

(19) 日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2020-52961

(P2020-52961A)

(43) 公開日 令和2年4月2日 (2020. 4. 2)

(51) Int.Cl.	F I	テーマコード (参考)
G06F 16/00 (2019.01)	G06F 17/30 340A	5B084
G06F 40/274 (2020.01)	G06F 17/27 665	5B091
G06F 40/186 (2020.01)	G06F 17/24 680	5B109
G06F 16/30 (2019.01)	G06F 17/30 170A	
G06F 13/00 (2006.01)	G06F 13/00 540A	
審査請求 未請求 請求項の数 17 O L (全 17 頁)		

(21) 出願番号 特願2018-184591 (P2018-184591)
 (22) 出願日 平成30年9月28日 (2018. 9. 28)

(71) 出願人 000001007
 キヤノン株式会社
 東京都大田区下丸子3丁目30番2号
 (74) 代理人 100125254
 弁理士 別役 重尚
 (72) 発明者 大志万 宗士
 東京都大田区下丸子3丁目30番2号 キ
 ヤノン株式会社内
 Fターム (参考) 5B084 AA01 AA12 AB06 AB31 BB01
 CA11 CC09 CD26 CF12
 5B091 CA01
 5B109 ND04 VA02

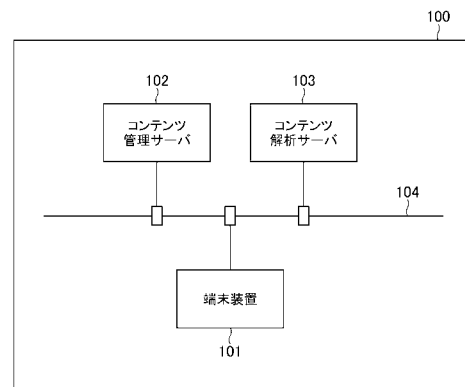
(54) 【発明の名称】 コンテンツ提供システム、コンテンツ提供方法、情報処理装置、及びプログラム

(57) 【要約】 (修正有)

【課題】 編集されている部分データの内容と類似するデータをユーザに提供することができるコンテンツ提供システム、コンテンツ提供方法、情報処理装置、及びプログラムを提供する。

【解決手段】 コンテンツ提供システム100は、コンテンツ管理サーバ102で管理されるドキュメントを構成する複数のページデータのうち、端末装置101でユーザが作業をしている表示ページデータが分類されるクラスに対応付けされたページデータを、端末装置101に提供する。

【選択図】 図1



【特許請求の範囲】**【請求項 1】**

ユーザが操作する情報処理装置に予め登録されたコンテンツを提供するコンテンツ提供システムであって、

前記登録されたコンテンツを構成する複数の部分データを解析する解析手段と、

各前記部分データを予め定められた複数のクラスタの何れかに対応付けして管理する管理手段と、

前記情報処理装置に表示された表示部分データが分類されるクラスタを判別するクラスタ判別手段と、

前記登録されたコンテンツを構成する複数の部分データのうち、前記判別したクラスタに対応付けされた部分データを前記情報処理装置に提供するコンテンツ提供手段とを備えることを特徴とするコンテンツ提供システム。

10

【請求項 2】

前記部分データは、複数ページからなるコンテンツを構成する各ページに対応するデータであることを特徴とする請求項 1 記載のコンテンツ提供システム。

【請求項 3】

前記部分データは、複数の章からなるコンテンツを構成する各章に対応するデータであることを特徴とする請求項 1 又は 2 記載のコンテンツ提供システム。

【請求項 4】

前記部分データは、複数の節からなるコンテンツを構成する各節に対応するデータであることを特徴とする請求項 1 又は 2 記載のコンテンツ提供システム。

20

【請求項 5】

前記部分データは、複数の段落からなるコンテンツを構成する各段落に対応するデータであることを特徴とする請求項 1 又は 2 記載のコンテンツ提供システム。

【請求項 6】

前記登録されたコンテンツを構成する複数の部分データのうち、前記判別したクラスタに対応付けされた部分データのサムネイル画像を前記情報処理装置に送信する画像送信手段を更に備えることを特徴とする請求項 1 乃至 5 のいずれか 1 項に記載のコンテンツ提供システム。

【請求項 7】

30

前記登録されたコンテンツを予め定められた複数のコンテンツクラスタの何れかに対応付けして管理する他の管理手段と、

前記情報処理装置に表示された表示部分データで構成される表示コンテンツが分類されるコンテンツクラスタを判別するコンテンツクラスタ判別手段とを更に備え、

前記判別したクラスタに基づいて選択された前記情報処理装置への提供候補を前記コンテンツクラスタに基づいて絞り込むことを特徴とする請求項 6 記載のコンテンツ提供システム。

【請求項 8】

前記コンテンツ提供システムは、前記登録されたコンテンツを管理するコンテンツ管理サーバと、コンテンツ解析サーバとを備え、

40

前記コンテンツ解析サーバは、前記解析手段、前記管理手段、前記クラスタ判別手段、前記画像送信手段、前記他の管理手段、及び前記コンテンツクラスタ判別手段を備えることを特徴とする請求項 7 記載のコンテンツ提供システム。

【請求項 9】

ユーザが操作する情報処理装置に予め登録されたコンテンツを提供するコンテンツ提供方法であって、

前記登録されたコンテンツを構成する複数の部分データを解析する解析ステップと、

各前記部分データを予め定められた複数のクラスタの何れかに対応付けして管理する管理ステップと、

前記情報処理装置に表示された表示部分データが分類されるクラスタを判別するクラス

50

タ判別ステップと、

前記登録されたコンテンツを構成する複数の部分データのうち、前記判別したクラスタに対応付けされた部分データを前記情報処理装置に提供するコンテンツ提供ステップとを有することを特徴とするコンテンツ提供方法。

【請求項 10】

登録されたコンテンツを管理するコンテンツ管理サーバ及び前記コンテンツを構成する複数の部分データを解析するコンテンツ解析サーバとデータ通信を行う情報処理装置であって、

ユーザによるドキュメントの操作を検出する検出手段と、

前記情報処理装置に表示された表示部分データを示す情報を含むドキュメント関連情報を前記コンテンツ解析サーバに送信する送信手段と、

前記コンテンツを構成する複数の部分データのうち前記ドキュメント関連情報に対応する部分データを示す画像を受信する受信手段と、

前記画像を表示する表示手段と、

前記コンテンツを構成する複数の部分データのうち前記画像に対応する部分データを取得する取得手段とを備えることを特徴とする情報処理装置。

【請求項 11】

前記部分データは、複数のページからなるコンテンツを構成する各ページに対応するデータであることを特徴とする請求項 10 記載の情報処理装置。

【請求項 12】

前記部分データは、複数の章からなるコンテンツを構成する各章に対応するデータであることを特徴とする請求項 10 又は 11 記載の情報処理装置。

【請求項 13】

前記部分データは、複数の節からなるコンテンツを構成する各節に対応するデータであることを特徴とする請求項 10 又は 11 記載の情報処理装置。

【請求項 14】

前記部分データは、複数の段落からなるコンテンツを構成する各段落に対応するデータであることを特徴とする請求項 10 又は 11 記載の情報処理装置。

【請求項 15】

前記送信手段は、前記ドキュメント関連情報を前記コンテンツ解析サーバに送信してから予め設定された所定の時間が経過した際に前記情報処理装置に表示される表示部分データを示す情報を含む別のドキュメント関連情報を前記コンテンツ解析サーバに送信することを特徴とする請求項 10 乃至 14 のいずれか 1 項に記載の情報処理装置。

