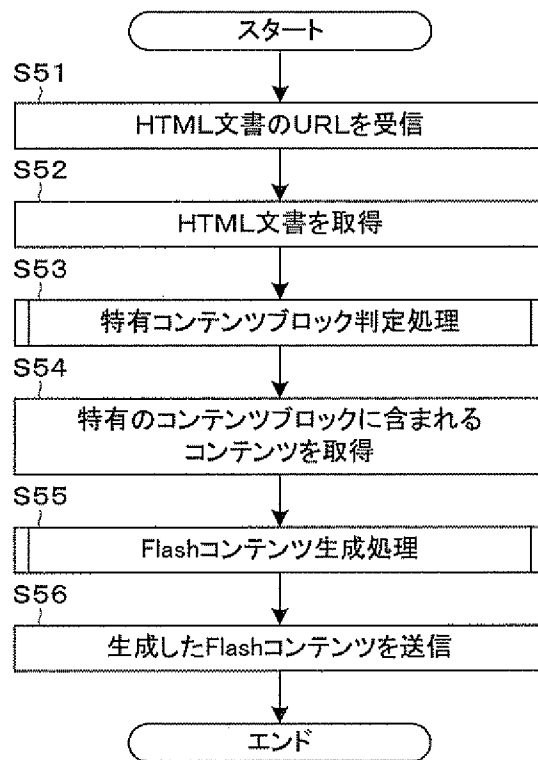
[図8]



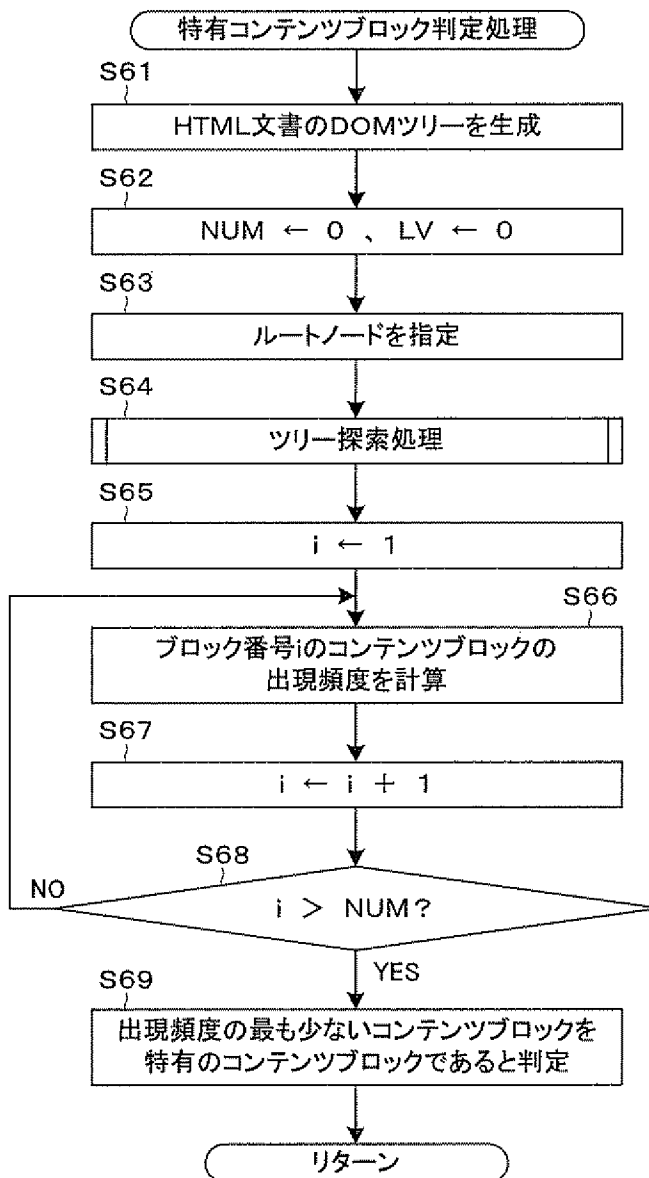
[図9]



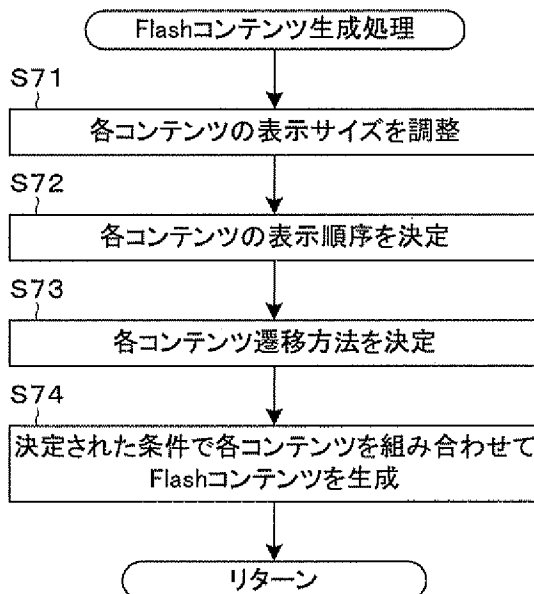
[図10]



