

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5494126号
(P5494126)

(45) 発行日 平成26年5月14日(2014.5.14)

(24) 登録日 平成26年3月14日(2014.3.14)

(51) Int. Cl. F 1
G 0 6 F 17/30 (2006.01)
 G 0 6 F 17/30 3 4 0 A
 G 0 6 F 17/30 3 5 0 C
 G 0 6 F 17/30 1 1 0 F

請求項の数 7 (全 21 頁)

<p>(21) 出願番号 特願2010-80217 (P2010-80217) (22) 出願日 平成22年3月31日 (2010.3.31) (65) 公開番号 特開2011-215679 (P2011-215679A) (43) 公開日 平成23年10月27日 (2011.10.27) 審査請求日 平成25年1月18日 (2013.1.18)</p>	<p>(73) 特許権者 000002897 大日本印刷株式会社 東京都新宿区市谷加賀町一丁目1番1号 (74) 代理人 100096091 弁理士 井上 誠一 (72) 発明者 小林 宣幸 東京都新宿区市谷加賀町一丁目1番1号 大日本印刷株式会社内 審査官 梅本 達雄 (56) 参考文献 特開2008-071303 (JP, A)) 特開2005-062987 (JP, A)) 最終頁に続く</p>
--	---

(54) 【発明の名称】 文書推薦システム、文書推薦装置、文書推薦方法、及びプログラム

(57) 【特許請求の範囲】

【請求項1】

ユーザ端末とサーバとがネットワークを介して通信接続され、複数の前記ユーザ端末と前記サーバとの間で文書に関する情報を共有する文書推薦システムであって、

前記ユーザ端末が、既読文書に対する評価として理解度及び有益度を付与し、サーバに登録する登録手段と、

前記サーバにおいて、前記登録手段により登録された文書全体から有益度が比較的高い文書を抽出し、理解度レベル別に推薦文書群を作成する推薦文書群抽出手段と、

前記サーバにおいて、各ユーザの前記既読文書に対する前記理解度に応じた推薦文書を、対応する理解度レベルの前記推薦文書群から検索する検索手段と、

前記サーバにおいて、前記検索手段により検索された推薦文書の情報を前記ユーザ端末に提示する提示手段と、を備え、

前記推薦文書群抽出手段は、前記文書全体から少なくとも平易な有益文書群と、難解な有益文書群とを抽出し、

前記検索手段は、ユーザの理解度の低い分野については前記平易な有益文書群から推薦文書を検索し、ユーザの理解度の高い分野については前記難解な有益文書群から推薦文書を検索することを特徴とする文書推薦システム。

【請求項2】

前記検索手段は、

前記推薦文書群の各文書の特徴を表現する文書ベクトルを生成する文書ベクトル生成手

段と、

前記既読文書の特徴を文書ベクトルの形式で表現する検索クエリを生成する検索クエリ生成手段と、

前記文書ベクトル生成手段によって生成された前記推薦文書群の各文書の文書ベクトルと、前記検索クエリ生成手段により生成された検索クエリとの類似度を算出する類似度算出手段と、

前記類似度算出手段によって算出された類似度に基づいて、推薦文書を決定する推薦文書決定手段と、

を備えることを特徴とする請求項 1 に記載の文書推薦システム。

【請求項 3】

ユーザ端末とサーバとがネットワークを介して通信接続され、複数の前記ユーザ端末と前記サーバとの間で文書に関する情報を共有する文書推薦システムであって、

前記ユーザ端末が、既読文書に対する評価として理解度及び有益度を付与し、サーバに登録する登録手段と、

前記サーバにおいて、前記登録手段により登録された文書全体から有益度が比較的高い文書を抽出し、理解度レベル別に推薦文書群を作成する推薦文書群抽出手段と、

前記サーバにおいて、各ユーザの前記既読文書に対する前記理解度に応じた推薦文書を、対応する理解度レベルの前記推薦文書群から検索する検索手段と、

前記サーバにおいて、前記検索手段により検索された推薦文書の情報を前記ユーザ端末に提示する提示手段と、を備え、

前記検索手段は、

前記推薦文書群の各文書の特徴を表現する文書ベクトルを生成する文書ベクトル生成手段と、

前記既読文書の特徴を文書ベクトルの形式で表現する検索クエリを生成する検索クエリ生成手段と、

前記文書ベクトル生成手段によって生成された前記推薦文書群の各文書の文書ベクトルと、前記検索クエリ生成手段により生成された検索クエリとの類似度を算出する類似度算出手段と、

前記類似度算出手段によって算出された類似度に基づいて、推薦文書を決定する推薦文書決定手段と、を備え、

前記検索クエリ生成手段は、全ユーザが前記文書に対して付与した理解度から該文書の理解度の基準値を求め、該基準値に対する相対的なユーザの理解度を重みとして算出する理解度重み算出手段を更に備え、

前記理解度重み算出手段により算出された重みを、前記既読文書の特徴を表現する文書ベクトルに乗じて、前記検索クエリを生成することを特徴とする文書推薦システム。

【請求項 4】

前記提示手段は、前記検索手段により検索された推薦文書が複数ある場合は、更に有益度の高いものから順に前記ユーザ端末に提示することを特徴とする請求項 1 から請求項 3 のいずれかに記載の文書推薦システム。

【請求項 5】

ネットワークを介してユーザ端末と通信接続され、ユーザ端末に対して文書の推薦情報を提示する文書推薦装置であって、

複数のユーザ端末から、既読文書に対する評価として理解度及び有益度の付与を受け付ける評価受付手段と、

前記評価受付手段によって評価の付与された文書全体から有益度が比較的高い文書を抽出し、理解度レベル別に推薦文書群を作成する推薦文書群抽出手段と、

各ユーザの前記既読文書に対する前記理解度に応じた推薦文書を、対応する理解度レベルの前記推薦文書群から検索する検索手段と、

前記検索手段により検索された推薦文書の情報を前記ユーザ端末に提示する提示手段と、を備え、

10

20

30

40

50

前記推薦文書群抽出手段は、前記文書全体から少なくとも平易な有益文書群と、難解な有益文書群とを抽出し、

前記検索手段は、ユーザの理解度の低い分野については前記平易な有益文書群から推薦文書を検索し、ユーザの理解度の高い分野については前記難解な有益文書群から推薦文書を検索することを特徴とする文書推薦装置。

【請求項6】

ユーザ端末とサーバとがネットワークを介して通信接続され、複数の前記ユーザ端末と前記サーバとの間で文書に関する情報を共有する文書推薦システムにおける文書推薦方法であって、

前記ユーザ端末が、既読文書に対する評価として理解度及び有益度を付与し、サーバに登録するステップと、

前記サーバにおいて、登録された文書全体から有益度が比較的高い文書を抽出し、理解度レベル別に推薦文書群を作成する推薦文書群抽出ステップと、

前記サーバにおいて、各ユーザの前記既読文書に対する前記理解度に応じた推薦文書を、対応する理解度レベルの前記推薦文書群から検索する検索ステップと、

前記サーバにおいて、検索された推薦文書の情報を前記ユーザ端末に提示する提示ステップと、を含み、

前記推薦文書群抽出ステップは、前記文書全体から少なくとも平易な有益文書群と、難解な有益文書群とを抽出し、

前記検索ステップは、ユーザの理解度の低い分野については前記平易な有益文書群から推薦文書を検索し、ユーザの理解度の高い分野については前記難解な有益文書群から推薦文書を検索することを特徴とする文書推薦方法。

【請求項7】

コンピュータを請求項5に記載の文書推薦装置として機能させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、ユーザに適した文書を推薦する文書推薦システム等に関する。

【背景技術】

【0002】

近年、情報技術の発達やインターネットの普及に伴い、情報源が多様化し、大量の情報が蓄積されている。そのため、大量の蓄積情報を整理したり、ユーザが望む情報を適切に入手したりするための情報フィルタリング技術が研究されている。また、情報フィルタリング技術を利用して、ユーザに適した情報を推薦するリコメンデーション技術（情報推薦）が提案されている。

例えば、特許文献1では、所定の専門分野に関連する語彙または情報を具備する専門知識リストを作り、情報提供サービスにおけるユーザの操作履歴と専門知識リスト内の情報の認知度に応じて各ユーザの信頼度を評価し、信頼度の高いユーザ間で認知されている情報を専門分野に関する情報として推薦することが提案されている。

