

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第4848397号
(P4848397)

(45) 発行日 平成23年12月28日 (2011.12.28)

(24) 登録日 平成23年10月21日 (2011.10.21)

(51) Int.Cl.

F I

G 0 6 F 17/30 (2006.01)

G 0 6 F 17/30 3 1 0 Z

G 0 6 F 17/30 3 3 0 C

請求項の数 8 (全 13 頁)

(21) 出願番号	特願2008-163265 (P2008-163265)	(73) 特許権者	500257300
(22) 出願日	平成20年6月23日 (2008.6.23)		ヤフー株式会社
(65) 公開番号	特開2010-3219 (P2010-3219A)		東京都港区赤坂9丁目7番1号
(43) 公開日	平成22年1月7日 (2010.1.7)	(74) 代理人	100106002
審査請求日	平成21年9月15日 (2009.9.15)		弁理士 正林 真之
		(72) 発明者	七里 康敏
			東京都港区六本木六丁目10番1号 ヤフー株式会社内
		(72) 発明者	富田 麻紀
			東京都港区六本木六丁目10番1号 ヤフー株式会社内
		(72) 発明者	前川 英之
			東京都港区六本木六丁目10番1号 ヤフー株式会社内

最終頁に続く

(54) 【発明の名称】 関連クエリ導出装置、関連クエリ導出方法及びプログラム

(57) 【特許請求の範囲】

【請求項1】

コンテンツに係る文章情報から、当該コンテンツに登場するキャラクタ名及び当該コンテンツのタイトル並びに当該コンテンツの属性情報を抽出する抽出手段と、

前記抽出手段により抽出したキャラクタ名及びタイトル並びに属性情報を、当該コンテンツを識別するコンテンツ識別情報に関連付けて記憶する記憶手段と、

文字列をクエリとして検索リクエストを受け付けた場合、当該クエリに前記記憶手段により記憶されたキャラクタ名、タイトル又は属性情報が含まれているか否かを判断する判断手段と、

前記判断手段により、当該クエリにキャラクタ名又はタイトルが含まれているが属性情報の少なくとも一部が含まれていないと判断されたときには、前記記憶手段において当該キャラクタ名又はタイトルに関連付けられている属性情報を、新たなクエリとして導出し、

10

当該クエリに属性情報の少なくとも一部が含まれているがキャラクタ名又はタイトルが含まれていないと判断されたときには、前記記憶手段において当該属性情報に関連付けられているキャラクタ名又はタイトルを、新たなクエリとして導出し、

又は当該クエリに属性情報の一部が含まれているが他の属性情報が含まれていないと判断されたときには、当該他の属性情報を、新たなクエリとして導出する導出手段と、を備える関連クエリ導出装置。

【請求項2】

20

前記コンテンツに含まれている音声情報を抽出する音声情報抽出手段と、
前記音声情報抽出手段により抽出された前記音声情報を文章情報に変換する変換手段と、
を更に備え、

前記抽出手段は、前記変換手段により変換された前記文章情報から前記キャラクタ名、
前記タイトル又は前記属性情報を抽出することを特徴とする請求項1記載の関連クエリ導出装置。

【請求項3】

前記コンテンツの中から、前記音声情報抽出手段により抽出するための時間範囲を特定する特定手段を更に備える請求項2記載の関連クエリ導出装置。

【請求項4】

前記特定手段は、前記コンテンツの中における第1のフレーズに対応する箇所から、第2のフレーズに対応する箇所までを、前記音声情報抽出手段により抽出するための時間範囲として特定することを特徴とする請求項3記載の関連クエリ導出装置。

【請求項5】

前記特定手段は、前記コンテンツと共に取得される特定情報に基づいて、前記音声情報抽出手段により抽出するための時間範囲を特定することを特徴とする請求項3又は請求項4記載の関連クエリ導出装置。

【請求項6】

前記検索リクエストとして受け付けたクエリを、前記導出手段により導出された前記新たなクエリに差し替えて検索を実行する検索手段を更に備える請求項1から請求項5のいずれかに記載の関連クエリ導出装置。

【請求項7】

コンテンツに係る文章情報から、当該コンテンツに登場するキャラクタ名及び当該コンテンツのタイトル並びに当該コンテンツの属性情報を抽出する抽出ステップと、

前記抽出ステップにより抽出したキャラクタ名及びタイトル並びに属性情報を、当該コンテンツを識別するコンテンツ識別情報に関連付けて記憶する記憶ステップと、

文字列をクエリとして検索リクエストを受け付けた場合、当該クエリに前記記憶ステップにより記憶されたキャラクタ名、タイトル又は属性情報が含まれているか否かを判断する判断ステップと、

前記判断ステップにより、当該クエリにキャラクタ名又はタイトルが含まれているが属性情報の少なくとも一部が含まれていないと判断されたときには、前記記憶ステップにおいて当該キャラクタ名又はタイトルに関連付けられている属性情報を、新たなクエリとして導出し、

当該クエリに属性情報の少なくとも一部が含まれているがキャラクタ名又はタイトルが含まれていないと判断されたときには、前記記憶ステップにおいて当該属性情報に関連付けられているキャラクタ名又はタイトルを、新たなクエリとして導出し、