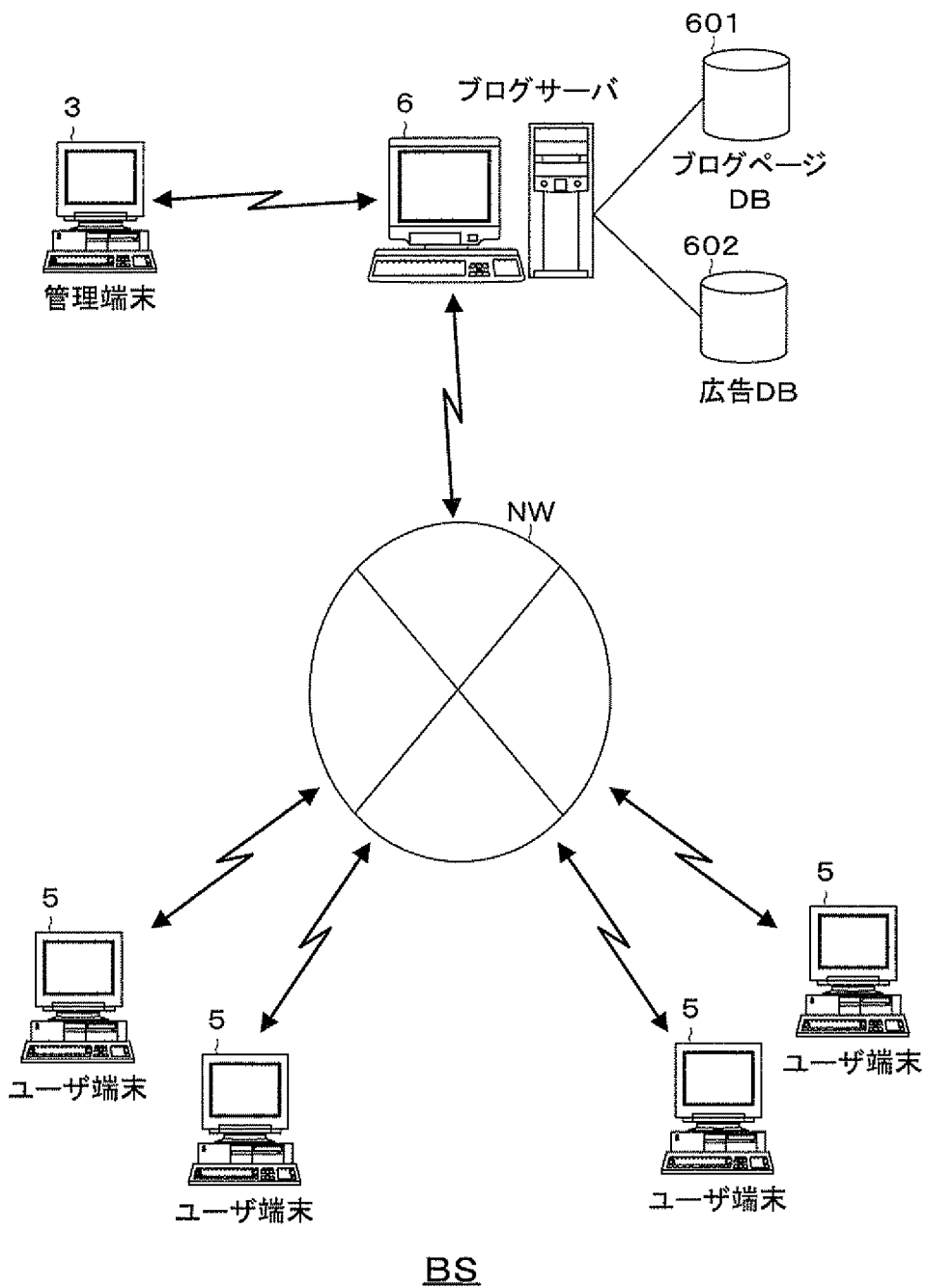
[図11]



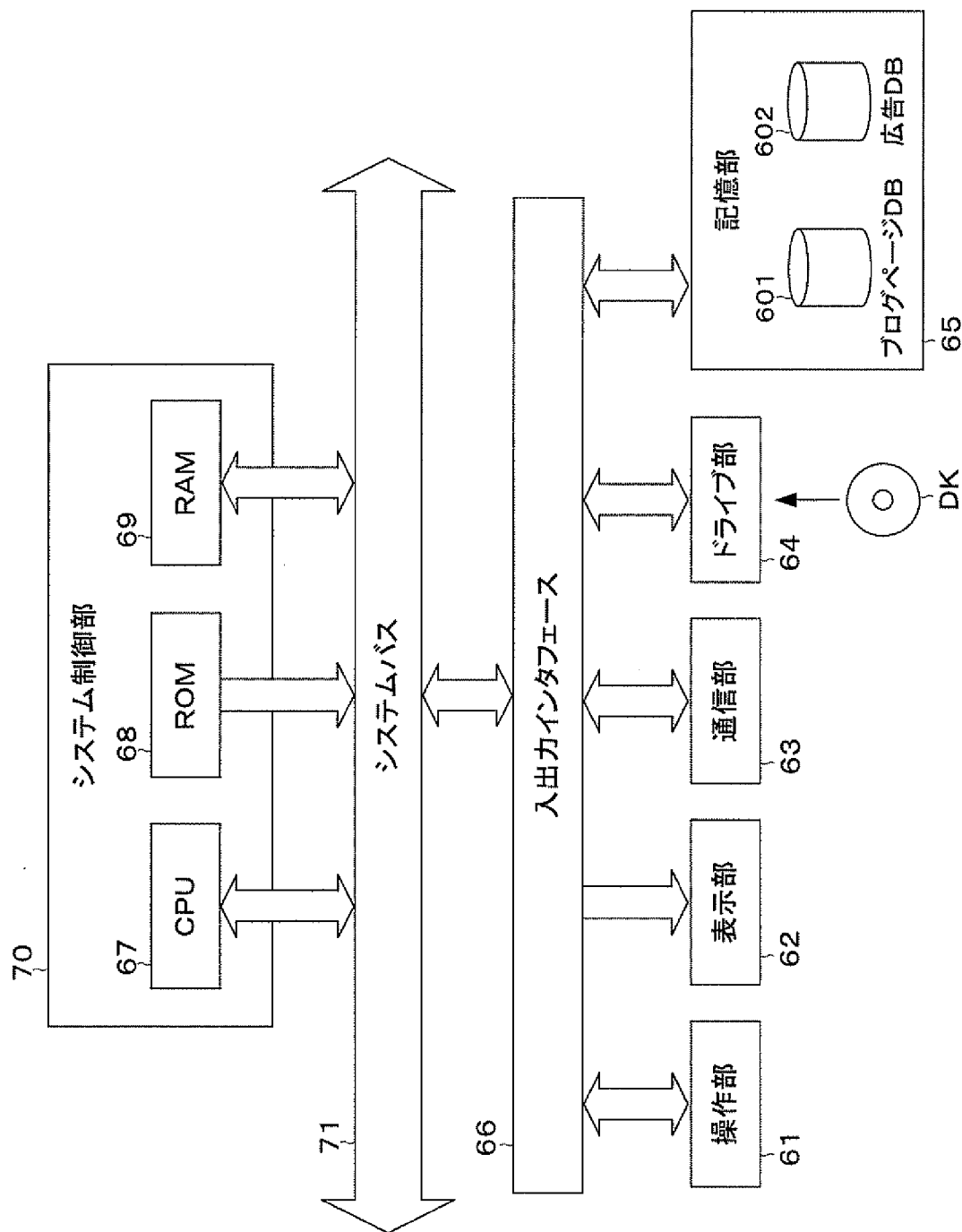
[図12]



