

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5048852号
(P5048852)

(45) 発行日 平成24年10月17日(2012.10.17)

(24) 登録日 平成24年7月27日(2012.7.27)

(51) Int.Cl.

G06F 17/30 (2006.01)

F 1

G06F 17/30 220Z
G06F 17/30 170A

請求項の数 12 (全 26 頁)

(21) 出願番号 特願2011-40259 (P2011-40259)
 (22) 出願日 平成23年2月25日 (2011.2.25)
 (65) 公開番号 特開2012-178026 (P2012-178026A)
 (43) 公開日 平成24年9月13日 (2012.9.13)
 審査請求日 平成24年6月14日 (2012.6.14)

早期審査対象出願

(73) 特許権者 399037405
 楽天株式会社
 東京都品川区東品川四丁目12番3号
 (74) 代理人 100088155
 弁理士 長谷川 芳樹
 (74) 代理人 100113435
 弁理士 黒木 義樹
 (74) 代理人 100144440
 弁理士 保坂 一之
 (72) 発明者 渡辺 一史
 東京都調布市調布ヶ丘一丁目5番地1 国立大学法人電気通信大学内
 (72) 発明者 岡部 誠
 東京都調布市調布ヶ丘一丁目5番地1 国立大学法人電気通信大学内

最終頁に続く

(54) 【発明の名称】検索装置、検索方法、検索プログラム、及びそのプログラムを記憶するコンピュータ読み取可能な記録媒体

(57) 【特許請求の範囲】

【請求項1】

所定のデータベースから抽出された文書に基づいて生成されたプロファイル・データであって、検索対象を特定する識別子と、該文書の作成時刻と、該文書内の語句を1以上のトピック別に分類することで求めた該トピック別の出現頻度とを含む該プロファイル・データに基づいて、ユーザにより指定された検索条件に対応する前記検索対象を抽出する検索手段と、

前記検索手段により抽出された検索対象を前記ユーザに提示するために出力する出力手段と

を備え、

前記検索条件が、第1の指定時期における、前記ユーザにより指定された1以上の指定トピック別の第1の出現頻度と、該第1の指定時期よりも後の第2の指定時期における該指定トピック別の第2の出現頻度とを含み、

前記検索手段が、第1の検索対象時期における前記プロファイル・データから得られた前記指定トピック別の出現頻度が前記第1の出現頻度と同一又は類似し、且つ、該第1の検索対象時期よりも後の第2の検索対象時期における前記プロファイル・データから得られた前記指定トピック別の出現頻度が前記第2の出現頻度と同一又は類似する前記検索対象を抽出する、

検索装置。

【請求項2】

10

20

前記検索条件が、更に、前記第1の指定時期と前記第2の指定時期との間における1以上の第3の指定時期のそれについての前記指定トピック別の第3の出現頻度を含み、

前記検索手段が、前記第1の検索対象時期における前記出現頻度が前記第1の出現頻度と同一又は類似し、前記第2の検索対象時期における前記出現頻度が前記第2の出現頻度と同一又は類似し、且つ、前記1以上の第3の指定時期と時間の前後関係が対応する1以上の第3の検索対象時期のそれにおいて、該第3の検索対象時期における前記プロファイル・データから得られた前記指定トピック別の出現頻度が、対応する前記第3の出現頻度と同一又は類似する前記検索対象を抽出する、

請求項1に記載の検索装置。

【請求項3】

10

前記検索条件が、隣接する前記指定時期間の時間間隔を更に含み、

隣接する前記検索対象時期間の時間間隔が、対応する前記指定時期間の時間間隔と同じである、

請求項1又は2に記載の検索装置。

【請求項4】

前記プロファイル・データで示される1以上の前記トピックが、前記所定のデータベースから抽出された前記文書に基づいて得られたトピック群から所定の不要トピックを除いた残りのトピックである、

請求項1～3のいずれか一項に記載の検索装置。

【請求項5】

20

前記指定トピックが、前記プロファイル・データで示される1以上のトピックを含む上位階層トピックである、

請求項1～4のいずれか一項に記載の検索装置。

【請求項6】

前記指定トピックが複数ある場合に、一の前記指定時における、前記指定トピック別の出現頻度の総和に対する一部の前記指定トピック別の出現頻度の比率が変更されると、該指定時における残りの前記指定トピック別の出現頻度の比率が該変更に応じて調整される、

請求項1～5のいずれか一項に記載の検索装置。

【請求項7】

30

前記一の指定時における前記一部の指定トピック別の出現頻度が変更された場合に、該一の指定時に対応する一の探索時における前記プロファイル・データから得られた前記一部の指定トピック別の出現頻度が、変更された前記一部の指定トピック別の出現頻度と同一又は類似し、且つ、他の前記指定時に対応する他の探索時における前記プロファイル・データから得られた前記指定トピック別の出現頻度が、該他の指定時における前記指定トピック別の出現頻度と同一又は類似する前記検索対象が抽出され、前記一の探索時における該検索対象の前記トピック別の出現頻度に基づいて前記残りの指定トピック別の出現頻度の比率が調整され、

前記一の探索時と前記他の探索時との前後関係が、前記一の指定時と前記他の指定時との前後関係と対応する、

請求項6に記載の検索装置。

40

【請求項8】

前記語句が複数の前記トピックに重複して出現する頻度に基づいて設定される、前記指定トピックそれぞれの間の関連の強さを示す重みと、前記一部の指定トピック別の出現頻度の変更量に基づいて、前記残りの指定トピック別の出現頻度の比率が調整される、

請求項6に記載の検索装置。

【請求項9】

検索手段が、更に、前記抽出された検索対象に対応する付加情報を、又は前記第1の検索対象時から前記第2の検索対象時までの間における該検索対象の前記指定トピック別の出現頻度に対応する付加情報を、追加のデータベースから抽出し、

50

前記出力手段が、更に、前記検索手段により抽出された付加情報を前記ユーザに提示するために出力する、

請求項 1 ~ 8 のいずれか一項に記載の検索装置。

【請求項 1 0】

検索装置により実行される検索方法であって、

所定のデータベースから抽出された文書に基づいて生成されたプロファイル・データであって、検索対象を特定する識別子と、該文書の作成時刻と、該文書内の語句を 1 以上のトピック別に分類することで求めた該トピック別の出現頻度とを含む該プロファイル・データに基づいて、ユーザにより指定された検索条件に対応する前記検索対象を抽出する検索ステップと、

10

前記検索ステップにおいて抽出された検索対象を前記ユーザに提示するために出力する出力ステップと

を含み、

前記検索条件が、第 1 の指定時期における、前記ユーザにより指定された 1 以上の指定トピック別の第 1 の出現頻度と、該第 1 の指定時期よりも後の第 2 の指定時期における該指定トピック別の第 2 の出現頻度とを含み、

前記検索ステップでは、第 1 の検索対象時期における前記プロファイル・データから得られた前記指定トピック別の出現頻度が前記第 1 の出現頻度と同一又は類似し、且つ、該第 1 の検索対象時期よりも後の第 2 の検索対象時期における前記プロファイル・データから得られた前記指定トピック別の出現頻度が前記第 2 の出現頻度と同一又は類似する前記検索対象が抽出される、

20

検索方法。

【請求項 1 1】

コンピュータを、

所定のデータベースから抽出された文書に基づいて生成されたプロファイル・データであって、検索対象を特定する識別子と、該文書の作成時刻と、該文書内の語句を 1 以上のトピック別に分類することで求めた該トピック別の出現頻度とを含む該プロファイル・データに基づいて、ユーザにより指定された検索条件に対応する前記検索対象を抽出する検索手段と、

前記検索手段により抽出された検索対象を前記ユーザに提示するために出力する出力手段と

30

して機能させ、

前記検索条件が、第 1 の指定時期における、前記ユーザにより指定された 1 以上の指定トピック別の第 1 の出現頻度と、該第 1 の指定時期よりも後の第 2 の指定時期における該指定トピック別の第 2 の出現頻度とを含み、

前記検索手段が、第 1 の検索対象時期における前記プロファイル・データから得られた前記指定トピック別の出現頻度が前記第 1 の出現頻度と同一又は類似し、且つ、該第 1 の検索対象時期よりも後の第 2 の検索対象時期における前記プロファイル・データから得られた前記指定トピック別の出現頻度が前記第 2 の出現頻度と同一又は類似する前記検索対象を抽出する、

40

検索プログラム。

【請求項 1 2】

コンピュータを、

所定のデータベースから抽出された文書に基づいて生成されたプロファイル・データであって、検索対象を特定する識別子と、該文書の作成時刻と、該文書内の語句を 1 以上のトピック別に分類することで求めた該トピック別の出現頻度とを含む該プロファイル・データに基づいて、ユーザにより指定された検索条件に対応する前記検索対象を抽出する検索手段と、

前記検索手段により抽出された検索対象を前記ユーザに提示するために出力する出力手段と

50

して機能させ、

前記検索条件が、第1の指定時期における、前記ユーザにより指定された1以上の指定トピック別の第1の出現頻度と、該第1の指定時期よりも後の第2の指定時期における該指定トピック別の第2の出現頻度とを含み、

前記検索手段が、第1の検索対象時期における前記プロファイル・データから得られた前記指定トピック別の出現頻度が前記第1の出現頻度と同一又は類似し、且つ、該第1の検索対象時期よりも後の第2の検索対象時期における前記プロファイル・データから得られた前記指定トピック別の出現頻度が前記第2の出現頻度と同一又は類似する前記検索対象を抽出する、

検索プログラムを記憶するコンピュータ読取可能な記録媒体。

10

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、蓄積された文書に基づく検索を実行する検索装置、検索方法、検索プログラム、及びそのプログラムを記憶するコンピュータ読取可能な記録媒体に関する。

【背景技術】

【0002】

近年、ブログやマイクロブログなどの電子文書による情報発信が一般化している。また、その電子文書の分析手法についても様々なものが考えられている。文書分析手法の一つとして、トピックモデルと呼ばれる文書生成モデルを利用した手法がある。例えば下記非特許文献1には、そのトピックモデルに時系列情報を取り入れることで、時間情報を持った文書から流行の移り変わりを視覚的に提示することが記載されている。

20

【先行技術文献】

【非特許文献】

【0003】

【非特許文献1】Xuerui Wang and Andrew McCallum. Topics over time: a non-markov continuous-time model of topical trends. In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '06, pp.424-433. New York, NY, USA, 2006. ACM.