【0003】

ところで、書籍や文書を推薦する場合は、推薦対象とするユーザの既読文書の傾向や知識レベルに応じた文書を推薦することが望まれる。

【先行技術文献】

【特許文献】

【0004】

【特許文献1】特開2004-192542号公報

【発明の概要】

【発明が解決しようとする課題】

【0005】

しかしながら、上述の特許文献1では情報推薦者の信頼度が評価されるのみで、推薦のターゲットとするユーザの知識レベルを評価していない。また、従来の書籍購入サイト等では、既読ユーザが商品に対する評価やレビューを登録できるようになっているものがあるが、これらの評価は評価者の知識レベルが専門レベルか入門レベルかが問われていない。そのため、評価やユーザの購入履歴に基づいてユーザの読書分野の傾向を判断し、適した書籍を推薦するとしても、ユーザの知識レベルに適合したものが推薦されるとは限らず、その結果、ユーザの知識レベルが高いにも関わらず入門レベルの書籍を提示してしまったり、知識レベルが低いにも関わらず専門レベルの書籍を提示してしまうことがあった。

【0006】

本発明は、このような課題に鑑みてなされたもので、その目的とするところは、ユーザの知識レベルに適合した文書を推薦する文書推薦システム等を提供することである。

【課題を解決するための手段】

【0007】

前述した課題を解決するため第1の発明は、ユーザ端末とサーバとがネットワークを介して通信接続され、複数の前記ユーザ端末と前記サーバとの間で文書に関する情報を共有する文書推薦システムであって、前記ユーザ端末が、既読文書に対する評価として理解度及び有益度を付与し、サーバに登録する登録手段と、前記サーバにおいて、前記登録手段により登録された文書全体から有益度が比較的高い文書を抽出し、理解度レベル別に推薦文書群を作成する推薦文書群抽出手段と、前記サーバにおいて、各ユーザの前記既読文書に対する前記理解度に応じた推薦文書を、対応する理解度レベルの前記推薦文書群から検索する検索手段と、前記サーバにおいて、前記検索手段により検索された推薦文書の情報を前記ユーザ端末に提示する提示手段と、を備え、前記推薦文書群抽出手段は、前記文書全体から少なくとも平易な有益文書群と、難解な有益文書群とを抽出し、前記検索手段は、ユーザの理解度の低い分野については前記平易な有益文書群から推薦文書を検索し、ユーザの理解度の高い分野については前記難解な有益文書群から推薦文書を検索することを特徴とする文書推薦システムである。

【0008】

第1の発明によれば、複数のユーザ端末とサーバとの間で文書に関する情報を共有する文書推薦システムにおいて、ユーザ端末が、既読文書に対する評価として理解度及び有益度を付与し、サーバに登録すると、サーバは、登録されている文書全体から有益度が比較的高い文書を抽出し、理解度レベル別に推薦文書群を作成する。また、サーバは、各ユーザの既読文書に対する理解度に応じた推薦文書を、対応する理解度レベルの推薦文書群から検索し、検索された推薦文書の情報をユーザ端末に提示する。

従って、サーバは、全ユーザの文書に対する評価を基に理解度レベルに応じた推薦文書群を抽出でき、また、その中から各ユーザの既読文書の理解度に応じた文書を検索して推薦できる。そのため、ユーザの理解度レベル(知識レベル)に適合した文書を推薦することが可能となる。その結果、文書に対するコンバージョン(閲覧、購買等)を向上させることが可能となる。

また、ユーザの理解度に応じて、ユーザの苦手分野に対しては易しくて有益な文書を推薦でき、ユーザの得意分野に対しては難しいが有益な文書を推薦できるようになる。

【0010】

また、前記検索手段は、前記推薦文書群の各文書の特徴を表現する文書ベクトルを生成する文書ベクトル生成手段と、前記既読文書の特徴を文書ベクトルの形式で表現する検索クエリを生成する検索クエリ生成手段と、前記文書ベクトル生成手段によって生成された前記推薦文書群の各文書の文書ベクトルと、前記検索クエリ生成手段により生成された検索クエリとの類似度を算出する類似度算出手段と、前記類似度算出手段によって算出された類似度に基づいて、推薦文書を決定する推薦文書決定手段と、を備えることが望ましい。

。

10

20

30

40

50

これにより、サーバによる推薦文書の検索の際に、文書の特徴を表現する文書ベクトルによりユーザの既読文書の特徴を求め、その特徴に類似する推薦文書を決定できる。その結果ユーザの既読文書の傾向に応じた適切な文書を推薦できるようになる。

【0011】

第2の発明は、ユーザ端末とサーバとがネットワークを介して通信接続され、複数の前記ユーザ端末と前記サーバとの間で文書に関する情報を共有する文書推薦システムであって、前記ユーザ端末が、既読文書に対する評価として理解度及び有益度を付与し、サーバに登録する登録手段と、前記サーバにおいて、前記登録手段により登録された文書全体から有益度が比較的高い文書を抽出し、理解度レベル別に推薦文書群を作成する推薦文書群抽出手段と、前記サーバにおいて、各ユーザの前記既読文書に対する前記理解度に応じた推薦文書を、対応する理解度レベルの前記推薦文書群から検索する検索手段と、前記サーバにおいて、前記検索手段により検索された推薦文書の情報を前記ユーザ端末に提示する提示手段と、を備え、前記検索手段は、前記推薦文書群の各文書の特徴を表現する文書ベクトルを生成する文書ベクトル生成手段と、前記既読文書の特徴を文書ベクトルの形式で表現する検索クエリを生成する検索クエリ生成手段と、前記文書ベクトル生成手段によって生成された前記推薦文書群の各文書の文書ベクトルと、前記検索クエリ生成手段により生成された検索クエリとの類似度を算出する類似度算出手段と、前記類似度算出手段によって算出された類似度に基づいて、推薦文書を決定する推薦文書決定手段と、を備え、前記検索クエリ生成手段は、全ユーザが前記文書に対して付与した理解度から該文書の理解度の基準値を求め、該基準値に対する相対的なユーザの理解度を重みとして算出する理解度重み算出手段を更に備え、前記理解度重み算出手段により算出された重みを、前記既読文書の特徴を表現する文書ベクトルに乗じて、前記検索クエリを生成することを特徴とする文書推薦システムである。

第2の発明によれば、複数のユーザ端末とサーバとの間で文書に関する情報を共有する文書推薦システムにおいて、ユーザ端末が、既読文書に対する評価として理解度及び有益度を付与し、サーバに登録すると、サーバは、登録されている文書全体から有益度が比較的高い文書を抽出し、理解度レベル別に推薦文書群を作成する。また、サーバは、各ユーザの既読文書に対する理解度に応じた推薦文書を、対応する理解度レベルの推薦文書群から検索し、検索された推薦文書の情報をユーザ端末に提示する。

従って、サーバは、全ユーザの文書に対する評価を基に理解度レベルに応じた推薦文書群を抽出でき、また、その中から各ユーザの既読文書の理解度に応じた文書を検索して推薦できる。そのため、ユーザの理解度レベル（知識レベル）に適合した文書を推薦することが可能となる。その結果、文書に対するコンバージョン（閲覧、購買等）を向上させることが可能となる。

また、サーバによる推薦文書の検索の際に、文書の特徴を表現する文書ベクトルによりユーザの既読文書の特徴を求め、その特徴に類似する推薦文書を決定できる。その結果ユーザの既読文書の傾向に応じた適切な文書を推薦できるようになる。

さらに、全ユーザの理解度を基準とする各ユーザの理解度を相対的に評価することが可能となり、ユーザの既読文書の傾向と、全体の中での個々のユーザの知識レベルとを解析して、ユーザに最適な文書を推薦することが可能となる。

【0012】

また、前記提示手段は、前記検索手段により検索された推薦文書が複数ある場合は、更に有益度の高いものから順に前記ユーザ端末に提示することが望ましい。

これにより、検索された推薦文書が複数ある場合は、更に有益度の高いものから順にユーザ端末に提示することが可能となり、ユーザにとってより有益な文書推薦システムを提供できる。

【0013】

第3の発明は、ネットワークを介してユーザ端末と通信接続され、ユーザ端末に対して文書の推薦情報を提示する文書推薦装置であって、複数のユーザ端末から、既読文書に対す

る評価として理解度及び有益度の付与を受け付ける評価受付手段と、前記評価受付手段によって評価の付与された文書全体から有益度が比較的高い文書を抽出し、理解度レベル別に推薦文書群を作成する推薦文書群抽出手段と、各ユーザの前記既読文書に対する前記理解度に応じた推薦文書を、対応する理解度レベルの前記推薦文書群から検索する検索手段と、前記検索手段により検索された推薦文書の情報を前記ユーザ端末に提示する提示手段と、を備え、前記推薦文書群抽出手段は、前記文書全体から少なくとも平易な有益文書群と、難解な有益文書群とを抽出し、前記検索手段は、ユーザの理解度の低い分野については前記平易な有益文書群から推薦文書を検索し、ユーザの理解度の高い分野については前記難解な有益文書群から推薦文書を検索することを特徴とする文書推薦装置である。