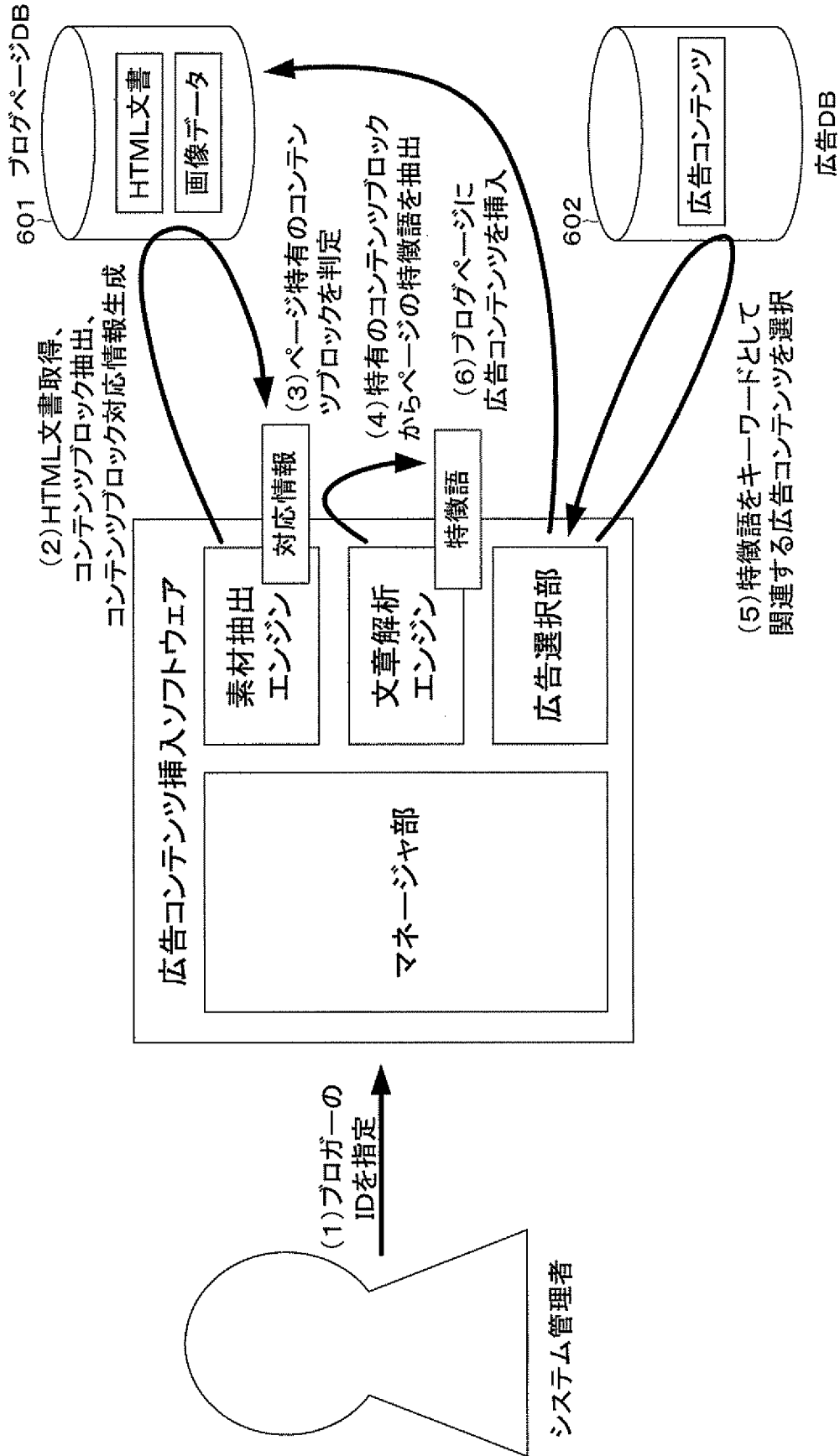
[図13]