又は当該クエリに属性情報の一部が含まれているが他の属性情報が含まれていないと判断されたときには、当該他の属性情報を、新たなクエリとして導出する導出ステップと、を含むことを特徴とする関連クエリ導出方法。

【請求項8】

請求項7記載の方法をコンピュータに実行させる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、ユーザにより入力されたクエリに基づいて、ユーザが所望する情報を提供する関連クエリ導出装置、関連クエリ導出方法及びプログラムに関する。

【背景技術】

【0002】

近年においては、インターネット上に無数のウェブページが存在しており、目的とするウェブページを検索することは容易ではない。そのため、目的とするウェブページの検索

10

20

30

40

50

を補助するために検索エンジンが提供されている。

【0003】

ユーザは、このような検索エンジンを利用して、様々なキーワードを組み合わせて入力し、目的とするウェブページを検索している。

【0004】

例えば、ユーザは、以前に視聴したテレビ番組の中（特に、ある特定のコーナー）で紹介された情報について、より詳細に知りたいとき等に検索エンジンを利用して所望する情報の検索を行う（例えば、引用文献1を参照。）。
【特許文献1】特開2003-219374号公報

【発明の開示】

【発明が解決しようとする課題】

【0005】

ここで、ユーザが適式なキーワード（例えば、正確なお店の名称等）を入力することができる場合には、検索エンジンを利用することによって目的とする情報を簡易に取得することができるが、このような適式なキーワードを入力することができない場合には、目的とする情報にたどり着くまでに、検索エンジンに入力するキーワードの組み合わせを工夫したり、複数のウェブページを経由する等の多大な労力を払う必要があり、また時間もかかってしまう。また、ユーザが所望するウェブページが存在するのにもかかわらず、このような検索のための労力や時間が原因となって、所望するウェブページへのアクセスをユーザが諦めてしまうこともある。

【0006】

また、以前に目にしたコンテンツ（例えば、小説等）における特定のシーンを覚えているがコンテンツのタイトルが思い出せない場合や、コンテンツのタイトルは覚えているが、その内容（例えば、名言等）を思い出せない場合には、詳しい人に直接尋ねるか、又は検索エンジンを利用して、複数回に及ぶサーチを実行する必要がある。

【0007】

本発明は、予め生成されている情報テーブルを参照して、任意のクエリに基づいて、コンテンツに係るキャラクタ、タイトル又は属性情報を導出することができる関連クエリ導出装置、関連クエリ導出方法及びプログラム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】

（1）コンテンツに係る文章情報から、当該コンテンツに登場するキャラクタ名及び当該コンテンツのタイトル並びに当該コンテンツの属性情報を抽出する抽出手段と、

前記抽出手段により抽出したキャラクタ名及びタイトル並びに属性情報を、当該コンテンツを識別するコンテンツ識別情報に関連付けて記憶する記憶手段と、

文字列をクエリとして検索リクエストを受け付けた場合、当該クエリに前記記憶手段により記憶されたキャラクタ名、タイトル又は属性情報が含まれているか否かを判断する判断手段と、

前記判断手段により、当該クエリにキャラクタ名又はタイトルが含まれているが属性情報の少なくとも一部が含まれていないと判断されたときには、前記記憶手段において当該キャラクタ名又はタイトルに関連付けられている属性情報を、新たなクエリとして導出し、

当該クエリに属性情報の少なくとも一部が含まれているがキャラクタ名又はタイトルが含まれていないと判断されたときには、前記記憶手段において当該属性情報に関連付けられているキャラクタ名又はタイトルを、新たなクエリとして導出し、

又は当該クエリに属性情報の一部が含まれているが他の属性情報が含まれていないと判断されたときには、当該他の属性情報を、新たなクエリとして導出する導出手段と、を備える関連クエリ導出装置。

【0009】

（1）に係る発明によれば、クエリにコンテンツのキャラクタ名とタイトルを明示的又

10

20

30

40

50

は暗示的に示す語句が含まれている場合には、その属性情報を導出し、クエリにコンテンツの属性情報を明示的又は暗示的に示す語句が含まれている場合には、そのキャラクタ名とタイトルを導出し、クエリにコンテンツの属性情報の一部を明示的又は暗示的に示す語句が含まれている場合には、その他の属性情報を導出する。

【0010】

このようにして、(1)に係る発明によれば、コンテンツに関する任意のクエリに基づいて、当該コンテンツに係るキャラクタ名、タイトル又は属性情報の不足を補って新たなクエリとして導出することができる。その結果、ユーザは、コンテンツに関する未知の情報や忘れてしまった情報等を容易に取得できる。更に、これらの情報(新たなクエリ)に基づいて検索を実行することにより、コンテンツに関する有用な情報を簡易に収集することができる。

10

【0011】

(2)前記コンテンツに含まれている音声情報を抽出する音声情報抽出手段と、前記音声情報抽出手段により抽出された前記音声情報を文章情報に変換する変換手段と、を更に備え、

