

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局

(43) 国際公開日
2016 年 11 月 17 日(17.11.2016)

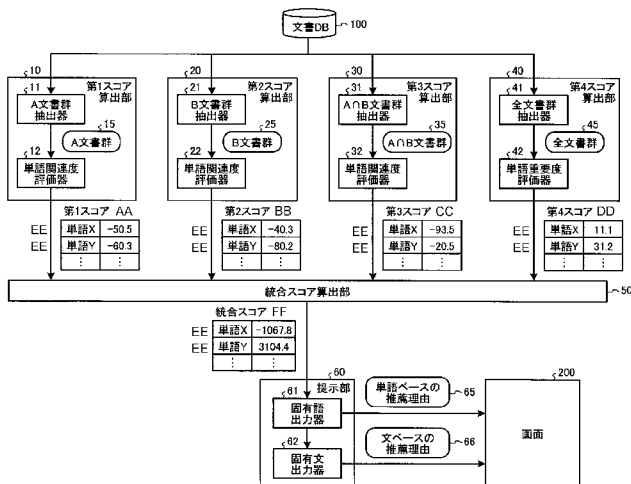


(10) 国際公開番号
WO 2016/181475 A1

- (51) 国際特許分類:
G06F 17/30 (2006.01)
- (21) 国際出願番号: PCT/JP2015/063532
- (22) 国際出願日: 2015 年 5 月 11 日(11.05.2015)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (71) 出願人: 株式会社東芝 (KABUSHIKI KAISHA TOSHIBA) [JP/JP]; 〒1058001 東京都港区芝浦一丁目 1 番 1 号 Tokyo (JP). 東芝ソリューション株式会社 (TOSHIBA SOLUTIONS CORPORATION) [JP/JP]; 〒2128585 神奈川県川崎市幸区堀川町 7 2 番地 3 4 Kanagawa (JP).
- (72) 発明者: 浜田 伸一郎 (HAMADA, Shinichiro); 〒2128585 神奈川県川崎市幸区堀川町 7 2 番地 3 4 東芝ソリューション株式会社内 Kanagawa (JP).
- (74) 代理人: 特許業務法人酒井国際特許事務所 (SAKAI INTERNATIONAL PATENT OFFICE); 〒1000013 東京都千代田区霞が関 3 丁目 8 番 1 号 虎の門三井ビルディング Tokyo (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).
- 添付公開書類:
— 国際調査報告 (条約第 21 条(3))

(54) Title: INFORMATION PRESENTATION DEVICE, INFORMATION PRESENTATION METHOD, AND PROGRAM

(54) 発明の名称: 情報提示装置、情報提示方法およびプログラム



(57) Abstract: An information presentation device according to an embodiment of the present invention is provided with: a first score calculation unit (10) for extracting an A-document group (15) from a document database (100) and calculating the first score of each of the words included therein; a second score calculation unit (20) for extracting a B-document group (25) from the document database (100) and calculating the second score of each of the words included therein; a third score calculation unit (30) for extracting an A∩B-document group (35) from the document database (100) and calculating the third score of each of the words included therein; an integral score calculation unit (50) for subtracting the first score and second score from the third score for each of the words included in the A∩B-document group and calculating an integral score; and a presentation unit (60) for presenting, as a reason for recommendation, one or more important words selected on the basis of the integral score and/or one or more items of text in the A∩B-document group that includes the important words.

(57) 要約:

[続葉有]

- | | | | |
|------------|-------------------------------|-----|-------------------------------------|
| 10 | First score calculation unit | 50 | Integral score calculation unit |
| 11 | A-document group extractor | 60 | Presentation unit |
| 12, 22, 32 | Word correlation evaluator | 61 | Proper word outputter |
| 15 | A-document group | 62 | Proper sentence outputter |
| 20 | Second score calculation unit | 65 | Word base recommendation reason |
| 21 | B-document group extractor | 66 | Sentence base recommendation reason |
| 25 | B-document group | 100 | Document database |
| 30 | Third score calculation unit | 200 | Screen |
| 31 | A∩B-document group extractor | AA | First score |
| 35 | A∩B-document group | BB | Second score |
| 40 | Fourth score calculation unit | CC | Third score |
| 41 | All-documents group extractor | DD | Fourth score |
| 42 | Word importance evaluator | EE | Word |
| 45 | All-documents group | FF | Integral score |



実施形態の情報提示装置は、文書DB（100）からA文書群（15）を抽出してそこに含まれる各単語の第1スコアを算出する第1スコア算出部（10）と、文書DB（100）からB文書群（25）を抽出してそこに含まれる各単語の第2スコアを算出する第2スコア算出部（10）と、文書DB（100）から $A \cap B$ 文書群（35）を抽出してそこに含まれる各単語の第3スコアを算出する第3スコア算出部（30）と、 $A \cap B$ 文書群（35）に含まれる各単語のそれぞれについて、第3スコアから第1スコアと第2スコアとを減算して統合スコアを算出する統合スコア算出部（50）と、統合スコアに基づいて選択された1以上の重要単語または重要単語を含む $A \cap B$ 文書群（35）中の1以上のテキストの少なくとも一方を推薦理由として提示する提示部（60）と、を備える。

明 細 書

発明の名称： 情報提示装置、情報提示方法およびプログラム

技術分野

[0001] 本発明の実施形態は、情報提示装置、情報提示方法およびプログラムに関する。

背景技術

[0002] 電子商取引（ＥＣ：Electronic Commerce）のサービスを提供するＥＣシステムの多くは、ユーザがある商品を参照しているときに、その商品に関連する別の商品を併せて提示する商品推薦機能を持つ。商品推薦機能は、ユーザが参照している商品（以下、「第１商品」という。）と同種の商品を購入の選択肢として提示する対立型推薦と、第１商品と調和する別商品（以下、「第２商品」という。）を紹介して、いわゆる“ついで買い”を促す調和型推薦に大別される。このうち調和型推薦は、多くの場合、統計的観点において第１商品と相関性の高い商品を第２商品として提示する仕組みによって実現されている。

[0003] 調和型推薦においては、第１商品と第２商品との組み合わせ効果をユーザが認識していることが重要である。つまり、第１商品と併せて単に第２商品を提示しただけでは、ユーザがこれらの商品の組み合わせ効果を認識していない場合、第２商品を“ついで買い”する動機付けが生まれない。例えば、秩父のＢ級グルメとして話題になった「味噌ポテト」を知らないユーザにとっては、「味噌」と併せて「ジャガイモ」が提示されただけでは、これらは奇妙な食材の組み合わせに感じられるだけで、「味噌」のついでに「ジャガイモ」を買おうとする購買意欲は生じない。このため、第２商品を提示する際には、第１商品と第２商品との組み合わせ効果に関する情報を含む推薦理由を併せて提示することが、調和型推薦による販売促進の効果を高める上で有効になると考えられる。

[0004] しかし、これまでのＥＣシステムでは、商品単体に関する推薦理由を提示

する仕組み（例えばレビュー表示機能など）は持っていますが、複数の商品の組み合わせ効果に関する情報を含む推薦理由を提示する仕組みは持っていない。このため、このような組み合わせ効果に関する情報を含む推薦理由を提示する仕組みの構築が求められている。

先行技術文献

特許文献

[0005] 特許文献1：特開2006-190127号公報

発明の概要

発明が解決しようとする課題

[0006] 本発明が解決しようとする課題は、第1商品と第2商品との組み合わせ効果に関する情報を含む推薦理由を適切に提示して、調和型推薦による販売促進の効果を高めることができる情報提示装置、情報提示方法およびプログラムを提供することである。

課題を解決するための手段

[0007] 実施形態の情報提示装置は、ユーザが参照している第1商品に調和する第2商品を推薦する際に、前記第1商品と前記第2商品との組み合わせ効果に関する情報を含む推薦理由を提示する情報提示装置であり、第1スコア算出部と、第2スコア算出部と、第3スコア算出部と、統合スコア算出部と、提示部と、を備える。第1スコア算出部は、検索対象の文書群から前記第1商品に関する第1文書群を抽出し、該第1文書群に含まれる各単語のそれぞれについて、前記第1商品との関連性を表す第1スコアを算出する。第2スコア算出部は、検索対象の文書群から前記第2商品に関する第2文書群を抽出し、該第2文書群に含まれる各単語のそれぞれについて、前記第2商品との関連性を表す第2スコアを算出する。第3スコア算出部は、検索対象の文書群から前記第1商品と前記第2商品との双方に関する第3文書群を抽出し、該第3文書群に含まれる各単語のそれぞれについて、前記第1商品および前記第2商品の双方との関連性を表す第3スコアを算出する。統合スコア算出

