(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2005-293213
(P2005-293213A)
(43)公開日　平成17年10月20日(2005.10.20)

(51)Int.Cl.[7]	F I		テーマコード（参考）
G06F 17/30	G06F 17/30	210A	5B075
	G06F 17/30	170J	
	G06F 17/30	320Z	
	G06F 17/30	340A	

審査請求 未請求　請求項の数 15　OL　（全 14 頁）

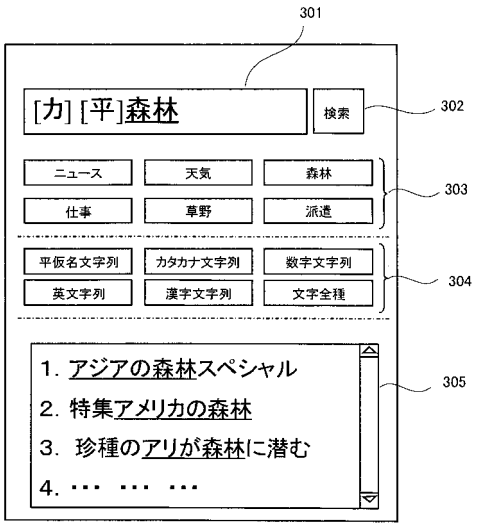
(21)出願番号　特願2004-107127 (P2004-107127)
(22)出願日　平成16年3月31日 (2004.3.31)

(71)出願人　000005049
シャープ株式会社
大阪府大阪市阿倍野区長池町２２番２２号
(74)代理人　100101823
弁理士　大前　要
(72)発明者　浜地　淳
大阪府大阪市阿倍野区長池町２２番２２号
シャープ株式会社内

Fターム(参考)　5B075 NK02　NK31　PP23　UU01

(54)【発明の名称】文字列抽出装置、コンテンツ検索装置、文字列抽出方法、プログラムおよび記録媒体

(57)【要約】
【課題】　ユーザが文字列を入力する際の手間を軽減することができるように改良された文字列抽出装置を提供することを主要な目的とする。
【解決手段】　この発明にかかる文字列抽出装置は、情報源から文字列を抽出するためのキーワードとなる第１の文字列を提示する第１文字列提示手段３０３を備える。文字列条件提示手段３０４は、文字列が満たす一定の条件を提示する。ユーザが、第１の文字列と、文字列条件提示手段３０４が提示した文字列が満たす一定の条件の少なくとも１つとの、組み合わせを指示した時に、組み合わせ提示手段３０１が、この組み合わせを提示する。第２文字列提示手段３０５は、上記情報源から、組み合わせ提示手段３０１が提示した組み合わせ文字列を含む第２の文字列を抽出し、提示する。
【選択図】図３



【特許請求の範囲】
【請求項１】
情報源から取り出したい文字列を抽出するためのキーワードとなる第１の文字列を提示する第１文字列提示手段と、
文字列が満たす一定の条件を提示する文字列条件提示手段と、
ユーザが、前記第１の文字列と、前記文字列条件提示手段が提示した前記文字列が満たす一定の条件の少なくとも１つとの、組み合わせを指示した時に、この組み合わせを提示する組み合わせ提示手段と、
前記情報源から、前記組み合わせ提示手段が提示した組み合わせ文字列を含む第２の文字列を抽出し、提示する第２文字列提示手段と、を備えた文字列抽出装置。
【請求項２】
前記第１文字列提示手段は、前記情報源として所定の文字列群が登録されている辞書である文字列群辞書を含み、さらに該文字列群辞書から前記第１の文字列を抽出する第１文字列抽出手段と、抽出された該第１の文字列を提示する手段とを含む、請求項１に記載の文字列抽出装置。
【請求項３】
前記第２文字列提示手段は、前記文字列群辞書から、前記第１の文字列と前記文字列が満たす一定の条件を含む第２の文字列を抽出し、提示する請求項２に記載の文字列抽出装置。
【請求項４】
前記第１文字列抽出手段は、前記第１の文字列が有する品詞情報に基づいて前記第１の文字列を抽出する請求項１から３のいずれか１項に記載の文字列抽出装置。
【請求項５】
前記第１文字列抽出手段は、ユーザー嗜好または履歴情報に基づいて前記第１の文字列を抽出する請求項１から４のいずれか１項に記載の文字列抽出装置。
【請求項６】
前記文字列条件提示手段は、少なくとも１種以上の、前記文字列が満たす一定の条件を提示する請求項１から５のいずれか１項に記載の文字列抽出装置。
【請求項７】
前記文字列条件提示手段は、前記文字列が満たす一定の条件の下位概念であって、階層化されたメニューで選択付与できる属性メニューをさらに提示する請求項１から６のいずれか１項に記載の文字列抽出装置。
【請求項８】
前記文字列条件提示手段は、前記属性メニューに加えて、文字列の並びを示したテンプレート情報メニューをも提示する請求項７に記載の文字列抽出装置。
【請求項９】
前記第２の文字列をユーザー嗜好または履歴情報により順位付けする手段と、
前記順位付けの情報を前記第２の文字列に付与する手段とを備えた請求項１から８に記載の文字列抽出装置。
【請求項１０】
請求項２から９に記載の文字列抽出装置のいずれかあるいは複数を備えたコンテンツ検索装置であって、
前記文字列群辞書はコンテンツに付与された文字列のリストを含み、
前記第１の文字列と前記文字列条件提示手段が提示した前記文字列が満たす一定の条件の少なくとも１つとの組み合わせに基づいて、前記コンテンツを検索するコンテンツ検索装置。
【請求項１１】
情報源から取り出したい文字列を抽出するためのキーワードとなる第１の文字列を提示する第１文字列提示工程と、
文字列が満たす一定の条件を提示する文字列条件提示工程と、