30

【発明の概要】

【発明が解決しようとする課題】

【0004】

ところで、電子文書の発信数は膨大であるため、個人が自分に必要な情報に出会うことは容易でない。上記非特許文献1に記載の手法は、影響力の高い筆者の発見や話題の抽出、コミュニティの抽出を主な目的としており、ユーザが探したい状況変化の情報を提供するものとはいえない。

【0005】

そこで本発明は、所望の状況変化に対応する情報を提示することが可能な検索装置、検索方法、検索プログラム、及びそのプログラムを記憶するコンピュータ読取可能な記録媒体を提供することを目的とする。

40

【課題を解決するための手段】

【0006】

本発明の検索装置は、所定のデータベースから抽出された文書に基づいて生成されたプロファイル・データであって、検索対象を特定する識別子と、該文書の作成時刻と、該文書内の語句を1以上のトピック別に分類することで求めた該トピック別の出現頻度とを含む該プロファイル・データに基づいて、ユーザにより指定された検索条件に対応する検索対象を抽出する検索手段と、検索手段により抽出された検索対象をユーザに提示するために出力する出力手段とを備え、検索条件が、第1の指定時期における、ユーザにより指定された1以上の指定トピック別の第1の出現頻度と、該第1の指定時期よりも後の第2の指定時期における該指定トピック別の第2の出現頻度とを含み、検索手段が、第1の検索

50

対象時期におけるプロファイル・データから得られた指定トピック別の出現頻度が第1の出現頻度と同一又は類似し、且つ、該第1の検索対象時期よりも後の第2の検索対象時期におけるプロファイル・データから得られた指定トピック別の出現頻度が第2の出現頻度と同一又は類似する検索対象を抽出する。

【0007】

本発明の検索方法は、検索装置により実行される検索方法であって、所定のデータベースから抽出された文書に基づいて生成されたプロファイル・データであって、検索対象を特定する識別子と、該文書の作成時刻と、該文書内の語句を1以上のトピック別に分類することで求めた該トピック別の出現頻度とを含む該プロファイル・データに基づいて、ユーザにより指定された検索条件に対応する検索対象を抽出する検索ステップと、検索ステップにおいて抽出された検索対象をユーザに提示するために出力する出力ステップとを含み、検索条件が、第1の指定時期における、ユーザにより指定された1以上の指定トピック別の第1の出現頻度と、該第1の指定時期よりも後の第2の指定時期における該指定トピック別の第2の出現頻度とを含み、検索ステップでは、第1の検索対象時期におけるプロファイル・データから得られた指定トピック別の出現頻度が第1の出現頻度と同一又は類似し、且つ、該第1の検索対象時期よりも後の第2の検索対象時期におけるプロファイル・データから得られた指定トピック別の出現頻度が第2の出現頻度と同一又は類似する検索対象が抽出される。10

【0008】

本発明の検索プログラムは、コンピュータを、所定のデータベースから抽出された文書に基づいて生成されたプロファイル・データであって、検索対象を特定する識別子と、該文書の作成時刻と、該文書内の語句を1以上のトピック別に分類することで求めた該トピック別の出現頻度とを含む該プロファイル・データに基づいて、ユーザにより指定された検索条件に対応する検索対象を抽出する検索手段と、検索手段により抽出された検索対象をユーザに提示するために出力する出力手段として機能させ、検索条件が、第1の指定時期における、ユーザにより指定された1以上の指定トピック別の第1の出現頻度と、該第1の指定時期よりも後の第2の指定時期における該指定トピック別の第2の出現頻度とを含み、検索手段が、第1の検索対象時期におけるプロファイル・データから得られた指定トピック別の出現頻度が第1の出現頻度と同一又は類似し、且つ、該第1の検索対象時期よりも後の第2の検索対象時期におけるプロファイル・データから得られた指定トピック別の出現頻度が第2の出現頻度と同一又は類似する検索対象を抽出する。20

【0009】

本発明のコンピュータ読取可能な記録媒体は、コンピュータを、所定のデータベースから抽出された文書に基づいて生成されたプロファイル・データであって、検索対象を特定する識別子と、該文書の作成時刻と、該文書内の語句を1以上のトピック別に分類することで求めた該トピック別の出現頻度とを含む該プロファイル・データに基づいて、ユーザにより指定された検索条件に対応する検索対象を抽出する検索手段と、検索手段により抽出された検索対象をユーザに提示するために出力する出力手段として機能させ、検索条件が、第1の指定時期における、ユーザにより指定された1以上の指定トピック別の第1の出現頻度と、該第1の指定時期よりも後の第2の指定時期における該指定トピック別の第2の出現頻度とを含み、検索手段が、第1の検索対象時期におけるプロファイル・データから得られた指定トピック別の出現頻度が第1の出現頻度と同一又は類似し、且つ、該第1の検索対象時期よりも後の第2の検索対象時期におけるプロファイル・データから得られた指定トピック別の出現頻度が第2の出現頻度と同一又は類似する検索対象を抽出する、検索プログラムを記憶する。30

【0010】

このような発明によれば、二つの時期における指定トピック別の出現頻度（第1及び第2の出現頻度）と同一又は類似する二つの出現頻度を有する検索対象が抽出される。そして、その検索対象がユーザに提示するために出力される。このように、指定されたトピックについての出現頻度の変化が検索条件と一致又は類似する検索対象を抽出することで、40

所望の状況変化に対応する情報をユーザに提示することができる。

【0011】

本発明の検索装置では、検索条件が、更に、第1の指定時期と第2の指定時期との間ににおける1以上の第3の指定時期のそれぞれについての指定トピック別の第3の出現頻度を含み、検索手段が、第1の検索対象時期における出現頻度が第1の出現頻度と同一又は類似し、第2の検索対象時期における出現頻度が第2の出現頻度と同一又は類似し、且つ、1以上の第3の指定時期と時間の前後関係が対応する1以上の第3の検索対象時期のそれにおいて、該第3の検索対象時期におけるプロファイル・データから得られた指定トピック別の出現頻度が、対応する第3の出現頻度と同一又は類似する検索対象を抽出してもよい。

10

【0012】

この場合には、3以上の時期における指定トピック別の出現頻度と同一又は類似する複数の出現頻度を有する検索対象が抽出され、その検索対象が出力される。これにより、状況変化の過程を考慮した検索を行うことができるので、所望の状況変化に対応する情報をユーザに提示することができる。

【0013】

本発明の検索装置では、検索条件が、更に、隣接する指定時間間の時間間隔を更に含み、隣接する検索対象時間間の時間間隔が、対応する指定時間間の時間間隔と同じであってもよい。

【0014】

この場合には、指定時間間の時間間隔が検索条件として更に指定されるので、所望の状況変化及びその変化の速度に対応する情報をユーザに提示することができる。

20

【0015】

本発明の検索装置では、プロファイル・データで示される1以上のトピックが、所定のデータベースから抽出された文書に基づいて得られたトピック群から所定の不要トピックを除いた残りのトピックであってもよい。

【0016】

この場合には、不要なトピックが予め取り除かれるので、検索の精度を上げることができる。

【0017】

本発明の検索装置では、指定トピックが、プロファイル・データで示される1以上のトピックを含む上位階層トピックであってもよい。

30

【0018】

この場合には、トピックを階層構造で表すことでプロファイル・データのトピックが上位階層トピックに集約され、その上位階層トピックを指定トピックとする検索が行われる。このような指定トピックを用いることで、検索操作の利便性が向上する。

【0019】

本発明の検索装置では、指定トピックが複数ある場合に、一の指定時期における、指定トピック別の出現頻度の総和に対する一部の指定トピック別の出現頻度の比率が変更されると、該指定時期における残りの指定トピック別の出現頻度の比率が該変更に応じて調整されてもよい。

40

【0020】

本発明の検索装置では、一の指定時期における一部の指定トピック別の出現頻度が変更された場合に、該一の指定時期に対応する一の探索時期におけるプロファイル・データから得られた一部の指定トピック別の出現頻度が、変更された一部の指定トピック別の出現頻度と同一又は類似し、且つ、他の指定時期に対応する他の探索時期におけるプロファイル・データから得られた指定トピック別の出現頻度が、該他の指定時期における指定トピック別の出現頻度と同一又は類似する検索対象が抽出され、一の探索時期における該検索対象のトピック別の出現頻度に基づいて残りの指定トピック別の出現頻度の比率が調整され、一の探索時期と他の探索時期との前後関係が、一の指定時期と他の指定時期との前後

50

関係と対応してもよい。

【0021】

本発明の検索装置では、語句が複数のトピックに重複して出現する頻度に基づいて設定される、指定トピックそれぞれの間の関連の強さを示す重みと、一部の指定トピック別の出現頻度の変更量とに基づいて、残りの指定トピック別の出現頻度の比率が調整されてもよい。

【0022】

これらの場合には、ユーザがある指定トピックの出現頻度を変えるとその変更に応じて他の指定トピックの出現頻度が調整されるので、検索操作の利便性が向上する。

【0023】

本発明の検索装置では、検索手段が、更に、抽出された検索対象に対応する付加情報を、又は第1の検索対象時期から第2の検索対象時期までの間ににおける該検索対象の指定トピック別の出現頻度に対応する付加情報を、追加のデータベースから抽出し、出力手段が、更に、検索手段により抽出された付加情報をユーザに提示するために出力してもよい。

【0024】

この場合には、抽出された検索対象に対応する付加情報も出力されるので、所望の状況変化に対応する情報だけでなく該付加情報もユーザに提示することができる。

【発明の効果】

【0025】

本発明によれば、指定されたトピックについての出現頻度の推移が検索条件と一致又は類似する検索対象が抽出されるので、所望の状況変化に対応する情報を提示することができる。