部は、前記第3文書群に含まれる各単語のそれぞれについて、前記第3スコアから前記第1スコアと前記第2スコアとを減算し、統合スコアを算出する。提示部は、前記統合スコアに基づいて所定の基準に従って選択された1以上の重要単語、または、該重要単語を含む前記第3文書群中の1以上のテキストの少なくとも一方を、前記推薦理由として提示する。

図面の簡単な説明

- [0008] [図1]図1は、第1実施形態の情報提示装置の構成例を示す図である。
- [図2]図2は、A文書群抽出器の処理手順を示すフローチャートである。
- [図3]図3は、単語の表現正規化に用いる同意語辞書の一例を示す図である。
- [図4]図4は、全文書群抽出器の処理手順を示すフローチャートである。
- [図5]図5は、単語関連度評価器の処理手順を示すフローチャートである。
- [図6]図6は、単語重要度評価器の処理手順を示すフローチャートである。
- [図7]図7は、統合スコア算出部の処理手順を示すフローチャートである。
- [図8]図8は、固有文出力器の処理手順を示すフローチャートである。
- [図9]図9は、第2実施形態の情報提示装置の構成例を示す図である。
- [図10]図10は、A文書群抽出器の処理手順を示すフローチャートである。
- [図11]図11は、 $A \cap B$ 文書群抽出器の処理手順を示すフローチャートである。
- [図12]図12は、 $A \cap B$ 文書群抽出器の判定例を説明する図である。
- [図13]図13は、単語関連度評価器の処理手順を示すフローチャートである。
- [図14]図14は、情報提示装置のハードウェア構成例を示すブロック図である。

発明を実施するための形態

- [0009] 以下、実施形態の情報提示装置、情報提示方法およびプログラムを、図面を参照して詳細に説明する。
- [0010] 実施形態の情報提示装置は、ユーザが参照している第1商品に調和する第2商品を推薦する際に、第1商品と第2商品との組み合わせ効果に関する情

報を含む推薦理由を提示する。このような推薦理由をあらゆる商品の組み合わせについて予め人手で作成することは困難である。しかし、商品の組み合わせ効果に関する情報は、例えば、各種のWebページ、SNS (Social Networking Service)、ブログなどの文書群に存在する。そこで、本実施形態では、このような文書群から両商品に関する文書群を見つけ出し、さらに商品の組み合わせ効果などの推薦理由としてふさわしい言及箇所を特定してユーザに提示する。なお、以下では説明の簡便化のため、第1商品を商品A、第1商品に関する記載のある文書をA文書、第2商品を商品B、第2商品に関する記載のある文書をB文書、第1商品と第2商品との双方に関する記載のある文書をA∩B文書と呼ぶ。

[0011] <第1実施形態>

まず、第1実施形態の情報提示装置について説明する。図1は、第1実施形態の情報提示装置の構成例を示す図である。本実施形態の情報提示装置は、図1に示すように、第1スコア算出部10と、第2スコア算出部20と、第3スコア算出部30と、第4スコア算出部40と、統合スコア算出部50と、提示部60とを備え、文書DB (Database) 100から取得した第1商品と第2商品との組み合わせ効果に関する情報を含む推薦理由を画面200に表示して、ECシステムのサービスを利用しているユーザに提示する。なお、本実施形態の情報処理装置は、ECシステムの機能の一部として実現されることを想定するが、これに限らず、例えばECシステムと連動して動作する独立のシステムあるいは装置として構成されてもよい。

[0012] 文書DB 100は、本実施形態において検索対象とする任意の文書群であり、各種のWebページ、SNS、ブログなどを想定している。また、画面200は、ECシステムのサービスを利用しているユーザの端末装置に表示される画面を想定しており、一般的にはWebブラウザを搭載する端末装置に表示されるWeb画面である。

[0013] 第1スコア算出部10は、A文書群抽出器11と、単語関連度評価器12とを含む。

- [0014] A文書群抽出器11は、文書DB100に対して単語ベースの検索を行い、文書DB100から商品Aに関する記載のあるA文書をすべて抽出してA文書群15を得る。
- [0015] 単語関連度評価器12は、A文書群15における各単語のヒストグラム（単語ごとの頻度を列挙したデータ）を作成し、各単語のそれぞれについて、A文書群15中の出現頻度に応じた第1スコアを算出する。ただし、各単語は辞書を用いて全半角・日英・送り仮名などの表記揺れを吸収する。また、各単語の出現頻度は総単語数で割ることで数値を正規化し、logスケールにしたものを第1スコアとする。したがって、第1スコアは負の値であり、A文書群15中の出現頻度が高い単語ほど、0に近い高い値の第1スコアが与えられる。
- [0016] 第2スコア算出部20は、B文書群抽出器21と、単語関連度評価器22とを含む。
- [0017] B文書群抽出器21は、文書DB100に対して単語ベースの検索を行い、文書DB100から商品Bについての記載のあるB文書をすべて抽出してB文書群25を得る。
- [0018] 単語関連度評価器22は、B文書群25における各単語のヒストグラムを作成し、各単語のそれぞれについて、B文書群25中の出現頻度に応じた第2スコアを算出する。ただし、各単語は辞書を用いて全半角・日英・送り仮名などの表記揺れを吸収する。また、各単語の出現頻度は総単語数で割ることで数値を正規化し、logスケールにしたものを第2スコアとする。したがって、第2スコアは負の値であり、B文書群25中の出現頻度が高い単語ほど、0に近い高い値の第2スコアが与えられる。
- [0019] 第3スコア算出部30は、 $A \cap B$ 文書群抽出器31と、単語関連度評価器32とを含む。
- [0020] $A \cap B$ 文書群抽出器31は、文書DB100に対して単語ベースの検索を行い、文書DB100から商品Aと商品Bとの双方についての記載のある $A \cap B$ 文書をすべて抽出して $A \cap B$ 文書群35を得る。

- [0021] 単語関連度評価器 32 は、 $A \cap B$ 文書群 35 における各単語のヒストグラムを作成し、各単語のそれぞれについて、 $A \cap B$ 文書群 35 中の出現頻度に応じた第 3 スコアを算出する。ただし、各単語は辞書を用いて全半角・日英・送り仮名などの表記揺れを吸収する。また、各単語の出現頻度は総単語数で割ることで数値を正規化し、 $\log$ スケールにしたものを第 3 スコアとする。したがって、第 3 スコアは負の値であり、 $A \cap B$ 文書群中の出現頻度が高い単語ほど、0 に近い高い値の第 3 スコアが与えられる。
- [0022] 第 4 スコア算出部 40 は、全文書群抽出器 41 と、単語重要度評価器 42 とを含む。
- [0023] 全文書群抽出器 41 は、文書 DB 100 から文書をすべて抽出して全文書群 45 を得る。
- [0024] 単語重要度評価器 42 は、全文書群 45 における各単語が含まれる文書数のヒストグラムを作成し、各単語のそれぞれについて、全文書群 45 中における当該単語を含む文書の出現頻度に応じた第 4 スコアを算出する。ただし、各単語は辞書を用いて全半角・日英・送り仮名などの表記揺れを吸収する。また、各単語を含む文書の出現頻度は総文書数で割ることで数値を正規化し、 $\log$ スケールにしてさらに正負を反転させたものを第 4 スコアとする。したがって、第 4 スコアは正の値であり、当該単語を含む文書の出現頻度が低いほど高い値の第 4 スコアが与えられる。
- [0025] 統合スコア算出部 50 は、 $A \cap B$ 文書群 35 に含まれる各単語のそれぞれについて、第 3 スコアと、第 1 スコアと、第 2 スコアと、第 4 スコアとを用いて、後述の式 (1) を用いた計算によって統合スコアを算出する。統合スコアは、商品 A と商品 B との双方に関する話題に対する固有性を表す指標であり、商品 A と商品 B との双方に関する話題に対する固有性が高い単語ほど、高い値の統合スコアが与えられる。
- [0026] 提示部 60 は、固有語出力器 61 と、固有文出力器 62 とを含む。
- [0027] 固有語出力器 61 は、統合スコアに基づいて、商品 A と商品 B との双方に関する話題に対する固有性が高い 1 以上の重要単語（固有語）を選択し、単

語ベースの推薦理由65として画面200に出力する。推薦理由が単語のみでよい場合は、この固有語出力器61が出力する単語ベースの推薦理由65が画面200に表示される。