10

【0014】

第3の発明によれば、文書推薦装置は、複数のユーザ端末から、既読文書に対する評価として理解度及び有益度の付与を受け付け、評価の付与された文書全体から有益度が比較的高い文書を抽出し、理解度レベル別に推薦文書群を作成する。また、各ユーザの既読文書に対する理解度に応じた推薦文書を、対応する理解度レベルの推薦文書群から検索し、ユーザ端末に提示する。

従って、文書推薦装置は、全ユーザの文書に対する評価を基に理解度レベルに応じた推薦文書群を抽出でき、また、その中から各ユーザの既読文書の理解度に応じた文書を検索して推薦できる。そのため、ユーザの理解度レベルに適合した文書を推薦することが可能となる。その結果、文書に対するコンバージョン（閲覧、購買等）を向上させることが可能となる。

20

また、ユーザの理解度に応じて、ユーザの苦手分野に対しては易しくて有益な文書を推薦でき、ユーザの得意分野に対しては難しいが有益な文書を推薦できるようになる。

【0015】

第4の発明は、ユーザ端末とサーバとがネットワークを介して通信接続され、複数の前記ユーザ端末と前記サーバとの間で文書に関する情報を共有する文書推薦システムにおける文書推薦方法であって、前記ユーザ端末が、既読文書に対する評価として理解度及び有益度を付与し、サーバに登録するステップと、前記サーバにおいて、登録された文書全体から有益度が比較的高い文書を抽出し、理解度レベル別に推薦文書群を作成する推薦文書群抽出ステップと、前記サーバにおいて、各ユーザの前記既読文書に対する前記理解度に応じた推薦文書を、対応する理解度レベルの前記推薦文書群から検索する検索ステップと、前記サーバにおいて、検索された推薦文書の情報を前記ユーザ端末に提示する提示ステップと、を含み、前記推薦文書群抽出ステップは、前記文書全体から少なくとも平易な有益文書群と、難解な有益文書群とを抽出し、前記検索ステップは、ユーザの理解度の低い分野については前記平易な有益文書群から推薦文書を検索し、ユーザの理解度の高い分野については前記難解な有益文書群から推薦文書を検索することを特徴とする文書推薦方法である。

30

【0016】

第4の発明の文書推薦方法によれば、複数のユーザ端末とサーバとの間で文書に関する情報を共有する文書推薦システムにおいて、ユーザ端末が、既読文書に対する評価として理解度及び有益度を付与し、サーバに登録すると、サーバは、登録されている文書全体から有益度が比較的高い文書を抽出し、理解度レベル別に推薦文書群を作成する。また、サーバは、各ユーザの既読文書に対する理解度に応じた推薦文書を、対応する理解度レベルの推薦文書群から検索し、検索された推薦文書の情報をユーザ端末に提示する。

40

従って、サーバは、全ユーザの文書に対する評価を基に理解度レベルに応じた推薦文書群を抽出でき、また、その中から各ユーザの既読文書の理解度に応じた文書を検索して推薦できる。そのため、ユーザの理解度レベルに適合した文書を推薦することが可能となる。その結果、文書に対するコンバージョン（閲覧、購買等）を向上させることが可能とな

50

る。

また、ユーザの理解度に応じて、ユーザの苦手分野に対しては易しくて有益な文書を推薦でき、ユーザの得意分野に対しては難しいが有益な文書を推薦できるようになる。

【 0 0 1 7 】

第5の発明は、コンピュータを第3の発明の文書推薦装置として機能させるためのプログラムである。

10

【発明の効果】

【 0 0 1 8 】

本発明によれば、ユーザの知識レベルに応じた適切な文書を推薦する文書推薦装置等を提供できる。

20

【図面の簡単な説明】

【 0 0 1 9 】

【図1】本発明に係る文書推薦システム1の全体構成図

【図2】文書推薦システム1を利用した書棚共有コミュニティサービスの概念図

【図3】書誌データベース3が保持する書誌情報31の内容を示す図

【図4】文書推薦サーバ5が保持するユーザ情報51、文書評価記事情報52、ユーザプロフィール53の内容を示す図

【図5】文書推薦サーバ5及びユーザ端末2として利用されるコンピュータのハードウェア構成図

30

【図6】事前処理について説明するフローチャート

【図7】事前処理における文書の分類処理について説明する図

【図8】文書推薦処理の流れを説明するフローチャート

【図9】ユーザによる文書リストの登録について説明する図

【図10】ユーザプロフィールの作成（苦手分野と得意分野の決定）について説明する図

【図11】ユーザの苦手分野検索クエリ53dについて説明する図

【図12】ユーザの得意分野検索クエリ53eについて説明する図

【図13】ユーザの「苦手分野」に適切な易しくて有益な文書の選定について説明する図

【図14】ユーザの「得意分野」に適切な難しいが有益な文書の選定について説明する図

40

【発明を実施するための形態】

【 0 0 2 0 】

以下、図面に基づいて本発明の好適な実施形態について詳細に説明する。

まず、図1～図5を参照して本発明の構成について説明する。

【 0 0 2 1 】

図1は、本発明に係る文書推薦システム1のシステム構成を示す図である。

図1に示すように、文書推薦システム1は、複数のユーザ端末2、書誌データベース3、及び文書推薦サーバ5がネットワーク9を介して互いに通信接続されて構成される。

本実施の形態では、複数ユーザ間でネットワーク9を介して互いに既読文書に関する情報を共有する書棚共有コミュニティサービスにて本発明に係る文書推薦システム1を利用

50

する例を示す。

【0022】

書棚共有コミュニティサービスとは、ネットワーク9を介して各ユーザ端末2から文書推薦サーバ5に登録された既読文書に関する情報を仮想的な書棚として扱い、複数のユーザ間で各自の書棚、既読文書、及び文書の評価等を交換し合うものである。図2に示すように、文書推薦サーバ5は各ユーザ端末2に個別のユーザ専用の書棚画面21を提供したり、他ユーザの書棚画面21を閲覧可能に提供したりする。

なお、本発明において、文書とは、書籍、雑誌、記事、論文、その他の書類を含むが、紙媒体に限定されず、電子化された書籍、雑誌、記事、論文、その他、インターネットで公開される記事コンテンツのような電子媒体であってもよい。本実施の形態では、説明を簡略に行うための一例として、紙媒体で流通している書籍の推薦について例示する。

10

【0023】

ユーザ端末2は、書棚共有コミュニティサービスの会員ユーザが使用するコンピュータであり、Webブラウザがインストールされている。ユーザ端末2はWebブラウザを介して書棚共有コミュニティサービスを提供するサイトのサーバ(文書推薦サーバ5)へのアクセスが可能となる。ユーザ端末2が文書推薦サーバ5へログインすると、各ユーザ専用の書棚画面21(図2参照)を受信し、ユーザ端末2の表示部に表示する。ユーザ端末2は、書棚画面21や後述する文書登録画面(図9参照)を利用して、文書推薦サーバ5との間で文書登録及び推薦文書の表示に関する種々の処理を実行する。

なお、ユーザ端末2は、汎用的なコンピュータに代えて、電子書籍閲覧端末、デジタルテレビ、携帯電話、モバイル端末、通信機能搭載のゲーム機などで実現することもできる。

20

【0024】

書誌データベース3は、書誌情報31を蓄積して記憶するデータベースである。

図3に書誌データベース3にて取り扱う書誌情報31の例を示す。

図3に示すように、書誌情報31とは、各文書の「文書ID」、「タイトル」、「カテゴリ」、「内容」、「著者」等の情報を含むものである。書誌情報31の「内容」には、文書の本文、目次、または説明文等の情報が含まれる。

なお、本実施の形態では、現実に流通している書籍に関する書誌情報が登録されているものとするが、上述のように、より広い概念の各種文書(書籍、雑誌、記事、論文、記事コンテンツ等)の書誌情報が登録されているものとしてもよい。

30

【0025】

文書推薦サーバ5は、書棚共有コミュニティサービスを提供するサイトのサーバ装置であり、書棚共有コミュニティサイトの会員ユーザに関するユーザ情報51や、各ユーザ端末2から登録された文書の評価記事情報52、全ユーザのユーザプロフィール53等を管理する。