【図面の簡単な説明】

【0026】

【図1】第1実施形態に係る検索システムの全体構成を示す図である。

【図2】図1に示す検索サーバの機能構成を示す図である。

【図3】図1に示す検索サーバのハードウェア構成を示す図である。

【図4】プロファイル・データの構成例を示す図である。

【図5】プロファイル・データの作成について説明するための図である。

【図6】プロファイル・データの作成について説明するための図である。

【図7】親トピックの設定について説明するための図である。

【図8】図1に示すユーザ端末で表示される検索画面の例を示す図である。

【図9】ユーザ操作に応じて将来モデルを調整する方法の一例を説明するための図である。

【図10】ユーザ操作に応じて将来モデルを調整する方法の別の例を説明するための図である。

【図11】図1に示す検索システムの動作を示すシーケンス図である。

【図12】第1実施形態に係る検索プログラムの構成を示す図である。

【図13】第2実施形態に係る検索システムの全体構成を示す図である。

【図14】図13に示す検索サーバの機能構成を示す図である。

【図15】第2実施形態に係る検索プログラムの構成を示す図である。

【発明を実施するための形態】

【0027】

以下、添付図面を参照しながら本発明の実施形態を詳細に説明する。なお、図面の説明において同一又は同等の要素には同一の符号を付し、重複する説明を省略する。

【0028】

(第1実施形態)

まず、図1～10を用いて、第1実施形態に係る検索サーバ10の機能及び構成を説明する。検索サーバ10はユーザにより指定された検索条件に対応する筆者(文書作成者)を抽出してそのユーザに提示する検索装置である。図1に示すように、検索サーバ10は

10

20

30

40

50

、通信ネットワークNを介して複数の文書データベース（所定のデータベース）D及び複数のユーザ端末Tと互いに通信可能に接続されている。図1に示す検索サーバ10、文書データベースD、及びユーザ端末Tは検索システムを構成している。

【0029】

検索サーバ10について説明する前に、文書データベースD及びユーザ端末Tについて説明する。まず、文書データベースDは、インターネット上に存在する文書（電子文書）を記憶するコンピュータである。文書の例としては、ブログやマイクロブログ、ソーシャル・ネットワーク・サービス（SNS）、オンライン・ショッピング・サイトなどで投稿された記事が挙げられるが文書の種類や内容はなんら限定されない。図1では文書データベースDを二つ示しているが、文書データベースDの台数は任意である。また、文書データベースDの設置箇所も任意であり、例えばブログ管理システムやSNSシステム、オンライン・ショッピング・システムに配置されてもよい。10

【0030】

ユーザ端末Tは、検索条件の入力を受け付け、その条件に対応する検索結果を表示するコンピュータである。ユーザ端末Tの例としては、高機能携帯電話機（スマートフォン）や携帯情報端末（PDA）、あるいは各種パーソナルコンピュータ（タブレットPC、デスクトップPC、ノートPCなど）が挙げられるが、ユーザ端末Tの例はこれらに限定されない。図1ではユーザ端末Tを3台示しているが、ユーザ端末Tの台数は任意である。

【0031】

以上を前提として検索サーバ10について説明する。図2に示すように、検索サーバ10は機能的構成要素としてプロファイル・データベース11、インターフェース提供部12、第1検索部13、モデル調整部14、及び第2検索部15を備えている。20

【0032】

この検索サーバ10は、図3に示すように、オペレーティングシステムやアプリケーション・プログラムなどを実行するCPU101と、ROM及びRAMで構成される主記憶部102と、ハードディスクなどで構成される補助記憶部103と、ネットワークカードなどで構成される通信制御部104と、キーボードやマウスなどの入力部105と、モニタなどの出力部106とで構成される。検索サーバ10の各機能は、CPU101や主記憶部102の上に所定のソフトウェアを読み込ませ、CPU101の制御の下で通信制御部104や入力部105、出力部106などを動作させ、主記憶部102や補助記憶部103におけるデータの読み出し及び書き込みを行うことで実現される。処理に必要なデータやデータベースは主記憶部102や補助記憶部103内に格納される。30

【0033】

なお、図3では検索サーバ10が1台のコンピュータで構成されているように示しているが、このサーバの機能を複数台のコンピュータに分散させてよい。例えば、プロファイル・データベース11を備えるコンピュータとそれ以外の機能的構成要素を備えるコンピュータとにより検索サーバ10を構成してもよい。

【0034】

図2に戻って、プロファイル・データベース11は、文書データベースDから抽出された文書に基づいて生成されたプロファイル・データを記憶する手段である。40

【0035】

プロファイル・データは、文書がどのようなトピック（話題）を扱っているか、言い換えれば、筆者がどのようなトピックについて言及しているかを示すデータであり、文書毎に作成される。図4に示すように、このプロファイル・データは、文書の筆者を一意に特定する筆者IDと、該文書自身を一意に特定する文書IDと、該文書の作成時刻と、所定の100個のトピック（トピック0～99）のそれぞれについて該トピックに割り当てられた単語の個数とを含んでいる。各トピックに割り当てられた単語の個数は、一の文書における該トピックの出現頻度とも言える。

【0036】

プロファイル・データは、文書の集合を分析することで複数のトピックを決定し、その50

トピック群と各文書とを照合することで生成される。この処理の詳細を説明する。

【0037】

具体的には、LDA (Latent Dirichlet Allocation) 及びギブスサンプリングを用いて、まず100個のトピックを決定する。まず、トピックを決定するために、所定の期間に作成された十分な数の文書を文書データベースDから収集する。例えば、最近1週間に作成された500万件の文書を収集する。収集する際には、所定のアプリケーション・プログラミング・インターフェース (API) を利用してもよい。例えば、Twitter (登録商標) 内の記事を収集する際には Streaming API を用いることができる。続いて、前処理として、本文に含まれるユーザ名やハッシュタグなどを除去し、形態素解析により名詞のみを文書から抽出する。

10

【0038】

続いて、抽出された名詞の集合に対してLDAを適用することで頻出単語とトピックとの対応表を生成し、100個のトピックを決める。この際には、LDA処理の収束のためにその処理を所定の回数だけ繰り返す。この繰り返し回数はサンプリング回数ともいうことができ、例えばこの回数を1000に設定してもよい。生成された対応表の例を図5に示す。図5では、商品名AやチームXなどの固有名詞を含む様々な名詞がトピック毎に集約されている。なお、図5の例では、便宜的に各トピックに「デジタル機器」のような名前を付している。

【0039】

次に、上記と同様に所定のAPIを用いて、検索サービスを提供するために必要な数のユーザのそれぞれについて、所定数の文書を文書データベースDから収集する。例えば、1ユーザあたり1000件の文書を収集する。ここで、この処理における収集対象期間はトピックを決定する際の期間（上記の例では最近1週間）とは異なり、検索サービスを提供するために必要な任意の期間（例えば3ヶ月、1年、3年など）である。続いて、各文書にLDA及びギブスサンプリングを適用することで該文書内の単語をトピック別に分類し、これにより各文書についてトピック別の出現頻度を求める。なお、この際には各文書についてのサンプリング回数を50に設定している。ギブスサンプリングではランダム性を持つ要素が存在するので、その50回1セットの処理を3回繰り返して得られた合計結果をトピック別の出現頻度としている。

20

【0040】

続いて、検索サービスの目的に応じて不要なトピックを削除する。本実施形態では、端末ユーザ（検索者）により指定された興味を持つ筆者を検索結果として該ユーザに示すので、筆者の趣味や嗜好とは関係ない情報については無視してよい。そこで、趣味や嗜好とはいえないトピックを削除する。例えば、図6に示すようなトピックを削除してもよい。なお、図6の例でも便宜的に各トピックに「時間」などのような名前を付している。このような削除処理により、例えば100個のトピックが40個に絞られる。このように不要なトピックを予め取り除くことで、筆者検索の精度を上げることができる。

30

【0041】

以上の処理により、図4に示すデータ構造を持つプロファイル・データが生成され、プロファイル・データベース11に格納される。なお、このとき、削除されたトピックに対応する出現頻度は0に初期化されている。

40

【0042】

ところで、いずれかの文書で出現するトピック数は比較的多い（例えば40個）ので、本実施形態では、ユーザの検索操作を容易にするためにこれらのトピックをいくつかのグループに集約する。例えば図7に示すように、上記処理により抽出されたトピック（削除処理後に残ったトピック）を5個の親トピック（IT、生活、政治・経済、エンターテインメント、及びゲーム・アニメ）のいずれかと関連付け、この対応関係を検索サーバ10内で保持しておく。ユーザには親トピックのみが示される。親トピックは、対応する抽出トピックの上位階層に位置する概念である。

【0043】

50

なお、このようなプロファイル・データの準備方法は上記の例に限定されず、下記のような変形が可能である。

【0044】

まず、文書の収集方法は限定されず、所定のクローリングにより文書を収集してもよいし、人手による作業もを利用して収集を行ってもよい。また、プロファイル・データは定期的に（動的に）行われてもよい。

【0045】

各処理において収集する文書の個数や範囲（時間的範囲および筆者数）、サンプリング回数、所定のサンプリング回数を1セットの処理とした場合の当該処理の繰り返し回数は上記の例に限定されず、任意に設定してよい。また、不要なトピックの削除や、抽出されたトピックの集約（親トピックの設定）は省略可能である。更に、本実施形態では単語とトピックとの対応を求めているが、語句は単語に限定されず、例えば複数の単語から成る句であってもよい。10

【0046】

図2に戻って、インターフェース提供部12は、検索処理用のユーザ・インターフェース、すなわち検索画面をユーザ端末Tに提供する手段である。検索サービスを利用しようとすると、端末ユーザが所定の操作を行うと、検索画面を要求する信号（例えばHTTPリクエスト）をユーザ端末Tが検索サーバ10に送信する。インターフェース提供部12はその信号に応じて、検索画面のデータ（例えばウェブページ）をユーザ端末Tに送信する。これにより、ユーザ端末T上に検索画面が表示される。なお、検索画面の実現方法は任意であり、例えばウェブアプリケーションではなくJava（登録商標）アプリケーションにより実装されてもよい。20

【0047】

検索画面の例を図8に示す。検索画面100は、検索条件（クエリ）を入力するための入力領域110と、検索結果を表示する表示領域120とに大きく分かれている。

【0048】

入力領域110は、ユーザ入力欄111、期間入力欄112、現在モデル欄113、将来モデル欄114、検索ボタン115、及び文書表示欄116を含んでいる。

【0049】

ユーザ入力欄111は、筆者検索の基準となるユーザ（基準ユーザ）を入力するためのインターフェースである。端末ユーザは誰を基準ユーザとして設定してもよく、例えば自分自身や他の筆者をユーザ入力欄111で指定してよい。30