[0028] 固有文出力器62は、 $A \cap B$ 文書群35から、固有語出力器61により選択された重要単語（固有語）を多く持つ1以上の文を選択し、文ベースの推薦理由66として画面200に出力する。推薦理由を文とすることが要求される場合は、この固有文出力器62が出力する文ベースの推薦理由66が画面200に表示される。また、固有語出力器61が出力する単語ベースの推薦理由65と、固有文出力器62が出力する文ベースの推薦理由66との双方を画面200に表示させるようにしてもよい。

[0029] なお、本実施形態では、固有文出力器62の処理単位を文としているが、固有文出力器62は、文ではなく、フレーズ、パッセージ、パラグラフなどを処理単位としてもよい。この場合も、固有文出力器62の処理単位が変わるだけで、同様の処理により所望のテキストを推薦理由として画面200に表示させることができる。

[0030] 次に、本実施形態の情報提示装置を構成する上述した各部による処理手順の詳細について説明する。

[0031] まず、A文書群抽出器11の処理手順を説明する。A文書群抽出器11の処理は、文書DB100からすべてのA文書を見つけ出すことが目的である。A文書の抽出は、例えば、従来的な方法を用いた単語ベースの検索により行うことができる。一般的な検索処理では、検索対象の文書群のインデックスをあらかじめ作成しておく処理方式を用いるのが一般的である。しかし、本実施形態では説明を簡単にするため、インデックスを作成しないで検索するgrep方式を用いるものとする。

[0032] 図2は、A文書群抽出器11の処理手順を示すフローチャートである。A文書群抽出器11は、まず、A商品に関するメタデータから商品名を取り出して、これを検索のクエリとする（ステップS101）。

[0033] 次に、A文書群抽出器11は、クエリの表現正規化を行う（ステップS1

02)。具体的には、A文書群抽出器11は、まずクエリの表記ゆれ（半全角、日英、送り仮名など）を吸収し、さらに図3に示すような同意語辞書を用いて、クエリ（ここでは商品Aの商品名）を代表的表現に置き換える。例えば、クエリ「スマホ」は「スマートフォン」に置換され、クエリ「パソコン」は「PC」に置換される。

[0034] 次に、A文書群抽出器11は、文書DB100から文書を1つ取り出す（ステップS103）。そして、A文書抽出器11は、ステップS103で取り出した文書に含まれる各単語に対して、ステップS102と同様の手法により表現正規化を行う（ステップS104）。

[0035] 次に、A文書群抽出器11は、ステップS104で単語の表現正規化が行われた文書内に、ステップS102で表現正規化が行われたクエリ（つまり商品Aの商品名）が含まれているか確認し、表現正規化されたクエリが含まれていれば、出力するA文書群15に当該文書を追加する（ステップS105）。

[0036] 次に、A文書群抽出器11は、文書DB100から取り出していない文書があるか否かを判定し（ステップS106）、文書DB100から取り出していない文書があれば（ステップS106：Yes）、ステップS103に戻って以降の処理を繰り返す。一方、文書DB100のすべての文書に対してステップS103～ステップS105の処理を行っていれば（ステップS106：No）、A文書群抽出器11は、A文書群15を出力し（ステップS107）、一連の処理を終了する。

[0037] B文書群抽出器21の処理は、文書DB100からすべてのB文書を見つけ出すことが目的である。B文書の抽出は、A文書の抽出と同様に単語ベースの検索により行う。B文書群抽出器21の処理は、検索に用いるクエリが商品Bの商品名に置き換わり、出力する文書群がB文書群25となるだけで、上述したA文書群抽出器11の処理と同様であるため、詳細な説明は省略する。

[0038] $A \cap B$ 文書群抽出器31の処理は、文書DB100からすべての $A \cap B$ 文

書を見つけ出すことが目的である。A $\cap$ B文書の抽出は、A文書やB文書の抽出と同様に単語ベースの検索により行う。A $\cap$ B文書群抽出器31の処理は、検索に用いるクエリが商品Aの商品名と商品Bの商品名のアンド条件となり、出力する文書群がA $\cap$ B文書群35となるだけで、上述したA文書群抽出器11やB文書群抽出器21の処理と同様であるため、詳細な説明は省略する。

[0039] 全文書群抽出器41の処理は、文書DB100から全文書を取り出して、後続処理のためにそれぞれの文書に含まれる各単語の表現正規化を行うことが目的である。

[0040] 図4は、全文書群抽出器41の処理手順を示すフローチャートである。全文書群抽出器41は、まず、文書DB100から文書を1つ取り出す（ステップS201）。そして、全文書抽出器41は、ステップS201で取り出した文書に含まれる各単語に対して、図2のステップS102と同様の手法により表現正規化を行い（ステップS202）、出力する全文書群45に当該文書を追加する（ステップS203）。

[0041] 次に、全文書群抽出器41は、文書DB100から取り出していない文書があるか否かを判定し（ステップS204）、文書DB100から取り出していない文書があれば（ステップS204：Yes）、ステップS201に戻って以降の処理を繰り返す。一方、文書DB100のすべての文書に対してステップS201～ステップS203の処理を行っていれば（ステップS204：No）、全文書群抽出器41は、全文書群45を出力し（ステップS205）、一連の処理を終了する。

[0042] 次に、単語関連度評価器12の処理手順を説明する。単語関連度評価器12の処理は、A文書群15に含まれる各単語について、商品Aとの関連性を表す第1スコアを算出することを目的とする。本実施形態では、A文書群15における各単語の出現数を総単語数で割りlogスケールに変換することで各単語の対数確率を求め、これを第1スコアとする。これは、単位テキスト量あたりの各単語の頻度を計測していることとなり、情報検索においてよ

く用いられる指標である $t f$ (term frequency) を正規化したものと同等と言える。

[0043] 図5は、単語関連度評価器12の処理手順を示すフローチャートである。単語関連度評価器12は、まず、各単語の出現数を集計するための集計用ヒストグラムを初期化する（ステップS301）。

[0044] 次に、単語関連度評価器12は、A文書群15から文書を1つ取り出す（ステップS302）。そして、単語関連度評価器12は、ステップS302で取り出した文書内に含まれる単語のヒストグラムを作成し（ステップS303）、得られたヒストグラムを集計用ヒストグラムに加算する（ステップS304）。

[0045] 次に、単語関連度評価器12は、A文書群15から取り出していない文書があるか否かを判定し（ステップS305）、A文書群15から取り出していない文書があれば（ステップS305：Yes）、ステップS302に戻って以降の処理を繰り返す。一方、A文書群15のすべての文書に対してステップS302～ステップS304の処理を行っていれば（ステップS305：No）、単語関連度評価器12は、集計用ヒストグラムから各単語の対数確率を割り出す（ステップS306）。具体的には、集計用ヒストグラムが示す各単語の頻度を x 、A文書群15中の総単語数を y とすると、対数確率は $\log(x/y)$ である。そして、単語関連度評価器12は、ステップS306で算出した各単語の対数確率を、各単語の第1スコアとして出力し（ステップS307）、一連の処理を終了する。なお、 $x=0$ の場合、対数確率は $-\infty$ となる。 ∞ や $-\infty$ は計算機では直接扱えないため、極端に大きな値や小さな値で代用する方法が考えられる。以降においても ∞ や $-\infty$ を扱う場合は同様の方法を用いればよい。

[0046] 単語関連度評価器22の処理は、B文書群25に含まれる各単語について、商品Bとの関連性を表す第2スコアを算出することを目的とする。第2スコアは、第1スコアと同様に、B文書群25に含まれる各単語の対数確率である。単語関連度評価器22の処理は、与えられる文書セットがB文書群2

5に置き換わり、B文書群25に含まれる各単語の対数確率を第2スコアとして出力するだけで、上述した単語関連度評価器12の処理と同様であるため、詳細な説明は省略する。

[0047] 単語関連度評価器32の処理は、 $A \cap B$ 文書群35に含まれる各単語について、商品Aおよび商品Bの双方との関連性を表す第3スコアを算出することを目的とする。第3スコアは、第1スコアや第2スコアと同様に、 $A \cap B$ 文書群35に含まれる各単語の対数確率である。単語関連度評価器32の処理は、与えられる文書セットが $A \cap B$ 文書群35に置き換わり、 $A \cap B$ 文書群35に含まれる各単語の対数確率を第3スコアとして出力するだけで、上述した単語関連度評価器12の処理と同様であるため、詳細な説明は省略する。