【請求項 16】

前記送信手段は、前記ユーザによる前記表示部分データの変更操作を検出した際に、変更後の表示部分データを示す情報を含むドキュメント関連情報を前記コンテンツ解析サーバに送信することを特徴とする請求項 10 乃至 15 のいずれか 1 項に記載の情報処理装置。

【請求項 17】

登録されたコンテンツを管理するコンテンツ管理サーバ及び前記コンテンツを構成する複数の部分データを解析するコンテンツ解析サーバとデータ通信を行う情報処理装置にインストールされたアプリケーションを実行するためのプログラムであって、

前記アプリケーションは、

ユーザによるドキュメントの操作を検出する制御を行い、

前記情報処理装置に表示された表示部分データを示す情報を含むドキュメント関連情報を前記コンテンツ解析サーバに送信する制御を行い、

前記コンテンツを構成する複数の部分データのうち前記ドキュメント関連情報に対応する部分データを示す画像を受信する制御を行い、

前記画像を表示する制御を行い、

前記コンテンツを構成する複数の部分データのうち前記画像に対応する部分データを取

10

20

30

40

50

得する制御を行うことを特徴とするプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、コンテンツ提供システム、コンテンツ提供方法、情報処理装置、及びプログラムに関する。

【背景技術】

【0002】

ユーザがOfficeソフト等によってドキュメントを編集している際に、編集の参考となる別のドキュメントを提供するコンテンツ提供システムが知られている。コンテンツ提供システムは、ユーザが入力したドキュメント（以下、「入力ドキュメント」という。）が分類されるクラスタを判別し、データベースに予め登録されたドキュメントの中から、判別したクラスタと類似度が高いドキュメントをユーザに提供する（例えば、特許文献1参照）。これにより、入力ドキュメントと内容が類似するドキュメントをユーザに提供してドキュメントの編集の手助けをすることができる。

10

【先行技術文献】

【非特許文献】

【0003】

【特許文献1】特開2008-158590号公報

20

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかしながら、従来のコンテンツ提供システムでは、分類されるクラスタの判別がドキュメント単位で行われるので、ユーザが編集しているページや章といった部分データの内容と類似するデータをユーザに提供することができない。

【0005】

本発明の目的は、編集されている部分データの内容と類似するデータをユーザに提供することができるコンテンツ提供システム、コンテンツ提供方法、情報処理装置、及びプログラムを提供することにある。

30

【課題を解決するための手段】

【0006】

上記目的を達成するために、本発明のコンテンツ提供システムは、ユーザが操作する情報処理装置に予め登録されたコンテンツを提供するコンテンツ提供システムであって、前記登録されたコンテンツを構成する複数の部分データを解析する解析手段と、各前記部分データを予め定められた複数のクラスタの何れかに対応付けして管理する管理手段と、前記情報処理装置に表示された表示部分データが分類されるクラスタを判別するクラスタ判別手段と、前記登録されたコンテンツを構成する複数の部分データのうち、前記判別したクラスタに対応付けされた部分データを前記情報処理装置に提供するコンテンツ提供手段とを備えることを特徴とする。

40

【発明の効果】

【0007】

本発明によれば、編集されている部分データの内容と類似するデータをユーザに提供することができる。

【図面の簡単な説明】

【0008】

【図1】本発明の実施の形態に係るコンテンツ提供システムの構成を概略的に示すブロック図である。

【図2】図1のコンテンツ解析サーバに設けられる制御装置及び端末装置に設けられる制御装置の各ハードウェア構成を概略的に示すブロック図である。

50

【図 3】図 1 のコンテンツ解析サーバ及び端末装置の各機能構成を示すブロック図である。

【図 4】図 1 の端末装置におけるレコメンド画像の表示について説明するための図である。

【図 5】図 3 のドキュメント解析部によって実行されるクラスタリング処理の手順を示すフローチャートである。

【図 6】図 5 の処理におけるページデータの特徴のベクトル化について説明するための図である。

【図 7】図 1 のコンテンツ解析サーバによって管理される部分データ情報管理テーブルの一例を示す図である。

10

【図 8】図 1 の端末装置によって実行される表示制御処理の手順を示すフローチャートである。

【図 9】図 1 のコンテンツ解析サーバによって実行されるレコメンド画像生成処理の手順を示すフローチャートである。

【図 10】図 9 のステップ S 9 0 3 におけるクラスタの判別を説明するための図である。

【図 11】図 9 のステップ S 9 0 4 におけるレコメンド対象の選択を説明するための図である。

【図 12】図 1 の端末装置におけるレコメンド画像の表示例を示す図である。

【図 13】図 5 のクラスタリング処理の変形例の手順を示すフローチャートである。

【図 14】図 1 のコンテンツ解析サーバによって管理されるドキュメント情報管理テーブルの一例を示す図である。

20

【図 15】図 9 のレコメンド画像生成処理の変形例の手順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】

【0009】

以下、本発明の実施の形態について図面を参照しながら詳述する。

【0010】

図 1 は、本発明の実施の形態に係るコンテンツ提供システム 100 の構成を概略的に示すブロック図である。図 1 において、コンテンツ提供システム 100 は、情報処理装置としての端末装置 101、コンテンツ管理サーバ 102、及びコンテンツ解析サーバ 103 を備える。なお、本実施の形態では、説明を容易にするために、端末装置 101、コンテンツ管理サーバ 102、及びコンテンツ解析サーバ 103 を 1 台ずつ備える構成について説明するが、各装置の台数はこれに限られない。例えば、コンテンツ提供システム 100 は、端末装置 101 やコンテンツ管理サーバ 102 を複数台備えても良い。端末装置 101、コンテンツ管理サーバ 102、及びコンテンツ解析サーバ 103 は、ネットワーク 104 を介してデータ通信可能である。ネットワーク 104 は、インターネット、有線 LAN、無線 LAN、又はこれらの組み合わせである。端末装置 101、コンテンツ管理サーバ 102、及びコンテンツ解析サーバ 103 は、ネットワーク 104 と直接、又は接続機器（不図示）を介して接続されている。接続機器は、例えば、ルータ、ゲートウェイ、及びプロキシサーバである。