前記抽出手段は、前記変換手段により変換された前記文章情報から前記キャラクタ名、前記タイトル又は前記属性情報を抽出することを特徴とする(1)記載の関連クエリ導出装置。

【0012】

(2)に係る発明によれば、コンテンツに含まれている音声情報を文章情報に変換し、変換された文章情報からキャラクタ名、タイトル又は属性情報を抽出する。

20

【0013】

このようにして、(2)に係る発明によれば、音声情報からキャラクタ名、タイトル又は属性情報を抽出することができるので、予め文章情報を用意することなく、音声として放送されたコンテンツに関する情報を蓄積することができる。

【0014】

(3)前記コンテンツの中から、前記音声情報抽出手段により抽出するための時間範囲を特定する特定手段を更に備える(2)記載の関連クエリ導出装置。

【0015】

(3)に係る発明によれば、コンテンツに含まれる音声のうち、特定の時間範囲を抽出するので、不必要な情報が蓄積されることを抑制し、有用な情報に絞り込める可能性がある。

30

【0016】

(4)前記特定手段は、前記コンテンツの中における第1のフレーズに対応する箇所から、第2のフレーズに対応する箇所までを、前記音声情報抽出手段により抽出するための時間範囲として特定することを特徴とする(3)記載の関連クエリ導出装置。

【0017】

(4)に係る発明によれば、時間的に連続するコンテンツの音声情報の中における第1のフレーズから第2のフレーズの間を時間範囲として特定する。

【0018】

このようにして、(4)に係る発明によれば、例えば、以前に視聴したコンテンツ(テレビ等)のうち、所定のフレーズにより特定される特定の場面における情報について、ユーザにより適式なキーワードによる検索が行われなかった場合でも簡易に提供することが可能となる。

40

【0019】

(5)前記特定手段は、前記コンテンツと共に取得される特定情報に基づいて、前記音声情報抽出手段により抽出するための時間範囲を特定することを特徴とする(3)又は(4)記載の関連クエリ導出装置。

【0020】

(5)に係る発明によれば、コンテンツの時間的に連続する音声のうち、特定情報(例

50

えば、CMの挿入を示す情報)に基づいて、時間範囲を特定する。

【0021】

このようにして、(5)に係る発明によれば、例えば、以前に視聴したコンテンツ(テレビ等)において、CM等の無関係な部分を排除して属性情報等を取得することができる。したがって、ユーザの所望するコンテンツに関連する情報のみを簡易に提供することが可能となる。

【0022】

(6)前記検索リクエストとして受け付けたクエリを、前記導出手段により導出された前記新たなクエリに差し替えて検索を実行する検索手段を更に備える(1)から(5)のいずれかに記載の関連クエリ導出装置。

10

【0023】

(6)に係る発明によれば、ユーザにより適式なキーワードによる検索が行われなかった場合でも、一定条件下、コンテンツに関する適式なキーワードに差し替えて検索を実行するため、ユーザが所望する情報を簡易に提供することができる。

【0024】

(7)コンテンツに係る文章情報から、当該コンテンツに登場するキャラクタ名及び当該コンテンツのタイトル並びに当該コンテンツの属性情報を抽出する抽出ステップと、

前記抽出ステップにより抽出したキャラクタ名及びタイトル並びに属性情報を、当該コンテンツを識別するコンテンツ識別情報に関連付けて記憶する記憶ステップと、

20

文字列をクエリとして検索リクエストを受け付けた場合、当該クエリに前記記憶ステップにより記憶されたキャラクタ名、タイトル又は属性情報が含まれているか否かを判断する判断ステップと、

前記判断ステップにより、当該クエリにキャラクタ名又はタイトルが含まれているが属性情報の少なくとも一部が含まれていないと判断されたときには、前記記憶ステップにおいて当該キャラクタ名又はタイトルに関連付けられている属性情報を、新たなクエリとして導出し、

当該クエリに属性情報の少なくとも一部が含まれているがキャラクタ名又はタイトルが含まれていないと判断されたときには、前記記憶ステップにおいて当該属性情報に関連付けられているキャラクタ名又はタイトルを、新たなクエリとして導出し、

又は当該クエリに属性情報の一部が含まれているが他の属性情報が含まれていないと判断されたときには、当該他の属性情報を、新たなクエリとして導出する導出ステップと、を含むことを特徴とする関連クエリ導出方法。

30

【0025】

(7)に係る発明によれば、当該方法を実行することにより、(1)と同様の効果が期待できる。

【0026】

(8)(7)記載の方法をコンピュータに実行させることを特徴とするプログラム。

【0027】

このような構成によれば、当該プログラムをコンピュータに実行させることにより、(1)と同様の効果が期待できる。

40

【発明の効果】

【0028】

本発明によれば、任意のクエリに基づいて、コンテンツに係るキャラクタ名、タイトル又は属性情報を導出することができる。

【発明を実施するための最良の形態】

【0029】

以下、本発明の実施形態について図を参照しながら説明する。

【0030】

[システム構成]