ユーザが、前記第１の文字列と、前記文字列条件提示工程で提示された前記文字列が満たす一定の条件の少なくとも１つとの、組み合わせを指示した時に、この組み合わせを提示する組み合わせ提示工程と、
前記情報源から、前記組み合わせ提示工程で提示された組み合わせ文字列を含む第２の文字列を抽出し、提示する第２文字列提示工程と、を備えた文字列抽出方法。
【請求項１２】
前記第１文字列提示工程は、前記情報源として所定の文字列群が登録されている辞書である文字列群辞書を読み込む工程と、該文字列群辞書から前記第１の文字列を抽出する第１文字列抽出工程と、抽出された該第１の文字列を提示する工程とを含む、請求項１１に記載の文字列抽出方法。
【請求項１３】
前記第２文字列提示工程は、前記文字列群辞書から、前記第１の文字列と前記文字列条件提示工程で提示された前記文字列が満たす一定の条件を含む第２の文字列を抽出し、提示する請求項１２に記載の文字列抽出方法。
【請求項１４】
情報源から取り出したい文字列を抽出するためのキーワードとなる第１の文字列を提示する第１文字列提示手順と、
文字列が満たす一定の条件を提示する文字列条件提示手順と、
ユーザが、前記第１の文字列と、前記文字列条件提示手順が提示した前記文字列が満たす一定の条件の少なくとも１つとの、組み合わせを指示した時に、この組み合わせを提示する組み合わせ提示手順と、
前記情報源から、前記組み合わせ提示手順において組み合わされた文字列を含む第２の文字列を抽出し、提示する第２文字列提示手順とをコンピュータに実行させるためのプログラム。
【請求項１５】
情報源から取り出したい文字列を抽出するためのキーワードとなる第１の文字列を提示する第１文字列提示手順と、
文字列が満たす一定の条件を提示する文字列条件提示手順と、
ユーザが、前記第１の文字列と、前記文字列条件提示手順が提示した前記文字列が満たす一定の条件の少なくとも１つとの、組み合わせを指示した時に、この組み合わせを提示する組み合わせ提示手順と、
前記情報源から、前記組み合わせ提示手順において組み合わされた文字列を含む第２の文字列を抽出し、提示する第２文字列提示手順とをコンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

【発明の詳細な説明】
【技術分野】
【０００１】
本発明は一般に文字列抽出装置に関するものであり、より特定的には、ユーザが文字列を入力する際の手間を軽減することができるように改良された文字列抽出装置に関する。この発明は、そのような文字列抽出装置を有するコンテンツ検索装置に関する。この発明は、ユーザが文字列を入力する際の手間を軽減することができるように改良された文字列抽出方法に関する。この発明はさらにそのような方法をコンピュータによって実現するプログラムに関する。この発明はさらにそのようなプログラムを記録したコンピュータ読み取り可能な記録媒体に関する。
【背景技術】
【０００２】
近年、画像の記録再生から伝送までにわたって広くデジタル技術を適用したデジタルテレビが普及してきている。デジタルテレビは、従来のアナログテレビに比べると、伝送路がデジタルなので、多重化によって静止画像や文字放送などの異なる形態の情報を同時に

伝送することが容易である。また、情報圧縮技術を利用することにより、数チャンネル分の画像をまとめて送ることができる。さらに、画像の解像度や形状を変更して送ることができる、双方向通信ができる、などといった利点を備えている。

【０００３】

そして、デジタルテレビの使用者は、プラットフォームと呼ばれる業者などを通じて、各種テレビ放送局のＥＰＧ（Electronic Programming Guide）のサービスを享受することが可能となっており、これにより各種テレビ放送局のテレビ番組表を画面上で見ることができる。

【０００４】

図８は、このようなＥＰＧのサービスを利用して番組情報を検索するシステムの全体構成を示すブロック図である（例えば、特許文献１参照）。図８を参照して、デジタルテレビ１１は、サーバ３２が提供する各種放送局３１ａ～３１ｚのＥＰＧのサービスを享受することができるものである。デジタルテレビ１１のリモコン２１は、日本語をはじめとする自然言語（平仮名、カタカナ、漢字、数字などの文字列）を入力することが可能である。検索エンジン１５は、リモコン２１により入力された自然言語を検索キーにして各種放送局のＥＰＧに含まれる番組情報の検索を行う。検索の結果、該当する一つ又は複数の番組情報が、ディスプレイ１２上において、検索のアクセス率の高い順、もしくは放送時間順に配列されて表示される。

【０００５】

図９は、他の従来例にかかる番組検索装置のブロック図である（例えば特許文献２参照）。受信装置は、ユーザの操作を入力する操作手段１と、文字等のグラフィック情報を表示するためのＣＲＴ、ＬＣＤディスプレイあるいは、映像出力端子よりなる画像表示手段２と、メモリやハードディスクからなる番組情報記憶手段３と、マイクロコンピュータ及びプログラム等が格納されたメモリからなる制御部４により構成される。

【０００６】

制御部４において番組情報記憶手段３の番組情報から時間及びチャンネルに従い番組名、番組属性を並べた番組表の作成や、特定の番組群の番組名、番組属性を並べた番組一覧を作成する。表示画面作成手段５は、特定の番組の詳細な情報の表示画面作成等の処理を行う。番組名取得手段６は、表示画面作成手段５により作成された画面において操作手段１により利用者が特定した番組の番組名を取り出す。文字削除手段７は、予め設定された種類の文字を削除するものである。番組検索手段８は、文字削除手段７により編集された文字列を検索キーとして、番組情報記憶手段３内の番組情報から、検索キーを含む番組を検索する。