【0026】

図4に文書推薦サーバ5にて取り扱うデータの例を示す。

図4(a)に示すように、ユーザ情報51は、「ユーザID」、「ユーザ名」、「ユーザ属性情報」、「文書評価記事ID」等の情報が含まれる。

40

「ユーザ属性情報」とは、ユーザの年齢、性別、職業等のユーザ固有の情報であり、「文書評価記事ID」とは、当該ユーザの書棚に登録されている文書評価記事のリストである。文書評価記事とは、文書に対するユーザの評価であり、有益度、理解度、コメント等の情報が含まれる。

【0027】

図4(b)に示すように、文書評価記事情報52は、各ユーザから登録された全ての文書評価記事情報を管理するものであり、「文書評価記事ID」、「文書ID」、「有益度」、「理解度」、「コメント」等の情報が含まれる。

「有益度」とは評価したユーザにとって文書が役に立った度合いを表現する指標であり、例えば、0~5の数値等で表される。

50

「理解度」とは評価したユーザにとって文書の理解しやすさを表現する指標であり、例えば、0～5の数値等で表される。

「コメント」とは評価対象（文書）に対する感想（文字列）である。

【0028】

図4(c)に示すように、ユーザプロフィール53とは、各ユーザから登録された文書リストや文書評価記事情報と、それらの情報に基づいて文書推薦サーバ5が作成した個々のユーザについてのプロフィール情報とを含むものである。すなわち、文書推薦サーバ5は、ユーザから登録された既読文書を文書リスト53aとしてユーザプロフィール53に登録する。また、文書リスト53aやユーザ情報51の文書評価記事IDに基づいて、ユーザの既読文書を苦手分野文書53b、得意分野文書53cに分類し、苦手分野検索クエリ53d、得意分野検索クエリ53eを作成し、ユーザプロフィール53として記憶する。詳細は後述する。

10

【0029】

次に、図5を参照して、ユーザ端末2、文書推薦サーバ5として利用されるコンピュータのハードウェア構成を説明する。

コンピュータは、図5に示すように、例えば、制御部11、記憶部12、メディア入出力部13、入力部14、表示部15、通信部16、周辺機器I/F部17等がバス18を介して接続されて構成される。

【0030】

制御部11は、CPU(Central Processing Unit)、ROM(Read Only Memory)、RAM(Random Access Memory)等により構成される。

20

CPUは、記憶部12、ROM、記録媒体等に格納されるプログラムをRAM上のワークメモリ領域に呼び出して実行し、バス18を介して接続された各部を駆動制御する。ROMは、コンピュータのブートプログラムやBIOS等のプログラム、データ等を恒久的に保持する。RAMは、ロードしたプログラムやデータを一時的に保持するとともに、制御部11が各種処理を行うために使用するワークエリアを備える。

【0031】

記憶部12は、HDD(ハードディスクドライブ)であり、制御部11が実行するプログラムや、プログラム実行に必要なデータ、OS(オペレーティング・システム)等が格納されている。これらのプログラムコードは、制御部11により必要に応じて読み出されてRAMに移され、CPUに読み出されて実行される。

30

【0032】

メディア入出力部13は、例えば、HD(ハードディスク)ドライブ、フロッピー(登録商標)ディスクドライブ、PDドライブ、CDドライブ、DVDドライブ、MOドライブ等のメディア入出力装置であり、データの入出力を行う。

入力部14は、例えば、キーボード、マウス等のポインティング・デバイス、テンキー等の入力装置であり、入力されたデータを制御部11へ出力する。

表示部15は、例えば液晶パネル、CRTモニタ等のディスプレイ装置と、ディスプレイ装置と連携して表示処理を実行するための論理回路(ビデオアダプタ等)で構成され、制御部11の制御により入力された表示情報をディスプレイ装置上に表示させる。

40

周辺機器I/F(インタフェース)部17は、コンピュータに周辺機器を接続させるためのポートであり、周辺機器I/F部17を介してコンピュータは周辺機器とのデータの送受信を行う。周辺機器I/F部17は、USBやIEEE1394やRS-232C等で構成されており、通常複数の周辺機器I/Fを有する。周辺機器との接続形態は有線、無線を問わない。

通信部16は、通信制御装置、通信ポート等を有し、ネットワーク9との通信を媒介する通信インタフェースであり、通信制御を行う。

バス18は、各装置間の制御信号、データ信号等の授受を媒介する経路である。

【0033】

50

次に、文書推薦サーバ5により実行される事前処理、及び文書推薦処理の流れを、図6～図14を参照して説明する。

【0034】

文書推薦サーバ5には、書棚共有コミュニティサービスを利用する複数のユーザ端末2から、既読文書の登録、文書評価記事の登録が行われているものとする。文書推薦サーバ5は、ユーザ端末2から既読文書の登録、文書評価記事の登録が行われると、ユーザ情報51の「文書評価記事ID」を更新し、文書評価記事情報52に新規情報を追加する。また、ユーザプロファイル53の文書リスト53aを更新する。

【0035】

文書推薦サーバ5の制御部11は、図6、図7に示す事前処理を行って、全ユーザから登録された全文書を整理する。

図6に示すように、事前処理において、文書推薦サーバ5の制御部11は、文書評価記事情報52に基づき、当該文書推薦サーバ5に登録されている文書から平易有益文書55と難解有益文書57とを抽出する(ステップS1)。

平易有益文書55とは、ユーザの「理解度」が所定の指標(例えば平均値)以上で、「有益度」が所定の指標(例えば平均値)以上の文書群である。

難解有益文書57とは、ユーザの「理解度」が所定の指標(例えば平均値)より小さく、「有益度」が所定の指標(例えば平均値)以上の文書群である。

【0036】

図7を参照して、ステップS1の文書の分類について説明する。

図7(a)に示すように、文書推薦サーバ5の制御部11は、記憶部12に記憶されている文書評価記事情報52から全文書の「理解度」及び「有益度」を集約し、図7(b)に示すように、全文書の「理解度」及び「有益度」の指標値(例えば平均値)を算出する。図7(b)の例では、文書推薦サーバ5に登録されている文書の平均理解度は「2.8」、平均有益度は「2.5」が算出される。

次に、文書推薦サーバ5の制御部11は、登録されている文書の「理解度」が所定の指標(例えば平均値)以上で、「有益度」が所定の指標(例えば平均値)以上の文書を平易有益文書55として抽出する。また、文書推薦サーバ5の制御部11は、登録されている文書の「理解度」が所定の指標(例えば平均値)より小さく、「有益度」が所定の指標(例えば平均値)以上の文書を難解有益文書57として抽出する。平易有益文書55の一例を図7(c)に、難解有益文書57の一例を図7(d)に示す。

なお、「理解度」及び「有益度」の指標値は、ここでは一例として平均値を用いているがこれに限定されず、中央値としてもよいし、その他の統計的に算出可能な数値としてもよい。

平易有益文書55、難解有益文書57は、ユーザに推薦する文書を決定する際に参照される。

【0037】

また、事前処理として、文書推薦サーバ5の制御部11は、平易有益文書55及び難解有益文書57に登録されている文書の内容の特徴を表す文書ベクトルを生成する(ステップS2)。文書ベクトルとは、文書から抽出されるキーワードをベクトルの次元(成分)とし、各次元の値をキーワードの特徴の程度を表現する数値(ここでは、TF-IDF)で表したものである。

具体的には、例えば、各文書の書誌情報51(タイトル、内容、著者)、及び文書評価記事情報53(コメント)の形態素解析を行い、抽出したキーワードのTF-IDF(Term Frequency - Inverse Document Frequency)を算出して文書ベクトルを生成する(ステップS2)。TF(Term Frequency)とは、ある文書にキーワードが出現する頻度であり、そのキーワードの文書における重要度を表すものである。DF(Document Frequency)とは、全文書中、キーワードが現れる文書数であり、多くの文書に含まれるキーワードは一般的な語で重要性が低いと考えるものである。TF-IDFにより、そのキーワードがある文書

10

20

30

40

50

の特徴を表す程度を数値化できる。ここで算出したTF-IDFは、文書ベクトルの各成分の値となる。

【0038】

次に、文書推薦サーバ5の実行する文書推薦処理の流れを図8～図14を参照して説明する。

【0039】

図8に示すように、まずユーザ端末2から文書推薦サーバ5に文書の登録が行われる(ステップS21)。

文書の登録について、図9を参照して説明する。

ユーザ端末2が文書推薦サーバ5にログインすると、文書推薦サーバ5は図9(a)に示すようなユーザ専用の書棚画面21を該当ユーザ端末2に送信する。書棚画面21にはユーザが登録した文書が書棚領域21aに一覧表示またはサムネイル表示される。また、書棚画面21には新規追加ボタン21bが設けられており、新規追加ボタン21bがクリック操作されると、文書推薦サーバ5は図9(b)に示す文書登録画面22をユーザ端末2に対して送信する。文書登録画面22には文書を検索するための検索ボタン22a、文書ID入力欄22b、理解度入力欄22c、有益度入力欄22d、コメント入力欄22e、登録ボタン22fが設けられる。ユーザ端末2にて各入力欄22b～22eに対して文書ID、理解度、有益度、コメントが入力され、登録ボタン22fがクリック操作されると、入力された内容が文書推薦サーバ5に送信される。文書推薦サーバ5は受信した文書ID、理解度、有益度、コメントを文書評価記事情報53として新規登録すると共に、該当ユーザのユーザプロフィール53の書棚(文書リスト53a)、ユーザ情報51の文書評価記事情報IDを更新する。