図1には、本実施形態に係るサーバ10と、放送局30と、ユーザ端末40とから構成

50

される情報処理システム１を示す。なお、図１においては、情報処理システム１は、サーバ１０と、放送局３０と、ユーザ端末４０とがそれぞれ１つずつで示されているが、これに限られず、それぞれ複数台で構成されていてもよい。

【００３１】

サーバ１０は、図２に示すように、制御部３００を構成するＣＰＵ（Ｃｅｎｔｒａｌ Ｐｒｏｃｅｓｓｉｎｇ Ｕｎｉｔ）３１０（マルチプロセッサ構成ではＣＰＵ３２０等複数のＣＰＵが追加されてもよい）、バスライン２００、通信Ｉ／Ｆ（Ｉ／Ｆ：インタフェース）３３０、メインメモリ３４０、ＢＩＯＳ（Ｂａｓｉｃ Ｉｎｐｕｔ Ｏｕｔｐｕｔ Ｓｙｓｔｅｍ）３５０、Ｉ／Ｏコントローラ３６０、ハードディスク３７０、光ディスクドライブ３８０及び半導体メモリ３９０を備える。なお、ハードディスク３７０、光ディスクドライブ３８０及び半導体メモリ３９０はまとめて記憶装置４１０と呼ばれる。

10

【００３２】

制御部３００は、サーバ１０を統括的に制御する部分であり、ハードディスク３７０に記憶された各種プログラムを適宜読み出して実行することにより、上述したハードウェアと協働し、本発明に係る各種機能を実現している。

【００３３】

通信Ｉ／Ｆ３３０は、サーバ１０が、ネットワークを介してユーザ端末４０等の他の装置と情報を送受信する場合のネットワーク・アダプタである。

【００３４】

ＢＩＯＳ３５０は、サーバ１０の起動時にＣＰＵ３１０が実行するブートプログラムや、サーバ１０のハードウェアに依存するプログラム等を記録する。

20

【００３５】

Ｉ／Ｏコントローラ３６０には、ハードディスク３７０、光ディスクドライブ３８０及び半導体メモリ３９０等の記憶装置４１０を接続することができる。

【００３６】

ハードディスク３７０は、本ハードウェアをサーバ１０として機能させるための各種プログラム、本発明の機能を実行するプログラム及び後述するテーブル等を記憶する。なお、サーバ１０は、外部に別途設けたハードディスク（図示せず）を外部記憶装置として利用することもできる。

【００３７】

光ディスクドライブ３８０としては、例えば、ＤＶＤ－ＲＯＭドライブ、ＣＤ－ＲＯＭドライブ、ＤＶＤ－ＲＡＭドライブ及びＣＤ－ＲＡＭドライブを使用することができる。この場合は各ドライブに対応した光ディスク４００を使用する。光ディスク４００から光ディスクドライブ３８０によりプログラム又はデータを読み取り、Ｉ／Ｏコントローラ３６０を介してメインメモリ３４０又はハードディスク３７０に提供することもできる。

30

【００３８】

なお、本発明でいうコンピュータとは、記憶装置、制御部等を備えた情報処理装置をいい、サーバ１０は、記憶装置４１０、制御部３００等を備えた情報処理装置により構成され、この情報処理装置は、本発明のコンピュータの概念に含まれる。

【００３９】

また、本発明に係るサーバ１０は、上述のような構成を有することにより、放送局３０により放送された番組（コンテンツ）の中において発せられた音声情報（例えば、会話等）から、頻出するフレーズや、音量の大きな部分の語句等、特徴的な語句を抽出する。そして、サーバ１０は、ユーザ端末４０から番組に関する検索リクエストが送信されてきたときに、先に抽出した語句を新たな検索式（クエリ）として提供し、更に、この新たなクエリに従って、関連するウェブページを検索する機能を有する。

40

【００４０】

〔機能構成〕

ここで、当該機能を発揮するための構成について、図３に示す機能ブロック図を用いて説明する。サーバ１０は、特定部１１（特定手段）と、音声情報抽出部１２（音声情報抽

50

出手段)と、変換部13(変換手段)と、特徴語抽出部14(抽出手段)と、記憶部15(記憶手段)と、判断部16(判断手段)と、導出部17(導出手段)と、検索部18(検索手段)と、コンテンツログ情報テーブルDB20と、を備える。

【0041】

特定部11は、時間的に連続するコンテンツの音声情報(例えば、テレビ放送を受信して得られる音声情報)に基づいて、抽出するための時間範囲を特定する。ここで、特定部11による時間範囲の特定方法について以下に例示する。

【0042】

特定部11は、音声を認識する機能により構成されており、例えば、時間的に連続する音声情報の中における第1のフレーズに対応する箇所から、第2のフレーズに対応する箇所までを、音声情報抽出部12により抽出するための時間範囲として特定する。具体的には、特定部11は、音声情報の中における第1のフレーズ(テレビ放送を例にとると、出演者による「次のコーナーは・・・」等の特定のコーナー(特集)を開始する旨の宣言)に対応する箇所から、第2のフレーズ(テレビ放送を例にとると、出演者による「以上・・・コーナーでした。」等の特定のコーナーの終了の宣言)に対応する箇所までを、音声を認識する機能により特定する。