文字削除手段７において削除される文字種として、記号と数字を設定しておくことで、放送番組に特有な再放送を示す記号や、ドラマなどの話数を表す数字が削除された番組名が検索キーとなり、その番組の再放送および、次回の番組を検索することが可能となる。

【０００７】

【特許文献１】特開２００３－１１０９６０号公報

【特許文献２】特開２００１－８６４２２号公報

【発明の開示】

【発明が解決しようとする課題】

【０００８】

しかしながら、前述の如く、特許文献１の開示技術では、検索を行うために何らかの手段を用いて検索キーワードをユーザが全て入力する必要があり、入力する際の手間が煩雑である。また、特許文献２の開示技術では、キーワードを取り出すために番組情報をユーザが選択後、同情報のテキスト内からキーワードを抜き出す手続きを踏む必要がある。また、適切なキーワードを含む番組情報が存在しない場合には、特許文献１と同様にキーワードを一から入力する必要があり、入力する際の手間が煩雑である。

【０００９】

この発明は上記のような問題点を解決するためになされたもので、ユーザが文字列を入力する際の手間を軽減することができるように改良された文字列抽出装置を提供することを目的とする。

【００１０】

この発明の他の目的は、そのような文字列抽出装置を有するコンテンツ検索装置を提供することにある。

【００１１】

この発明のさらに他の目的は、ユーザが文字列を入力する際の手間を軽減することができるように改良された文字列抽出方法を提供することにある。

【００１２】

この発明のさらに他の目的は、そのような方法をコンピュータによって実現するプログラムを提供することにある。

【００１３】

この発明のさらに他の目的は、そのようなプログラムを記録したコンピュータ読み取り可能な記録媒体を提供することにある。

【課題を解決するための手段】

【００１４】

この発明にかかる文字列抽出装置は、情報源から取り出したい文字列を抽出するためのキーワードとなる第１の文字列を提示する第１文字列提示手段と、文字列が満たす一定の条件を提示する文字列条件提示手段と、ユーザが、上記第１の文字列と、上記文字列条件提示手段が提示した上記文字列が満たす一定の条件の少なくとも１つとの、組み合わせを指示した時に、この組み合わせを提示する組み合わせ提示手段と、上記情報源から、上記組み合わせ提示手段が提示した組み合わせ文字列を含む第２の文字列を抽出し、提示する第２文字列提示手段とを備える。

【００１５】

この発明によれば、ユーザは、上記第１の文字列と、上記文字列条件提示手段が提示した上記文字列が満たす一定の条件の少なくとも１つとの、組み合わせを指示するだけで、取り出したい第２の文字列を取得することができるので、ユーザが入力する際の打鍵数を減らすことができ、文字列を入力する際の手間を軽減することができる。

【００１６】

この発明の好ましい実施態様によれば、上記第１文字列提示手段は、上記情報源として所定の文字列群が登録されている辞書である文字列群辞書を含み、さらに該文字列群辞書から上記第１の文字列を抽出する第１文字列抽出手段と、抽出された該第１の文字列を提示する手段とを含む。上記第２文字列提示手段は、上記文字列群辞書から、上記第１の文字列と上記文字列が満たす一定の条件を含む第２の文字列を抽出し、提示する。

【００１７】

このように構成することにより、各種テレビ放送局のＥＰＧのサービスを享受することができるようになる。

【００１８】

上記第１文字列抽出手段は、上記第１の文字列が有する品詞情報に基づいて上記第１の文字列を抽出するのが好ましい。品詞情報に基づいて文字列を抽出することにより、情報の絞込みが可能となる。品詞としては、例えば、名詞、動詞、形容詞などがある。

【００１９】

この発明のさらに好ましい実施態様によれば、上記第１文字列抽出手段は、ユーザー嗜好または履歴情報に基づいて上記第１の文字列を抽出する。このように構成することにより、情報がより一層絞込まれる。ここで言うユーザー嗜好とは、ユーザーの履歴情報から、そのユーザーと似た行動を取っているユーザーの嗜好情報をもとに、ユーザの嗜好を推測するシステム（協調フィルタリング）など、ユーザーの利用履歴を利用した任意の嗜好類推アルゴリズムを用いて類推されるものも含む。

【００２０】

上記文字列条件提示手段は、少なくとも１種以上の、上記文字列が満たす一定の条件を提示する。条件としては、文字列が例えば全て平仮名で構成されるという条件、全てカタカナで構成されるという条件、全て数字で構成されるという条件、全てアルファベットで構成されるという条件、全て漢字で構成されるという条件、あらゆる文字で構成されるという条件などがある。

【００２１】

上記文字列条件提示手段は、上記文字列が満たす一定の条件の下位概念であって、階層化されたメニューで選択付与できる属性メニューをさらに提示するのが好ましい。単なる文字種の情報だけでなく、属性（国名、人名等）を付与することで、より効率的な検索を行うことができる。属性メニューを、階層化して提示することとしたのは、属性の数が増えるとメニューから選択しづらくなるので、ユーザに選択しやすくするためである。

【００２２】

上記文字列条件提示手段は、上記属性メニューに加えて、文字列の並びを示したテンプレート情報メニューをも提示するようにしてもよい。単なる属性情報だけではなく、文字列の並びを示したテンプレートを情報として付与することにより、より細かい条件を指定することが可能となる。

【００２３】

この発明のさらに好ましい実施態様によれば、上記第２の文字列をユーザー嗜好または履歴情報により順位付けする手段と、上記順位付けの情報を上記第２の文字列に付与する手段とをさらに備える。順位付けする手段を設けることにより、検索結果のさらなる絞込みが行われる。

【００２４】

この発明の他の局面に従うコンテンツ検索装置は、上述の文字列抽出装置のいずれかあるいは複数を備え、上記文字列群辞書はコンテンツに付与された文字列のリストを含み、上記第１の文字列と、上記文字列条件提示手段が提示した上記文字列が満たす一定の条件の少なくとも１つとの、組み合わせに基づいて、上記コンテンツを検索する。