【0050】

期間入力欄112は、後述する現在モデルを生成するために収集する文書の期間（現在からどれだけ遡って文書を収集するかを示す期間）を入力するためのインターフェースである。以下では、この期間を「収集期間」という。

【0051】

現在モデル欄113は、指定された収集期間において基準ユーザにより作成された文書のプロファイル・データを集計して得られた親トピック別の出現頻度（第1の出現頻度）を百分率などの割合で表示する領域である。親トピック別の出現頻度は「興味モデル」や「出現頻度分布」ということもできる。以下では、この欄に示される親トピック別の出現頻度を「現在モデル」あるいは「現在の興味モデル」という。本実施形態では、現在モデルは後述する第1検索部13により取得され、図8に示すようにヒストグラムで表示される。40

【0052】

将来モデル欄114は、端末ユーザによる親トピック別の出現頻度（第2の出現頻度）の入力を受け付ける領域である。以下では、この欄に示される親トピック別の出現頻度を「将来モデル」あるいは「将来の興味モデル」という。本実施形態では、将来モデルも割合で示され、ヒストグラムで表示される。

【0053】

10

20

30

40

50

将来モデルの初期値として、現在モデルと同じ値が設定及び表示され、端末ユーザはこの欄内に表示されている各棒を上げ下げすることで、将来モデル（各親トピックの出現頻度の大きさ）を変えることができる。各親トピックの出現頻度は比率で示されるので、端末ユーザがある親トピックについての出現頻度の大きさを変えると、他のトピックの出現頻度が自動的に調整される。具体的には、ユーザ端末Tが、現在モデルと、変更された親トピックを示す変更情報と、該親トピックの変更後の出現頻度を含む未調整将来モデル（調整前の将来モデル）のデータとを検索サーバ10に送信する。その後、ユーザ端末Tは検索サーバ10のモデル調整部14（後述）において調整された将来モデルのデータを受信して将来モデル欄114に表示する。このように、一部の指定トピックの出現頻度が変更されるとその変更に応じて他の指定トピックの出現頻度が調整されるので、検索操作の利便性が向上する。

【0054】

検索ボタン115は、入力領域110において指定された検索条件に基づく検索処理を検索サーバ10に指示するためのインターフェースである。

【0055】

文書表示欄116は、現在モデルを得るために用いた文書を端末ユーザの指示入力に応じて表示する領域である。この欄に表示される文書は、筆者IDが基準ユーザIDであり、且つ作成時期が収集期間に含まれる文書であり、文書データベースDから読み出される。

【0056】

表示領域120は、筆者選択欄121、変化前モデル欄122、変化後モデル欄123、及び文書表示欄124を含んでいる。

【0057】

筆者選択欄121は、検索された1以上の筆者（検索条件と同一又は類似する興味モデルの変化を示した筆者）のうちの一人を選択するためのインターフェースである。

【0058】

変化前モデル欄122は、筆者選択欄で選択された筆者についての、現在モデルと同一又は類似する興味モデル（以下では「変化前モデル」あるいは「変化前の興味モデル」という）を表示する領域である。

【0059】

変化前モデル欄122は、筆者選択欄で選択された筆者についての、将来モデルと同一又は類似する興味モデル（以下では「変化後モデル」あるいは「変化後の興味モデル」という）を表示する領域である。

【0060】

文書表示欄124は、筆者選択欄121で選択された筆者の文書を端末ユーザの指示入力に応じて表示する領域である。この欄に表示される文書は、筆者IDが選択された筆者のものであり、且つ作成時期が変化前モデル又は変化後モデルに対応する期間に含まれる文書であり、文書データベースDから読み出される。

【0061】

なお、興味モデルの表現方法はヒストグラムに限定されず、任意に決めてよい。例えば、円グラフを用いて興味モデルを表現してもよい。

【0062】

図2に戻って、第1検索部13は、プロファイル・データベース11内のプロファイル・データに基づいて現在モデルを生成する手段である。検索画面100において基準ユーザ及び収集期間が入力されると、ユーザ端末Tは該基準ユーザのユーザIDと該収集期間とを示す検索条件（以下では「第1の検索条件」ともいう）を検索サーバ10に送信する。第1検索部13はその検索条件を受信すると、筆者IDが基準ユーザのユーザIDであり、且つ作成時刻が収集期間に含まれるプロファイル・データをプロファイル・データベース11から読み出す。続いて、第1検索部13は抽出したプロファイル・データで示されるトピック0～99の出現頻度と、図7に示すような対応表とに基づいて、親トピック

10

20

30

40

50

別に出現頻度を集計する。続いて、第1検索部13は各親トピックの集計結果を割合に換算し、換算後の結果を興味モデルとしてユーザ端末Tに送信する。これにより、現在モデルがヒストグラムで現在モデル欄113に表示される。

【0063】

モデル調整部14は、検索画面100上での将来モデルの変更操作に応じて、該将来モデルにおける各出現頻度を調整する手段である。モデル調整部14は、ユーザ端末Tから受信した現在モデル、変更情報、及び未調整将来モデルに基づいて将来モデルの各出現頻度を調整する。調整の方法は任意に決めてよいが、ここでは二つの調整方法を説明する。

【0064】

第1の方法は、他の筆者の興味モデルの推移を参照して、指定された変化に最も近い興味モデルを調整後の将来モデルとする類似興味モデル参照法である。図9を用いてこの方法を説明する。

【0065】

現在モデル及び未調整将来モデルにおけるi番目の親トピックの出現頻度をそれぞれ T_i^C, T_i^D とすると、現在モデルはベクトル $T^C = \{T_0^C, T_1^C, T_2^C, T_3^C, T_4^C\}^t$ で表され、未調整将来モデル（希望の将来モデル）はベクトル $T^D = \{T_0^D, T_1^D, T_2^D, T_3^D, T_4^D\}^t$ で表される。なお、tは転置を表す。このとき、一部の親トピックについてのみ出現頻度が変更された場合には、モデル調整部14は、現在興味モデルと、出現頻度が変更された親トピックにのみ注目した未調整将来モデル T^D に基づいてプロファイル・データベース11を検索する。

10

20

【0066】

例えば図9に示すように、3番目及び4番目の親トピックについての出現頻度のみが矢印A1, A2のようにユーザにより変更された場合には、モデル調整部14は未調整将来モデル $T^D = \{T_0^D, T_1^D, T_2^D, T_3^D, T_4^D\}^t$ を $T^D = \{*, *, T_2^D, T_3^D, *\}^t$ と置き換える。ここで、「*」はワイルドカードである。そして、モデル調整部14は、任意の期間（他の探索時期）における第1の興味モデルが現在モデルと同一又は類似し、且つ、該期間よりも後の期間（一の探索時期）における第2の興味モデル T^F が置換後の未調整将来モデルと同一又は類似する筆者を特定する。ただし、 $T^F = \{T_0^F, T_1^F, T_2^F, T_3^F, T_4^F\}^t$ である。ここで、同一性及び類似性を判定する際には、モデル調整部14は下記第2検索部15と同様にマンハッタン距離という概念を利用する。

30

【0067】

置換後の未調整将来モデルと同一又は類似する第2の興味モデル T^F を検索する際には、モデル調整部14は変更された出現頻度についてのみ（上記の例では T_2^D, T_3^D についてのみ）マンハッタン距離を求める。なお、各モデルにおける出現頻度の総和は一定である必要があり、これは下記式(1)のように表される。

【数1】

$$\sum_i T_i^C = \sum_i T_i^D = \sum_i T_i^F = const. \quad \dots (1)$$

【0068】

そして、モデル調整部14は最も類似する筆者の第2の興味モデル T^F を将来モデルとしてユーザ端末Tに送信する。したがって、検索画面100では、現在モデル $T^C = \{T_0^C, T_1^C, T_2^C, T_3^C, T_4^C\}^t$ が将来モデル $T^F = \{T_0^F, T_1^F, T_2^F, T_3^F, T_4^F\}^t$ へと推移するよう表示される。

40

【0069】

このような第1の方法において、モデル調整部14は、最も類似する興味モデル T^F 以外を将来モデルとして設定してもよい。例えば、モデル調整部14はマンハッタン距離が小さい順に抽出した所定数の興味モデルの平均を将来モデルとしてユーザ端末Tに送信してもよい。また、モデル調整部14は、マンハッタン距離が所定値以下となる興味モデルの平均を将来モデルとしてユーザ端末Tに送信してもよい。

50

【0070】

第2の方法は、親トピック間の関連性を考慮して将来モデルの各出現頻度を調整するトピック伝播法である。LDAにより各単語がトピックと対応付けられる際には、一つの単語が複数のトピックに所属する場合がある。このような単語の重複の頻度によりトピック間に関連が生まれ、したがって、親トピック間にも関連が生まれる。トピック伝播法では、親トピックそれぞれの間の関連の強さを示す重みを利用する。図10を用いてこの方法を説明する。

【0071】

図10に示すように、5個の親トピックをノードn0～n4で表し、二つのトピック間の重みを $w_{i,j}$ (i,jは互いに隣接するノードの番号)で表すものとする。なお、重みwは予め検索サーバ10内の所定の記憶手段に格納されており、モデル調整部14により参照されるものとする。10

【0072】

これを前提として、例えば、ノードn2に対応する出現頻度がユーザ操作によりvだけ変更されたとする。ノードn2に関連する重みは $w_{0,2}, w_{1,2}, w_{2,3}, w_{2,4}$ であるから、ノードn2の出現頻度を変えるためには、これらの重みを考慮して、他ノードn0,n1,n3,n4との間で移動させる出現頻度の量を決めればよい。 $w_T = w_{0,2} + w_{1,2} + w_{2,3} + w_{2,4}$ とすると、モデル調整部14は他ノードとノードn2との間の移動量を以下のように決定する。

- ・ノードn0との間では、 $v(w_{0,2}/w_T)$
- ・ノードn1との間では、 $v(w_{1,2}/w_T)$
- ・ノードn3との間では、 $v(w_{2,3}/w_T)$
- ・ノードn4との間では、 $v(w_{2,4}/w_T)$