10

【0043】

また、特定部11は、例えば、時間的に連続する音声情報と共に取得されるコンテンツ中の時間位置に関する特定情報(例えば、テレビ放送波に含まれているCMの挿入を示す信号)に基づいて、音声情報抽出部12により抽出するための時間範囲を特定する。具体的には、特定部11は、あるCMが終了して番組に移行するときから、当該番組が終了してCMに移行するときまでを、CMの挿入を示す信号に基づいて特定する。

20

【0044】

なお、テレビ放送の番組としては、ドラマや、映画や、バラエティや、ドキュメント等複数のものが考えられる。

【0045】

音声情報抽出部12は、特定部11により特定された時間範囲における音声情報を抽出する。

【0046】

変換部13は、音声を認識し、当該認識に基づいて文字情報(テキスト)に変換する機能(音声認識エンジン)により構成されている。なお、変換部13は、いわゆる音響モデルや言語モデル等を利用して音声認識を行う。

30

【0047】

なお、音声を文字情報に変換する際には、音量や、声質等に基づくキャラクタを識別する情報や、音楽(BGM等)の有無等を、変換された文字情報と関連付けて抽出することが好ましい。これらは、コンテンツの特徴部分や、複数の属性情報の関連性(例えば、同一キャラクタの発言)等を判断する材料となる。

【0048】

ここで、上述の特定部11は、放送局30から受信する音声情報や特定情報に基づいて時間範囲を特定することとしたが、これには限られず、変換部13により変換された文字情報に基づいて範囲を特定してもよい。すなわち、文字情報を解析することにより、例えば、所定のフレーズを目印にして、特定の範囲の文字情報を抽出することができる。

40

【0049】

特徴語抽出部14は、変換部13により変換された文字情報に基づいて当該コンテンツの特徴語を抽出する。ここで、特徴語抽出部14は、例えば、tf-idf(Term Frequency - Inverse Document Frequency、出現頻度に基づいて文章中の特徴的な語句を抽出するためのアルゴリズム)を利用して文字情報から複数の語句(例えば、頻出する施設名や、フレーズ等)を抽出する。

【0050】

また、例えば、音量の大きなフレーズや、BGMの始まる前後のフレーズ等、出現頻度

50

に関わらず特徴的な語句を抽出してもよい。

【 0 0 5 1 】

特徴語抽出部 1 4 は、このようにして抽出した語句を、文章の係り受け解析や、変換部 1 3 により識別されたキャラクタ識別情報等に基づいて、キャラクタに対して関連付けを行う。

【 0 0 5 2 】

記憶部 1 5 は、特徴語抽出部 1 4 により抽出した特徴語を、コンテンツを識別するコンテンツ識別情報 (コンテンツ I D) に関連付けて、属性情報として、コンテンツログ情報テーブル D B 2 0 に記憶する。

【 0 0 5 3 】

具体的には、コンテンツログ情報テーブル D B 2 0 に、図 4 に示すようなコンテンツ情報テーブルが格納される。コンテンツ情報テーブルでは、コンテンツごとにコンテンツ I D が付されており、コンテンツのタイトル及び登場するキャラクタ名並びに各種属性情報が関連付けて記憶される。

【 0 0 5 4 】

ここで、特徴語抽出部 1 4 によりキャラクタと関連付けられた属性情報については、それぞれキャラクタ名ごとのレコードとして記憶され、キャラクタと関連付けられなかった属性情報は、更に別のレコードとして記憶される。なお、コンテンツ情報テーブルの形態はこれには限られず、例えば、適宜複数のテーブルに分割されてもよい。

【 0 0 5 5 】

また、コンテンツ情報テーブルに記憶される情報は、特徴語抽出部 1 4 により抽出された語句には限られない。サーバ 1 0 は、放送される音声情報とは別に、番組表や解説文、あるいはブログ等のウェブページ等、様々な文章情報を受け付けることとしてよい。この場合、サーバ 1 0 は、受け付けた文章情報に基づいて、特徴語抽出部 1 4 によりコンテンツ情報テーブルを生成する。その後、音声情報や、更に別の文章情報を受信することにより、コンテンツ情報テーブルのデータを追加、更新する。

【 0 0 5 6 】

判断部 1 6 は、ユーザ端末 4 0 から、任意の文字列をクエリとして検索リクエストを受け付けた場合、コンテンツ情報テーブル (図 4) を参照し、このクエリにタイトル又はキャラクタ名が含まれているか否か、更に、このクエリに属性情報が含まれているか否かを判断する。この判断の基準は、正確に語句が一致する場合の他、所定以上の類似度である場合、類義語を定義した辞書に基づいて類似と判断される場合等も、含むこととする。

【 0 0 5 7 】

導出部 1 7 は、判断部 1 6 により、検索リクエストのクエリにキャラクタ名又はタイトルが含まれているが属性情報の少なくとも一部が含まれていないと判断されたときには、このキャラクタ名又はタイトルに関連付けられている属性情報を、新たなクエリとして導出する。

