転記考学

【特許番号】特許第３６５６９８６号

【出願日】】平成１２年２月１４日

【出願人】日本電信電話株式会社

【代理人】川久保 新一

提出日：２０１１年１１月１２日

綾木健一郎

【書類名】特許請求の範囲

【請求項１】

　複数の文書情報コンテンツを、蓄積する文書情報コンテンツ蓄積手段と；

　上記文書情報コンテンツの全文書を形態素解析し、出現頻度の高い名詞を上位からＮ個選び、次いで、各文書について、上記Ｎ個の名詞のそれぞれの出現頻度を表すＮ次元のベクトルを構成してコンテンツプロファイルを生成するコンテンツプロファイル生成手段と；

　上記文書情報コンテンツを表す代表画像であるサムネイルを提示するとともに、所定の閲覧利用者によってサムネイルがクリックされると、対応するコンテンツの詳細文書情報を提示する提示手段と；

　複数の文書情報コンテンツに対応する複数のサムネイルが閲覧利用者によって順次クリックされると、閲覧利用者によって閲覧された文書情報コンテンツに対応するコンテンツプロファイルの最新の所定回数の履歴から、コンテンツプロファイルの要素毎の平均値を各要素とするＮ次元の、閲覧利用者の興味を表すユーザプロファイルを演算するユーザプロファイル生成手段と；

　蓄積された複数の文書情報コンテンツに対し、２つのコンテンツプロファイルの間の関連の度合いをベクトルの距離尺度を用いて表し、コンテンツプロファイルの間の関連の度合いを、多次元尺度構成法を用いて２次元平面空間上の距離尺度として表現し、２次元平面上の文書情報コンテンツの配置座標を計算する第１のサムネイル配置演算手段と；

　上記ユーザプロファイルと各コンテンツプロファイルとの間の類似度を、ベクトルの大きさを正規化し、対応するベクトル要素間の距離の和に基づいてそれぞれ演算し、３次元的な提示手段を用いて画面垂直方向に演算された類似度が高いコンテンツに対応するサムネイル程、手前に配置するように反映して上記文書情報コンテンツの配置座標を演算する第２のサムネイル配置演算手段と；

　を有することを特徴とする関連発見型情報閲覧システム。

【書類名】明細書

【発明の名称】関連発見型情報閲覧システム

【発明の属する技術分野】

　【０００１】

　本発明は、ＨＴＭＬを始めとするマルチメディア文書の蓄積・検索等の分野において、文書情報コンテンツ間の関連と、利用者の興味と、コンテンツとの間の関連を同時に構造的に提示し、しかも、閲覧利用者の興味の変化に追従して、常に適切な文書情報コンテンツを構造的に提示することによって、漠然とした検索意図を持った閲覧利用者が、キーワード等を用いて明示的に自分の興味を表現できない場合でも、関連発見的に興味にあった文書情報コンテンツとの出会いへ導くための文書情報コンテンツ閲覧に関する。

【０００２】

【従来の技術】

　【０００３】

　検索は、質問を記述することによって、大量の情報の中から該当する情報を抽出する手法である。一般の利用者を想定した場合、質問は自然言語表現を用いるのが自然であるが、現状では、システム側の自然言語理解能力が実用レベルに至っていないので、質問には単数または複数の検索キーワードを使用することが多い。また、複数の検索キーワードを使用する場合には、ａｎｄ／ｏｒ等、上記複数のキーワード間の結合形式を表す論理記号を使用する。

【０００４】

　大量の検索対象コンテンツに対して、利用者が、明確な意図を持ち、また、上記意思をシステム側に理解させるだけの能力を持っている場合、言語的な検索キーは、効率的であり、全文検索型であり、しかも極めて強力な検索エンジンが多数存在している。たとえば、Ｗｅｂ上のＨＴＭＬドキュメントの検索システムとしては、ｇｏｏやＩｎｆｏＳｅｅｋ等が存在する。