20

【0073】

ただし、 $v > 0$ 、すなわちノードn2について出現頻度を増やした場合には、上記計算により出現頻度が負になる他ノードが現れることがある。その場合には、モデル調整部14は該他ノードについての出現頻度が0に戻るまでノードn2から該他ノードに出現頻度を戻す。そして、モデル調整部14は、ノードn2にとって不足分の量 v' を他ノード(出現頻度が0に戻されたノードは除く)から移す処理を上記と同様に実行する。なお、ユーザ操作により複数のトピックの出現頻度が変更された場合には、モデル調整部14は各ノード(各トピック)の変更量 v_k (kはノード番号)について上記の調整を行えばよい。30。モデル調整部14は調整後の将来モデルをユーザ端末Tに送信する。これにより、検索画面100に調整後の将来モデルが表示される。

【0074】

図2に戻って、第2検索部15は、検索画面100の入力領域110で指定された検索条件(以下では「第2の検索条件」ともいう)に対応する筆者を検索する手段である。第2検索部15は受信部15a、抽出部(検索手段)15b、及び送信部(出力手段)15cを備えている。

30

【0075】

受信部15aは、ユーザ端末Tから検索条件を受信する手段である。検索画面100上の検索ボタン115が端末ユーザによりクリックされると、ユーザ端末Tが検索条件を検索サーバ10に送信し、受信部15aが該検索条件を受信する。受信部15aは受信した検索条件を抽出部15bに出力する。40

【0076】

この検索条件(第2の検索条件)には、端末ユーザのユーザID、基準ユーザのユーザID、収集期間、該収集期間に対応する現在モデル、及び将来モデルが含まれている。収集期間に対応する現在モデルは、第1の指定時期(収集期間)における親トピック別の第1の出現頻度に相当する。また、将来モデルは、第2の指定時期(将来のある時点又は期間)における親トピック別の第2の出現頻度に相当する。本実施形態では5個の親トピックが指定トピックとして扱われている。50

【0077】

抽出部15bは、プロファイル・データベース11内のプロファイル・データに基づいて、入力された検索条件に対応する筆者を抽出する手段である。抽出部15bは、端末ユーザ及び基準ユーザとは異なるユーザを検索対象の筆者として設定する。すなわち、抽出部15bは入力された二つのユーザIDのいずれとも異なる筆者IDを持つプロファイル・データに対して検索処理を実行する。したがって、検索対象を特定する識別子は筆者IDである。

【0078】

抽出部15bは、検索対象の各筆者について次のような検索処理を実行する。すなわち、抽出部15bは入力された収集期間に対応する任意の期間（以下では「開始期間」ともいう。第1の検索対象時期に相当する）に対応するプロファイル・データをプロファイル・データベース11から読み出す。ここで、開始期間と収集期間とでは、期間の始点及び終点は互いに異なるが、期間の長さは同じである。ただし、厳密な長さは異なっていてもよい。例えば、収集期間が直近の1ヶ月であってそれが4月1日～4月30日の30日間である場合に、開始期間が3月1日～3月31日の31日間であったり2月1日～2月28日の28日間であったりしてもよい。

10

【0079】

続いて、抽出部15bは読み出した1以上のプロファイル・データで示される出現頻度を親トピック別に集計することで、開始期間における興味モデルを生成する。続いて、抽出部15bは生成された興味モデルが現在モデルと同一又は類似するか否かを判定する。

20

【0080】

抽出部15bは、興味モデルの同一性及び類似性を判定する際に、下記式(1)で示されるマンハッタン距離 $d_1(x, y)$ を用いる。

【数2】

$$d_1(x, y) = \sum_{k=1}^n |x_k - y_k| \quad \dots(2)$$

ここで、左辺の x は検索条件として設定された興味モデルのベクトル、左辺の y は検索対象の筆者についての興味モデルのベクトルである。また、 n は親トピックの個数であり、本実施形態では $n = 5$ である。右辺の x_k は検索条件として設定された興味モデルにおける親トピック k の出現頻度（比率）であり、右辺の y_k は検索対象の筆者についての興味モデルにおける親トピック k の出現頻度（比率）である。ベクトル x, y はそれぞれ、 $x = \{x_1, x_2, x_3, x_4, x_5\}$ 、 $y = \{y_1, y_2, y_3, y_4, y_5\}$ と表される。

30

【0081】

抽出部15bは、上記マンハッタン距離が所定の閾値以下である場合に、開始期間における興味モデルが現在モデルと同一又は類似すると判定する。なお、閾値の決定方法は任意である。マンハッタン距離が0であれば、現在モデルと同一の興味モデルが得られたことになる。

【0082】

また、抽出部15bは、期間の長さが収集期間と同じであり、且つ始点が開始期間よりも後である任意の期間（以下では「終了期間」ともいう。第2の検索対象時期に相当する）に対応するプロファイル・データをプロファイル・データベース11から読み出す。ここで、終了期間と収集期間との間における始点、終点、及び期間の長さの関係は、開始期間に関して述べたことと同様である。続いて、抽出部15bは読み出した1以上のプロファイル・データで示される出現頻度を親トピック別に集計することで、終了期間における興味モデルを生成する。続いて、抽出部15bは生成された興味モデルが将来モデルと同一又は類似するか否かを判定する。この判定を行う場合も、抽出部15bは上記式(1)で示されるマンハッタン距離 $d_1(x, y)$ を用いて開始期間の場合と同様に同一性及び類似性を判定する。

40

50

【0083】

抽出部15bは、一の筆者について開始期間及び終了期間を変えながら、これら二つの期間においてマンハッタン距離が所定の閾値以下となる興味モデルの組合せを抽出する。そして、この組合せに係る、開始期間及び終了期間に対応する二つの興味モデルを、それぞれ変化前モデル及び変化後モデルとして保持する。また、抽出部15bは変化前後の興味モデルに対応する開始期間及び終了期間も保持する。なお、検索条件と同一又は類似する変化前及び変化後の興味モデルの組合せが複数得られた場合には、抽出部15bは任意の方法で一の組合せを選択すればよい。例えば、抽出部15bはマンハッタン距離の和又は平均値が最小の組合せを選択してもよい。

【0084】

10

抽出部15bは、各筆者についてこのような処理を行い、入力された検索条件に合致する1以上の筆者についての筆者ID、変更前モデル、変更後モデル、並びに、これらのモデルに対応する開始期間及び終了期間を取得する。そして、抽出部15bは取得したこれらのデータを検索結果として送信部15cに出力する。

【0085】

送信部15cは、入力された検索結果を端末ユーザに提示するために、該検索結果をユーザ端末Tに送信する手段である。これにより、ユーザ端末Tに表示されている検索画面100の表示領域120に変更前モデル及び変更後モデルが表示される。

【0086】

20

次に、図11を用いて、検索システム（特に検索サーバ10）の動作を説明するとともに本実施形態に係る検索方法について説明する。

【0087】

30

以下では、ユーザ端末T上に検索画面100が表示されたことを前提として説明する（ステップS11）。端末ユーザにより現在モデルを得るために第1の検索条件、すなわち基準ユーザ及び収集期間が指定されると（ステップS12）、ユーザ端末Tが該第1の検索条件を検索サーバ10に送信する（ステップS13）。検索サーバ10では、第1検索部13がその検索条件に基づいてプロファイル・データベース11からプロファイル・データを抽出し、そのプロファイル・データに基づいて現在モデルを生成する（ステップS14）。そして、第1検索部13はその現在モデルをユーザ端末Tに送信する（ステップS15）。ユーザ端末Tでは、その現在モデルが検索画面100に表示される（ステップS16）。この際に、本実施形態では現在モデルと同じ値が将来モデルの初期値として設定及び表示される。

【0088】

40

その後、端末ユーザによりその将来モデル内の出現頻度が変更されると（ステップS17）、ユーザ端末Tがその変更を示す未調整将来モデルを検索サーバ10に送信する（ステップS18）。検索サーバ10では、モデル調整部14がその未調整将来モデルに対して類似興味モデル参照法やトピック伝播法などの手法を適用することでその将来モデルを調整し（ステップS19）、調整された将来モデルをユーザ端末Tに送信する（ステップS20）。ユーザ端末Tでは調整後の将来モデルが検索画面100に表示される（ステップS21）。ステップS17～S21の処理は、端末ユーザの操作に応じて繰り返し実行され得る。

【0089】

その後、端末ユーザが検索ボタン115をクリックして検索を指示すると（ステップS22）、ユーザ端末Tが、筆者を検索するための条件、すなわち、現在及び将来の興味モデルを含む第2の検索条件を検索サーバ10に送信する（ステップS23）。

【0090】

50

検索サーバ10では、受信部15aがその検索条件を受信する。続いて、抽出部15bがその検索条件に基づいてプロファイル・データベース11からプロファイル・データを抽出し、そのプロファイル・データに基づいて筆者を検索する（ステップS24、検索ステップ）。具体的には、抽出部15bは、マンハッタン距離という概念を用いて、開始期

間における興味モデルが現在モデル同一又は類似し、且つ終了期間における興味モデルが将来モデルと同一又は類似する筆者を抽出する。この際に抽出部 15 b は、生成及び判定した興味モデルを変化前モデル及び変化後モデルとして保持する。そして、送信部 15 c が抽出された筆者をその変化前後の興味モデルと共に検索結果としてユーザ端末 T に送信する（ステップ S 25、送信ステップ）。ユーザ端末 T では、その検索結果が検索画面 100 に表示される（ステップ S 26）。これにより、端末ユーザは検索結果を得ることができる。

【0091】

次に、図 12 を用いて、コンピュータを検索サーバ 10 として機能させるための検索プログラム P 1 を説明する。

10

【0092】

検索プログラム P 1 は、メインモジュール P 10、プロファイル記憶モジュール P 11、インターフェース提供モジュール P 12、第 1 検索モジュール P 13、モデル調整モジュール P 14、及び第 2 検索モジュール P 15 を備えている。第 2 検索モジュール P 15 は、受信モジュール P 15 a、抽出モジュール P 15 b、及び送信モジュール P 15 c を備えている。

【0093】

メインモジュール P 10 は、検索機能を統括的に制御する部分である。プロファイル記憶モジュール P 11、インターフェース提供モジュール P 12、第 1 検索モジュール P 13、モデル調整モジュール P 14、及び第 2 検索モジュール P 15 を実行することにより実現される機能はそれぞれ、上記のプロファイル・データベース 11、インターフェース提供部 12、第 1 検索部 13、モデル調整部 14、及び第 2 検索部 15 の機能と同様である。受信モジュール P 15 a、抽出モジュール P 15 b、及び送信モジュール P 15 c を実行することにより実現される機能はそれぞれ、上記の受信部 15 a、抽出部 15 b、及び送信部 15 c の機能と同様である。