30

【0011】

40

端末装置 101 は、ユーザが直接操作する端末である。ユーザは、端末装置 101 を操作して、Office ソフト等によってドキュメントを編集する。コンテンツ管理サーバ 102 は、登録された複数のコンテンツを管理する。コンテンツ管理サーバ 102 は、例えば、複数ページからなるドキュメント、複数の章からなるドキュメント、複数の節からなるドキュメント、及び複数の段落からなるドキュメントのように様々なデータ構造のコンテンツを管理する。コンテンツ解析サーバ 103 は、コンテンツ管理サーバ 102 で管理されたドキュメントや、端末装置 101 から送信されたドキュメントを解析する。コンテンツ提供システム 100 では、コンテンツ管理サーバ 102 で管理されたドキュメントのうち、端末装置 101 でユーザが作業しているドキュメントと類似性が高いドキュメントが端末装置 101 に提供される。以下では、端末装置 101 に提供されるために選択さ

50

れるデータをレコメンドデータと称する。

【0012】

図2は、図1のコンテンツ解析サーバ103に設けられる制御装置200及び端末装置101に設けられる制御装置210の各ハードウェア構成を概略的に示すブロック図である。図2(a)は、コンテンツ解析サーバ103の制御装置200のハードウェア構成を示す。図2(b)は、端末装置101の制御装置210のハードウェア構成を示す。

【0013】

図2(a)において、制御装置200は、CPU201、ROM202、RAM203、記憶装置204、ネットワークI/F205、ディスプレイI/F206、操作入力I/F207、及び外部I/O208を備える。CPU201、ROM202、RAM203、記憶装置204、ネットワークI/F205、ディスプレイI/F206、操作入力I/F207、及び外部I/O208はシステムバス209を介して互いに接続されている。

10

【0014】

制御装置200は、コンテンツ解析サーバ103全体を統括的に制御する。CPU201は、ROM202に格納されたプログラムを実行することで、各種処理を制御する。ROM202は、CPU201が実行するプログラムや設定データを格納する。RAM203は、CPU201の作業領域として、また、各データの一時格納領域として用いられる。記憶装置204は、後述する図3(a)の各モジュールを制御するためのプログラム等を格納する。ネットワークI/F205は、ネットワーク104を介して接続された外部装置、例えば、端末装置101やコンテンツ管理サーバ102とのデータ通信を制御する。ディスプレイI/F206には、液晶ディスプレイ等の外部ディスプレイ（不図示）が接続される。操作入力I/F207には、キーボード、マウス、及びタッチパネル等の操作入力機器（不図示）が接続される。外部I/O208には、USBメモリや外付けの記憶デバイス等が接続される。

20

【0015】

図2(b)において、制御装置210は、CPU211、ROM212、RAM213、記憶装置214、ネットワークI/F215、ディスプレイI/F216、操作入力I/F217、及び外部I/O218を備える。CPU211、ROM212、RAM213、記憶装置214、ネットワークI/F215、ディスプレイI/F216、操作入力I/F217、及び外部I/O218はシステムバス219を介して互いに接続されている。

30

【0016】

制御装置210は、端末装置101全体を統括的に制御する。CPU211は、ROM212に格納されたプログラムを実行することで、各種処理を制御する。ROM212は、CPU211が実行するプログラムや設定データを格納する。RAM213は、CPU211の作業領域として、また、各データの一時格納領域として用いられる。記憶装置214は、後述する図3(b)の各モジュールを制御するためのプログラム等を格納する。ネットワークI/F215は、ネットワーク114を介して接続された外部装置、例えば、コンテンツ管理サーバ102やコンテンツ解析サーバ103とのデータ通信を制御する。ディスプレイI/F216には、液晶ディスプレイ等の外部ディスプレイ（不図示）が接続される。操作入力I/F217には、キーボード、マウス、及びタッチパネル等の操作入力機器（不図示）が接続される。外部I/O218には、USBメモリや外付けの記憶デバイス等が接続される。

40

【0017】

図3は、図1のコンテンツ解析サーバ103及び端末装置101の各機能構成を示すブロック図である。図3(a)は、コンテンツ解析サーバ103の機能構成を示す。図3(b)は、端末装置101の機能構成を示す。

【0018】

図3(a)において、コンテンツ解析サーバ103は、モジュールとして、データ生成

50

部 3 0 1、ドキュメント解析部 3 0 2、制御部 3 0 3、通信部 3 0 4、ドキュメントクラスタ DB 3 0 5、及びページクラスタ DB 3 0 6 を備える。上述した各モジュールの処理は、CPU 2 0 1 が ROM 2 0 2 や記憶装置 2 0 4 に格納されたプログラムを実行することによって行われる。

【0019】

データ生成部 3 0 1 は、レコメンドデータを示す画像を端末装置 1 0 1 に表示させるためのレコメンド表示用データを生成する。レコメンド表示用データは、レコメンドデータのサムネイル画像（以下、「レコメンド画像」という。）、レコメンドデータのページ番号、及びレコメンドデータの格納場所を示すアドレスを含む。ドキュメント解析部 3 0 2 は、ドキュメントの構造を解析する。例えば、ドキュメント解析部 3 0 2 は、コンテンツ管理サーバ 1 0 2 で管理された全てのドキュメントのページ情報を解析する。また、ドキュメント解析部 3 0 2 は、ユーザが端末装置 1 0 1 を用いて編集しているドキュメントの構造を解析する。制御部 3 0 3 は、制御装置 2 0 0 及び当該制御装置 2 0 0 に接続された機器を制御する。また、制御部 3 0 3 は、上述したコンテンツ解析サーバ 1 0 3 の各モジュールの処理の実行を制御する。通信部 3 0 4 は、ネットワーク 1 0 4 に接続された外部装置とのデータ通信を制御する。ドキュメントクラスタ DB 3 0 5 は、後述する図 1 4 のドキュメント情報管理テーブル 1 4 0 0 を管理する。ページクラスタ DB 3 0 6 は、後述する図 7 の部分データ情報管理テーブル 7 0 0 を管理する。

【0020】

図 3（b）において、端末装置 1 0 1 は、モジュールとして、通信部 3 1 1、表示部 3 1 2、操作部 3 1 3、制御部 3 1 4、アプリケーション実行部 3 1 5、操作検出部 3 1 6、及びレコメンド実行部 3 1 7 を備える。上述した端末装置 1 0 1 の各モジュールの処理は、CPU 2 1 1 が ROM 2 1 2 や記憶装置 2 1 4 に格納されたプログラムを実行することによって行われる。

【0021】

通信部 3 1 1 は、ネットワーク 1 0 4 に接続された外部装置とのデータ通信を制御する。例えば、通信部 3 1 1 は、コンテンツ解析サーバ 1 0 3 から後述するレコメンド表示用データを受信する（受信手段）。また、通信部 3 1 1 は、コンテンツ管理サーバ 1 0 2 からレコメンドデータを取得する（取得手段）。表示部 3 1 2 は、端末装置 1 0 1 のディスプレイ（不図示）の表示制御を行う。操作部 3 1 3 は、端末装置 1 0 1 に接続されたキーボード、マウス、及びタッチパネル等の操作入力機器（不図示）で入力された指示を受け付ける。制御部 3 1 4 は、制御装置 2 1 0 及び当該制御装置 2 1 0 に接続された機器を制御する。また、制御部 3 1 4 は、上述した端末装置 1 0 1 の各モジュールの処理について制御する。アプリケーション実行部 3 1 5 は、端末装置 1 0 1 にインストールされたアプリケーションを実行する。操作検出部 3 1 6 は、上記操作入力機器で受け付けた指示や、アプリケーション実行部 3 1 5 によるアプリケーションの実行状況等に基づいてユーザによる端末装置 1 0 1 への操作を検出する。レコメンド実行部 3 1 7 は、後述する図 8 の表示制御処理を実行する。