[0048] 次に、単語重要度評価器42の処理手順を説明する。単語重要度評価器42の処理は、文書DB100内の各単語が持つ一般的な重要性を表す第4スコアを算出することを目的とする。本実施形態では、単語の重要性の指標として情報検索などでよく用いられるidf (Inverse Document Frequency) を求めて、これを各単語の第4スコアとする。ある単語のidfは、当該単語を含む文書の負の対数確率である。つまり、当該単語を含む文書数を x 、全文書数を y とすると、 $idf = -\log(x/y)$ である。一般的に、めったに出現しない単語（すなわち出現確率の低い単語）は、出現した際に読者に与える情報量が多く重要であると考えられるが、この場合、idfは高い値を示す。

[0049] 図6は、単語重要度評価器42の処理手順を示すフローチャートである。単語重要度評価器42は、まず、各単語の出現数を集計するための集計用ヒストグラムを初期化する（ステップS401）。

[0050] 次に、単語重要度評価器42は、全文書群45から文書を1つ取り出す（ステップS402）。そして、単語重要度評価器42は、ステップS402で取り出した文書内に含まれる単語の2値ヒストグラムを作成し（ステップS403）、得られたヒストグラムを集計用ヒストグラムに加算する（ステ

ップS 4 0 4)。2値ヒストグラムは、1か0の頻度値しか持たないヒストグラムであり、文書内に出現する単語に対して、出現数にかかわらず1が与えられる。

[0051] 次に、単語重要度評価器42は、全文書群45から取り出していない文書があるか否かを判定し（ステップS 4 0 5）、全文書群45から取り出していない文書があれば（ステップS 4 0 5：Y e s）、ステップS 4 0 2に戻って以降の処理を繰り返す。一方、全文書群45のすべての文書に対してステップS 4 0 2～ステップS 4 0 4の処理を行っていれば（ステップS 4 0 5：N o）、単語重要度評価器42は、集計用ヒストグラムから各単語を含む文書の負の対数確率を割り出す（ステップS 4 0 6）。具体的には、集計用ヒストグラムが示す各単語の頻度 x 、全文書群45の総文書数を y とすると、負の対数確率は $-\log(x/y)$ である。そして、単語重要度評価器42は、各単語のそれぞれについて、ステップS 4 0 6で算出した当該単語を含む文書の負の対数確率を、各単語の第4スコアとして出力し（ステップS 4 0 7）、一連の処理を終了する。

[0052] 次に、統合スコア算出部50の処理手順を説明する。統合スコア算出部50の処理は、 $A \cap B$ 文書群35内の各単語について、商品Aと商品Bとの双方に関する話題に対する固有性（つまり、 $A \cap B$ 文書群35にのみ顕著に出現する単語であるかどうかの度合い）を表す指標となる統合スコアを算出することを目的とする。これにより、商品Aと商品Bとの組み合わせに関する説明にふさわしい単語を見つけることができるようになる。

[0053] 本実施形態では、統合スコアの計算に下記式（1）を用いるものとする。ただし、下記式（1）の w は単語、 $ntf(w)$ は与えられた文書セットにおける単語 w の対数確率、 idf は全文書群45における単語 w を含む文書の負の対数確率である。

[数1]

$$(ntf_{A \cap B}(w) \cdot 2 - ntf_A(w) - ntf_B(w)) * idf(w) \quad \dots (1)$$

- [0054] 式(1)の第1項は、 $A \cap B$ 文書群35における単語 w の対数確率を示しており、単語関連度評価器32が出力する第3スコアに相当する。この第1項の値(第3スコア)が高いほど、当該単語 w が、 $A \cap B$ 文書群35において多く出現していることを示している。
- [0055] 式(1)の第2項は、A文書群15における単語 w の対数確率を示しており、単語関連度評価器12が出力する第1スコアに相当する。この第2項の値(第1スコア)が高いほど、当該単語 w がA文書群15において多く出現していることを示している。
- [0056] 式(1)の第3項は、B文書群25における単語 w の対数確率を示しており、単語関連度評価器22が出力する第2スコアに相当する。この第3項の値(第2スコア)が高いほど、当該単語 w がB文書群25において多く出現していることを示している。
- [0057] 式(1)の第4項は、全文書群45における単語 w の希少性を示しており、単語重要度評価器42が出力する第4スコアに相当する。この第4項の値(第4スコア)が高いほど、当該単語 w は希少性があり、出現したときの情報量が多く重要な単語であることを示している。
- [0058] 式(1)は、第1項から第2項および第3項を減算して統合スコアを求める計算式となっている。これにより、 $A \cap B$ 文書群35において多く出現し、かつ、A文書群15やB文書群25ではあまり出現していない単語に対し、高い値の統合スコアが与えられることとなる。このことから、統合スコアは、商品Aや商品Bの個別説明ではなく、両商品にまたがる説明にふさわしい度合いを示していると考えられる。なお、第1項を2倍しているのは、第1項から減算している項が2つあるためである。 $A \cap B$ 文書群35、A文書群15、B文書群25のそれぞれで同頻度で出現する単語の固有性は0と考えられるが、式(1)のように第1項を2倍しておくことで、この場合の統合スコアを0とすることができる。ただし、第1項を2倍することは必須ではなく、第1項を2倍せずに第2項および第3項を減算してもよい。
- [0059] また、式(1)は、第1項から第2項および第3項を減算した値に、さら

に第4項を乗算して統合スコアを求める計算式となっている。これにより、各単語の一般的な観点での重要性を加味した統合スコアが得られる。つまり、A文書群15の文書数、B文書群25の文書数、および $A \cap B$ 文書群35の文書数が十分でない場合、第4項を乗算せずに各単語の統合スコアを算出すると統合スコアが過適応してしまうリスクがあるが、第4項を乗算することで、このリスクを回避できる。ただし、第4項の乗算は必須ではなく、第4項を乗算せずに統合スコアを算出してもよい。

- [0060] 図7は、統合スコア算出部50の処理手順を示すフローチャートである。統合スコア算出部50は、まず、 $A \cap B$ 文書群35から単語を1つ取り出す（ステップS501）。
- [0061] 次に、統合スコア算出部50は、ステップS501で取り出した単語について、単語関連度評価器32が出力した第3スコアの値を、式（1）の第1項に当てはめる（ステップS502）。
- [0062] 次に、統合スコア算出部50は、ステップS501で取り出した単語について、単語関連度評価器12が出力した第1スコアの値を、式（1）の第2項に当てはめる（ステップS503）。
- [0063] 次に、統合スコア算出部50は、ステップS501で取り出した単語について、単語関連度評価器22が出力した第2スコアの値を、式（1）の第3項に当てはめる（ステップS504）。
- [0064] 次に、統合スコア算出部50は、ステップS501で取り出した単語について、単語重要度評価器42が出力した第4スコアの値を、式（1）の第4項に当てはめる（ステップS505）。
- [0065] 次に、統合スコア算出部50は、式（1）を用いて、ステップS501で取り出した単語の統合スコアを算出する（ステップS506）。
- [0066] 次に、統合スコア算出部50は、 $A \cap B$ 文書群35から取り出していない単語があるか否かを判定し（ステップS507）、 $A \cap B$ 文書群35から取り出していない単語があれば（ステップS507：Yes）、ステップS501に戻って以降の処理を繰り返す。一方、 $A \cap B$ 文書群35に含まれるす

すべての単語に対してステップS501～ステップS506の処理を行って
いれば（ステップS507：No）、統合スコア算出部50は、各単語の統合
スコアを出力し（ステップS508）、一連の処理を終了する。

[0067] 次に、固有語出力器61の処理手順を説明する。固有語出力器61の処理
は、 $A \cap B$ 文書群35に含まれる単語のうち、商品Aと商品Bとの双方に関
する話題に対する固有性の高い単語（固有語）を重要単語として選択して出
力することを目的とする。本実施形態では、 $A \cap B$ 文書群35に含まれる単
語のうち、統合スコアが高い上位k個の単語を重要単語として出力するもの
とする。

[0068] すなわち、固有語出力器61は、統合スコア算出部50から出力された統
合スコアを値が高い順にソートし、統合スコアの値が高い順に上記k個の単
語を重要単語として選択して出力する。B商品の推薦理由が単語のみでよい
場合は、この固有語出力器61が出力する重要単語が、単語ベースの推薦理
由65として画面200に表示される。また、推薦理由を文とすることが要
求される場合は、固有語出力器61が出力する重要単語が、固有文出力器6
2に渡される。