【００２５】

この発明によれば、上記コンテンツを検索する場合において、ユーザが文字列を入力する際の手間を軽減することができる。

【００２６】

この発明の他の局面にかかる文字列抽出方法は、情報源から取り出したい文字列を抽出するためのキーワードとなる第１の文字列を提示する第１文字列提示工程と、文字列が満たす一定の条件を提示する文字列条件提示工程と、ユーザが、上記第１の文字列と、上記文字列条件提示工程で提示された上記文字列が満たす一定の条件の少なくとも１つとの、組み合わせを指示した時に、この組み合わせを提示する組み合わせ提示工程と、上記情報源から、上記組み合わせ提示工程で提示された組み合わせ文字列を含む第２の文字列を抽出し、提示する第２文字列提示工程とを備える。

【００２７】

この発明によれば、ユーザは、上記第１の文字列と、上記文字列条件提示工程で提示された上記文字列が満たす一定の条件の少なくとも１つとの、組み合わせを指示するだけでよいので、ユーザが入力する際の打鍵数を減らすことができ、文字列を入力する際の手間を軽減することができる。

【００２８】

この発明の好ましい実施態様によれば、上記第１文字列提示工程は、上記情報源として所定の文字列群が登録されている辞書である文字列群辞書を読み込む工程を含み、さらに該文字列群辞書から上記第１の文字列を抽出する第１文字列抽出工程と、抽出された該第１の文字列を提示する工程とを含む。

【００２９】

上記第２文字列提示工程は、上記文字列群辞書から、上記第１の文字列と上記文字列条件提示工程で提示された上記文字列が満たす一定の条件を含む第２の文字列を抽出し、提

示するのが好ましい。

【００３０】

この発明の他の局面に従うプログラムは、情報源から取り出したい文字列を抽出するためのキーワードとなる第１の文字列を提示する第１文字列提示手順と、文字列が満たす一定の条件を提示する文字列条件提示手順と、ユーザが、上記第１の文字列と、上記文字列条件提示手順が提示した上記文字列が満たす一定の条件の少なくとも１つとの、組み合わせを指示した時に、この組み合わせを提示する組み合わせ提示手順と、上記情報源から、上記組み合わせ提示手順において組み合わされた文字列を含む第２の文字列を抽出し、提示する第２文字列提示手順とをコンピュータに実行させるためのプログラムにかかる。

【００３１】

この発明のさらに他の局面にかかるコンピュータ読み取り可能な記録媒体は、情報源から取り出したい文字列を抽出するためのキーワードとなる第１の文字列を提示する第１文字列提示手順と、文字列が満たす一定の条件を提示する文字列条件提示手順と、ユーザが、上記第１の文字列と、上記文字列条件提示手順が提示した上記文字列が満たす一定の条件の少なくとも１つとの、組み合わせを指示した時に、この組み合わせを提示する組み合わせ提示手順と、上記情報源から、上記組み合わせ提示手順において組み合わされた文字列を含む第２の文字列を抽出し、提示する第２文字列提示手順とをコンピュータに実行させるためのプログラムを記録してなる。

【発明の効果】

【００３２】

この発明によれば、ユーザが入力する際の打鍵数を減らすことができ、文字列を入力する際の手間を軽減することができるとともに、必要な情報を効率的に検索することができるという優れた効果を奏し得る。

【発明を実施するための最良の形態】

【００３３】

文字列を入力する際の手間を軽減するという目的を、ユーザが、キーワードとなる文字列と、文字列が満たす一定の条件の少なくとも１つとの、組み合わせを指示するだけで、取り出したい情報を取得するということにより実現した。

【００３４】

図１は、本発明にかかる文字列抽出装置のブロック図である。

【００３５】

図１を参照して、コンテンツ読込部１０１でコンテンツ源からコンテンツの読み込みを行う。読み込まれた各コンテンツに付与されている文字列から、キーワード抽出部１０２でキーワードの抽出を行う。後述する実施例では、キーワードは単語となる。必要に応じて、抽出されたキーワードを、キーワード絞込み部１０３で、品詞情報例えば名詞、固有名詞等、ユーザー嗜好、履歴情報などを用いて設定された数に絞り込むなどの処理を行う。また、キーワードを出現数が多いものから所定の数だけ抽出し、絞り込むなどの処理を行ってもよい。なお、ここで言うユーザー嗜好とは、ユーザーの履歴情報から、そのユーザーと似た行動を取っているユーザーの嗜好情報をもとに、ユーザの嗜好を推測するシステム（協調フィルタリング）など、ユーザーの利用履歴を利用した任意の嗜好類推アルゴリズムを用いて類推されるものも含む。

【００３６】

抽出されたキーワード１０５は、表示部１０８により、候補としてユーザーに提示される。ユーザーは、提示されたキーワードを選んで入力部１１０から入力する。

【００３７】

文字列条件提示パターン１０７では、文字列が満たす一定の条件が候補として提示される。条件としては、文字列が例えば全て平仮名で構成されるという条件、全てカタカナで構成されるという条件、全て数字で構成されるという条件、全てアルファベットで構成されるという条件、全て漢字で構成されるという条件、あらゆる文字で構成されるという条件などがある。ユーザーは、文字列条件提示パターン１０７の中から選んで、入力部１１

０から入力する。

【００３８】

ユーザによって入力された情報は、それぞれ、キーワード１０５、文字列条件提示パターン１０７に送られる。キーワード１０５、文字列条件提示パターン１０７にそれぞれ送られた情報は、キーワードメニュー・文字列条件提示パターンの生成部１０６で、組み合わせられることにより検索文字列が生成し、コンテンツ検索部１０４が利用可能な形態に変換され、その後、コンテンツ検索部１０４に情報として送られる。