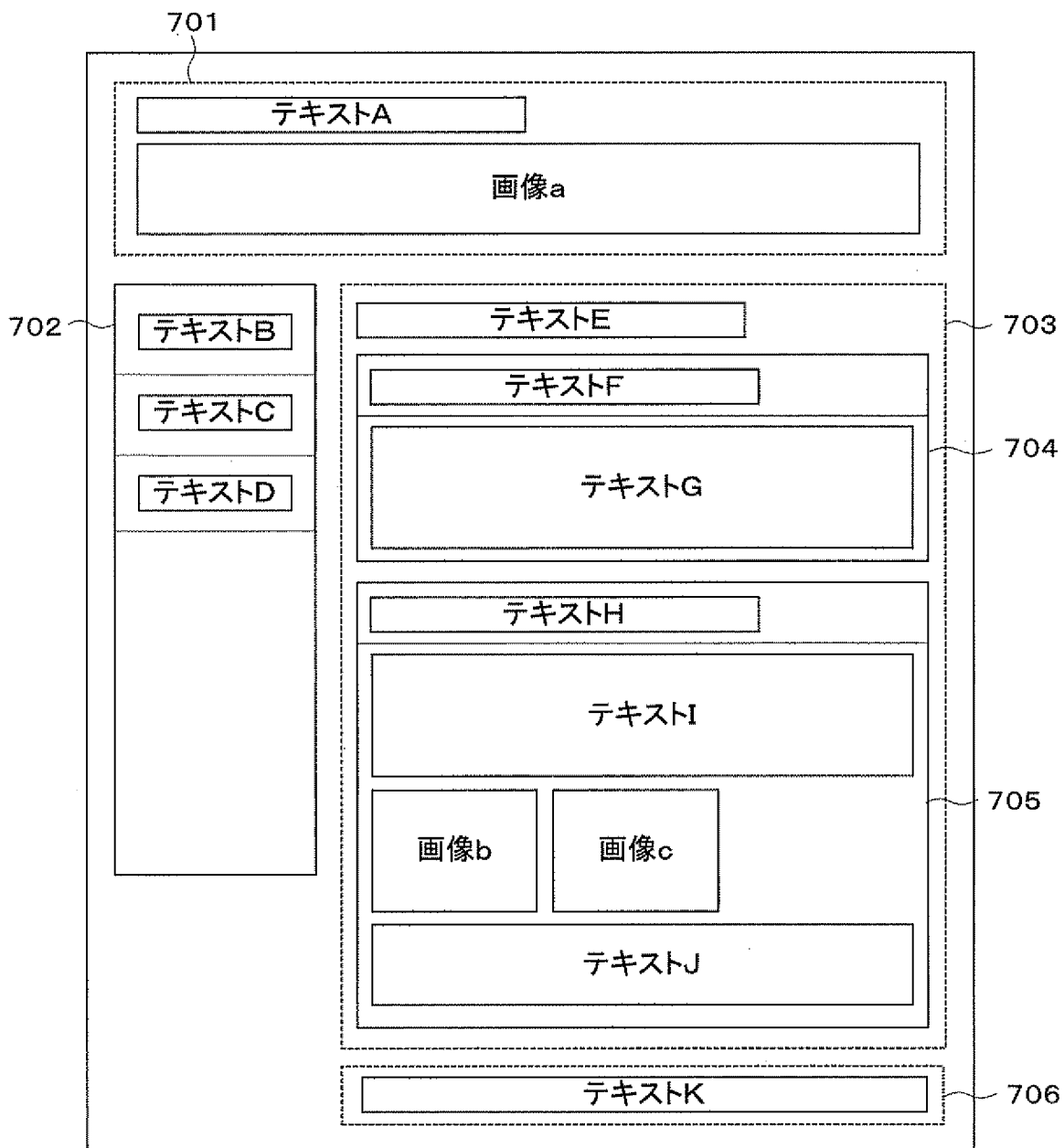
[図14]