【 0 0 5 8 】

また、導出部 1 7 は、検索リクエストのクエリに属性情報の少なくとも一部が含まれているがキャラクタ名又はタイトルが含まれていないと判断されたときには、この属性情報に関連付けられているキャラクタ名又はタイトルを、新たなクエリとして導出する。

【 0 0 5 9 】

更に、導出部 1 7 は、検索リクエストのクエリに属性情報の一部が含まれているが他の属性情報が含まれていないと判断されたときには、検索リクエストのクエリに含まれない他の属性情報を、新たなクエリとして導出する。

【 0 0 6 0 】

このようにして、導出部 1 7 は、ユーザ端末 4 0 から受信したクエリに対して、関連する新たなクエリを導出する。その結果、タイトルやキャラクタ名からコンテンツの内容に関する詳細キーワードや、コンテンツの内容からタイトルやキャラクタ名をユーザ端末 4 0 に対して送信することができる。

10

20

30

40

50

【 0 0 6 1 】

検索部 18 は、導出部 17 により導出された新たなクエリに基づいて、検索を実行する。すなわち、ユーザ端末 40 から受信したクエリを、関連する新たなクエリに差し替えてウェブページを検索する。なお、この検索は、送信した新たなクエリが選択されることにより、ユーザ端末 40 から明示的な要求指示を受け付けたことに応じて実行されることとしてよい。

【 0 0 6 2 】

このような構成によれば、本実施形態は、例えば、ユーザが以前に目にしたコンテンツ（例えば、小説や映画等）における、特定のシーンに基づいて検索リクエストを受け付けた場合に、タイトルやキャラクタ名を新たなクエリとして情報提供し、更にこのタイトルやキャラクタ名から、詳細情報を検索することができる。また、コンテンツのタイトルに基づいて検索リクエストを受け付けた場合には、このコンテンツに関する属性情報（例えば、名言、地名、商品等）を提供し、更にこの属性情報から、詳細情報を検索することができる。このように、サーバ 10 が関連情報を提供することにより、ユーザは、所望の情報に簡易にアクセスすることができる。

10

【 0 0 6 3 】

〔 処理手順 〕

ここで、本発明を適用した場合において実現され得る具体的な処理手順について、図 5 に示すフローチャートを参照して説明する。なお、以下に示す処理手順は、一例であってこれ以外にも実現され得る処理手順は無数に存在する。

20

【 0 0 6 4 】

なお、本発明に係るサーバ 10 は、予め、単一の又は複数の放送局において放送された番組（ジャンルを問わず全ての番組を対象にしてもよいし、特定のジャンル（映画等）の番組のみを対象にしてもよい。）の音声情報（例えば、会話による話し言葉）と、その他に受け付けた文章情報とに基づいて作成された文字情報を分析し、この分析結果に基づいてコンテンツに関する特徴語を、コンテンツのタイトル及びキャラクタ名に関連付けて、コンテンツログ情報テーブル DB 20 に格納されているコンテンツ情報テーブルに登録しておく（図 4 を参照。）。

【 0 0 6 5 】

ユーザが、過去に視聴した番組に関して、詳細な情報、あるいは忘れてしまった情報を取得したいと考えた場合、あいまいな記憶に基づいて、ウェブページの検索を試みる。

30

【 0 0 6 6 】

ステップ S 1 において、サーバ 10 は、ユーザ端末 40 から、ネットワークを介して検索キーワードを受信する。ここで、ユーザ端末 40 を利用するユーザは、検索キーワードをサーバ 10 に送信する際に、例えば、映画のタイトルを覚えていない場合、特定のシーンや映画の種類等に関する情報から検索キーワードを指定する。具体的には、例えば、図 6 に示すように、検索フィールド X に「アニメ 学園祭の前日 夢 亀」を入力して、検索を試みる。

【 0 0 6 7 】

ステップ S 2 において、サーバ 10 は、ステップ S 1 にて受信した検索キーワードが、コンテンツ情報テーブル（図 4）に記憶された、いずれかのコンテンツに関するデータ内に存在するか否かを判定する。この判定が YES の場合はステップ S 3 に移り、判定が NO の場合はステップ S 4 に移る。なお、ステップ S 1 において、検索キーワード「アニメ 学園祭の前日 夢 亀」を受け付けた場合は、コンテンツ情報テーブル（図 4）との一致度に基づいて、コンテンツ ID「0001」が特定されてステップ S 3 に移る。

40

【 0 0 6 8 】

ステップ S 3 において、サーバ 10 は、ステップ S 2 にて特定されたコンテンツに関する情報を、コンテンツ情報テーブル（図 4）から抽出し、新たなキーワードとしてユーザ端末 40 に送信する。具体的には、ステップ S 2 において、検索キーワードがコンテンツの属性情報であると判断された場合には、タイトルとキャラクタ名や、その他の属性情報