【0022】

図 4 は、図 1 の端末装置 1 0 1 におけるレコメンド画像の表示について説明するための図である。

【0023】

図 4（a）の画面 4 0 0 は、端末装置 1 0 1 のディスプレイ（不図示）上に表示される画面を模式的に表している。端末装置 1 0 1 では、レコメンドデータを取得するためのレコメンドデータ取得アプリケーションが起動すると、画面 4 0 0 にウィンドウ 4 0 1 が表示される。ウィンドウ 4 0 1 は、端末装置 1 0 1 上で実行され且つドキュメントの表示や編集を実施可能なアプリケーションソフトウェアのウィンドウである。ユーザは、ウィンドウ 4 0 1 を介してドキュメントの閲覧や編集を行う。以下では、閲覧や編集を行うためにウィンドウ 4 0 1 に表示されるドキュメントを表示ドキュメント（表示コンテンツ）と称する。ユーザがドキュメントを開く操作を行うと、画面 4 0 0 は、ウィンドウ 4 0 1 を

10

20

30

40

50

表示する領域402と、レコメンド画像404～407を表示する領域403とに分割される。レコメンド画像404～407は、コンテンツ管理サーバ102で管理されたドキュメントを構成する複数のページに対応するデータ（以下、「ページデータ」という。）のうち、ウィンドウ401に表示されているページデータ（以下、「表示ページデータ」という。）（表示部分データ）と類似性が高いページデータのサムネイル画像である。領域403には、複数のレコメンド画像が表示され、マウス（不図示）等の操作でスクロールすることで領域403に収まらないレコメンド画像を表示することも可能である。

【0024】

図4（b）は、ユーザがマウス等によりレコメンド画像405を選択した状態を表している。選択されたレコメンド画像405の枠線は、例えば、図4（b）に示すように、強調表示される。ウィンドウ408は、ユーザがレコメンド画像405を選択した後、レコメンド画像405に対応するページデータ（レコメンドデータ）を表示するためのウィンドウである。このように、本実施の形態では、ユーザがレコメンド画像を選択することで、選択したレコメンド画像に対応するページデータ（レコメンドデータ）を画面400に表示することができる。ユーザは、表示ページデータの編集の参考資料又は素材資料として、レコメンドデータを使用する。

【0025】

図5は、図3のドキュメント解析部302によって実行されるクラスタリング処理の手順を示すフローチャートである。図5の処理は、CPU201がROM202や記憶装置204に格納されたプログラムを実行することによって行われる。また、図5の処理は、例えば、コンテンツ管理サーバ102に新たなドキュメントが登録された際、若しくは予め設定された所定の時刻に達した際に実行される。

【0026】

図5において、まず、ドキュメント解析部302は、コンテンツ管理サーバ102で管理された全てのドキュメントのページ情報を解析する（ステップS501）。具体的に、ドキュメント解析部302は、各ドキュメントのページ情報をドキュメントの構造情報から取得し、各ページデータのテキストデータを抽出する。また、ドキュメント解析部302は、抽出したテキストデータに基づいて各ページデータの特徴をベクトル化する。本実施の形態では、Doc2Vec等を用いて各ページデータの特徴のベクトル化が行われる。図6は、ベクトル化した各ページデータの特徴を特徴空間にプロットした様子を模式的に表す図である。なお、特徴空間は、N次元（Nは整数）の基底ベクトルを軸として定義される空間であるが、本実施の形態では、説明を容易にするために、特徴量1、2の2次元として説明する。図6において、ベクトル601等の白丸が各ページデータの特徴をベクトル化した特徴ベクトルである。ページデータとドキュメントとの対応関係は、図7の部分データ情報管理テーブル700で管理される。部分データ情報管理テーブル700は、ベクトルID701、ドキュメントID702、ドキュメントアドレス703、ページ番号704、及びクラスタID705で構成される。ベクトルID701には、各特徴ベクトルを識別するための識別子が記録される。ドキュメントID702には、コンテンツ管理サーバ102によって管理された各ドキュメントを識別するための識別子が記録される。ドキュメントアドレス703には、コンテンツ管理サーバ102によって管理された各ドキュメントの格納場所を示すアドレスが記録される。ページ番号704には、ドキュメントのページ番号が記録される。クラスタID705には、ステップS502におけるクラスタリングの結果、具体的に、ページ番号に対応するページデータが対応付けされたクラスタを識別するための識別子が記録される。

【0027】

次いで、ドキュメント解析部302は、ステップS501でベクトル化して得られた各ページデータの特徴ベクトルをクラスタリングする（ステップS502）。クラスタリングには、k-means法、X-means法、最短距離法、及びワード法等の方法が用いられる。図6において、枠602～604がクラスタを表しており、例えば、枠602内の特徴ベクトルは同じクラスタに属する。クラスタリングした結果が、管理テーブル

10

20

30

40

50

701のクラスタID705に記録される。このようにして、本実施の形態では、コンテンツ管理サーバ102で管理されるドキュメントの各ページデータが予め定められた複数のクラスタの何れかに対応付けされる。その後、ドキュメント解析部302は、本処理を終了する。

【0028】

図8は、図1の端末装置101によって実行される表示制御処理の手順を示すフローチャートである。図8の処理は、CPU211がROM212や記憶装置214に格納されたプログラムを実行することによって行われる。

【0029】

図8において、CPU211は、操作検出部316により、ユーザによるドキュメントに対する操作（以下、「ドキュメント操作」という。）を検出したか否かを判別する（ステップS801）。ドキュメント操作は、具体的に、ドキュメントを開く操作である。このような操作に関する情報は、操作部313から制御部314にリアルタイムで通知され、通知を受信した制御部314は、ドキュメント操作が行われた旨を操作検出部316に通知する。この通知に基づいて操作検出部316がドキュメント操作を検出すると（ステップS801でYES）、CPU211は、ドキュメント操作が検出された表示ドキュメントに関連する情報（以下、「ドキュメント関連情報」という。）を通信部311を介してコンテンツ解析サーバ103へ送信する（ステップS802）。ドキュメント関連情報には、表示ドキュメント、及び表示ページデータのページ番号を示す情報が含まれる。ドキュメント関連情報を受信したコンテンツ解析サーバ103は、後述する図9のレコメンド画像生成処理を実行する。レコメンド画像生成処理では、コンテンツ解析サーバ103は、表示ページデータの特徴量と類似性が高いページデータのレコメンド画像を生成し、レコメンド画像を含むレコメンド表示用データを端末装置101に送信する。レコメンド表示用データには、レコメンド画像の他に、レコメンドデータのページ番号、及びレコメンドデータの格納場所を示すアドレスが含まれる。