【０００５】

　また、ＷＷＷは、ＨＴＭＬ文書中にＵＲＬで示されるリンクを持たせることによって、世界中のＷＷＷサーバに分散する多種多様な情報を閲覧することができる。

【０００６】

【発明が解決しようとする課題】

　【０００７】

　情報の候補数を減少させ、情報を横断的に扱う手段として、キーワード検索が広く使われている。

【０００８】

　しかし、キーワード検索で得られる全ての情報がユーザの興味に深く関連するとは限らないという問題がある。また、ユーザが自分の興味をキーワードとしてうまく表現できない場合には、利用が難しいという問題がある。

【０００９】

　特に、明確な意図を持たずに、カタログやガイドブック等を眺め見るように、ある種の情報ブラウジングの過程を支援するためのツールとして、上記キーワード検索の手法を用いることは、次の理由から、必ずしも効率的とは言えない。

　（１）漠然とした興味や意図をキーワードで表現することが難しい。

　（２）ユーザがその分野に詳しくない場合には、適切なキーワードを知らない可能性がある。

　【００１０】

　（３）興味の対象や視点が刻一刻と移り変わるような場合、これに追従するような検索を実現することが難しい。

【００１１】

　また、ＷＷＷは、文書中にＵＲＬで示されるリンクを持たせることによって、世界中のＷＷＷサーバに分散する多種多様な情報を閲覧できる。しかし、ここで表現されるリンクは、文書作成者の立場に基づいた、コンテンツ依存性の強いリンクであり、利用者の興味を反映させるようには構成されていないので、リンクを辿った場合、ユーザの興味に応じた情報を閲覧することは必ずしも容易ではない。

【００１２】

　つまり、リンクは、文書の著者によって明示的に張られたものであり、互いに関連する複数の文書同士であっても、著者がその関連を意識しなければ、互いに無関係の文書であるという扱いになる。したがって、興味や視点の変化は、著者によって張られた範囲でのみ、実現されるという問題がある。

【００１３】

　本発明は、特に明確な意図を持たずに、カタログやガイドブック等を眺め見るように、文書情報コンテンツを、直感的に提示し、閲覧することができる関連発見型情報閲覧システム、関連発見型情報閲覧方法および記録媒体を提供することを目的とするものである。

【００１４】

　また、本発明は、特に明確な意図を持たずに、カタログやガイドブック等を眺め見るように、文書情報コンテンツを、直感的に提示し、閲覧することができ、しかも、閲覧ユーザの興味に深く関連すると思われる情報文書情報コンテンツを、情報間の関連や構造とともに、直感的に提示し、閲覧することができる関連発見型情報閲覧システム、関連発見型情報閲覧方法および記録媒体を提供することを目的とするものである。

【００１５】

【課題を解決するための手段】

　【００１６】

　本発明は、複数の文書情報コンテンツを、蓄積する文書情報コンテンツ蓄積手段と、上記文書情報コンテンツの全文書を形態素解析し、出現頻度の高い名詞を上位からＮ個選び、次いで、各文書について、上記Ｎ個の名詞のそれぞれの出現頻度を表すＮ次元のベクトルを構成してコンテンツプロファイルを生成するコンテンツプロファイル生成手段と、上記文書情報コンテンツを表す代表画像であるサムネイルを提示するとともに、所定の閲覧利用者によってサムネイルがクリックされると、対応するコンテンツの詳細文書情報を提示する提示手段と、複数の文書情報コンテンツに対応する複数のサムネイルが閲覧利用者によって順次クリックされると、閲覧利用者によって閲覧された文書情報コンテンツに対応するコンテンツプロファイルの最新の所定回数の履歴から、コンテンツプロファイルの要素毎の平均値を各要素とするＮ次元の、閲覧利用者の興味を表すユーザプロファイルを演算するユーザプロファイル生成手段と、蓄積された複数の文書情報コンテンツに対し、２つのコンテンツプロファイルの間の関連の度合いをベクトルの距離尺度を用いて表し、コンテンツプロファイルの間の関連の度合いを、多次元尺度構成法を用いて２次元平面空間上の距離尺度として表現し、２次元平面上の文書情報コンテンツの配置座標を計算する第１のサムネイル配置演算手段と、上記ユーザプロファイルと各コンテンツプロファイルとの間の類似度を、ベクトルの大きさを正規化し、対応するベクトル要素間の距離の和に基づいてそれぞれ演算し、３次元的な提示手段を用いて画面垂直方向に演算された類似度が高いコンテンツに対応するサムネイル程、手前に配置するように反映して上記文書情報コンテンツの配置座標を演算する第２のサムネイル配置演算手段とを有することを特徴とする関連発見型情報閲覧システムである。