20

【0094】

検索プログラム P 1 は、例えば、CD-ROM や DVD-ROM、半導体メモリ等の有形の記録媒体に固定的に記録された上で提供される。また、検索プログラム P 1 は、搬送波に重畠されたデータ信号として通信ネットワークを介して提供されてもよい。

【0095】

30

以上説明したように、本実施形態によれば、二つの時期における指定トピック別の出現頻度（現在及び将来の興味モデル）と同一又は類似する二つの出現頻度（変化前及び変化後の興味モデル）を有する筆者が抽出される。そして、抽出された筆者が検索結果としてユーザ端末 T に送信され、検索画面 100 上に表示される。このように、指定されたトピックについての出現頻度の変化が検索条件と一致又は類似する筆者を抽出することで、ユーザに指定されたように興味が変化した筆者の情報をそのユーザに提示することができる。なお、筆者の興味の変化は状況変化の一種である。

【0096】

（第 2 実施形態）

次に、図 13, 14 を用いて、第 2 実施形態に係る検索サーバ 10 A の機能及び構成を説明する。この検索サーバ 10 A は、筆者検索だけでなく、検索された筆者により購入された商品や該筆者の興味モデルの変化に対応する商品の情報をユーザ端末に提供する点で、第 1 実施形態における検索サーバ 10 と異なる。以下では、第 1 実施形態と同一又は同等の事項については説明を省略する。

40

【0097】

図 13 に示すように、本実施形態の検索システムは、追加のデータベースである商品データベース P 及び購入履歴データベース R を更に備えている。商品データベース P は、商品を特定する商品 ID と、その商品の名前やカテゴリ、価格、商品画像などの任意の商品属性とを含む商品データを記憶する手段である。また、購入履歴データベース R は、商品を購入したユーザを特定するユーザ ID と、その商品の商品 ID と、購入日時とを含む購

50

入データを記憶する手段である。商品データベース P 及び購入履歴データベース R の設置箇所としては、例えばオンライン・ショッピング・システム内が挙げられるが、これらのデータベースはどこに設けられてもよく、同一のシステム内に設けられる必要もない。また、図 1 では商品データベース P 及び購入履歴データベース R を一つのみ示しているが、これらのデータベースの台数も何ら限定されない。検索サーバ 10 A は、商品データベース P 及び購入履歴データベース R とも通信ネットワーク N を介して通信可能である。

【0098】

図 14 に示すように、検索サーバ 10 A は機能的構成要素としてプロファイル・データベース 11、インターフェース提供部 12、第 1 検索部 13、モデル調整部 14、及び第 2 検索部 15 A を備えている。以下では、第 1 実施形態とは異なる第 2 検索部 15 A について説明する。10

【0099】

第 2 検索部 15 A は、受信部 15 a、抽出部 15 b、レコメンド部（検索手段）15 d、及び送信部 15 c を備えている。受信部 15 a 及び送信部 15 c の機能は第 1 実施形態で説明したものと同様である。また、抽出部 15 b は、検索結果をレコメンド部 15 d に出力するという点では第 1 実施形態と異なるが、該検索結果を得る機能の詳細は第 1 実施形態で説明したものと同様である。したがって、以下では受信部 15 a、抽出部 15 b、及び送信部 15 c の説明を省略し、レコメンド部 15 d についてのみ説明する。

【0100】

レコメンド部 15 d は、端末ユーザに提示するレコメンド商品（推奨商品）を抽出する手段である。ここでは、二つの抽出方法を説明する。20

【0101】

第 1 の方法は、検索された筆者に関する商品を抽出する方法である。レコメンド部 15 d は、入力された検索結果で示される各筆者について次の処理を実行する。

【0102】

レコメンド部 15 d は、一の筆者の変化前モデル及び変化後モデルに対応する開始期間及び終了期間を参照して、これら二つの期間の時間間隔を推移期間として取得する。なお、推移期間の始点は開始期間の始点又は終点であり、推移期間の終点は終了期間の始点又は終点である。

【0103】

続いて、レコメンド部 15 d はユーザ ID が筆者のユーザ ID であり、且つ購入日時がその推移期間に含まれる購入データを購入履歴データベースから読み出すことで、筆者が該推移期間に購入した商品の商品 ID を取得する。30

【0104】

このときに 1 以上の商品 ID を取得できた場合には、レコメンド部 15 d は該商品 ID に対応する商品データを商品データベース P から読み出し、商品名や価格、商品画像などを含むレコメンド情報（付加情報）を生成する。なお、この際には、レコメンド部 15 d は商品データで示される商品のカテゴリと 5 個の親トピックとを比較して、いずれかの親トピックに対応するカテゴリに属する商品のみについてレコメンド情報を生成してもよい。40

【0105】

例えば、レコメンド部 15 d は、変化後モデルの中で出現頻度が最も高い親トピックに対応するカテゴリに属する商品のみを抽出してもよい。また、レコメンド部 15 d は、変化後モデルの中で出現頻度が所定値以上である親トピックに対応するカテゴリに属する商品のみを抽出してもよい。更に、レコメンド部 15 d は、変化前モデルから変化後モデルに推移する過程で出現頻度が最も増大したトピックに対応するカテゴリに属する商品のみを抽出してもよい。トピックと商品カテゴリとの対応は任意に定めてよく、レコメンド部 15 d はその対応を示すデータを予め内部に保持している。

【0106】

このようにレコメンド情報を生成すると、レコメンド部 15 d はそのレコメンド情報を50

入力された検索結果に含めて送信部 15c に出力する。

【0107】

一方、商品 ID を一つも取得できなかった場合には、レコメンド部 15d は入力された検索結果をそのまま送信部 15c に出力する。

【0108】

第 2 の方法は、変化前モデルから変化後モデルにどのように推移したかを示す推移パターンに基づいて、これらのモデルで示されるトピックに対応するカテゴリの商品を抽出する方法である。

【0109】

変化前モデルから変化後モデルに達するまでに時間 s を要するとし、変化前モデルの時点を $t = 0$ 、変化後モデルの時点を $t = s$ とする。そして、変化前及び変化後の興味モデル及びをそれぞれ $K(0)$ 、 $K(s)$ と表し、時刻 t における興味モデルを示す関数 $K(t) = \{K_0^t, K_1^t, K_2^t, K_3^t, K_4^t\}^T$ と表す。なお、 T は転置を示す。このとき、値 $K(t)$ は任意の関数 $f(t)$ を用いて下記式 (3) のように規定することができる。
10

$$K(t) = K(0) + (K(s) - K(0)) \cdot f(t)$$

【0110】

ここで、関数 $f(t)$ は単純な線形関数 (t/s) であってもよいし、指數関数的あるいは対数関数的に推移する関数であってもよい。例えば、関数 $f(t)$ が線形関数であれば、関数 $K(t)$ は「一定のスピードで増大する」推移モデル（線形モデル）を示すといえる。また、関数 $f(t)$ が指數関数的なものであれば、関数 $K(t)$ は「当初は緩やかに増大して、その後急激に増大する」推移モデル（指數関数的モデル）を示すといえる。また、関数 $f(t)$ が対数関数的なものであれば、関数 $K(t)$ は「当初は急激に増大して、その後緩やかに増大する」推移モデル（対数関数的モデル）を示すといえる。なお、レコメンド部 15d は複数の関数 $K(t)$ を取得してもよい。
20

【0111】

続いて、レコメンド部 15d は、取得した推移モデルに対応する商品を商品データベース P から読み出す。ここで、この処理の前提として、商品データには、商品のカテゴリだけでなく、対応する推移モデルの種類（例えば、線形モデル、指數関数的モデル、対数関数的モデル）や、該推移モデルのどの時点（例えば前半、中盤、後半など）に商品が対応するかを示すデータも含まれているものとする。例えば、レコメンド部 15d は、取得した推移モデル（関数 $K(t)$ ）毎に、該推移モデルで示される親トピックに対応するカテゴリを有し、且つその関数 $K(t)$ に対応する推移モデルの種類が指定されている商品を商品データベース P から読み出す。1 以上の商品データを取得できた場合には、レコメンド部 15d はその商品データに基づいて各推移モデルについてのレコメンド情報（付加情報）を生成する。そして、レコメンド部 15d は生成したレコメンド情報を入力された検索結果に含めて送信部 15c に出力する。
30

【0112】

レコメンド商品情報を含む検索結果はユーザ端末 T により受信及び表示される。この際に、レコメンド商品情報をどのように表示させるかは任意に決めてよい。例えば、レコメンド商品を表示するための領域を検索画面内に設けたり、別の画面にレコメンド商品を表示させてもよい。いずれにせよ、端末ユーザは、検索された筆者の興味が変化する課程で該筆者がどのような商品を購入したかを知ることができる。
40

【0113】

次に、図 15 を用いて、コンピュータを検索サーバ 10A として機能させるための検索プログラム P 2 を説明する。

【0114】

検索プログラム P 2 は、第 2 検索モジュール P 15 に代えて、レコメンドモジュール P 15d を含む第 2 検索モジュール P 15A を備える点で、第 1 実施形態における検索プログラム P 1 と異なる。第 2 検索モジュール P 15A を実行することにより実現される機能
50

は第2検索部15Aの機能と同様である。レコメンドモジュールP15dを実行することにより実現される機能は上記レコメンド部15dの機能と同様である。検索プログラムP2も検索プログラムP1と同様に提供される。

【0115】

本実施形態においても、第1実施形態と同様の効果を得ることができる。加えて、本実施形態では、筆者あるいは推移パターンに対応するレコメンド情報も付加情報としてユーザ端末Tに送信される。したがって、ユーザは、検索条件で指定したように興味モデルが変化した筆者の情報だけでなく、その筆者あるいは推移パターンに関連する情報も得ることができる。

【0116】

以上、本発明をその実施形態に基づいて詳細に説明した。しかし、本発明は上記実施形態に限定されるものではない。本発明は、その要旨を逸脱しない範囲で様々な変形が可能である。