【0030】

次いで、CPU211は、コンテンツ解析サーバ103からレコメンド表示用データを受信し（ステップS803）、当該レコメンド表示用データに含まれるレコメンド画像を画面400の領域403に表示する（ステップS804）。領域403に表示された当該レコメンド画像がユーザに選択されると、CPU211は、レコメンド表示用データに含まれるアドレスにアクセスして、当該アドレスが示すレコメンドデータを取得する。また、CPU211は、取得したレコメンドデータが表示された新たなウィンドウ、例えば、ウィンドウ408を領域402に表示する。次いで、CPU211は、表示ドキュメントを閉じる操作を検出したか否かを判別する（ステップS805）。

【0031】

ステップS805の判別の結果、表示ドキュメントを閉じる操作を検出しないとき、CPU211は、ステップS802でドキュメント関連情報を送信してから予め設定された所定の時間が経過したか否かを判別する（ステップS806）。所定の時間は、例えば、数分程度である。

【0032】

ステップS806の判別の結果、ステップS802でドキュメント関連情報を送信してから所定の時間が経過しないとき、CPU211は、ステップS805の処理に戻る。ステップS806の判別の結果、ステップS802でドキュメント関連情報を送信してから所定の時間が経過したとき、CPU211は、ステップS802の処理に戻る。すなわち、本実施の形態では、ドキュメント関連情報がコンテンツ解析サーバ103に送信されてから予め設定された所定の時間が経過した際に画面400に表示された表示ページデータに関する情報を含む別のドキュメント関連情報がコンテンツ解析サーバ103に送信される。

【0033】

ステップS805の判別の結果、表示ドキュメントを閉じる操作を検出したとき、CP

10

20

30

40

50

U 2 1 1 は、本処理を終了する。

【0034】

図9は、図1のコンテンツ解析サーバ103によって実行されるレコメンド画像生成処理の手順を示すフローチャートである。図9の処理は、CPU201がROM202や記憶装置204に格納されたプログラムを実行することによって行われる。

【0035】

図9において、CPU201は、ステップS802で端末装置101から送信されたドキュメント関連情報を受信する（ステップS901）。次いで、CPU201は、ドキュメント関連情報を解析する（ステップS902）。具体的に、CPU201は、ドキュメント解析部302により、受信したドキュメント関連情報に含まれるページ番号から特定した表示ページデータのテキストデータを抽出し、抽出したテキストデータに基づいて表示ページデータの特徴をベクトル化する。なお、CPU201は、ステップS501と同様の方法でベクトル化を行う。次いで、CPU201は、部分データ情報管理テーブル700に基づいて表示ページデータが分類されるクラスタを判別する（ステップS903）（クラスタ判別手段）。例えば、表示ページデータの特徴ベクトルが図10のベクトル1001である場合、CPU201は、図10のようにベクトル1001を含むクラスタ1002を表示ページデータが分類されるクラスタと判別する。また、表示ページデータの特徴ベクトルがクラスタ1002～1004の何れにも含まれないベクトル1005である場合、CPU201は、各クラスタ1002～1004の中心までの距離に基づいて表示ページデータが分類されるクラスタを判別する。この場合、CPU201は、クラスタ1002～1004のうち、中心からベクトル1005までの距離が最も短いクラスタ1002を表示ページデータが分類されるクラスタと判別する。

【0036】

次いで、CPU201は、コンテンツ管理サーバ102で管理されるドキュメントを構成する複数のページデータのうち、判別したクラスタに対応付けされたページデータをレコメンド対象として選択する（ステップS904）。ステップS904では、例えば、図11において、判別したクラスタ1101内のベクトル1102～1110に対応する全てのページデータをレコメンド対象として選択する。若しくは、判別したクラスタ1101内のベクトル1102～1110のうち、表示ページデータのベクトル1111を中心とする同心円領域1112内のベクトル1108～1110に対応するページデータをレコメンド対象として選択する。ベクトル1108～1110に対応するページデータは、表示ページデータとの関連性が極めて高いページデータである。

【0037】

次いで、CPU201は、レコメンド対象のサムネイル画像であるレコメンド画像を生成する（ステップS905）。具体的に、CPU201は、データ生成部301により、選択したレコメンド対象のアドレス及びページ番号を部分データ情報管理テーブル700から取得する。CPU201は、データ生成部301により、コンテンツ管理サーバ102で管理されるドキュメントを構成する複数のページデータのうち、取得したアドレスが示すページデータをサムネイル化して、レコメンド画像を生成する。次いで、CPU201は、レコメンド画像、レコメンドデータのページ番号、及びレコメンドデータの格納場所を示すアドレスを含むレコメンド表示用データを端末装置101に送信し（ステップS906）（画像送信手段）、本処理を終了する。

【0038】

上述した実施の形態によれば、コンテンツ管理サーバ102で管理されるドキュメントを構成する複数のページデータのうち、表示ページデータが分類されるクラスタに対応付けされたページデータが端末装置101に提供される。これにより、編集されている表示ページデータの内容と類似するレコメンドデータをユーザに提供することができる。

【0039】

また、上述した実施の形態では、コンテンツ管理サーバ102で管理されるドキュメントを構成する複数のページデータのうち、表示ページデータが分類されるクラスタに対応

付けされたレコメンドデータのサムネイル画像であるレコメンド画像が端末装置 101 に送信される。これにより、表示されたレコメンド画像から編集の参考資料として適したレコメンドデータをユーザに容易に選択させることができる。

【0040】

上述した実施の形態では、端末装置 101 は、コンテンツ管理サーバ 102 で管理されるドキュメントを構成する複数のページデータのうち、表示ページデータに関する情報を含むドキュメント関連情報に対応するページデータのレコメンド画像を表示し（例えば、図 4（a）のレコメンド画像 404～407 を参照）、当該レコメンド画像に対応するページデータ（レコメンドデータ）を取得する。これにより、編集されている表示ページデータの内容と類似するレコメンドデータをユーザに提供することができる。

10

【0041】

また、上述した実施の形態では、ドキュメント関連情報がコンテンツ解析サーバ 103 に送信されてから予め設定された所定の時間が経過した際にウインドウ 401 に表示される表示ページデータを示す情報を含む別のドキュメント関連情報がコンテンツ解析サーバ 103 に送信される。これにより、時間の経過に応じて変更された表示ページデータと類似性が高いレコメンドデータをユーザに提供することができる。

【0042】

以上、本発明について、上述した実施の形態を用いて説明したが、本発明は上述した実施の形態に限定されるものではない。例えば、表示ページデータのベクトルを生成した際に、コンテンツ管理サーバ 102 で管理される全てのドキュメントのページデータのクラスタリングと表示ページデータのクラスタリングとを実施しても良い。

20

【0043】

また、上述した実施の形態では、ステップ S801 で検出されるドキュメント操作は、ドキュメントを開く操作に限られず、例えば、ページを捲る操作や編集する操作といった表示ページデータの変更操作であっても良い。このような操作を検出すると、CPU 211 は、ステップ S805 の判別の結果、表示ドキュメントを閉じる操作を検出しないとき、ステップ S806 の処理を行わず、ステップ S801 の処理に戻る。このように処理することで、端末装置 101 において、表示ページデータの変更操作の検出に応じて、変更後の表示ページデータと類似性が高いページデータをユーザに提供することができる。