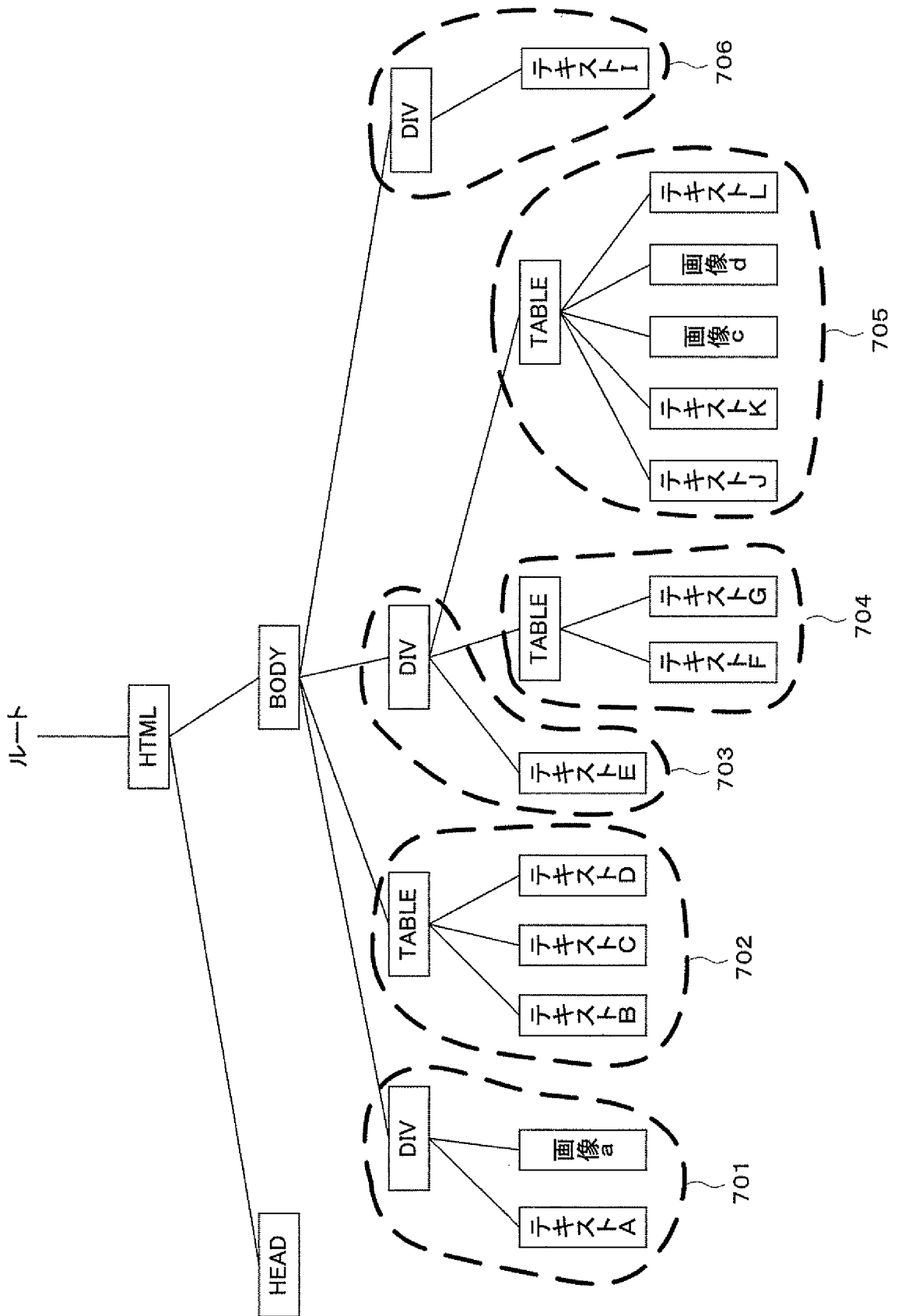
[図15]



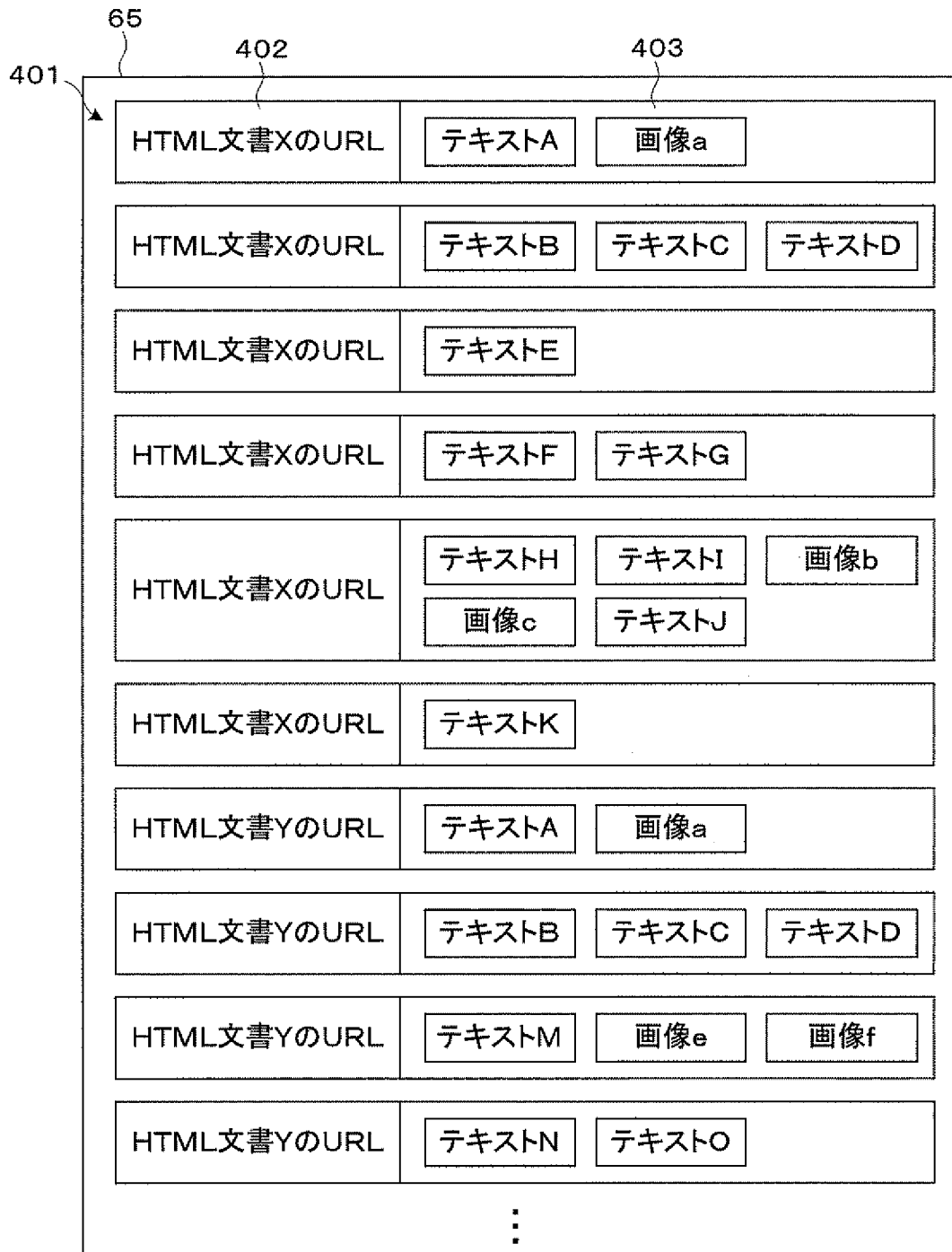
[図16]