検索ボタン22aがクリック操作された場合は、検索画面(不図示)に移行する。検索画面には、例えば文書のタイトル、著者等のキーワードを入力する入力欄が設けられるものとする。図9(c)に示すように、書誌データベース3に登録されている書誌情報31から書棚に登録する文書が検索されるものとするが、書誌データベース3に相当する別のデータベースから検索されるものとしてもよい。

【0040】

次に、図8のステップS22に示すように、文書推薦サーバ5は各ユーザのユーザプロフィール53を作成する(ステップS22)。すなわち、文書推薦サーバ5は各ユーザの登録した文書リスト53aからユーザの苦手分野と得意分野とをそれぞれ表現する検索クエリを作成する。検索クエリは文書ベクトルの形式で算出するものとする。

まず、苦手分野文書、得意分野文書の抽出について、図10を参照して説明する。

【0041】

図10(a)に示すように、文書推薦サーバ5は各ユーザが登録した文書リスト53a(既読文書)と文書の評価記事情報52とを取得する。文書評価記事情報には、文書IDとその文書の有益度、理解度、及びコメントが格納されている。

文書推薦サーバ5は、ユーザの既読文書の理解度と、その文書の全ユーザの平均理解度(図7(a)参照)とを比較する。

図10(a)の例では、ユーザ「User001」の文書リスト53aには文書ID「doc001」、「doc002」、「doc006」、「doc007」が含まれ、これらの各文書についてのユーザの理解度はそれぞれ「4.5」、「4.0」、「3.0」、「2.0」である。一方、事前処理にて算出された全ユーザを対象とした各文書「doc001」、「doc002」、「doc006」、「doc007」の平均理解度は、それぞれ「3.9」、「2.5」、「4.5」、「3.7」であるとする。

文書推薦サーバ5は、全ユーザを対象とした平均理解度と、ユーザ「User001」の理解度とを数値比較し、ユーザの理解度が平均理解度より小さいものを苦手分野文書53bとし、ユーザの理解度が平均理解度以上のものを得意分野文書53cとする。図10(b)に示すように、ユーザ「User001」の苦手分野文書53bは「doc006」、「doc007」となり、図10(c)に示すように、得意分野文書53cは「do

10

20

30

40

50

c 0 0 1」、「doc 0 0 2」となる。

【0042】

次に、文書推薦サーバ5は、ユーザの苦手分野文書53b及び得意分野文書53cから検索クエリを作成する。

図11を参照して、苦手分野検索クエリ53dの作成について説明する。

【0043】

図11(a)に示すように、文書推薦サーバ5は、苦手分野文書53bに含まれる各文書からキーワードを抽出する。具体的には、例えば、書誌情報(タイトル、内容、著者)と文書評価情報(コメント)に対して形態素解析ソフトウェアであるChasen(茶筌; <http://chasen-legacy.sourceforge.jp/>)を用いて形態素解析を行う。この場合、文書推薦サーバ5は、名詞、未知語、記号-アルファベットを抽出し、品詞毎に設定したキーワード生成ルール(例えば、連続して出現した名詞、未知語、記号-アルファベットは連結する等)を適用してキーワードを抽出する。このとき、予め登録されている辞書が参照され、辞書の登録語句に基づいて名詞、未知語、記号-アルファベットが判断される。また、例えば、予め用意されたキーワード候補の辞書を用意し、書誌情報(タイトル、内容、著者)と文書評価情報(コメント)に対して、完全一致検索を行って、一致したキーワードを抽出してもよい。

10

次に、文書推薦サーバ5は、各文書から抽出されるキーワードについてTF-IDFを算出する。図11(b)に示す例では、苦手分野文書53bに分類された文書「doc006」について各キーワードを抽出し、TF-IDFを算出する。例えば、文書「doc006」からはキーワードとして、「投資」、「金融工学」、「為替」、「株式」、「決算」等が抽出され、そのTF-IDFはそれぞれ「0.5」、「0.4」、「0.33」、「0.33」、「0.2」であるとする。重要なキーワードほど、TF-IDFの値が大きな値となる。

20

【0044】

更に、文書推薦サーバ5は、各文書についてユーザの理解度の重みを計算する。図11(c)に示すように、ユーザの理解度と全ユーザの平均理解度とから、理解度の重みを計算する。理解度の重みとは、全ユーザの理解度を基準に該当ユーザがどれだけ理解しているかを示すものであり、該当ユーザの理解度を相対的に表すものである。すなわち、理解度の重みによって、該当ユーザの全ユーザ内での知識レベルが表されることとなる。理解度の重みは以下の式(1)にて表される。

30

【0045】

【数1】

$$WN[006] = \frac{(\text{全ユーザのdoc006理解度平均}) - (\text{User001のdoc006理解度})}{\sum \{ (\text{全ユーザのdocN理解度平均}) - (\text{User001のdocN理解度}) \}} \dots (1)$$

* docNは、User001の苦手分野文書に含まれる文書

40

(この場合はdoc006、doc007)を指す。

【0046】

文書推薦サーバ5は、図11(d)に示すように、各文書から抽出された各キーワードのTF-IDFに理解度の重みを乗じ、更に苦手分野文書53b内の全文書でそれらの和をとって各キーワードのスコアを算出する。このとき、苦手分野文書に登場するキーワードの数と同じ次元数の文書ベクトルを新たに作り、これを苦手分野検索クエリ53dとする。苦手分野検索クエリ53dの各次元のキーワードがそれぞれ上述のスコアをもつこととなる。

【0047】

50

得意分野文書検索クエリ 5 3 e も図 1 1 に示す手順と同様に作成される (図 1 2 参照)。すなわち、文書推薦サーバ 5 は、得意分野文書 5 3 c からキーワードを抽出し、その TF - IDF を算出する。図 1 2 (b) に示す例では、得意分野文書 5 3 c に分類された文書「doc 0 0 1」から、キーワードとして、「C 言語」、「プログラミング」、「サーバ」、「Web」、「開発」等が抽出され、その TF - IDF はそれぞれ「0 . 5」、「0 . 4」、「0 . 3 3」、「0 . 3 3」、「0 . 2」であるとする。更に、図 1 2 (c) に示すように、文書推薦サーバ 5 は、各文書についてユーザの理解度の重みを計算し、図 1 2 (d) に示すように、各文書から抽出された各キーワードの TF - IDF に理解度の重みを乗じ、更に得意分野文書 5 3 c 内の全文書でそれらの和をとって各キーワードのスコアを算出する。このとき、得意分野文書に登場するキーワードの数と同じ次元数の文書ベクトルを新たに作り、これを得意分野検索クエリ 5 3 e とする。得意分野検索クエリ 5 3 e の各次元のキーワードがそれぞれ上述のスコアをもつこととなる。

10

【 0 0 4 8 】

図 8 のステップ S 2 2 のユーザプロファイル 5 3 の作成が完了すると、次に文書推薦サーバ 5 は、苦手分野検索クエリ 5 3 d を使用して平易有益文書 5 5 (図 7 (c)) 内を検索し、苦手分野対応文書 5 3 f を抽出する (ステップ S 2 3)。また、得意分野検索クエリ 5 3 e を使用して難解有益文書 5 7 (図 7 (d)) 内を検索し、得意分野対応文書 5 3 g を抽出する (ステップ S 2 4)。

図 1 3、図 1 4 を参照して、苦手分野対応文書 5 3 f の抽出、得意分野対応文書 5 3 g の抽出について説明する。

20

【 0 0 4 9 】

図 1 3 に示すように、文書推薦サーバ 5 は、作成した苦手分野検索クエリ 5 3 d を使用して平易有益文書 5 5 を検索する。検索方法としては、例えばベクトル空間モデルを用いる。

ベクトル空間モデルを用いた検索とは、書誌データベース 3 に蓄積されている文書 (書誌情報 3 1) 全体における各文書の文書ベクトルと、上述の苦手分野検索クエリ 5 3 d または得意分野検索クエリ 5 3 e との類似度を算出するという意味である。