【００３９】

キーワードメニュー・文字列条件提示パターンの生成部１０６より送られた情報を用いて、コンテンツ検索部１０４で、コンテンツ読込部１０１より読み込まれたコンテンツから該当するものの検索を行う。検索結果の一覧は、各コンテンツごとに付与されている情報で順位付け、絞込みが行われる。各コンテンツごとに付与される情報としては、該当コンテンツの参照回数（ＴＶ番組なら視聴率）、コンテンツ提供側が予め付与した人気度情報、コンテンツの製造時間、コンテンツに情報が付与されていない場合は、各コンテンツが読み込んだ時間、コンテンツのサイズなどである。これらは、コンテンツ源から読み込まれる際に付与されることになる。

【００４０】

なお、検索結果処理部１０９で、必要に応じて予めユーザーが入力した嗜好情報などで順位情報を付与して、絞込みを行ってもよい。コンテンツのリスト一覧は、表示部１０８を通じてユーザーに提示される。

【００４１】

次に、フローチャートを用いて、本発明の実施の形態をさらに詳細に説明する。

【００４２】

図２を参照して、最初に、ステップ２０１で、新たにキーワードの生成を行うか否かの判定を行い、以下のそれぞれの流れに従う。

【００４３】

新たにキーワードの生成を行う必要がある場合には、ステップ２０２で、まず、コンテンツの読み込みを行う。続いて、ステップ２０３で、コンテンツに付与されている文字列からキーワードを抽出する。抽出したキーワードは、ステップ２０４でメインメモリ上かストレージなどに一時的に保存される。ステップ２０５で、更にキーワードの抽出に用いるコンテンツがまだ存在するようであれば、ステップ２０２に戻り、再びコンテンツの読み込みを行い、存在しない場合は、ステップ２０６で、抽出されたキーワードの中から品詞が「名詞」、「固有名詞」のものを抽出する。このとき、ユーザー嗜好・履歴情報などから予め設定された数まで絞込み、順位付けを行ってもよい。ステップ２０７で、キーワードを出現数が多いものから所定の数だけ抽出する。最後に、ステップ２０８で、得られたキーワードをキーワードメニューとしてユーザーに提示する。

【００４４】

新たにキーワードの生成を行う必要がない場合には、ステップ２０９で、ユーザーがキーワードと、文字列条件提示パターンとをそれぞれ選んで、それぞれ入力する。この選ばれた組合わせに基づいて、ステップ２１０で、コンテンツの検索を行う。ステップ２１１で、該当するコンテンツが存在するか否かを判定する。該当するコンテンツが存在する場合は、ステップ２１２で、同コンテンツをメインメモリー上か、ストレージなどに一時的に保存する。

【００４５】

該当するコンテンツがもはや存在しない場合は、ステップ２１３で、検索結果の一覧を各コンテンツごとに付与されている情報で順位付けを行う。このとき、先に一時保存したコンテンツをユーザー嗜好・履歴情報などから絞り込み・順位付けを行ってもよい。その後に、ステップ２１４で、検索結果の一覧をユーザーに対し表示する。