[0069] 次に、固有文出力器62の処理手順を説明する。固有文出力器62の処理
は、 $A \cap B$ 文書群35から重要単語を多く含む文を見つけ出し、文ベースの
推薦理由66として画面200に出力することを目的とする。本実施形態で
は、重要単語を最も多く含む $A \cap B$ 文書群35中の文をベスト文として見つ
け出し、文ベースの推薦理由66として画面200に出力するものとする。
なお、上述したように、文の代わりにフレーズ、パッセージ、パラグラフな
どを推薦理由として画面200に表示させるようにしてもよい。

[0070] 図8は、固有文出力器62の処理手順を示すフローチャートである。固有
文出力器62は、まず、ベスト文およびベストスコアを初期化する（ステッ
プS601）。つまり、文ベースの推薦理由66として最終的に出力するベ
スト文を空文とし、そのベスト文に含まれる各単語の統合スコアの合計値で
あるベストスコアを $-\infty$ にする。

- [0071] 次に、固有文出力器62は、 $A \cap B$ 文書群35から文を1つ取り出す（ステップS602）。そして、固有文出力器62は、ステップS602で取り出した文に含まれる各単語の統合スコアを合計したものを当該文のスコアとする（ステップS603）。
- [0072] 次に、固有文出力器62は、ステップS603で求めた文のスコアがベストスコアを上回っているか確認し、ベストスコアを上回っていれば、ベスト文およびベストスコアを、当該文とそのスコアで置き換える（ステップS604）。
- [0073] 次に、固有文出力器62は、 $A \cap B$ 文書群35から取り出していない文があるか否かを判定し（ステップS605）、 $A \cap B$ 文書群35から取り出していない文があれば（ステップS605：Yes）、ステップS602に戻って以降の処理を繰り返す。一方、 $A \cap B$ 文書群35に含まれるすべての文に対してステップS602～ステップS604の処理を行っていれば（ステップS605：No）、固有文出力器62は、ベスト文を文ベースの推薦理由66として出力し（ステップS606）、一連の処理を終了する。
- [0074] 以上、具体的な例を挙げながら説明したように、本実施形態の情報提示装置によれば、商品Aと商品Bとの双方に関する話題に対する固有が高い単語、あるいはその単語を含む文を特定して単語ベースの推薦理由65、あるいは文ベースの推薦理由66として画面200に表示させる。したがって、この情報提示装置を用いることによって、ECシステムを利用するユーザに対して、商品Aと商品Bとの組み合わせ効果に関する情報を含む推薦理由を適切に提示して、調和型推薦による販売促進の効果を高めることができる。すなわち、ECシステムを利用するユーザにとっては、本実施形態の情報提示装置により提示される推薦理由を参照することでB商品を購入する動機付けが生まれ、新体験を伴う商品購入がしやすくなり、店舗にとっては販売機会を増やすことができる。

[0075] ＜第2実施形態＞

次に、第2実施形態の情報提示装置について説明する。本実施形態では、

ECシステムを利用したユーザによるレビュー記事など、ある商品について記載されていることが事前に予測される文書を、検索対象の文書群として用いる。ECシステムは、商品ページごとにユーザによるレビュー記事を管理していることが多い。このようなレビュー記事は、それぞれの商品に対する感想などを記載した文書であるため、推薦理由を見つけ出す対象として有効に利用できる。ただし、各レビュー記事は、レビュー対象の商品ID（商品識別情報）およびレビュー記事を記載したユーザの購入ログがメタデータとして紐付けられているとする。以下、商品IDおよび購入ログと紐付けられたレビュー記事をラベル付き文書と呼ぶ。

[0076] 第1実施形態では、一般的な文書を検索対象としていたため、A文書、B文書、 $A \cap B$ 文書を検索する手がかりとして、文書内に商品名が含まれているかどうかを用いていた。これに対し本実施形態では、検索対象とする各文書に付与されたレビュー対象の商品ID（レビュー記事に商品名が紐付けられている場合は商品名でもよい）を用いて検索する方法を取る。このため、文書検索エラーを排除できる（第1実施形態では表現揺れなどによるエラーのリスクがある）ほか、単に「おいしかった！また買います」などのように商品名が含まれていない文書であっても、メタデータを用いることで簡単に仕分けを行うことができるメリットがある。ただし、文書に紐付けられている商品IDは1つであるため、 $A \cap B$ 文書を判定するには工夫が必要である。そこで、本実施形態では、近いタイミングで商品Aと商品Bの両商品を購入したユーザがこれらの商品の購入から近いタイミングで記載したレビュー記事は、両商品への言及を含むレビュー記事である可能性が高いという仮説に基づいて、 $A \cap B$ 文書を特定するようにしている。

[0077] 図9は、第2実施形態の情報提示装置の構成例を示す図である。第2実施形態の情報提示装置は、図9に示すように、第1実施形態の第1スコア算出部10、第2スコア算出部20および第3スコア算出部30（図1参照）に代えて、第1スコア算出部70、第2スコア算出部80および第3スコア算出部90を備えている。また、第2実施形態の情報提示装置は、検索対象の

文書集合として、第1実施形態の文書DB100（図1参照）に代えて、ラベル付き文書DB300を用いる。ラベル付き文書DB300は、上述したように、例えばECシステムを利用したユーザによるレビュー記事の集合であり、各レビュー記事は商品IDおよび購入ログ400と紐付けられている。なお、第2実施形態の情報提示装置におけるその他の構成は、上述した第1実施形態と同様であるため、以下、第1実施形態と共通の構成要素については同一の符号を付して、重複した説明を適宜省略する。

[0078] 第1スコア算出部70は、A文書群抽出器71と、単語関連度評価器12とを含む。A文書群抽出器71は、商品Aの商品IDを用いてラベル付き文書DB300に対する検索を行い、ラベル付き文書DB300からA文書をすべて抽出してA文書群15を得る。単語関連度評価器12は、第1実施形態と共通である。

[0079] 第2スコア算出部80は、B文書群抽出器81と、単語関連度評価器22とを含む。B文書群抽出器81は、商品Bの商品IDを用いてラベル付き文書DB300に対する検索を行い、ラベル付き文書DB300からB文書をすべて抽出してB文書群25を得る。単語関連度評価器22は、第1実施形態と共通である。

[0080] 第3スコア算出部90は、 $A \cap B$ 文書群抽出器91と、単語関連度評価器92とを含む。

[0081] $A \cap B$ 文書群抽出器91は、商品Aの商品IDおよび商品Bの商品IDを用いてラベル付き文書DB300に対する検索を行い、ラベル付き文書DB300から $A \cap B$ 文書を抽出して確信度付き $A \cap B$ 文書群95を得る。ここでラベル付き文書DB300から抽出される $A \cap B$ 文書は、上述した仮説に基づいて抽出されるレビュー記事などのラベル付き文書であり、その文書に商品Aと商品Bとの双方に関する記述が含まれていることの確信度が与えられたものである。

[0082] 単語関連度評価器92は、確信度付き $A \cap B$ 文書群95に含まれる各単語のそれぞれについて、第1実施形態の関連度評価器32と同様に、出現頻度

に応じた第3スコアを算出する。ただし、本実施形態では、 $A \cap B$ 文書のそれぞれに商品Aと商品Bとの双方に関する記述が含まれていることの確信度が与えられており、各単語の頻度が、その単語が出現する文書の確信度を用いて計算される点が第1実施形態とは異なる。

[0083] 次に、本実施形態の情報提示装置において、第1実施形態とは異なる部分の処理手順の詳細について説明する。

[0084] まず、A文書群抽出器71の処理手順を説明する。A文書群抽出器71の処理は、ラベル付き文書DB300からすべてのA文書を見つけ出すことが目的である。

[0085] 図10は、A文書群抽出器71の処理手順を示すフローチャートである。A文書群抽出器71は、まず、A商品に関するメタデータからA商品の商品IDを取り出して、これを検索のクエリとする（ステップS701）。

[0086] 次に、A文書群抽出器71は、ラベル付き文書DB300から文書を1つ取り出す（ステップS702）。そして、A文書抽出器71は、ステップS701で取り出した文書のラベルがクエリの商品IDと一致するか確認し、一致していれば、出力するA文書群15に当該文書を追加する（ステップS703）。

[0087] 次に、A文書群抽出器71は、ラベル付き文書DB300から取り出していない文書があるか否かを判定し（ステップS704）、ラベル付き文書DB300から取り出していない文書があれば（ステップS704：Yes）、ステップS702に戻って以降の処理を繰り返す。一方、ラベル付き文書DB300のすべての文書に対してステップS702およびステップS703の処理を行っていれば（ステップS704：No）、A文書群抽出器71は、A文書群15を出力し（ステップS705）、一連の処理を終了する。