【００１７】

【発明の実施の形態および実施例】

　【００１８】

　図１は、本発明の実施の形態である装置の構成例を示すブロック図である。

【００１９】

　図２は、提示コンテンツの決定・配置のためのアルゴリズムを図式的に説明する為の図である。

【００２０】

　図３は、上記実施例の動作を示すフローチャートである。

【００２１】

　本装置は、文書情報コンテンツを蓄積する文書コンテンツ蓄積部１と、コンテンツの検索のための書誌情報であるコンテンツプロファイルを検出するコンテンツプロファイル生成部２と、閲覧利用者の興味・指向をあらわすユーザプロファイルを検出するユーザプロファイル生成部３と、コンテンツプロファイルおよびユーザプロファイルを用いて情報コンテンツ提示・配置を演算・決定するための提示コンテンツ決定・配置部４と、閲覧利用者に対する情報コンテンツの提示と情報コンテンツに対する利用者の閲覧操作を可能とする情報コンテンツ提示部５とで構成されている。

【００２２】

　まず、閲覧対象である大量の文書情報コンテンツを文書コンテンツ蓄積部１に記録蓄積する。

【００２３】

　あるいは、文書情報コンテンツ自体は本システム内部に必ずしも記録蓄積されている必要はなく、ネットワーク上に分散している情報コンテンツに対しても、コンテンツプロファイルと文書情報コンテンツに対するリンクとを文書コンテンツ蓄積部１に蓄積記録しておいても、本発明の目的は達成される。

【００２４】

　次に、コンテンツの検索のための書誌情報であるコンテンツプロファイルを検出するコンテンツプロファイル生成部２では、文書を形態素解析にかけ名詞を抽出、各名詞の文書中における出現数を求め、各名詞とそれらの出現数をコンテンツプロファイルとして記録保持する。コンテンツプロファイル並びにその検出手法の例としてたとえば次のような方法がある。

【００２５】

　まず、蓄積対象の文書全てを形態素解析にかけ名詞を抽出する。次いで、全ての文書にわたって出現頻度の高い名詞を上位からＮ個選び出す。次いで、各々の文書において、Ｎ個の名詞それぞれの出現数を表す次のようなＮ次元のキーワードベクトルを構成してコンテンツプロファイルとする。この例を図４に示してある。

【００２６】

　図４は、上記実施例において、選ばれたＮ種類の名詞と、文書中の出現数（キーワードベクトル）との対応例を示す図である。

【００２７】

　次に、閲覧利用者が情報コンテンツの閲覧を開始したら、閲覧利用者の興味・指向をあらわすユーザプロファイルを検出するユーザプロファイル生成部３では、閲覧利用者が閲覧した情報コンテンツの履歴をもとにユーザプロファイルを逐次生成する。ユーザプロファイルおよびその生成手法の例としてたとえば次のような方法がある。

【００２８】

　閲覧利用者が次々と情報コンテンツを閲覧している状況において、現時点から遡り過去Ｍ個の閲覧コンテンツのコンテンツプロファイル（キーワードベクトル）の要素毎の平均値を各要素とするようなＮ次元のベクトルを構成し、これをユーザプロファイルとする。たとえばＭ＝３の場合には、図５に示すようになる。

【００２９】

　図５は、上記実施例において、時刻ｔ－２に閲覧したコンテンツのプロファイルと、時刻ｔ－１に閲覧したコンテンツのプロファイルと、現時点ｔに閲覧したコンテンツのプロファイルと、現時点ｔにおけるユーザプロファイルとの対応例を示す図である。

　このとき、コンテンツプロファイルおよびユーザプロファイルを用いて情報コンテンツ提示・配置を演算決定するための提示コンテンツ決定・配置部４では、複数の情報コンテンツ間の関連の度合いを、コンテンツプロファイルをもとに計算する。また、ユーザの興味・指向と情報コンテンツとの間の関連の度合いをコンテンツプロファイルとユーザプロファイルをもとに計算する。たとえば、コンテンツ間のプロファイルの関連の度合いを次のようなベクトルの距離尺度を用いてあらわすことができる。