【0117】

上記各実施形態では、検索条件が現在モデル及び将来モデルという二つの興味モデルを含んでいたが、検索条件に含める興味モデル（指定トピック別の出現頻度）の個数は3以上でもよい。すなわち、現在モデルと最終到達時点での将来モデルだけでなく、現在から該最終到達時点までの間に設けられた1以上の間点での興味モデル（第3の指定時期における指定トピック別の第3の出現頻度。以下では「中間モデル」という）も検索条件に含めてよい。

【0118】

上記各実施形態では、現在モデルに対応する収集期間（第1の指定時期）が所定の過去の時点から現在までの期間であり、将来モデルに対応する期間（第2の指定時期）が将来の一時点又は一期間であったが、これら指定時期は任意に設定してよい。例えば、現在から3~2ヶ月前までの1ヶ月間を第1の指定時期とし、現在から2~1ヶ月前までの1ヶ月間を第2の指定時期としてもよい。すなわち、検索条件として設定する指定時期がいずれも過去の時期であってもよい。これとは逆に、各指定時期がいずれも将来の時期であってもよい。また、各指定時期は、幅を持った時間でなくともよく、ある一時点（時刻）であってもよい。これに応じて、検索結果で示される各興味モデルに対応する期間（検索対象期間）も、時間ではなくある一時点（時刻）であってもよい。

【0119】

この場合、抽出部15bは一の筆者について開始期間、終了期間、及びこれらの期間の間に位置する1以上の間点（第3の検索対象時期）を変えながら、これら複数の期間においてマンハッタン距離が所定の閾値以下となる興味モデルの組合せを抽出する。間点における興味モデルの同一性及び類似性の判定方法は、上述した変化前後の興味モデルに対して行われる方法と同様である。そして、抽出部15bは抽出した組合せに係る、開始期間、終了期間、及び1以上の間点についての複数の興味モデルを、それぞれ変化前モデル、変化後モデル、及び1以上の間点結果モデルとして保持する。そして、抽出部15bは検索した興味モデルをその対応期間と共に検索結果として送信部15cに出力する。

【0120】

例えば、現在モデル及び将来モデルに加えて、時点Ta, Tb, Tc（ただし、Ta < Tb < Tc）に対応する三つの中間モデルMa, Mb, Mcが検索条件として抽出部15bに入力されたとする。この場合には、抽出部15bはこれら3モデルと同一又は類似する三つの中間結果モデルMa', Mb', Mc'を抽出する。このとき、各中間結果モデルに対応する検索対象時期がそれぞれTa', Tb', Tc'とすれば、Ta' < Tb' < Tc'の関係が成り立つ。すなわち、検索対象となる複数の中間期間は、検索条件で指定された中間点の前後関係を維持している。

【0121】

このように3以上の興味モデルを用いて検索することで、状況変化の過程を考慮した検

10

20

30

40

50

索を行うことができるので、ユーザに指定されたように興味モデルが徐々に変化した筆者の情報をそのユーザに提示することができる。

【0122】

上記各実施形態において、現在モデルから将来モデルに移るまでの期間（推移期間）を検索条件に含めてよい。この場合には、抽出部15bは、開始期間から終了期間までの時間間隔が推移期間と一致することも条件に変化前後の興味モデルを抽出し、マンハッタン距離を用いて興味モデルの同一性及び類似性を判定する。これにより、指定された時間範囲内で指定されたように興味モデルが変化した筆者の情報をユーザに提示することができる。

【0123】

上記各実施形態では複数のトピックを含む興味モデルを用いたが、1トピックのみの出現頻度から成るモデルを用いてもよい。この場合には、モデルの推移は、各出現頻度の比率の変化ではなく、一つの出現頻度の増加率または減少率で表される。例えば、特定の資格試験のテストの評価値（例えば点数や偏差値）に関する1トピックから成るモデルを用いてもよい。1トピックのみのモデルを扱う場合には、上述したモデル調整部14は不要となるが、筆者検索の手法自体は上記実施形態と同様である。

【0124】

上記各実施形態では検索画面100の将来モデル欄114で一部の出現頻度がユーザ操作により変更されると他の出現頻度が自動的に調整されたが、このような調整機能を省略し、ユーザに調整を任せてもよい。この場合には、モデル調整部14を省略することができる。また、現在モデル欄113内の出現頻度もユーザ操作により変更可能としてもよい。この場合には、モデル調整部14は未調整現在モデルと将来モデルとに対して上記類似興味モデル参照法を適用することで現在モデルを調整してもよいし、未調整現在モデルに対して上記トピック伝播法を適用することで現在モデルを調整してもよい。上記のように中間モデルを検索条件に含める場合にも同様に、モデル調整部14はユーザ操作に応じてその中間モデルを調整することができる。

【0125】

上記各実施形態では文書の筆者を検索対象としたが、検索対象は何ら限定されない。例えば、商品やサービスなどを検索対象としてもよい。この場合には、文書データベースから収集した口コミ情報や商品レビューなどに基づいて、その検索対象に適したトピックを有するプロファイル・データを生成してもよい。これによりユーザは、例えば、最近人気が出てきた商品や、以前よりも評判が良いホテルなどを見つけることが可能になる。

【0126】

上記各実施形態では、端末ユーザの目的は所望の興味変化を経験した筆者を検索することにあったが、検索の目的は任意である。これに関連して、設定するトピックも任意に設定可能である。また、トピック別の出現頻度、すなわちモデルは「興味」以外の要素を示すものであってよい。例えば、学力の推移を把握するのに適したトピックを有するプロファイル・データを用意しておき、教師が生徒に対して目標設定を指導する目的や、ある個人が自身で目標を設定する目的で上記のような検索システムを活用してもよい。

【0127】

上記各実施形態では、マンハッタン距離を用いて興味モデルの同一性及び類似性を判定したが、ユークリッド距離などのような他の尺度や指標を用いてその同一性及び類似性を判定してもよい。

【0128】

上記各実施形態ではプロファイル・データベース11を検索サーバ10, 10A内に設けたが、プロファイル・データベースは検索サーバの外部や、検索サーバを含むシステムの外部に設けられていてもよい。

【0129】

上記各実施形態では、ユーザ端末Tと検索サーバ10, 10Aとにより、いわゆるクライアント・サーバ型の検索システムが構築されていたが、上述した検索サーバ10, 10

10

20

30

40

50

Aの機能をユーザ端末に組み込んでもよい。この場合には、ユーザ端末内で抽出された検索結果が所定のモニタ上に出力（表示）される。

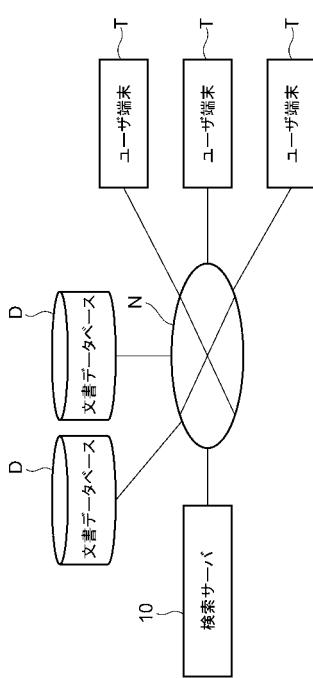
【符号の説明】

【0130】

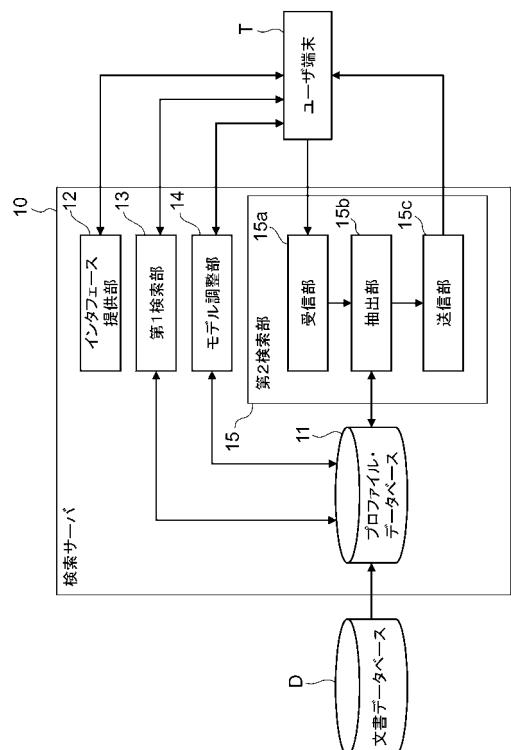
10, 10A...検索サーバ（検索装置）、11...プロファイル・データベース、12...インタフェース提供部、13...第1検索部、14...モデル調整部、15, 15A...第2検索部、15a...受信部、15b...抽出部（検索手段）、15c...送信部（出力手段）、15d...レコメンド部（検索手段）、D...文書データベース（所定のデータベース）、P...商品データベース（追加のデータベース）、R...購入履歴データベース（追加のデータベース）、P1, P2...検索プログラム、P10...メインモジュール、P11...プロファイル記憶モジュール、P12...インタフェース提供モジュール、P13...第1検索モジュール、P14...モデル調整モジュール、P15, P15A...検索モジュール、P15a...受信モジュール、P15b...抽出モジュール、P15c...送信モジュール、P15d...レコメンドモジュール、T...ユーザ端末。

10

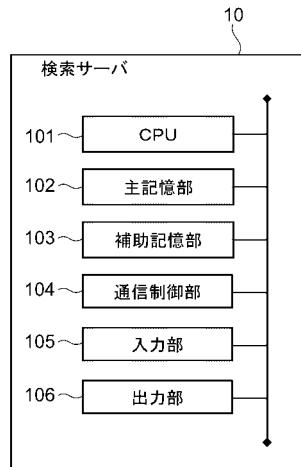
【図1】



【図2】



【図3】



【図4】

プロファイル・データ	
筆者ID	
文書ID	
作成時刻	
トピック0に割り当てられた単語の個数(トピック0の出現頻度)	
トピック1に割り当てられた単語の個数(トピック1の出現頻度)	
トピック2に割り当てられた単語の個数(トピック2の出現頻度)	
⋮	
トピック98に割り当てられた単語の個数(トピック98の出現頻度)	
トピック99に割り当てられた単語の個数(トピック99の出現頻度)	