文書全体における各文書の文書ベクトルは、図 6 の事前処理のステップ S 2 で算出される。

【 0 0 5 0 】

30

図 1 3 に示すように、文書推薦サーバ 5 は、苦手分野検索クエリ 5 3 d に対応する文書ベクトル (ユーザ苦手ベクトル) を $d [i]$ 、平易有益文書 5 5 内のある文書の文書ベクトルを $d [j]$ とし、これらのベクトル $d [i]$ 、 $d [j]$ の内積を算出して、類似度 $sim (d [i] , d [j])$ を求める。

具体的には、 $d [i]$ と $d [j]$ で共通する成分 (キーワード) のスコアの積を計算し、その総和を算出するという手順で文書ベクトル $d [i]$ と $d [j]$ の内積の値を求めることができる。

類似度 $sim (d [i] , d [j])$ は次式 (2) で表現される。式 (2) は、ベクトル $d [i]$ 、 $d [j]$ の内積を正規化した形式としている。

【 0 0 5 1 】

40

【 数 2 】

$$sim(d[i], d[j]) = \frac{d[i] \cdot d[j]}{\|d[i]\| \|d[j]\|} \dots (2)$$

【 0 0 5 2 】

文書推薦サーバ 5 は、平易有益文書 5 5 内のある全ての文書と、苦手分野検索クエリ 5 3 d との類似度を算出し、類似度の値が大きいものを苦手分野対応文書 5 3 f とする。図

50

13では、文書「doc014」、文書「doc009」、文書「doc015」の3つの文書が、苦手分野対応文書53fとして検索されている。なお、苦手分野対応文書53fは何冊であってもよい。

【0053】

また、得意分野対応文書53gについても図13と同様の手法で算出する。

図14に示すように、文書推薦サーバ5は、作成した得意分野検索クエリ53eを使用して難解有益文書57（図7（d））内をベクトル空間モデルを用いて検索する。

文書推薦サーバ5は、得意分野検索クエリ53eに対応する文書ベクトル（ユーザ得意ベクトル）を $d[i]$ 、難解有益文書57内のある文書の文書ベクトルを $d[j]$ とし、これらのベクトル $d[i]$ 、 $d[j]$ の内積を算出して、上述の式（2）で表現される類似度 $sim(d[i], d[j])$ を求める。具体的には、 $d[i]$ と $d[j]$ で共通する成分（キーワード）のスコアの積を計算し、その総和を算出するという手順で文書ベクトル $d[i]$ と $d[j]$ の内積の値を求めることができる。

文書推薦サーバ5は、難解有益文書57内のある全ての文書と、得意分野検索クエリ53eとの類似度を算出し、類似度の値が大きいものを得意分野対応文書53gとする。図14では、文書「doc012」、文書「doc003」、文書「doc005」の3つの文書が、得意分野対応文書53gとして検索されている。なお、得意分野対応文書53gは何冊であってもよい。

【0054】

ステップS23、ステップS24によって苦手分野対応文書53f、得意分野対応文書53gが検索されると、文書推薦サーバ5は、検索した苦手分野対応文書53f及び得意分野対応文書53gを推薦文書29として、該当するユーザ端末2に送信する。このとき、更に、文書推薦サーバ5は、有益度を用いて降順にソートし、上位のものを推薦文書29としてもよい。

ユーザ端末2は、文書推薦サーバ5から送信された苦手分野対応文書53f及び得意分野対応文書53gを書棚画面21内に提示する（ステップS25）。

【0055】

図2（b）に推薦文書29が提示された書棚画面21の一例を示す。図2（b）に示すように、ユーザ「User001」の書棚画面21の一部に、ユーザの苦手分野に対する推薦文書29と得意分野に対する推薦文書29とが一覧表示またはサムネイル表示される。

【0056】

以上説明したように、文書推薦システム1は、ユーザ端末2と文書推薦サーバ5とがネットワーク9を介して通信接続され、複数のユーザ端末2と文書推薦サーバ5との間で文書に関する情報を共有する。ユーザ端末2が、既読文書に対して理解度及び有益度の評価を付与して文書推薦サーバ5の各自の書棚（文書リスト53a）に登録すると、文書推薦サーバ5は、全ユーザから登録された文書リスト53aから有益度が比較的高い文書を抽出するとともに、理解度レベル別に推薦文書を分類する。例えば平易有益文書55と難解有益文書57として推薦すべき文書を分類する。また、文書推薦サーバ5は、各ユーザの文書リスト53aからユーザ別のユーザプロファイル53bを作成する。すなわち、個々のユーザの既読文書の理解度に応じた検索クエリを作成する。そして、文書推薦サーバ5は、苦手分野文書検索クエリ53dを用いて平易有益文書55を検索し、苦手分野に対して推薦すべき文書（苦手分野対応文書53f）を決定して、ユーザ端末2に提示する。同様に、文書推薦サーバ5は、得意分野文書検索クエリ53eを用いて難解有益文書57を検索し、得意分野に対して推薦すべき文書（得意分野対応文書53g）を決定してユーザ端末2に提示する。

【0057】

従って、ユーザの理解度に応じて、苦手分野に対しては易しくて有益な文書を推薦でき、得意分野に対しては難しいが有益な文書を推薦できるようになる。その結果、ユーザの知識レベルに適合した文書を提示でき、文書に対するコンバージョン（閲覧、購買等）を

10

20

30

40

50

向上させることが可能となる。

【 0 0 5 8 】

なお、上述の実施の形態では、文書に対する理解度をユーザ自身が評価して数値入力するものとしたが、より客観的に文書の理解度を判定するため、ユーザ端末 2 または文書推薦サーバ 5 のいずれかに理解度判定手段を更に備えるようにしてもよい。

例えば、電子書籍の場合であれば、書籍の頁を読み進める速さをユーザ端末 2 の電子書籍ブラウザにて計測し、速さが速いものは理解度が高く、遅いものは理解度が低いとみなす。具体的には、電子書籍ブラウザは閲覧頁数と閲覧時間とを記録し、1 頁当たりの読む速度を計測する。速度が速ければ理解度が高いとみなし、速度が遅ければ理解度が低いとみなす。例えば、ビジネス書の場合は、速度が 1 頁 / 分であれば、理解度「 5 . 0 0 」、1 頁 / 分未満 ~ 0 . 5 頁 / 分以上であれば、理解度「 4 . 5 0 」、・・・ 0 . 0 1 頁 / 分以下であれば、理解度「 0 . 5 0 」のように理解度を決定する。また、書籍のジャンル毎に速度の基準値を設け、基準値とユーザの速度とを比較することにより理解度を決定してもよい。

10

【 0 0 5 9 】

また、電子書籍の場合は、書籍閲覧中にユーザ端末 2 の電子書籍ブラウザの辞書機能を利用した回数を計測して理解度を算出するようにしてもよい。辞書機能を利用しない書籍は理解度が高く、辞書機能を多く利用した書籍は理解度が低いものとする。

具体的には、電子書籍ブラウザはユーザが書籍を閲覧中に、何回辞書機能を起動したかを計測する。辞書機能を起動した回数について、ジャンル毎に基準値を設け、基準値と比較して理解度を算出する。例えば、ビジネス書の場合は、基準値よりも起動回数が少なければ理解度「 5 . 0 0 」、基準値との差が 1 回以上 ~ 5 回未満であれば理解度「 4 . 5 0 」、・・・ 1 0 0 回以上であれば理解度「 0 . 5 0 」のように理解度を決定する。

20

【 0 0 6 0 】

また、文書推薦サーバ 5 はユーザが入力したコメントを評価し、コメント内の専門語の質及び数に基づいてユーザの理解度を算出するようにしてもよい。コメント中の語彙の数が多く語彙が専門的である場合は理解度が高いと判定し、コメント中の語彙の数が少なく語彙が専門的でない場合は理解度が低いと判定する。

具体的には、書籍のジャンル毎にキーワードとキーワードの重み (ポイント) を設定する。例えば、技術書の場合、キーワード「 C 言語」の重みは「 + 1 」、キーワード「参照渡し」の重みは「 + 3 」、キーワード「ポインタ」の重みは「 + 2 」、・・・のように予めキーワードとキーワードの重み (ポイント) を決定しておき、文書推薦サーバ 5 はユーザのコメントについて、上述のキーワードが含まれるかをカウントし、ポイントの合計を算出する。更に、文書のジャンル毎に基準を設けて基準に対するポイントの値から理解度を算出する。例えば、技術書の場合は、基準に対するポイントの値が 1 0 0 ポイントの場合は理解度「 5 . 0 0 」、9 0 ポイントの場合は理解度「 4 . 5 0 」、・・・、0 ポイントの場合は理解度「 0 . 5 0 」のように決定する。