[0088] B文書群抽出器81の処理は、ラベル付き文書DB300からすべてのB文書を見つけ出すことが目的である。B文書群抽出器81の処理は、検索に用いるクエリが商品Bの商品IDに置き換わり、出力する文書群がB文書群25となるだけで、上述したA文書群抽出器71の処理と同様であるため、

詳細な説明は省略する。

[0089] 次に、 $A \cap B$ 文書群抽出器 91 の処理手順を説明する。 $A \cap B$ 文書群抽出器 91 の処理は、ラベル付き文書 DB 300 から $A \cap B$ 文書を見つけ出すことが目的である。ラベル付き文書 DB 300 内の各ラベル付き文書は 1 つの商品 ID にしか結びついていないため、そのラベル付き文書が商品 A と商品 B との双方に関する記述を含んでいるかどうかをメタデータだけから判定することはできない。ここで視点を変えて、商品 A と商品 B とを同時あるいは近いタイミングで購入したユーザは、両商品の組合せに意図を持っており、そのようなユーザがそれに近いタイミングで記載したレビュー文書には、両商品の組合せに関する記述が含まれている可能性が高いと考えられる。そこで、本実施形態では、購入ログ 400 を用いてこの仮説に適合するユーザを選び、このユーザが記載したレビュー記事から、この仮説に適合するレビュー記事を、 $A \cap B$ 文書として抽出する。さらに、このように抽出された $A \cap B$ 文書群に対し、商品 A と商品 B との双方に関する記述が含まれていることの確信度を与えて、確信度付き $A \cap B$ 文書群 95 を得る。

[0090] 図 11 は、 $A \cap B$ 文書群抽出器 91 の処理手順を示すフローチャートである。 $A \cap B$ 文書群抽出器 91 は、まず、購入ログ 400 から 1 人のユーザを選択する（ステップ S801）。

[0091] 次に、 $A \cap B$ 文書群抽出器 91 は、ステップ S801 で選択したユーザが、所定の第 1 期間内に商品 A と商品 B を購入していることを示す購入ログのペアをすべて抜き出す（ステップ S802）。このときの判定例を図 12（a）に示す。上記の第 1 期間を 2 日とすると、図 12（a）の判定例 1 のように、ユーザ X の購入ログのうち、「11/7 15:20 商品 A 購入」と「11/7 18:20 商品 B 購入」のペアは、両商品を購入した時間差が 2 日以内のため、ステップ S802 の処理で抜き出される。一方、「11/7 18:20 商品 B 購入」と「11/10 9:50 商品 A 購入」のペアは、両商品を購入した時間差が 2 日を超えるため、ステップ S802 の処理では抜き出されない。この購入ログのペアの購入時刻の時間差を以

下では「購入時間差」と呼ぶ。

[0092] 次に、 $A \cap B$ 文書群抽出器91は、ステップS802で抜き出した購入ログのペアを1つ取り出す（ステップS803）。そして、 $A \cap B$ 文書群抽出器91は、ラベル付き文書DB300から、ステップS801で選択したユーザによって、ステップS803で取り出した購入ログのペアが示す購入時刻のうちの遅い方の購入時刻から所定の第2期間内に記載された、商品Aまたは商品Bの商品IDをラベルとして持つ文書（レビュー記事）をすべて取り出す（ステップS804）。

[0093] このときの判定例を図12（b）に示す。上記の第2期間を3日とすると、図12（b）の判定例2のように、ユーザXが記載したレビュー記事のうち、「11/9 12:00 商品Aレビュー記事」は、「11/7 18:20 商品B購入」の購入ログの購入時刻から3日以内に記載されたレビュー記事であるため、ステップS804の処理で取り出される。一方、「11/11 12:00 商品Aレビュー記事」は、「11/7 18:20 商品B購入」の購入ログの購入時刻から3日経過した後に記載されたレビュー記事であるため、ステップS804の処理では取り出されない。この購入ログの購入時刻とレビュー記載時刻の時間差を以下では「レビュー時間差」と呼ぶ。

[0094] 次に、 $A \cap B$ 文書群抽出器91は、ステップS803で取り出した購入ログのペアの購入時間差に応じた確信度を、ステップS804で取り出した各文書に対して割り当てる（ステップS805）。例えば、購入ログのペアが同じセッションでの購入の場合の確信度を100%、1時間以内の購入の場合の確信度を90%、2時間以内の購入の場合の確信度を80%、同日購入の場合の確信度を50%といったように、購入時間差が大きいほど低い値となる確信度を与える。なお、本実施形態では、ラベル付き文書DB300から取り出した文書に対し、その文書を取り出す要因となった購入ログのペアの購入時間差に応じた確信度を与えるようにしているが、確信度を与える方法はこれに限らない。例えば、ラベル付き文書DB300から取り出した文

書に対して、レビュー時間差が大きくなるほど低い値となる確信度を与えるようにしてもよいし、購入時間差とレビュー時間差との双方を考慮した確信度を与えるようにしてもよい。

[0095] 次に、 $A \cap B$ 文書群抽出器 91 は、ステップ S 805 の処理により得られた確信度付き文書を、出力する確信度付き $A \cap B$ 文書群 95 に追加する（ステップ S 806）。

[0096] 次に、 $A \cap B$ 文書群抽出器 91 は、ステップ S 803 で取り出していない購入ログのペアがあるか否かを判定し（ステップ S 807）、取り出していない購入ログのペアがあれば（ステップ S 807 : Yes）、ステップ S 803 に戻って以降の処理を繰り返す。一方、すべての購入ログのペアに対してステップ S 803 ～ステップ S 806 の処理を行っていれば（ステップ S 807 : No）、 $A \cap B$ 文書群抽出器 91 は、ステップ S 801 で選択していないユーザがいるか否かを判定し（ステップ S 808）、選択していないユーザがいれば（ステップ S 808 : Yes）、ステップ S 801 に戻って以降の処理を繰り返す。

[0097] 一方、購入ログに含まれるすべてのユーザを選択してステップ S 802 ～ステップ S 806 の処理を行っていれば（ステップ S 808 : No）、 $A \cap B$ 文書群抽出器 91 は、確信度付き $A \cap B$ 文書群 95 を出力し（ステップ S 809）、一連の処理を終了する。

[0098] 次に、単語関連度評価器 92 の処理手順を説明する。単語関連度評価器 92 の処理は、確信度付き $A \cap B$ 文書群 95 に含まれる各単語のそれぞれについて、第 1 実施形態の単語関連度評価器 32 と同様に、商品 A および商品 B の双方との関連性を表す第 3 スコアを算出することを目的とする。ただし、 $A \cap B$ 文書には確信度が与えられているため、それに伴う処理が第 1 実施形態の単語関連度評価器 32 とは異なる。

[0099] 図 13 は、単語関連度評価器 92 の処理手順を示すフローチャートである。単語関連度評価器 92 は、まず、各単語の出現数を集計するための集計用ヒストグラムおよび総単語数を初期化する（ステップ S 901）。総単語数

は、後述のように確信度付き $A \cap B$ 文書群 95 に含まれる総単語数を文書の確信度に応じて調整した値である。

[0100] 次に、単語関連度評価器 92 は、確信度付き $A \cap B$ 文書群 95 から文書を 1 つ取り出す（ステップ S 902）。そして、単語関連度評価器 92 は、ステップ S 902 で取り出した文書内に含まれる単語のヒストグラムを作成する（ステップ S 903）。ただし、この際、各単語に与えられる頻度は、実際の頻度に確信度を積算したものとする。例えば確信度が 50% の文書において、単語 A が 10 回、単語 B が 6 回、単語 C が 4 回出現したとすると、単語 A に与える頻度は 5 回、単語 B に与える頻度は 3 回、単語 C に与える頻度は 2 回となる。

[0101] 次に、単語関連度評価器 92 は、ステップ S 903 で得られたヒストグラムを集計用ヒストグラムに加算する（ステップ S 904）。また、単語関連度評価器 92 は、当該文書の単語数に確信度を積算した値を総単語数に加算する（ステップ S 905）。例えば、当該文書の単語数が 1000、確信度が 50% であれば、加算する単語数は 500 となる。