【００３０】

　コンテンツｉに対するコンテンツプロファイルベクトル

　Ｖi＝（ｃi,1，ｃi,2，…，ｃi,N）

　コンテンツｉとコンテンツｊとの関連度合い

【００３１】

【数１】

　【００３２】

　次に提示コンテンツ決定・配置部４では、全ての情報コンテンツの間の関連度合いを多次元尺度構成法を用いて２次元平面空間上の距離尺度として表現し、２次元平面上のコンテンツの配置（座標）を計算しておく。

【００３３】

　また、たとえば、ユーザプロファイルとコンテンツプロファイルとの類似度を次の計算によって求めることができる。

【００３４】

　コンテンツプロファイルベクトルＶc＝（ｃ1，ｃ2…）

　ユーザプロファイルベクトルＶu＝（ｕ1，ｕ2…）

　ユーザプロファイルとコンテンツプロファイルとの類似度

【００３５】

【数２】

　【００３６】

　閲覧利用者に対する情報コンテンツの提示と情報コンテンツに対する利用者の閲覧操作を可能とする情報コンテンツ提示部５では、提示コンテンツ決定・配置部４において計算された、情報コンテンツ間の関連と、利用者の興味・指向とコンテンツの間の関連を同時に構造的に提示し、かつ、閲覧利用者の興味・指向の変化に追従して常に適切な情報コンテンツ（サムネイル）を構造的に提示する。たとえば、図１の左側画面のように、コンテンツ間の関連、すなわち多次元尺度構成法を用いて計算された２次元平面上の位置、を画面並行平面方向に提示する（関連の大きい文書のサムネイル同士は平面方向近く、関連の薄い文書のサムネイル同士は離れて提示する）。また、閲覧利用者の興味・指向を反映したユーザプロファイルと各々のコンテンツプロファイルとの類似度をＶＲＭＬ等のような３次元的な提示手段を用いて画面垂直方向に提示する（利用者の興味・指向に近いコンテンツのサムネイルほど手前に提示する）。

【００３７】

　図２は、提示コンテンツの決定・配置のためのアルゴリズムを図式的に説明するための図である。図２において、もっとも右側の平面は、提示コンテンツ決定・配置部４で、式（１）を用いて計算した全ての情報コンテンツの間の関連度合いを、多次元尺度構成法を用いて２次元平面空間上の距離尺度として表現し、２次元平面上のコンテンツの配置（座標）を計算して作成した平面である。

【００３８】

　提示コンテンツ決定・配置部４ではさらに、ユーザプロファイルベクトルと各々のコンテンツとの類似度を式（２）を用いて計算し、類似度が図２の中ほどの提示閾値を超えたコンテンツのサムネイルを情報コンテンツ提示部５に提示する。このときの閲覧利用者の視点からの距離は、図２のように式（２）による類似度をもとに提示位置が決定される。したがって、閲覧利用者の興味・指向に近いコンテンツほど画面上手前に提示されることになる。

【００３９】

　情報コンテンツ提示部５では、閲覧利用者がコンテンツをあらわすサムネイルをクリックすることにより、情報コンテンツの詳細が提示されるとともに、閲覧履歴がユーザプロファイル生成部３に送信される。

【００４０】

　閲覧履歴の送信によって、ユーザプロファイル生成部３内部で式（２）によるユーザプロファイルの更新が行われ、情報コンテンツ提示部５に提示されるコンテンツサムネイルおよびその配置（奥行き方向の配置）が逐次変更される。

【００４１】

　また、文書コンテンツ蓄積部１は、複数の文書情報コンテンツを、蓄積する文書情報コンテンツ蓄積手段の例である。

【００４２】

　コンテンツプロファイル生成部２は、文書情報コンテンツの文書から、その文書内容を表すベクトル型データであるコンテンツプロファイルを生成するコンテンツプロファイル生成手段の例である。