30

【 0 0 6 1 】

また、文書推薦サーバ 5 が、ユーザから登録された文書の本文データを使用して文書に関する問題を作成し、ユーザに問題の回答を入力させ、回答結果に応じて理解度を判定してもよい。正解が多い場合は理解度が高いと判定し、不正解が多い場合は理解度が低いと判定する。

40

具体的には、文書推薦サーバ 5 は、文書の本文のデータに基づいて問題 (例えば穴埋め問題や多肢選択問題) を作成し、元の文書に紐付けて記憶しておく。ユーザが書籍を閲覧した後、文書推薦サーバ 5 は、ユーザ端末 2 に対し、文書に対応する問題を送信し、ユーザ端末 2 にて回答を入力させ、ユーザ端末 2 から回答を受信して点数 (ポイント) を算出する。更に文書推薦サーバ 5 は、文書のジャンル毎に基準を設け、基準に対するポイントの値から理解度を算出する。例えば、ビジネス書の場合は、基準に対するポイントの値が 1 0 0 点の場合は理解度「 5 . 0 0 」、9 0 点の場合は理解度「 4 . 5 0 」、・・・、0 点の場合は理解度「 0 . 5 0 」のように決定する。

50

【 0 0 6 2 】

以上のように、文書に対する理解度の判定をユーザ端末 2 または文書推薦サーバ 5 のいずれかで行なうことにより、ユーザの理解度を文書推薦サーバ 5 にて客観的に収集することが可能となり、より公正に文書の難易度を求めることができるようになる。

【 0 0 6 3 】

また、上述の実施の形態では、ユーザの理解度レベルに応じて文書を苦手分野文書 5 3 b と得意分野文書 5 3 c とに分類したが、分類は得意と苦手の 2 種類に限定されず、更に細かく理解度レベルを設定し、分類してもよい。この場合は、全ユーザの書棚の分類も、平易有益文書 5 5 と難解有益文書 5 7 の 2 種類に限定されず、対応する理解度レベルに分類しておき、ユーザの理解度レベルに対応した推薦文書を、対応する理解度レベルの推薦文書群から検索すればよい。また、ユーザの既読文書が苦手分野文書 5 3 b または得意分野文書 5 3 c のいずれか一方のみに該当することとなってもよい。

10

【 0 0 6 4 】

また、上述の実施の形態では、各文書の文書ベクトルを算出する際、書誌情報 3 1 や文書評価記事情報 5 2 のコメントからキーワードを抽出するものとしたが、書誌情報 3 1 の蓄積されていない例えばインターネットに公開されている文書等の文書ベクトルを生成する場合は、文書の本文そのものからキーワードを抽出するようにしてもよい。

【 0 0 6 5 】

その他、当業者であれば、本願で開示した技術的思想の範疇内において、各種の変更例または修正例に想到し得ることは明らかであり、それらについても当然に本発明の技術的範囲に属するものと了解される。

20

【 符号の説明 】

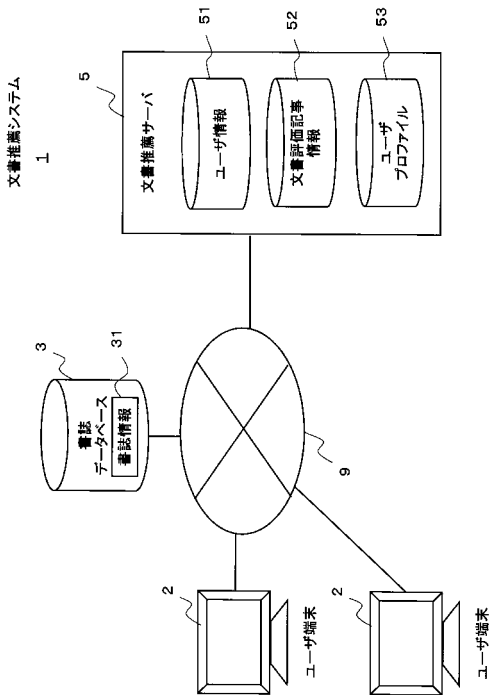
【 0 0 6 6 】

- 1 . . . 文書推薦システム
- 2 . . . ユーザ端末
- 3 . . . 書誌データベース
- 5 . . . 文書推薦サーバ
- 9 . . . ネットワーク
- 2 1 . . . 書棚画面
- 2 2 . . . 文書登録画面
- 2 9 . . . 推薦文書
- 3 1 . . . 書誌情報
- 5 1 . . . ユーザ情報
- 5 2 . . . 文書評価記事情報
- 5 3 . . . ユーザプロフィール
- 5 3 a . . . 書棚 (文書リスト)
- 5 3 b . . . 苦手分野文書
- 5 3 c . . . 得意分野文書
- 5 3 d . . . 苦手分野検索クエリ
- 5 3 e . . . 得意分野検索クエリ
- 5 3 f . . . 苦手分野対応文書
- 5 3 g . . . 得意分野対応文書
- 5 5 . . . 平易有益文書
- 5 7 . . . 難解有益文書

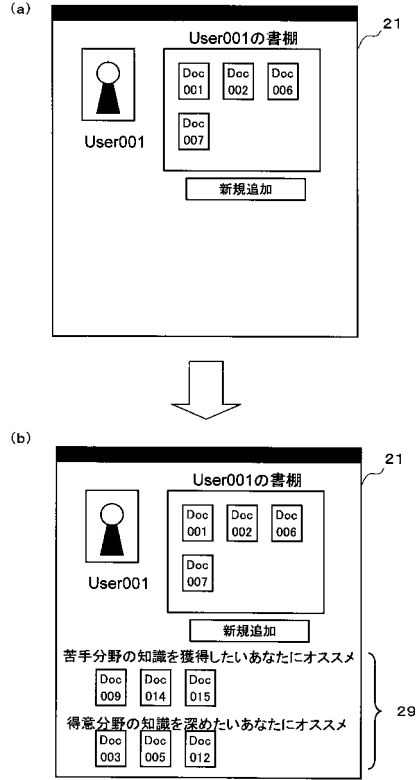
30

40

【図1】



【図2】



【図3】

雑誌情報 31

文書ID	文書の識別子
タイトル	文書のタイトル
カテゴリ	文書が所属するカテゴリ
内容	文書の文、目次もしくは説明文
著者	文書の作者

【図4】

(a) ユーザ情報 51

ユーザID	ユーザの識別子
ユーザ名	ユーザの名前
ユーザ属性情報	ユーザの年齢、性別、職業などユーザ固有の情報
文書評価記事ID	書棚に登録した文書評価記事のリスト(複数指定可)

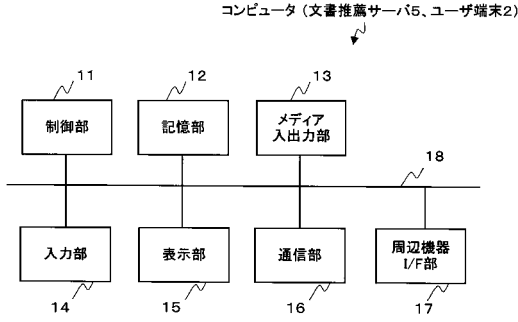
(b) 文書評価記事情報 52

文書評価記事ID	文書評価記事の識別子
文書ID	評価対象となる文書の識別子
有益度	評価したユーザにとって文書が役に立った度合いを表現する指標(数値)
理解度	評価したユーザにとって文書の理解しやすさを表現する指標(数値)
コメント	評価対象に対する感想を表現する文字列

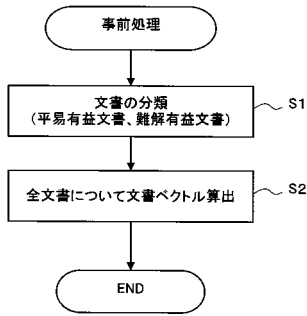
(c) ユーザプロフィール 53

User001の プロフィール	User002の プロフィール	...
書棚(文書リスト) 53a	書棚(文書リスト) 53a	
苦手分野文書 53b	苦手分野文書 53b	
得意分野文書 53c	得意分野文書 53c	
苦手分野検索クエリ 53d	苦手分野検索クエリ 53d	
得意分野検索クエリ 53e	得意分野検索クエリ 53e	