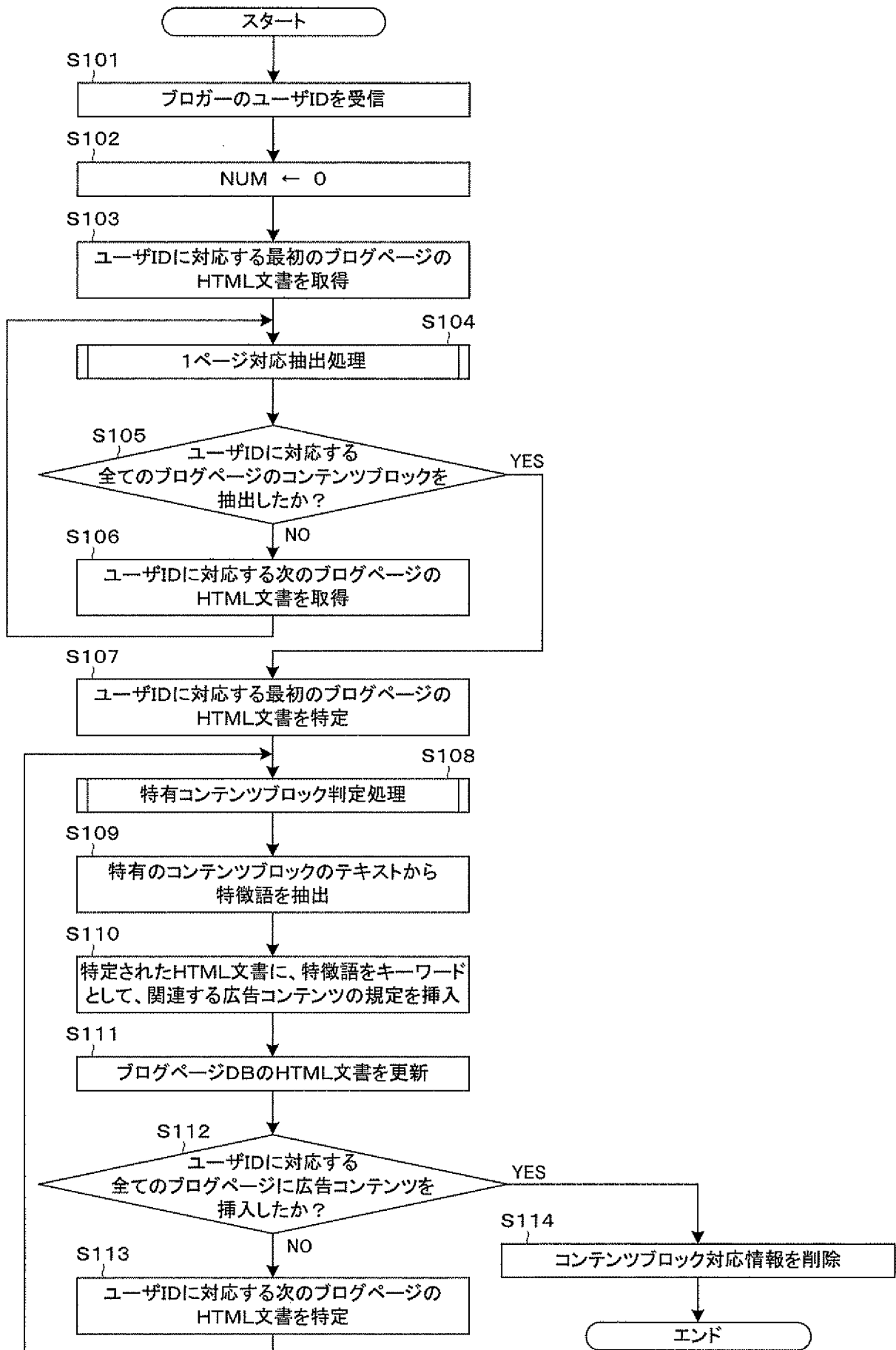
[図17]



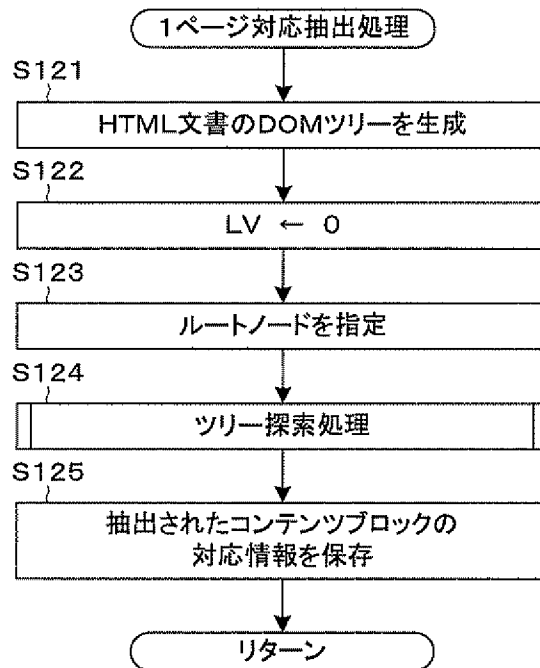
[図18]