【0044】

上述した実施の形態では、ページデータの特徴をベクトル化する処理量を極力減らして処理速度を上げるために、各ページデータのテキストデータに基づいてページデータの特徴をベクトル化したが、これに限られない。例えば、ページデータを構成する少なくとも一部の画像情報に基づいてページデータの特徴をベクトル化しても良い。画像情報を用いる場合、コンテンツ解析サーバ 103 は、画像特徴量を取得することでページデータをベクトル化する。

30

【0045】

また、上述した実施の形態では、クラスタリング及びレコメンドの対象をページ単位として説明したが、ドキュメントを構成する文章の章、節、段落等の文章構造の単位でクラスタリング及びレコメンドを行っても良く、また、ページ及び文章構造の両方を用いてクラスタリング及びレコメンドを行ってもよい。文章構造の単位でクラスタリング及びレコメンドが行われる場合、部分データ情報管理テーブル 700 では、ページ番号 704 の代わりに、文章構造に関する情報が記録される。

40

【0046】

上述した実施の形態では、例えば、複数ページからなる章のデータがレコメンド対象に選択された場合、レコメンド対象が複数ページからなるデータであることを示すレコメンド画像が端末装置 101 に表示されても良い。例えば、図 12（a）に示すように、ページ数分のデータが重なっているような画像 1201 が表示される。図 12（b）に示すように、各ページデータの縮小サムネイル画像が並べて表示される。図 12（c）に示すように、章の先頭ページのサムネイル画像 1203 に重畳してレコメンド対象のページ数を

50

示す画像 1204 が表示される。このような表示により、レコメンド対象が複数ページからなるデータであることをユーザに知らせることができる。

【0047】

上述した実施の形態では、コンテンツ提供システムの構成に限られず、端末装置 101 がコンテンツ解析サーバ 103 の機能を備え、図 5 及び図 9 の各処理を端末装置 101 が実行しても良い。

【0048】

また、上述した実施の形態では、ページ単位でクラスタリングした結果に基づいて選択したレコメンド対象（提供候補）をドキュメント単位でクラスタリングした結果に基づいて絞り込んでも良い。

【0049】

例えば、レコメンド対象の選択にページ単位でクラスタリングした結果を用いると、編集の参考資料として適していないデータ、例えば、表示ドキュメントと関連性が薄いドキュメントのページデータがレコメンド対象として選択されてしまうことがある。

【0050】

これに対応して、本実施の形態では、ページ単位でクラスタリングした結果に基づいて選択したレコメンド対象をドキュメント単位でクラスタリングした結果に基づいて絞り込む。

【0051】

図 13 は、図 5 のクラスタリング処理の変形例の手順を示すフローチャートである。図 13 の処理も、CPU 201 が ROM 202 や記憶装置 204 に格納されたプログラムを実行することによって行われる。また、図 13 の処理も、例えば、コンテンツ管理サーバ 102 に新たなドキュメントが登録された際、若しくは予め設定された所定の時刻に達した際に実行される。

【0052】

図 13 において、ドキュメント解析部 302 は、ステップ S501、S502 の処理を行う。次いで、ドキュメント解析部 302 は、ドキュメントの全体の特徴をベクトル化する。具体的に、ドキュメント解析部 302 は、ドキュメントを構成する全てのテキストデータを取得し、取得した全てのテキストデータに基づいてステップ S502 と同様の方法でベクトル化する。次いで、ドキュメント解析部 302 は、上記ドキュメントのクラスタリングを行う（ステップ S1301）。クラスタリングした結果は、図 14 のドキュメント情報管理テーブル 1400 で管理される。ドキュメント情報管理テーブル 1400 は、ベクトル ID 1401、ドキュメント ID 1402、ドキュメントアドレス 1403、及びクラスタ ID 1404 で構成される。ベクトル ID 1401 には、各特徴ベクトルを識別するための識別子が記録される。ドキュメント ID 1402 は、部分データ情報管理テーブル 700 のドキュメント ID 702 と対応し、コンテンツ管理サーバ 102 によって管理された各ドキュメントを識別するための識別子が記録される。ドキュメントアドレス 1403 には、コンテンツ管理サーバ 102 によって管理された各ドキュメントの格納場所を示すアドレスが記録される。クラスタ ID 1404 には、コンテンツ管理サーバ 102 によって管理された各ドキュメントが対応付けされたコンテンツクラスタを識別するための識別子が記録される。なお、本実施の形態では、コンテンツクラスタには、ステップ S502 で各ページデータが対応付けされるクラスタと区別可能な識別子が割り当てられる。例えば、各ページデータが対応付けされるクラスタには、図 7 に示すように、頭文字を「C」とした通し番号が識別子として割り当てられ、コンテンツクラスタには、図 14 に示すように、頭文字を「CD」とした通し番号が識別子として割り当てられる。

【0053】

図 15 は、図 9 のレコメンド画像生成処理の変形例の手順を示すフローチャートである。図 15 の処理も、CPU 201 が ROM 202 や記憶装置 204 に格納されたプログラムを実行することによって行われる。

【0054】

図15において、CPU201は、ステップS901～S904の処理を行う。次いで、CPU201は、ドキュメント解析部302により、表示ドキュメントが分類されるコンテンツクラスタを判別する（ステップS1501）（コンテンツクラスタ判別手段）。ステップS1501では、ステップS903で表示ページデータに対して行った処理と同様の処理を、表示ドキュメントに対して行う。次いで、CPU201は、ドキュメント解析部302により、ステップS1501で判別した結果に基づいてステップS904で選択したレコメンド対象を絞り込む（ステップS1502）。例えば、ステップS903において表示ページデータが分類されるクラスタがC004と判別された場合、部分データ情報管理テーブル700に基づいて、レコメンド対象として、ベクトルID（ドキュメントID）がP00001（D00001）、P00003（D00002）、及びP00006（D00003）に対応するページデータが選択される。これに対し、ステップS1501において表示ドキュメントが分類されるコンテンツクラスタがCD03と判別された場合、ドキュメント情報管理テーブル1400に基づいてレコメンド対象がP00006（D00003）に対応するページデータに絞り込まれる。なお、ステップS1501で判別されたコンテンツクラスタがドキュメント情報管理テーブル1400に含まれていない場合、例えば、レコメンド対象の絞り込みが行われない。若しくは、レコメンド対象を対応付けされたドキュメント数が最も多いコンテンツクラスタに属するドキュメントに絞り込む。次いで、CPU201は、ステップS905以降の処理を行う。

10

【0055】

上述した実施の形態では、ページ単位でクラスタリングした結果に基づいて選択したレコメンド対象がドキュメント単位でクラスタリングした結果に基づいて絞り込まれる。これにより、編集の参考資料としてより適したレコメンドデータをユーザに提供することができる。