【００４６】

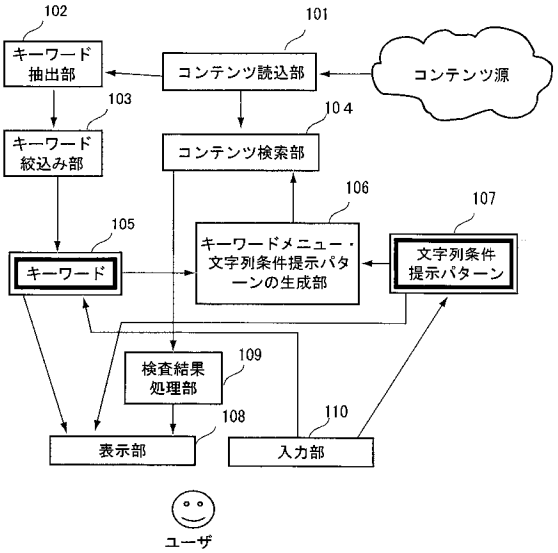
次に具体的実施例を用いて本発明を説明する。

【実施例１】
【００４７】
本実施例では、ＥＰＧの番組データベースから、取り出したい文字列を検索する例を示す。
【００４８】
図３は、本実施例にかかる検索キーワードの簡易入力機能を説明するための表示画面である。この例では、「コンテンツ」が「ＴＶ番組」となり、「文字列群辞書/コンテンツに付与された文字列のリスト」が「ＥＰＧデータ」となる。
【００４９】
図３を参照して、第１文字列提示手段であるキーワードメニュー３０３は、ＥＰＧデータより、選択可能なタイトル、サブタイトル、出演者を形態素解析技術などを用いて、単語単位に分割し、名詞・固有名詞候補を得て、各単語の出現する回数が多い(少ない)ものの上位の候補をメニューとして表示する。各キーワードメニューを押すと、組み合わせ提示手段である選択文字列入力枠３０１に対応する文字列が入力される。
【００５０】
文字列条件提示手段である文字列条件パターンメニュー３０４は、キーワードメニューだけでは表現しきれない文字列をカバーするためのものである。文字列条件パターンメニュー３０４は、少なくとも１種以上の、文字列が満たす一定の条件を提示する。本実施例では、６種類、提示されている。条件としては、例えば、内容は特定されないが文字列が全て平仮名で構成されるという条件、全てカタカナで構成されるという条件、全て数字で構成されるという条件、全てアルファベットで構成されるという条件、全て漢字で構成されるという条件、あらゆる文字で構成されるという条件などがある。
【００５１】
次に動作について説明する。
【００５２】
前処理として、所定期間内の「ＥＰＧデータ」から、形態素解析処理を用いて、キーワードメニュー３０３として利用するキーワードを抽出する。
【００５３】
さて、ユーザーは「アジアの森林」を入力したいとする。しかし、キーワードメニュー３０３には、入力したいキーワードに含まれる単語が「森林」しかない。そこで、「アジア」の代わりに「文字列条件パターンメニュー３０４」のカタカナ文字列を[カ]と、「の」の変わりに平仮名文字列を示す[平]を文字列条件パターンメニュー３０４で選択することにより、選択文字列入力枠３０１に入力する。続けて、キーワードメニュー３０３にある「森林」を選択することで、選択文字列入力枠３０１に、番組の検索を行うための文字列を入力する。続いて検索開始ボタン３０２を押すことにより、検索が開始される。検索結果は、第２文字列提示手段である候補リスト３０５に一覧として表示される。なお、このリストは、ユーザー嗜好・履歴情報によりその表示順位が決定される。ユーザは候補リスト３０５の一覧から視聴したい番組を選び、その番組の視聴予約、録画予約等を行うことになる。
【００５４】
次に、ハードウェアの構成例について説明する。
【００５５】
本発明における端末は、図４のような構成のコンピュータを用いて実現できる。
【００５６】
これらを動作させるプログラムはＣＤ－ＲＯＭドライブ４０３やネットワーク４０７を通じて供給され、ハードディスク４０２に登録される。あるいは、最初からハードディスク４０２に登録されていても構わない。また、ハードディスクの変わりにフラッシュメモリやＲＯＭを利用してもよい。プログラムを動作させる際には、一度ハードディスク４０２からメモリ４０６に読み出され、それをＣＰＵ４０４が順次処理することとなる。ユーザは、入力装置４０１から入力し、検索結果は表示装置４０５によって表示される。

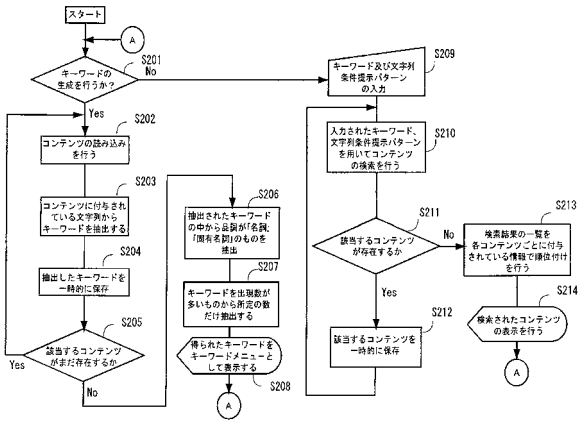
【実施例２】
【００５７】
図５は、実施例２にかかる検索キーワードの簡易入力機能を説明するための表示画面である。本実施例では、文字列条件パターン５０２に下位概念としての属性が付与されている。属性は、階層化されたメニューで各種属性を選択・付与することができる。この例では、「国名」が選択されている。
【００５８】
属性付の文字列条件パターン５０１を用いることにより、単なる文字種の情報だけでなく、属性（国名、人名等）を付与することで、より効率的な検索を行うことができる。この例では、「カタカナ」でかつ「国名」、属性無しの「平仮名」、文字列「森林」が提示されている。候補リスト５０３では、実施例１の候補リストとは異なり、「カタカナ」でマッチングが取れた箇所は全て「国名」となっている。
【実施例３】
【００５９】
図６は、実施例３にかかる検索キーワードの簡易入力機能を説明するための表示画面である。本実施例では、文字列条件パターンメニュー６０２が階層化されている。属性の数が増えるとメニューから選択しづらくなる。そこで、メニューを階層化することにより、ユーザに選択しやすくしている。
【００６０】
分類された属性付の文字列条件パターン６０１を用いることにより、より効率的な検索を行うことができる。この例では、「漢字」でかつ「国名」、属性無しの「平仮名」、文字列「森林」が提示されている。候補リスト６０３では、実施例１の候補リストとは異なり、「漢字」でマッチングが取れた箇所は全て「国名」となっている。なお、固有名詞全般を指定したい場合は、「固有名詞全般」を選択する。
【実施例４】
【００６１】
図７は、実施例４にかかる検索キーワードの簡易入力機能を説明するための表示画面である。本実施例では、階層化された文字列条件パターンメニュー７０２において、属性に加えて、文字列の並びを示したテンプレート情報が追加されている。テンプレート指定を行わない場合は、「テンプレート無し」を選ぶ。
【００６２】
テンプレート情報を付与した属性付の文字列条件パターン７０１を用いることにより、より細かい条件を指定することが可能となる。この例では、「アラビア数字」のテンプレート情報「○○時」が提示されている。「３時４５分～～」を検索したい場合は、「[数：ア：○○時][数：ア：○○分]～～」のように指定する。時刻の内、テンプレート全てにマッチする候補（あるいは検索）を指定する場合は「時刻全般」を選択する。
【００６３】
なお、ここでは、数字を例示したが、図６を再び参照して、「漢字文字列」の「人名」で「＊＊氏」「＊＊様」、「地域名」で「＊＊市」といったテンプレートも考えられる。
【００６４】
なお、上記実施例では、ＥＰＧデータの検索に本発明を応用する場合を例示したが、この発明はこれに限られるものでなく、インターネットの検索、携帯電話のメール検索、あるいは文字列生成装置（ワープロ等）を用いて作成して蓄積したデータの検索にも応用することができる。
【００６５】
今回開示された実施の形態はすべての点で例示であって制限的なものではないと考えられるべきである。本発明の範囲は上記した説明ではなくて特許請求の範囲によって示され、特許請求の範囲と均等の意味および範囲内でのすべての変更が含まれることが意図される。
【産業上の利用可能性】