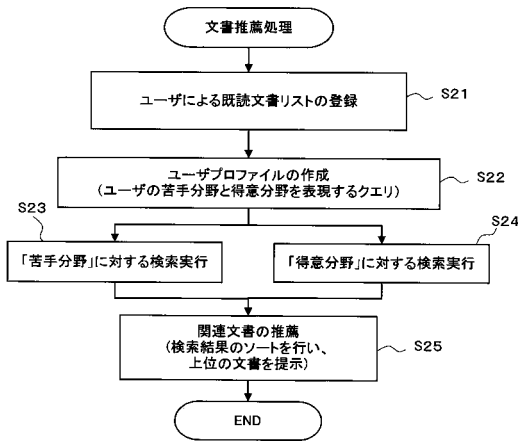
【図5】



【図6】



【図8】



【図7】

(a) 文書評価記事情報(一部) 52

文書ID	タイトル	理解度(平均)	有益度(平均)
doc001	C言語の基礎	3.9	2.4
doc002	Javaの応用	2.5	2.8
doc003	マスタリングC言語	2.2	4.2
doc004	Perl入門	4.2	3.8
doc005	Perlを極める	2.6	4.4
doc006	株式入門	4.5	3.1
doc007	投資の理論	3.7	3.5
doc008	金融危機入門	3.8	3.4
doc009	金融危機理論	3.3	4.2
doc009	投資理論論文集	1.8	3.8
...
doc020	日本の歴史入門	3.2	3.8

(b)

	理解度	有益度
全文書の平均	2.8	2.5

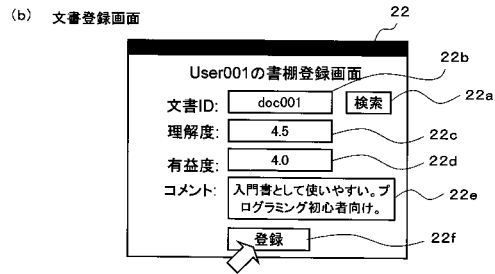
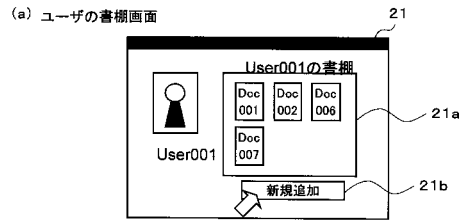
(c) 平易有益文書 55

文書ID	タイトル	理解度(平均)	有益度(平均)
doc004	Perl入門	4.2	3.8
doc006	株式入門	4.5	3.1
doc007	投資の理論	3.7	3.5
doc008	金融危機入門	3.8	3.4
doc009	金融危機理論	3.3	4.2
...

(d) 難解有益文書 57

文書ID	タイトル	理解度(平均)	有益度(平均)
doc002	Javaの応用	2.5	2.8
doc003	マスタリングC言語	2.2	4.2
doc005	Perlを極める	2.6	4.4
doc010	投資理論論文集	1.8	3.8
...

【図9】



(c) 書誌データベースに登録されている書誌情報31から検索

文書ID	タイトル	カテゴリ	内容	著者	価格
doc001	C言語の基礎	IT	プログラムの基礎、C言語を学ぶ。	Author001	¥3,500
doc002	Javaの応用	IT	サーバー向けプログラムの定書。Javaを学ぼう。	Author002	¥2,500
doc003	マスタリングC言語	IT	C言語をマスターしないあなたに。究極のプログラム本。	Author003	¥2,000
doc004	Perl入門	IT	Perlを1から学ぼう。今日からできるプログラム開発。	Author004	¥1,800
doc005	Perlを極める	IT	Perlを極めたいあなたに。実践的なテクニックが満載。	Author005	¥3,500
...
doc100	日本の歴史	歴史	日本の歴史について	Author100	¥1,500

【図10】

(a)

User001の書棚(文書リスト) 53a

文書ID	有益度	理解度	コメント
doc001	4.0	4.5	入門書として使いやすい。プログラミング初心者向け。
doc002	5.0	4.0	C言語の定書。これを持っていないと仕事ができないほど愛用しています。
doc006	3.5	3.0	投資について勉強しようと思ったのだが、なかなか難しい。勉強せねば。
doc007	2.5	2.0	金融工学の知識が必要で全然理解できなかった。いつか改めて読み直してみようと思う。

全ユーザの書棚の平均

文書ID	理解度
doc001	3.9
doc002	2.5
doc006	4.5
doc007	3.7

数値比較

(b)

文書ID	有益度	理解度	コメント
doc006	3.5	3.0	投資について勉強しようと思ったのだが、なかなか難しい。勉強せねば。
doc007	2.5	2.0	金融工学の知識が必要で全然理解できなかった。いつか改めて読み直してみようと思う。

User001の「苦手分野文書」 53b

(c)

文書ID	有益度	理解度	コメント
doc001	4.0	4.5	入門書として使いやすい。プログラミング初心者向け。
doc002	5.0	4.0	C言語の定書。これを持っていないと仕事ができないほど愛用しています。

User001の「得意分野文書」 53c

【図11】

(a) 各文書からキーワード抽出

(b) 各文書のTF-IDFを計算

苦手分野文書 doc006

tf-idfの計算

キーワード	TF	IDF	TF-IDF
投資	2	4	0.5
金融工学	2	5	0.4
為替	1	3	0.33
株式	1	3	0.33
決算	1	5	0.2
...

(c) 理解度の重みを計算

$$WN[006] = \frac{(\text{全ユーザのdoc006理解度平均}) - (\text{User001のdoc006理解度})}{\sum ((\text{全ユーザのdocN理解度平均}) - (\text{User001のdocN理解度}))}$$

* docNは、User001の苦手分野文書に含まれる文書(この場合はdoc006、doc007)を指す。

(d) 苦手分野検索クエリ53dを作成

苦手分野文書 doc006

苦手分野文書 doc007

53d

キーワード	スコア
投資	0.56
金融工学	0.33
為替	0.22
株式	0.12
決算	0.06
...	...

* キーワードのスコアは、全ユーザの理解度レベルを考慮したユーザの理解度を意味する。

【図12】

(a) 各文書からキーワード抽出

(b) 各文書のTF-IDFを計算

得意分野文書 doc001

tf-idfの計算

キーワード	TF	IDF	TF-IDF
C言語	2	4	0.5
プログラミング	2	5	0.4
サーバ	1	3	0.33
Web	1	3	0.33
開発	1	5	0.2
...

(c) 理解度の重みを計算

$$WP[001] = \frac{(\text{User001のdoc001理解度}) - (\text{全ユーザのdoc001理解度平均})}{\sum ((\text{User001のdocN理解度}) - (\text{全ユーザのdocN理解度平均}))}$$

* docNは、User001の得意分野文書に含まれる文書(この場合はdoc001、doc002)を指す。

(d) 得意分野検索クエリ53eを作成

得意分野文書 doc001

得意分野文書 doc002

53e

キーワード	スコア
C言語	0.71
プログラミング	0.44
サーバ	0.28
Web	0.22
開発	0.12
...	...

* キーワードのスコアは、全ユーザの理解度レベルを考慮したユーザの理解度を意味する。

【図13】

苦手分野検索クエリを使用して「平易有益文書」内を検索

苦手分野検索クエリ53d

平易有益文書 55

キーワード	スコア
投資	0.56
金融工学	0.33
為替	0.22
株式	0.12
決算	0.06
...	...

文書ID	タイトル	理解度(平均)	有益度(平均)
doc004	Per入門	4.2	3.8
doc006	株式入門	4.5	3.1
doc007	投資の理論	3.7	3.5
doc008	金融危機入門	3.8	3.4
doc009	金融危機理論	3.3	4.2
...

類似度

$$\text{sim}(d[i], d[j]) = \frac{d[i] \cdot d[j]}{|d[i]| |d[j]|}$$

苦手分野対応文書 53f

- doc014
- doc009
- doc015

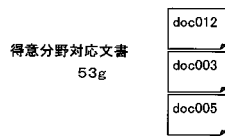
【 図 1 4 】

得意分野検索クエリを使用して「平易有益文書」内を検索

得意分野検索クエリ 53e		難解有益文書 57			
キーワード	スコア	文書ID	タイトル	理解度(平均)	有益度(平均)
C言語	0.71	doc002	Javaの応用	4.2	3.8
プログラミング	0.44	doc003	マシナリクC言語	4.5	3.1
サーバー	0.28	doc005	Perlを極める	3.7	3.5
Web	0.22	doc010	投資理論文庫	3.8	3.4
開発	0.12				
...	...				

類似度

$$\text{sim}(d[i], d[j]) = \frac{d[i] \cdot d[j]}{|d[i]| |d[j]|}$$



フロントページの続き

(58)調査した分野(Int.Cl. , DB名)

G 0 6 F 1 7 / 3 0