20

【0056】

本発明は、上述の実施の形態の1以上の機能を実現するプログラムをネットワーク又は記憶媒体を介してシステム又は装置に供給し、該システム又は装置のコンピュータにおける1つ以上のプロセッサがプログラムを読み出して実行する処理でも実現可能である。また、本発明は、1以上の機能を実現する回路（例えば、ASIC）によっても実現可能である。

30

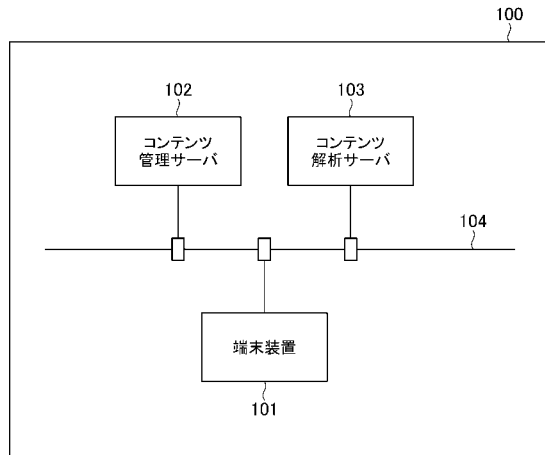
【符号の説明】

【0057】

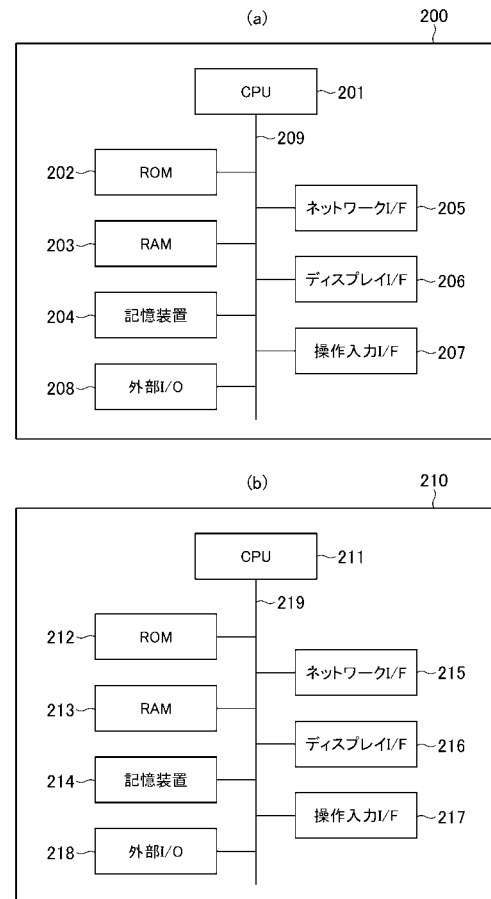
100 コンテンツ提供システム
101 端末装置
102 コンテンツ管理サーバ
103 コンテンツ解析サーバ
200, 210 制御装置
201, 211 CPU
302 ドキュメント解析部
404～407 レコメンド画像
700 部分データ情報管理テーブル
1400 ドキュメント情報管理テーブル