【図5】

頻出単語	トピック
…	…
用 パソコン 形態 商品名A 商品名B 型	トピックn1(デジタル機器)
試合 応援 犬 選手 チーム 選手 チーム チームY	トピックn2(野球)
風邪 今年 秋 度 体調 冬 大事 季節 気候 気	トピックn3(風邪)
風邪 月 心 緋麗 海 空 花 水 星 光 夜	トピックn4(自然)
猫 犬 山 駒歩 うち 川 クマ 鳥 森 虫	トピックn5(ペット)
…	…

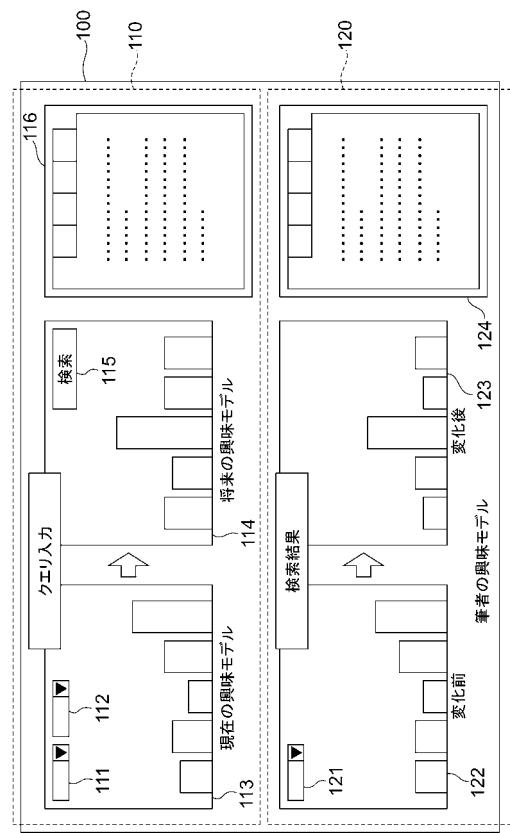
【図6】

頻出単語	トピック
…	…
時 分 度 時間 午後 沖 朝 時半 お知らせ	トピックm1(時間)
(△) (○) (×) (*) *	トピックm2(アスキー文字)
今日 雨 朝 天気 昨日 今朝 日 傘	トピックm3(天気)
そう 無理 絶対 ダメ 嫌 口 これ	トピックm4(感情)
金 ⇒ 簡単 方法 パック 完全 成功	トピックm5(広告)
…	…

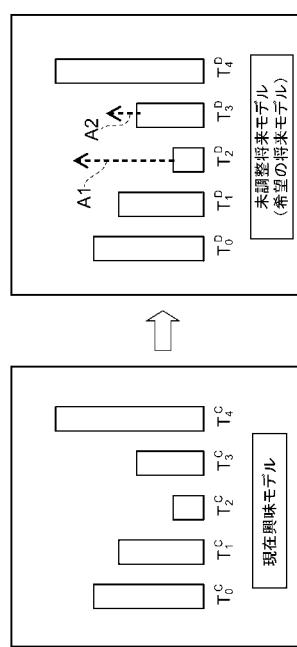
【図7】

類トピック	抽出されたトピック
IT	デジタル機器、モバイル、...
生活	ファッション、飲食、運動、部屋、育児、...
政治・経済	政治、経済、国際、...
エンターテインメント	野球、芸能、自然、ライフ音樂、...
ゲーム・アニメ	ゲーム、アニメーション、本、配信、...

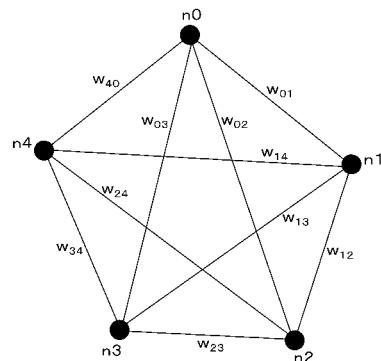
【図8】



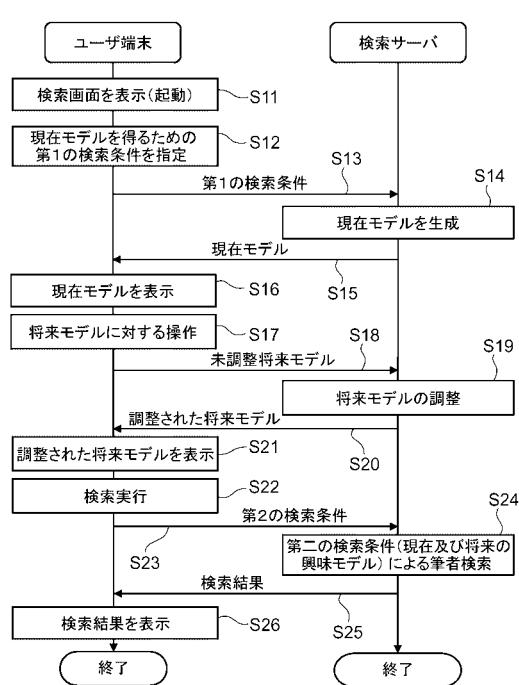
【図9】



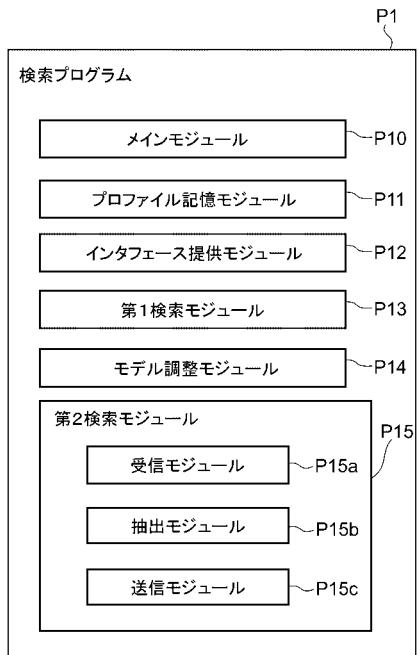
【図10】



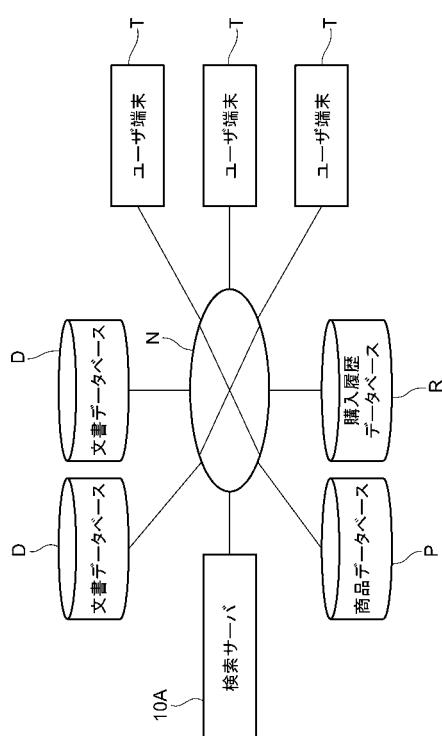
【図11】



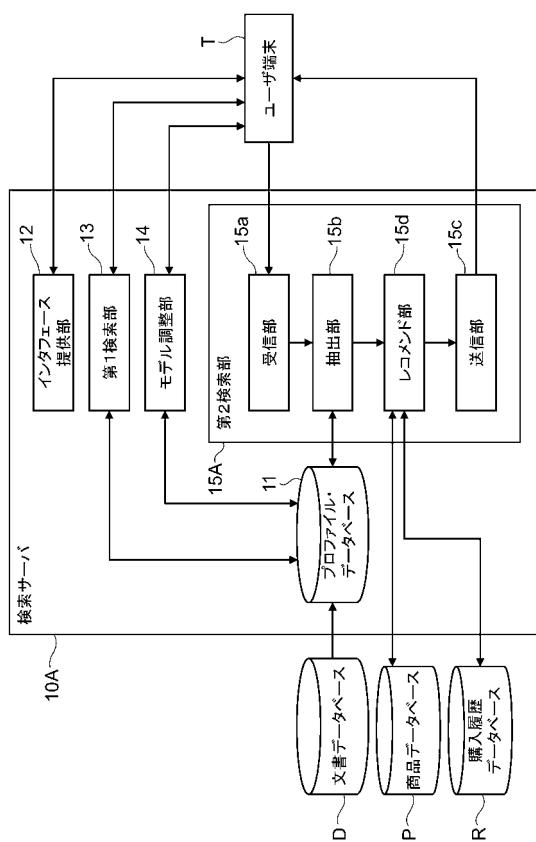
【図12】



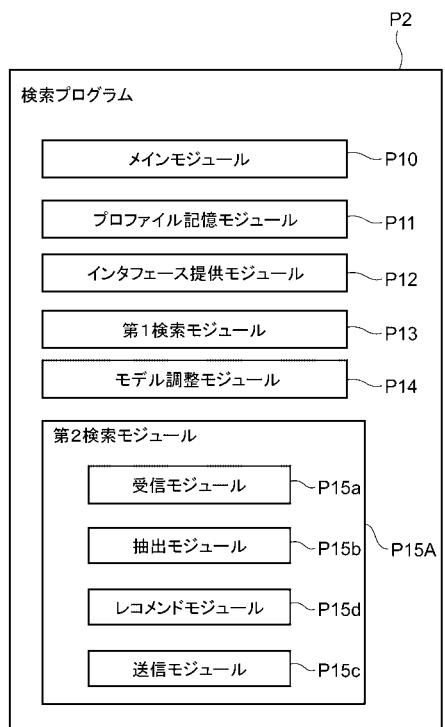
【図13】



【図14】



【図15】



フロントページの続き

(72)発明者 尾内 理紀夫
東京都調布市調布ヶ丘一丁目5番地1 国立大学法人電気通信大学内

(72)発明者 三條 正裕
東京都品川区東品川四丁目12番3号 楽天株式会社内

(72)発明者 平野 廣美
東京都品川区東品川四丁目12番3号 楽天株式会社内

審査官 吉田 誠

(56)参考文献 国際公開第2010/035455 (WO, A1)
特開2009-187395 (JP, A)
特開2001-216311 (JP, A)
国際公開第2009/060888 (WO, A1)

(58)調査した分野(Int.Cl., DB名)

G 06 F 17 / 30