50

が抽出される。また、検索キーワードがタイトルやキャラクタ名であると判断された場合には、コンテンツの属性情報、特にキャラクタに関する属性情報が抽出される。

【 0 0 6 9 】

これにより、ユーザ端末 4 0 では、ユーザにより入力された検索キーワードとは別に、新たなキーワードがリスト Z として表示される。ユーザは、このリスト Z からいずれかを選択することにより、検索キーワードを差し替える。そして、ユーザが検索ボタン Y を押下することにより、ユーザ端末 4 0 はサーバ 1 0 に対して検索リクエストを送信する。

【 0 0 7 0 】

ステップ S 4 において、サーバ 1 0 は、ユーザ端末 4 0 から、検索キーワードの指定と共に検索要求（検索リクエスト）があったか否かを判定する。この判定が Y E S の場合は

10

ステップ S 5 に移り、判定が N O の場合は処理を終了する。

【 0 0 7 1 】

ステップ S 5 において、サーバ 1 0 は、ユーザ端末 4 0 から検索リクエストを受信したので、指定されたキーワード（ステップ S 1 にて受信した検索キーワード、もしくは、ステップ S 3 にて送信した関連キーワード）に基づいて、ウェブページの検索を実行する。

【 0 0 7 2 】

以上、本発明の実施形態について説明したが、本発明は上述した実施形態に限るものではない。また、本発明の実施形態に記載された効果は、本発明から生じる最も好適な効果を列挙したに過ぎず、本発明による効果は、本発明の実施形態に記載されたものに限定されるものではない。

20

【 0 0 7 3 】

例えば、本実施形態では、サーバ 1 0 により抽出されたキーワードを、検索実行の前に、クエリ候補としてユーザ端末 4 0 にてリスト表示したが、これには限られない。例えば、サーバ 1 0 は、ユーザ端末 4 0 にて検索ボタン Y が押下され検索が実行されたことに応じて、クエリ候補としての新たなキーワードを送信して表示させてもよい。

【 0 0 7 4 】

また、サーバ 1 0 は、初期に入力された検索キーワードによる検索リクエストを受信したことに応じて、抽出した新たなキーワードに基づく検索結果（ウェブページへのリンク等）を、ユーザ端末 4 0 に送信し表示させてもよい。

【 0 0 7 5 】

30

また、図 6 における検索フィールド X は、タイトル、キャラクタ名及び属性情報それぞれに専用の複数フィールドに分割してもよい。これにより、サーバ 1 0 は、コンテンツ情報テーブル（図 4 ）とのマッチング精度を向上できる可能性がある。

【 0 0 7 6 】

また、本実施形態では、コンテンツとして、放送局 3 0 から放送される番組について説明したが、これには限られない。例えば、音声情報を含まない小説や漫画等であってもよく、予め、これらのコンテンツに関する文章情報を受け付けることにより、上述と同様の効果が期待できる。

【 0 0 7 7 】

なお、本実施形態においては、サーバ 1 0 は、ハードディスク 3 7 0 及び光ディスクドライブ 3 8 0 を有する構成として説明したが、これに限られず、これらの駆動系を有さない構成、いわゆるゼロスピンドルによる構成であってもよい。このような構成の場合には、ハードディスク 3 7 0 に記憶される内容は、大容量の半導体メモリ 3 9 0 に記憶される。