40

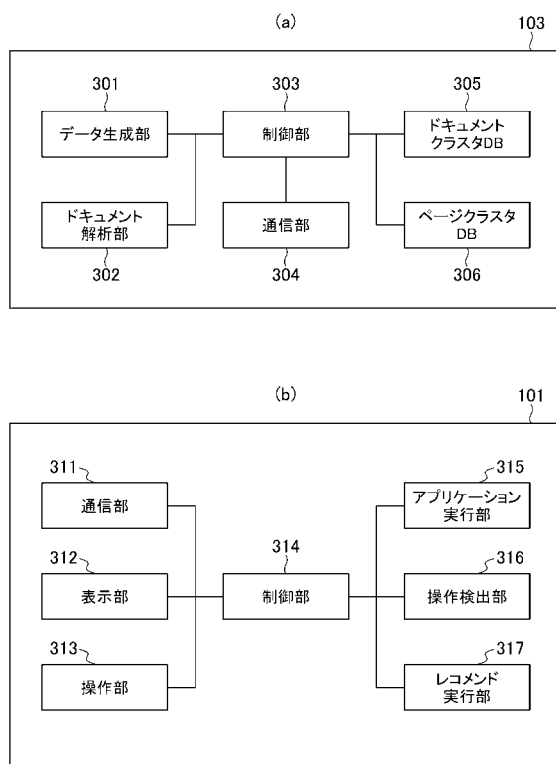
【図 1】



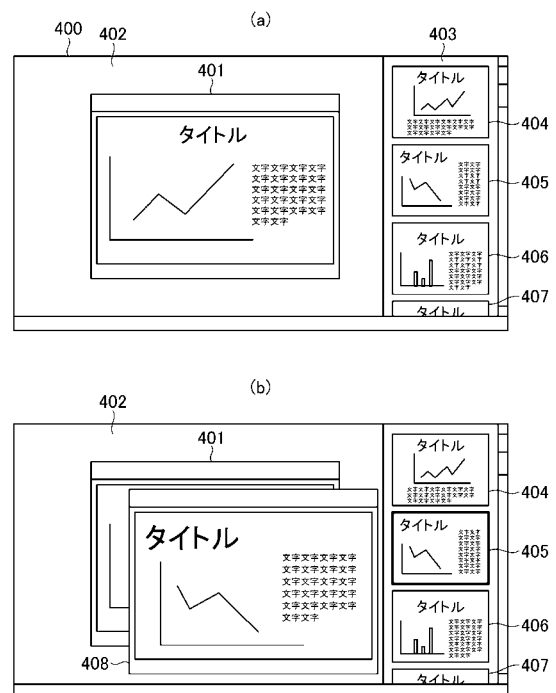
【図 2】



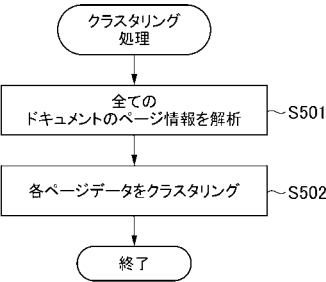
【図 3】



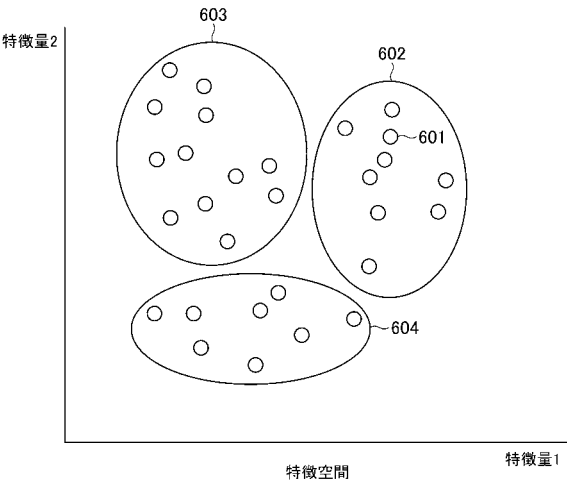
【図 4】



【図 5】



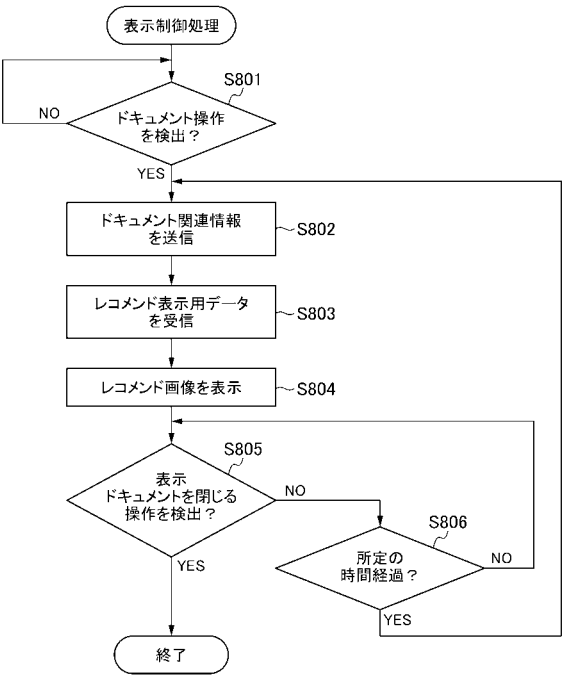
【図 6】



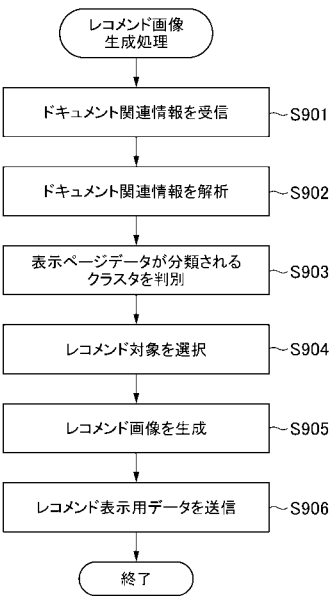
【図 7】

700	701	702	703	704	705
	ベクトルID	ドキュメントID	ドキュメントアドレス	ページ番号	クラスターID
	P000001	D00001	/xxx/yyy/zzz/lll.pptx	1	C004
	P000002	D00001	/xxx/yyy/zzz/lll.pptx	2	C001
	P000003	D00002	/xxx/yyy/zz/ll.pptx	1	C004
	P000004	D00002	/xxx/yyy/zz/ll.pptx	2	C003
	P000005	D00002	/xxx/yyy/zz/ll.pptx	3	C001
	P000006	D00003	/qq/kkkk.pptx	1	C004

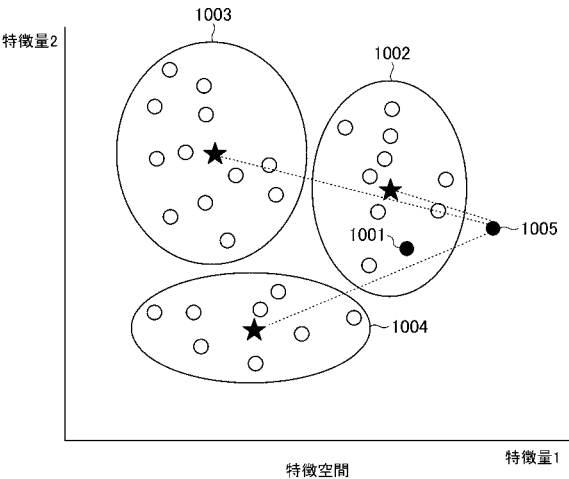
【図 8】



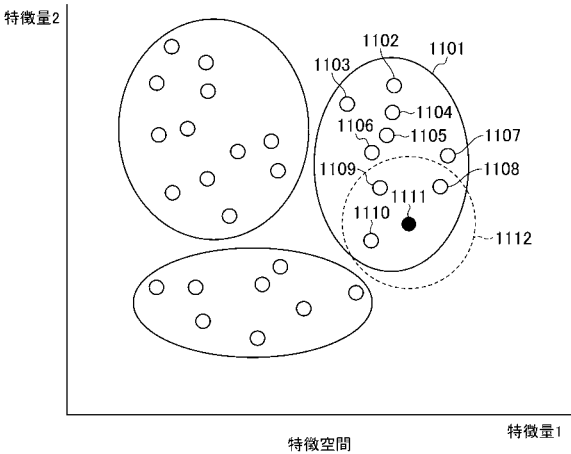
【図 9】



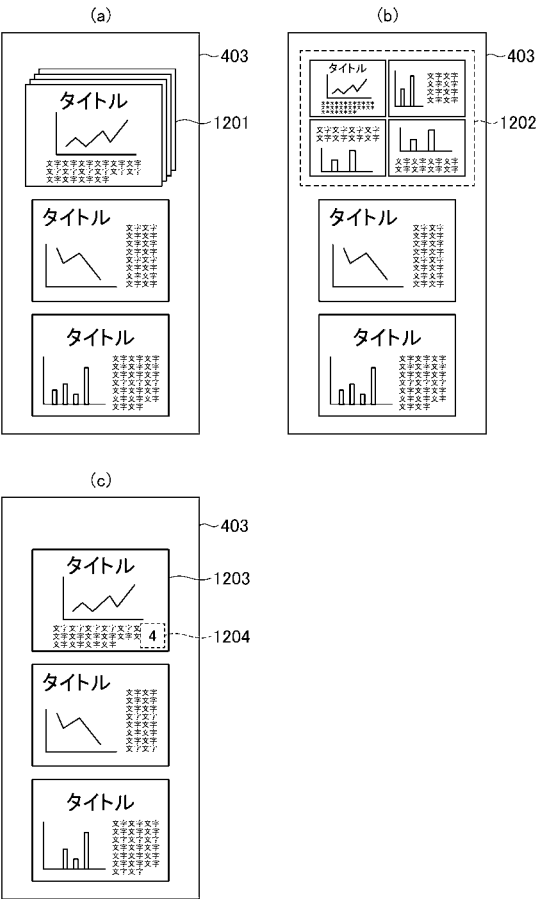
【図 1 0】



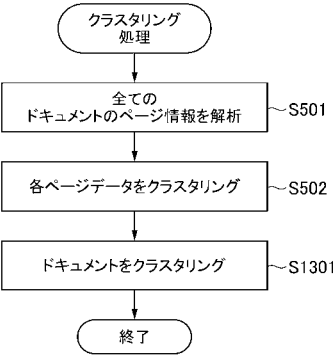
【図 1 1】



【図 1 2】



【図 1 3】



【図 1 4】

ベクトルID	ドキュメントID	ドキュメントアドレス	クラスターID
B00001	D00001	/xxx/yyy/zzz/lll.pptx	CD01
B00002	D00002	/xxx/yyy/zz/ll.pptx	CD01
B00003	D00003	/qqq/kkkkk.pptx	CD03
B00004	D00004	/qqq/kkkkk2.pptx	CD03
B00005	D00005	/xxx/yyy/zz/jjkk.pptx	CD02
B00006	D00006	/xxx/yyy/zzz/xyz.pptx	CD04

⋮

【図 15】