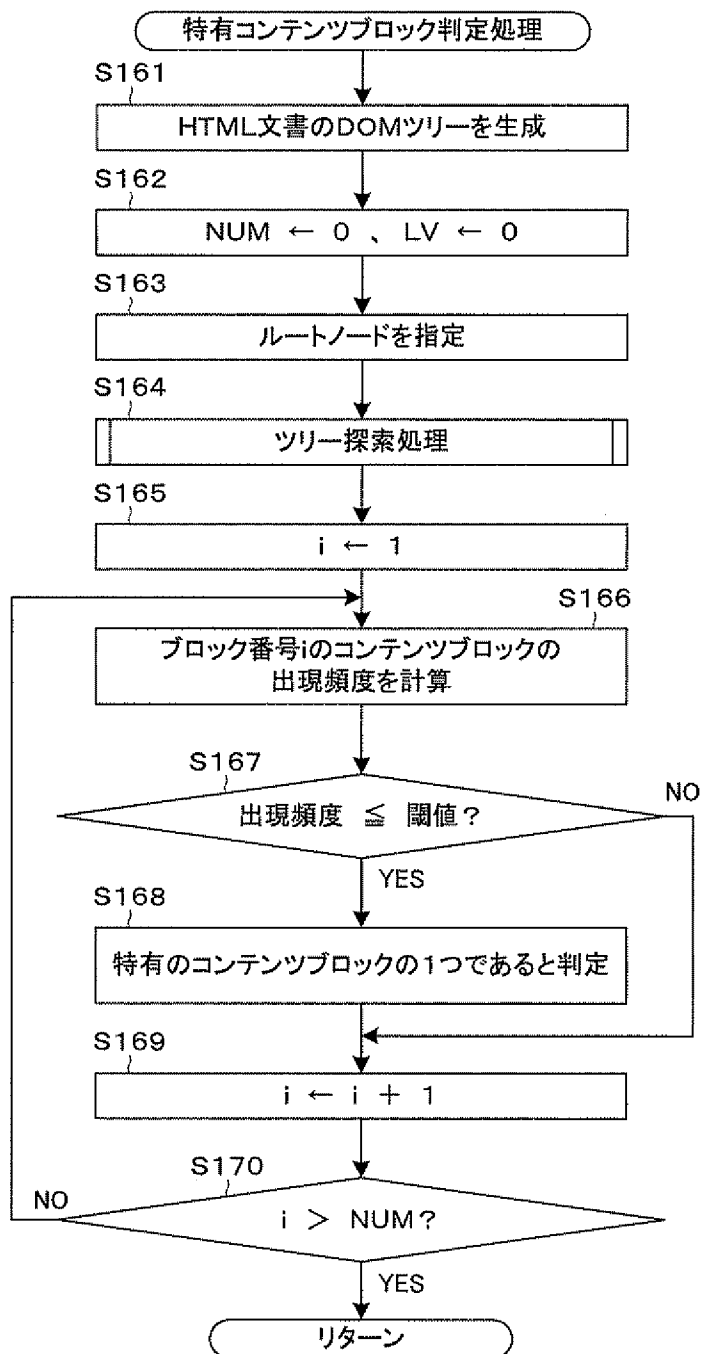
[図19]



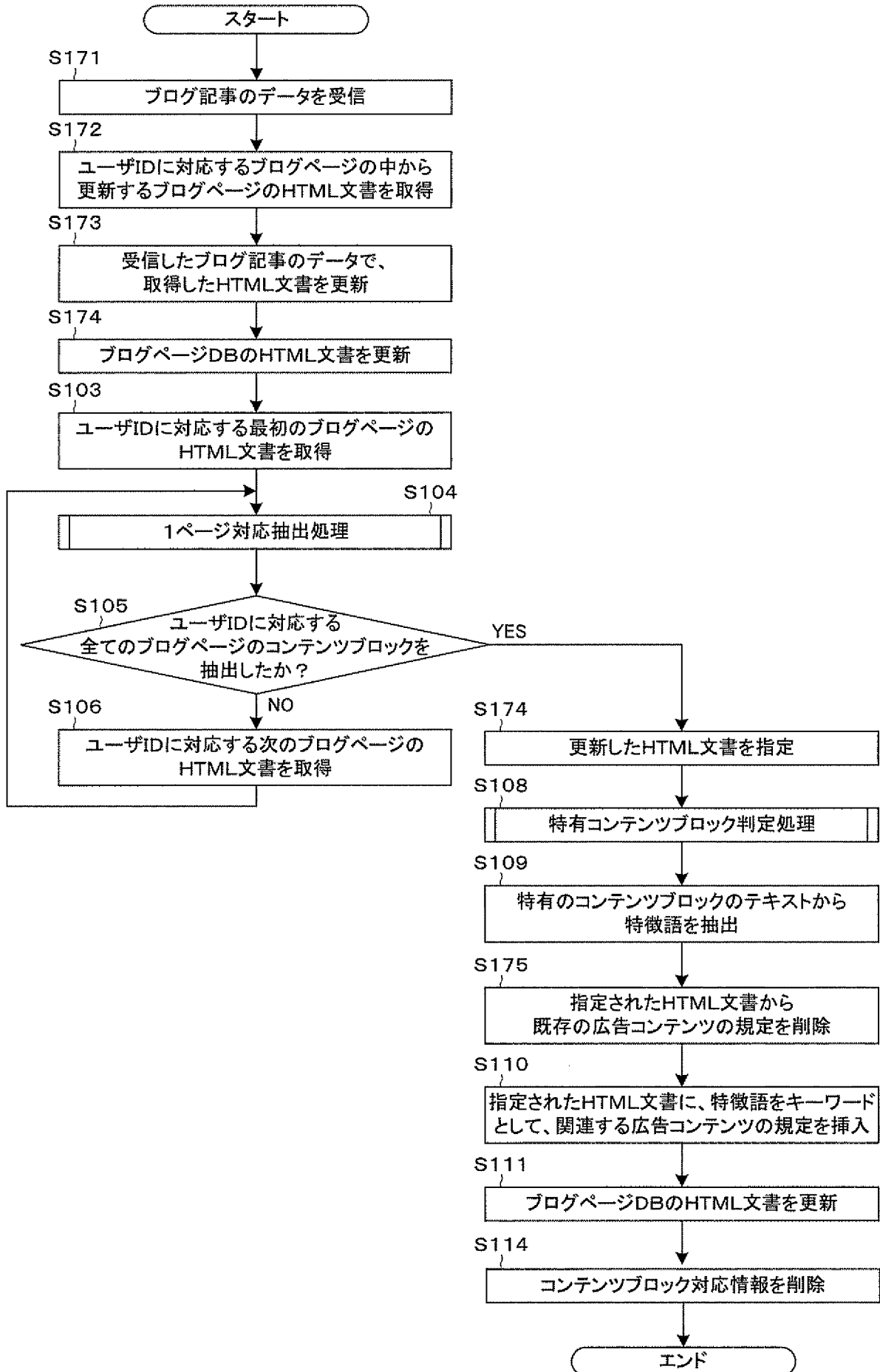
[図20]