【００６６】
本発明はユーザが入力する際の文字列入力の手間を軽減することが可能となるので、情報産業の分野において有用である。
【図面の簡単な説明】
【００６７】
【図１】本発明にかかる文字列抽出装置のブロック図
【図２】本発明にかかるプログラムによって行われる情報処理動作を示すフローチャート図
【図３】実施例１にかかる検索キーワードの簡易入力機能を説明するための表示画面
【図４】実施例１にかかるプログラムを実行するためのハードウエアの構成を示す図
【図５】実施例２にかかる検索キーワードの簡易入力機能を説明するための表示画面
【図６】実施例３にかかる検索キーワードの簡易入力機能を説明するための表示画面
【図７】実施例４にかかる検索キーワードの簡易入力機能を説明するための表示画面
【図８】従来の番組情報を検索するシステムの全体構成を示すブロック図
【図９】他の従来例かかる番組検索装置のブロック図
【符号の説明】
【００６８】
１０１　コンテンツ読み込み部
１０２　キーワード抽出部
１０３　キーワード絞込み部
１０４　コンテンツ検索部
１０５　キーワード
１０６　キーワードメニュー・文字列条件提示パターンの生成部
１０７　文字列条件提示パターン
１０８　表示部
１０９　検索結果処理部
１１０　入力部

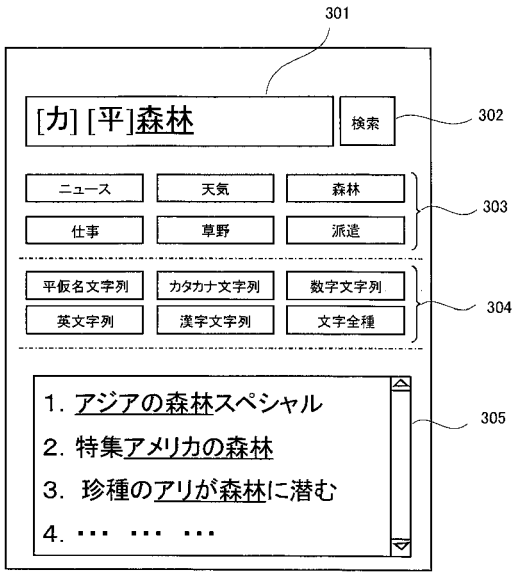
【図１】


102
キーワード抽出部
101
コンテンツ読込部
コンテンツ源
103
キーワード絞込み部
104
コンテンツ検索部
106
キーワードメニュー・文字列条件提示パターンの生成部
107
文字列条件提示パターン
105
キーワード
109
検査結果処理部
108
表示部
110
入力部
ユーザ

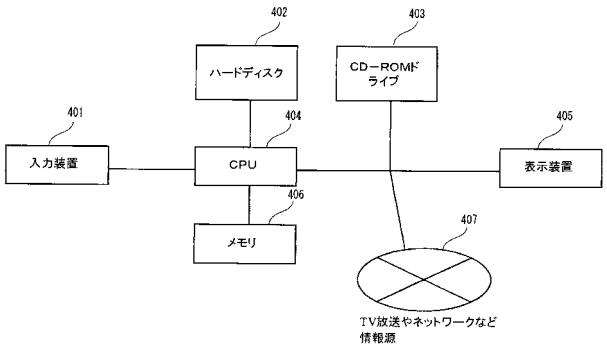
【図２】


スタート
A
S201
キーワードの生成を行うか？
No
Yes
S202
コンテンツの読み込みを行う
S203
コンテンツに付与されている文字列からキーワードを抽出する
S204
抽出したキーワードを一時的に保存
S205
該当するコンテンツがまだ存在するか
Yes
No
S206
抽出されたキーワードの中から品詞が「名詞」「固有名詞」のものを抽出
S207
キーワードを出現数が多いものから所定の数だけ抽出する
S208
得られたキーワードをキーワードメニューとして表示する
A
S209
キーワード及び文字列条件提示パターンの入力
S210
入力されたキーワード、文字列条件提示パターンを用いてコンテンツの検索を行う
S211
該当するコンテンツが存在するか
No
Yes
S212
該当するコンテンツを一時的に保存
S213
検索結果の一覧を各コンテンツごとに付与されている情報で順位付けを行う
S214
検索されたコンテンツの表示を行う
A

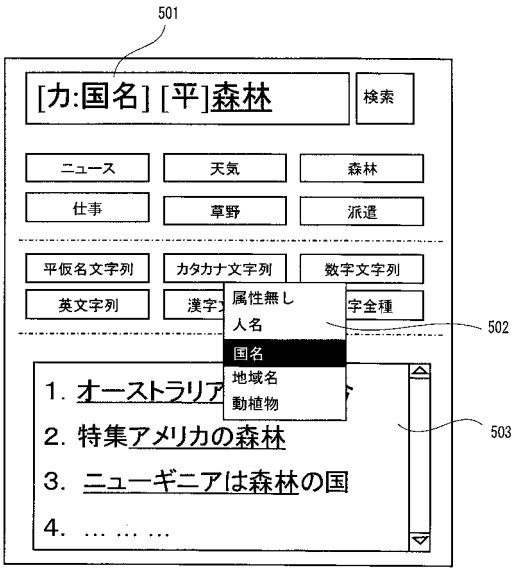
【図３】


301
[力] [平]森林
検索
302
ニュース
天気
森林
仕事
草野
派遣
303
平仮名文字列
カタカナ文字列
数字文字列
英文字列
漢字文字列
文字全種
304
1. アジアの森林スペシャル
2. 特集アメリカの森林
3. 珍種のアリが森林に潜む
4. ・・・ ・・・ ・・・
305