【００４３】

　ユーザプロファイル生成部３は、複数の文書情報コンテンツに対応する複数のサムネイルが閲覧利用者によって順次クリックされると、閲覧利用者によって閲覧された情報コンテンツに対応するコンテンツプロファイルの履歴から、閲覧利用者の興味を表すユーザプロファイルを演算するユーザプロファイル生成手段の例である。

【００４４】

　図１において、「イ鮨」、「ロすし」、「ハ寿司」、「竹そば」等がサムネイルの例である。

【００４５】

　提示コンテンツ決定・配置部４は、蓄積された複数の情報コンテンツに対応する複数のコンテンツプロファイルの間の類似度を演算することと、上記ユーザプロファイルとコンテンツプロファイルとの間の類似度を演算することにより、上記提示手段上へのサムネイルの配置を演算するサムネイル配置演算手段の例である。

【００４６】

　文書情報コンテンツ提示部５は、文書情報コンテンツを表す代表画像であるサムネイルを提示するとともに、所定の閲覧利用者によってサムネイルがクリックされると、対応するコンテンツの詳細文書情報を提示する提示手段の例である。

【００４７】

　また、上記サムネイル配置演算手段は、上記文書コンテンツに対する閲覧利用者の興味の度合いが高いコンテンツに対応するサムネイルほど、手前に配置するように演算する手段であり、上記閲覧利用者が所定のサムネイルをクリックする毎に、上記提示手段に提示する上記サムネイルの配置を演算しなおす手段であり上記文書情報コンテンツ同士の関連が強いほど、互いの平面距離を短くするように、サムネイルの配置を演算する手段である。

【００４８】

　なお、上記実施例を、記録媒体の発明として把握することができる。つまり、上記実施例は、複数の文書情報コンテンツを、蓄積する文書情報コンテンツ蓄積手順と、上記文書情報コンテンツの文書から、その文書内容を表すベクトル型データであるコンテンツプロファイルを生成するコンテンツプロファイル生成手順と、上記文書情報コンテンツを表す代表画像であるサムネイルを提示するとともに、所定の閲覧利用者によってサムネイルがクリックされると、対応するコンテンツの詳細文書情報を提示する提示手順と、複数の文書情報コンテンツに対応する複数のサムネイルが閲覧利用者によってクリックされると、閲覧利用者によって閲覧された情報コンテンツに対応するコンテンツプロファイルの履歴から、閲覧利用者の興味を表すユーザプロファイルを演算するユーザプロファイル生成手順と、蓄積された複数の情報コンテンツに対応する複数のコンテンツプロファイルの間の類似度を演算することと、上記ユーザプロファイルとコンテンツプロファイルとの間の類似度を演算することにより、提示手段上へのサムネイルの配置を演算するサムネイル配置演算手順とをコンピュータに実行させる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の例である。

【００４９】

　また、上記記録媒体の例としては、ＦＤ、ＣＤ、ＤＶＤ、ＨＤ、半導体メモリが考えられる。

【００５０】

【発明の効果】

　【００５１】

　本発明によれば、閲覧ユーザの興味に深く関連すると思われる文書情報コンテンツを、直感的に提示し、閲覧させることができるという効果を奏する。

【図面の簡単な説明】

　【００５２】

【図１】本発明の実施の形態である装置の構成例を示すブロック図である。

【図２】提示コンテンツの決定・配置のためのアルゴリズムを図式的に説明するための図である。

【図３】上記実施例の動作を示すフローチャートである。

【図４】上記実施例において、選ばれたＮ種類の名詞と、文書中の出現数（キーワードベクトル）との対応例を示す図である。

【図５】上記実施例において、時刻ｔ－２に閲覧したコンテンツのプロファイルと、時刻ｔ－１に閲覧したコンテンツのプロファイルと、現時点ｔに閲覧したコンテンツのプロファイルと、現時点ｔにおけるユーザプロファイルとの対応例を示す図である。

【符号の説明】

　【００５３】

　１００…関連発見型情報閲覧システム、

　１…文書コンテンツ蓄積部、

　２…コンテンツプロファイル生成部、

　３…ユーザプロファイル生成部、

　４…提示コンテンツ決定・配置部、

　５…文書情報コンテンツ提示部。