40

【図面の簡単な説明】

【 0 0 7 8 】

【図 1】本発明の実施形態に係るサーバと、放送局と、ユーザ端末とから構成される情報処理システムを示す図である。

【図 2】本発明の実施形態に係るサーバの構成を示すブロック図である。

【図 3】本発明の実施形態に係るサーバの機能的な構成を示す機能ブロック図である。

50

【図 4】本発明の実施形態に係るコンテンツ情報テーブルの構成を示す図である。

【図 5】本発明の実施形態に係るサーバによる処理手順を示すフローチャートである。

【図 6】本発明の実施形態に係るユーザ端末における検索の手順についての説明に供する図である。

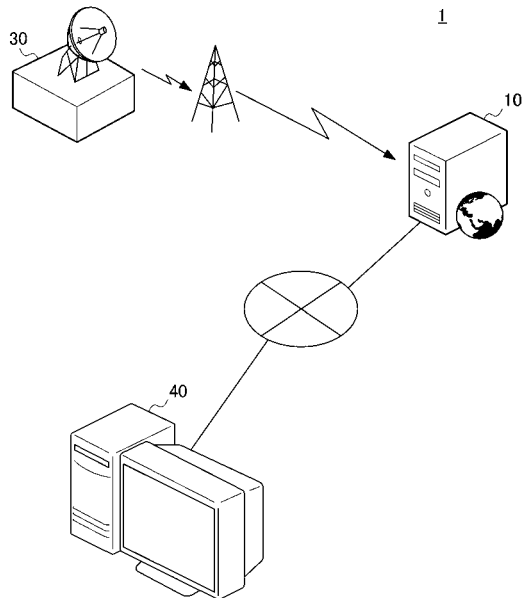
【符号の説明】

【 0 0 7 9 】

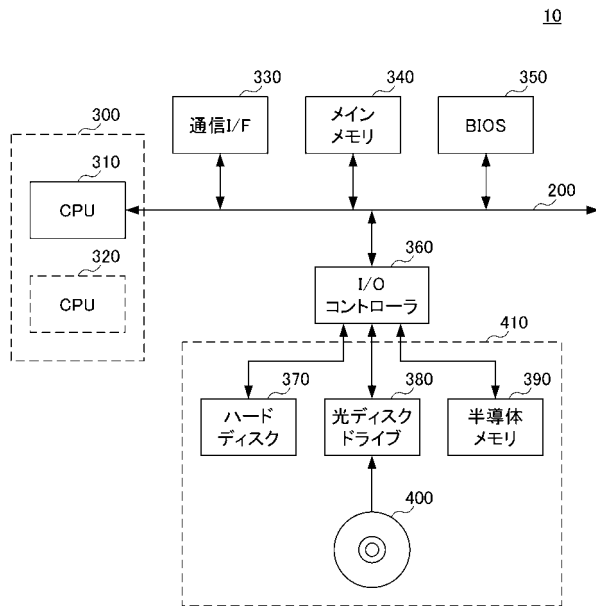
- 1 情報処理システム
- 10 サーバ
- 11 特定部（特定手段）
- 12 音声情報抽出部（音声情報抽出手段）
- 13 変換部（変換手段）
- 14 特徴語抽出部（特徴語抽出手段）
- 15 記憶部（記憶手段）
- 16 判断部（判断手段）
- 17 導出部（導出手段）
- 18 検索部（検索手段）
- 20 コンテンツログ情報テーブル DB（記憶手段）
- 30 放送局
- 40 ユーザ端末

10

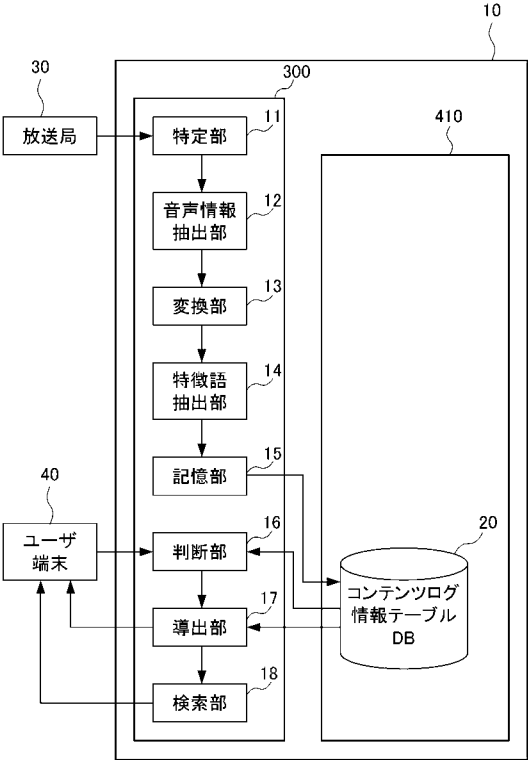
【図 1】



【図 2】



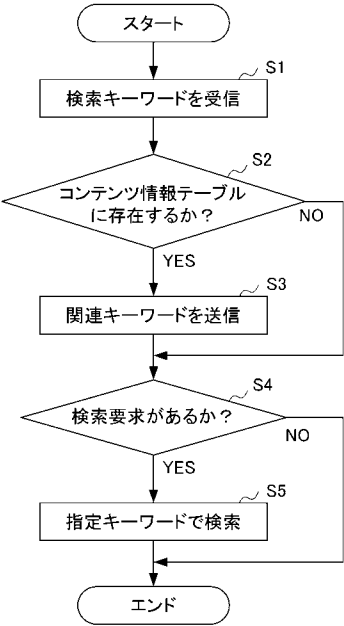
【図 3】



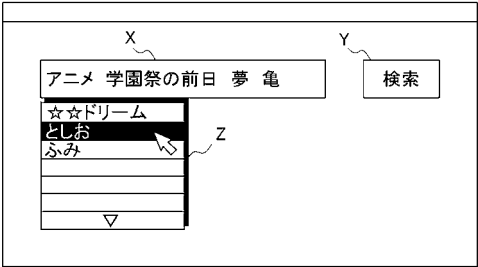
【図 4】

コンテンツID	タイトル	キャラクター名	属性1	属性2	属性3	属性4	属性5	属性6	...
0001	☆☆ドリーム	としお	SF	ギャグ	アニメ	大安市	学園祭の前日	夢	...
...	...	ふみ	高校生	不運	ガールハント	...	...	...	...
...	...	...	宇宙人	鬼	虎柄	...	...	...	...

【図 5】



【図 6】



フロントページの続き

- (72)発明者 ▲高▼林 貴仁
東京都港区六本木六丁目10番1号 ヤフー株式会社内
- (72)発明者 荻野 明子
東京都港区六本木六丁目10番1号 ヤフー株式会社内
- (72)発明者 中野 信
東京都港区六本木六丁目10番1号 ヤフー株式会社内

審査官 岩田 淳

- (56)参考文献 国際公開第2007/060968(WO,A1)
特開2007-52731(JP,A)

- (58)調査した分野(Int.Cl., DB名)
G06F 17/30
JSTPlus/JMEDPlus/JST7580(JDreamII)