[0102] 次に、単語関連度評価器 92 は、確信度付き $A \cap B$ 文書群 95 から取り出していない文書があるか否かを判定し（ステップ S 906）、確信度付き $A \cap B$ 文書群 95 から取り出していない文書があれば（ステップ S 906 : Yes）、ステップ S 902 に戻って以降の処理を繰り返す。一方、確信度付き $A \cap B$ 文書群 95 のすべての文書に対してステップ S 902～ステップ S 905 の処理を行っていれば（ステップ S 906 : No）、単語関連度評価器 92 は、集計用ヒストグラムから各単語の対数確率を割り出す（ステップ S 907）。具体的には、集計用ヒストグラムが示す各単語の頻度を x 、確信度付き $A \cap B$ 文書群 95 の総単語数（ステップ S 905 で加算された総単語数）を y とすると、対数確率は $\log(x/y)$ である。そして、単語関連度評価器 92 は、ステップ S 907 で算出した各単語の対数確率を、各単語の第 3 スコアとして出力し（ステップ S 908）、一連の処理を終了する。

[0103] なお、単語関連度評価器 9 2 において、上述した購入時間差やレビュー時間差に応じた確信度に基づく処理を行う方法を用いる場合、 $A \cap B$ 文書群抽出器 9 1 において $A \cap B$ 文書群を抽出する際に、第 1 期間や第 2 期間を用いた閾値処理を必ずしも行わなくてもよい。 $A \cap B$ 文書群抽出器 9 1 において閾値処理を行わなければ、非常に大きな購入時間差やレビュー時間差を持つレビュー記事も抽出されるが、そのようなレビュー記事には非常に小さな確信度が与えられるためである。閾値処理を行わなければ抽出されるレビュー記事が増大するため計算量が増えるが、閾値処理によるレビュー記事の取りこぼしを回避することができる。

[0104] 本実施形態の情報提示装置におけるその他の処理は、上述した第 1 実施形態と同様である。つまり、本実施形態の情報提示装置においても、統合スコア算出部 5 0 により、確信度付き $A \cap B$ 文書群 9 5 に含まれる各単語について統合スコアが算出され、固有語出力器 6 1 により、統合スコアが高い重要単語が単語ベースの推薦理由 6 5 として画面 2 0 0 に出力され、固有文出力器 6 2 により、重要単語を多く含む文が文ベースの推薦理由 6 6 として画面 2 0 0 に出力される。

[0105] したがって、本実施形態の情報提示装置を用いることによって、EC システムを利用するユーザに対して、商品 A と商品 B との組み合わせ効果に関する情報を含む推薦理由を適切に提示して、調和型推薦による販売促進の効果を高めることができる。すなわち、EC システムを利用するユーザにとっては、本実施形態の情報提示装置により提示される推薦理由を参照することで B 商品を購入する動機付けが生まれ、新体験を伴う商品購入がしやすくなり、店舗にとっては販売機会を増やすことができる。

[0106] 以上説明した第 1 実施形態または第 2 実施形態の情報提示装置における上述した各機能は、例えば、情報提示装置において所定のプログラムを実行することにより実現することができる。この場合、情報提示装置は、例えば図 1 4 に示すように、CPU (Central Processing Unit) 5 1 0 などのプロセッサ、ROM (Read Only Memory) 5 2 0 や RAM (Random Access M

emory) 530などの記憶装置、表示器や各種操作デバイスが接続される入出力I/F540、ネットワークに接続して通信を行う通信I/F550、各部を接続するバス560などを備えた、通常のコンピュータを利用したハードウェア構成とすることができる。

[0107] 上述した情報提示装置で実行されるプログラムは、例えば、インストール可能な形式又は実行可能な形式のファイルでCD-ROM (Compact Disk Read Only Memory)、フレキシブルディスク (FD)、CD-R (Compact Disk Recordable)、DVD (Digital Versatile Disc) 等のコンピュータで読み取り可能な記録媒体に記録されてコンピュータプログラムプロダクトとして提供される。

[0108] また、上述した情報提示装置で実行されるプログラムを、インターネットなどのネットワークに接続されたコンピュータ上に格納し、ネットワーク経由でダウンロードさせることにより提供するように構成してもよい。また、本実施形態の情報提示装置で実行されるプログラムをインターネットなどのネットワーク経由で提供または配布するように構成してもよい。

[0109] また、上述した情報提示装置で実行されるプログラムを、ROM520などに予め組み込んで提供するように構成してもよい。

[0110] 上述した情報提示装置で実行されるプログラムは、情報提示装置の各処理部（第1スコア算出部10, 70、第2スコア算出部20, 80、第3スコア算出部30, 90、第4スコア算出部40、統合スコア算出部50および提示部60）を含むモジュール構成となっており、実際のハードウェアとしては、例えば、CPU510（プロセッサ）が上記記録媒体からプログラムを読み出して実行することにより、上述した各処理部がRAM530（主記憶）上にロードされ、上述した各処理部がRAM530（主記憶）上に生成されるようになっている。なお、実施形態の情報提示装置は、上述した各処理部の一部または全部を、ASIC (Application Specific Integrated Circuit) やFPGA (Field-Programmable Gate Array) などの専用のハードウェアを用いて実現することも可能である。

[0111] 以上、本発明の実施形態を説明したが、ここで説明した実施形態は、例として提示したものであり、発明の範囲を限定することは意図していない。ここで説明した新規な実施形態は、その他の様々な形態で実施されることが可能であり、発明の要旨を逸脱しない範囲で、種々の省略、置き換え、変更を行うことができる。ここで説明した実施形態やその変形は、発明の範囲や要旨に含まれるとともに、請求の範囲に記載された発明とその均等の範囲に含まれる。

請求の範囲

[請求項1]

ユーザが参照している第1商品に調和する第2商品を推薦する際に、前記第1商品と前記第2商品との組み合わせ効果に関する情報を含む推薦理由を提示する情報提示装置であって、

検索対象の文書群から前記第1商品に関する第1文書群を抽出し、該第1文書群に含まれる各単語のそれぞれについて、前記第1商品との関連性を表す第1スコアを算出する第1スコア算出部と、

検索対象の文書群から前記第2商品に関する第2文書群を抽出し、該第2文書群に含まれる各単語のそれぞれについて、前記第2商品との関連性を表す第2スコアを算出する第2スコア算出部と、

検索対象の文書群から前記第1商品と前記第2商品との双方に関する第3文書群を抽出し、該第3文書群に含まれる各単語のそれぞれについて、前記第1商品および前記第2商品の双方との関連性を表す第3スコアを算出する第3スコア算出部と、

前記第3文書群に含まれる各単語のそれぞれについて、前記第3スコアから前記第1スコアと前記第2スコアとを減算し、統合スコアを算出する統合スコア算出部と、

前記統合スコアに基づいて所定の基準に従って選択された1以上の重要単語、または、該重要単語を含む前記第3文書群中の1以上のテキストの少なくとも一方を、前記推薦理由として提示する提示部と、を備える情報提示装置。

[請求項2]

前記第1スコア算出部は、検索対象の文書群から前記第1商品を表す記述を含む前記第1文書群を抽出し、該第1文書群に含まれる各単語のそれぞれについて、前記第1文書群における該単語の出現頻度が高いほど高い値となる前記第1スコアを算出し、

前記第2スコア算出部は、検索対象の文書群から前記第2商品を表す記述を含む前記第2文書群を抽出し、該第2文書群に含まれる各単語のそれぞれについて、前記第2文書群における該単語の出現頻度が

高いほど高い値となる前記第 2 スコアを算出し、

前記第 3 スコア算出部は、検索対象の文書群から前記第 1 商品を表す記述と前記第 2 商品を表す記述との双方を含む前記第 3 文書群を抽出し、該第 3 文書群に含まれる各単語のそれぞれについて、前記第 3 文書群における該単語の出現頻度が高いほど高い値となる前記第 3 スコアを算出する、請求項 1 に記載の情報提示装置。

[請求項3]

検索対象の文書群に含まれる各単語のそれぞれについて、検索対象の文書群における該単語を含む文書の出現頻度が低いほど高い値となる第 4 スコアを算出する第 4 スコア算出部をさらに備え、

前記統合スコア算出部は、前記第 3 文書群に含まれる各単語のそれぞれについて、前記第 3 スコアから前記第 1 スコアと前記第 2 スコアとを減算した値に、さらに前記第 4 スコアを積算または加算して、前記統合スコアを算出する、請求項 2 に記載の情報提示装置。

[請求項4]