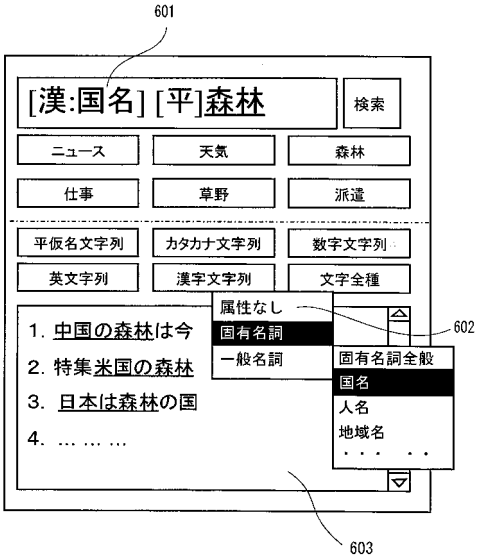
【図４】


402
ハードディスク
403
CD－ROMドライブ
401
入力装置
404
CPU
405
表示装置
406
メモリ
407
TV放送やネットワークなど情報源

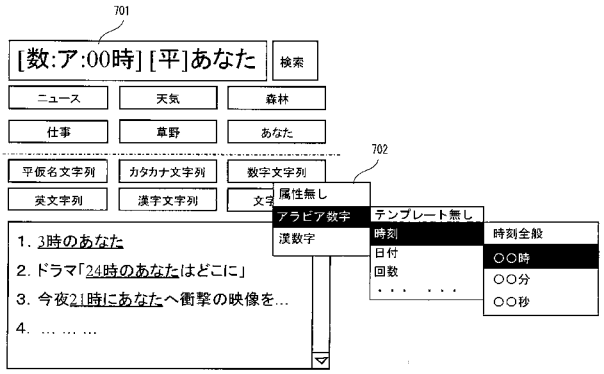
【図５】


501
[カ:国名] [平]森林
検索
ニュース
天気
森林
仕事
草野
派遣
平仮名文字列
カタカナ文字列
数字文字列
英文字列
漢字文字列
文字全種
属性無し
人名
国名
地域名
動植物
502
1. オーストラリア
2. 特集アメリカの森林
3. ニューギニアは森林の国
4. … … …
503

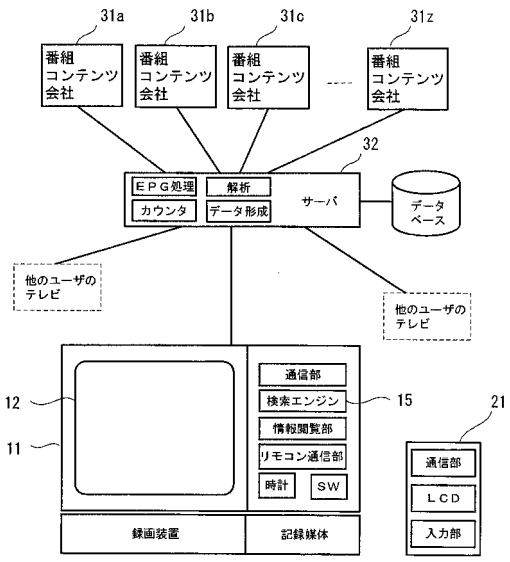
【図６】


601
[漢:国名] [平]森林
検索
ニュース
天気
森林
仕事
草野
派遣
平仮名文字列
カタカナ文字列
数字文字列
英文字列
漢字文字列
文字全種
属性なし
固有名詞
一般名詞
602
固有名詞全般
国名
人名
地域名
1. 中国の森林は今
2. 特集米国の森林
3. 日本は森林の国
4. … … …
603

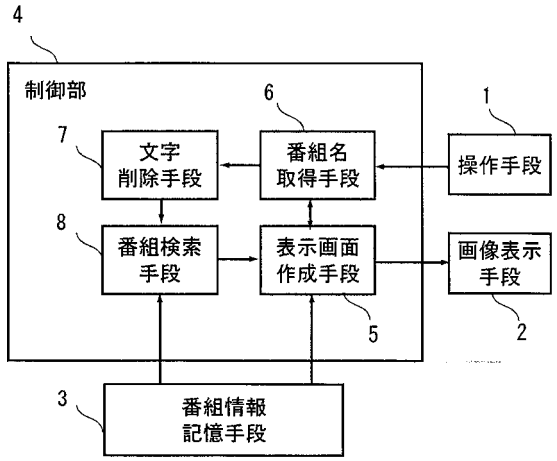
【図７】


701
[数:ア:00時] [平]あなた
検索
ニュース
天気
森林
仕事
草野
あなた
平仮名文字列
カタカナ文字列
数字文字列
英文字列
漢字文字列
属性無し
アラビア数字
漢数字
702
テンプレート無し
時刻
日付
回数
時刻全般
○○時
○○分
○○秒
1. 3時のあなた
2. ドラマ「24時のあなたはどこに」
3. 今夜21時にあなたへ衝撃の映像を…
4. … … …

【図８】


31a
31b
31c
31z
番組コンテンツ会社
番組コンテンツ会社
番組コンテンツ会社
番組コンテンツ会社
32
EPG処理
解析
カウンタ
データ形成
サーバ
データベース
他のユーザのテレビ
他のユーザのテレビ
12
11
通信部
検索エンジン
情報閲覧部
リモコン通信部
時計
SW
15
21
通信部
LCD
入力部
録画装置
記録媒体

【図９】


4
制御部
6
1
7
文字
削除手段
番組名
取得手段
操作手段
8
番組検索
手段
表示画面
作成手段
画像表示
手段
5
2
3
番組情報
記憶手段