[図21]



[図22]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2010/068820

A. CLASSIFICATION OF SUBJECT MATTER

G06Q50/00 (2006.01) i

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06Q50/00

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	1996-2010
Kokai Jitsuyo Shinan Koho	1971-2010	Toroku Jitsuyo Shinan Koho	1994-2010

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.
X	JP 2009-205499 A (NEC Corp.), 10 September 2009 (10.09.2009), entire text; fig. 1 to 20 (Family: none)	1-8, 12, 14-23, 27 9-11, 13, 24-26, 28
Y	JP 2006-146506 A (Image Corp.), 08 June 2006 (08.06.2006), entire text; fig. 1 to 12 (Family: none)	9-11, 24-26
Y	WO 2008/108515 A1 (NR Systems, Inc.), 12 September 2008 (12.09.2008), entire text; fig. 1 to 9 & US 2010/0121711 A & EP 2118826 A & KR 10-0780265 B1 & CN 101689265 A	13, 28

 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
08 December, 2010 (08.12.10)Date of mailing of the international search report
21 December, 2010 (21.12.10)Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2010/068820

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2007-080061 A (University of Tsukuba), 29 March 2007 (29.03.2007), entire text; fig. 1 to 7 (Family: none)	1-28
A	JP 2008-130032 A (Sharp Corp.), 05 June 2008 (05.06.2008), entire text; fig. 1 to 26 (Family: none)	1-28
A	JP 2006-259965 A (Sony Corp.), 28 September 2006 (28.09.2006), entire text; fig. 1 to 7 (Family: none)	8-11, 13, 21, 24-26, 28
A	JP 2003-308461 A (Toyo Kitchen & Living Co., Ltd.), 31 October 2003 (31.10.2003), entire text; fig. 1 to 37 (Family: none)	9-11, 21, 24-26
A	JP 2009-199513 A (NEC Corp.), 03 September 2009 (03.09.2009), entire text; fig. 1 to 6 (Family: none)	13, 28

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int.Cl. G06Q50/00(2006.01)i

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

Int.Cl. G06Q50/00

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2010年
日本国実用新案登録公報	1996-2010年
日本国登録実用新案公報	1994-2010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
X	JP 2009-205499 A (日本電気株式会社) 2009.09.10, 全文, 第1-20 図 (ファミリーなし)	1-8, 12, 14-23 , 27
Y		9-11, 13, 24-2 6, 28
Y	JP 2006-146506 A (株式会社イマージュ) 2006.06.08, 全文, 第1-12 図 (ファミリーなし)	9-11, 24-26

 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、一般的な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

08.12.2010

国際調査報告の発送日

21.12.2010

国際調査機関の名称及びあて先

日本国特許庁 (ISA/J P)
郵便番号100-8915
東京都千代田区霞が関三丁目4番3号

特許庁審査官 (権限のある職員)

小林 正和

5 L

4172

電話番号 03-3581-1101 内線 3562

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
Y	WO 2008/108515 A1 (エヌアールシステムズ インコーポレイテッド) 2008.09.12, 全文, 第1-9 図 & US 2010/0121711 A & EP 2118826 A & KR 10-0780265 B1 & CN 101689265 A	13, 28
A	JP 2007-080061 A (国立大学法人 筑波大学) 2007.03.29, 全文, 第1-7 図 (ファミリーなし)	1-28
A	JP 2008-130032 A (シャープ株式会社) 2008.06.05, 全文, 第1-26 図 (ファミリーなし)	1-28
A	JP 2006-259965 A (ソニー株式会社) 2006.09.28, 全文, 第1-7 図 (フ ァミリーなし)	8-11, 13, 21, 2 4-26, 28
A	JP 2003-308461 A (トーヨーキッチンアンドリビング株式会社) 2003.10.31, 全文, 第1-37 図 (ファミリーなし)	9-11, 21, 24-2 6
A	JP 2009-199513 A (日本電気株式会社) 2009.09.03, 全文, 第1-6 図 (ファミリーなし)	13, 28