検索対象の文書群は、商品の識別情報と関連付けられた文書群であり、

前記第 1 スコア算出部は、検索対象の文書群から前記第 1 商品の識別情報に関連付けられた前記第 1 文書群を抽出し、該第 1 文書群に含まれる各単語のそれぞれについて、前記第 1 文書群における該単語の出現頻度が高いほど高い値となる前記第 1 スコアを算出し、

前記第 2 スコア算出部は、検索対象の文書群から前記第 2 商品の識別情報に関連付けられた前記第 2 文書群を抽出し、該第 2 文書群に含まれる各単語のそれぞれについて、前記第 2 文書群における該単語の出現頻度が高いほど高い値となる前記第 2 スコアを算出し、

前記第 3 スコア算出部は、検索対象の文書群から、前記第 1 商品と前記第 2 商品の双方を購入したユーザによって記載された、前記第 1 商品の識別情報または前記第 2 商品の識別情報に関連付けられた前記第 3 文書群を抽出し、該第 3 文書群に含まれる各単語のそれぞれについて、前記第 3 文書群における該単語の出現頻度が高いほど高い値と

なる前記第3スコアを算出する、請求項1に記載の情報提示装置。

[請求項5] 前記第3スコア算出部は、検索対象の文書群から、前記第1商品と前記第2商品の双方を所定の第1期間内に購入したユーザによって、前記第1商品または前記第2商品の購入時から所定の第2期間内に記載された、前記第1商品の識別情報または前記第2商品の識別情報に関連付けられた前記第3文書群を抽出し、該第3文書群に含まれる各単語のそれぞれについて、前記第3文書群における該単語の出現頻度が高いほど高い値となる前記第3スコアを算出する、請求項4に記載の情報提示装置。

[請求項6] 前記第3スコア算出部は、前記第3文書群に含まれる各文書について、前記第1商品と前記第2商品の購入時間の差、あるいは、前記第1商品または前記第2商品の購入時から該文書が記載された時刻までの時間差をもとに、該文書に前記第1商品と前記第2商品の双方についての記述が含まれていることの確信度を設定し、前記第3文書群に含まれる各単語のそれぞれについて、前記第3文書群における該単語の出現頻度に応じたスコアに対して、該単語を含む文書に設定したスコアを積算または加算し、前記第3スコアを算出する、請求項4または5に記載の情報提示装置。

[請求項7] ユーザが参照している第1商品に調和する第2商品を推薦する際に、前記第1商品と前記第2商品との組み合わせ効果に関する情報を含む推薦理由を提示する情報提示装置により実行される情報提示方法であって、

前記情報提示装置が、検索対象の文書群から前記第1商品に関する第1文書群を抽出し、該第1文書群に含まれる各単語のそれぞれについて、前記第1商品との関連性を表す第1スコアを算出する工程と、

前記情報提示装置が、検索対象の文書群から前記第2商品に関する第2文書群を抽出し、該第2文書群に含まれる各単語のそれぞれについて、前記第2商品との関連性を表す第2スコアを算出する工程と、

前記情報提示装置が、検索対象の文書群から前記第 1 商品と前記第 2 商品との双方に関する第 3 文書群を抽出し、該第 3 文書群に含まれる各単語のそれぞれについて、前記第 1 商品および前記第 2 商品の双方との関連性を表す第 3 スコアを算出する工程と、

前記情報提示装置が、前記第 3 文書群に含まれる各単語のそれぞれについて、前記第 3 スコアから前記第 1 スコアと前記第 2 スコアとを減算し、統合スコアを算出する工程と、

前記情報提示装置が、前記統合スコアに基づいて所定の基準に従って選択された 1 以上の重要単語、または、該重要単語を含む前記第 3 文書群中の 1 以上のテキストの少なくとも一方を、前記推薦理由として提示する工程と、を含む情報提示方法。

[請求項 8]

コンピュータに、

検索対象の文書群からユーザが参照している第 1 商品に関する第 1 文書群を抽出し、該第 1 文書群に含まれる各単語のそれぞれについて、前記第 1 商品との関連性を表す第 1 スコアを算出する機能と、

検索対象の文書群から前記第 1 商品に調和する第 2 商品に関する第 2 文書群を抽出し、該第 2 文書群に含まれる各単語のそれぞれについて、前記第 2 商品との関連性を表す第 2 スコアを算出する機能と、

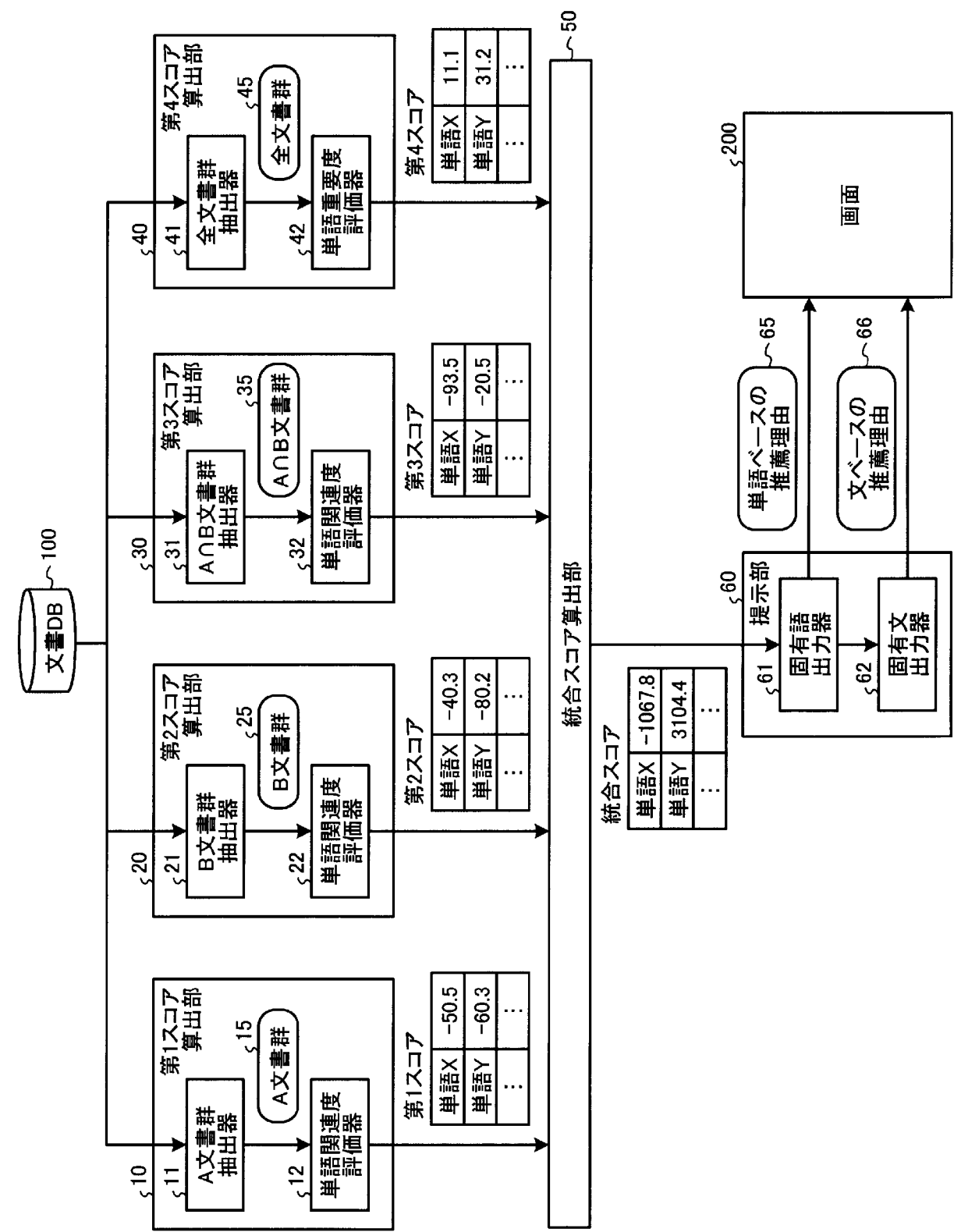
検索対象の文書群から前記第 1 商品と前記第 2 商品との双方に関する第 3 文書群を抽出し、該第 3 文書群に含まれる各単語のそれぞれについて、前記第 1 商品および前記第 2 商品の双方との関連性を表す第 3 スコアを算出する機能と、

前記第 3 文書群に含まれる各単語のそれぞれについて、前記第 3 スコアから前記第 1 スコアと前記第 2 スコアとを減算し、統合スコアを算出する機能と、

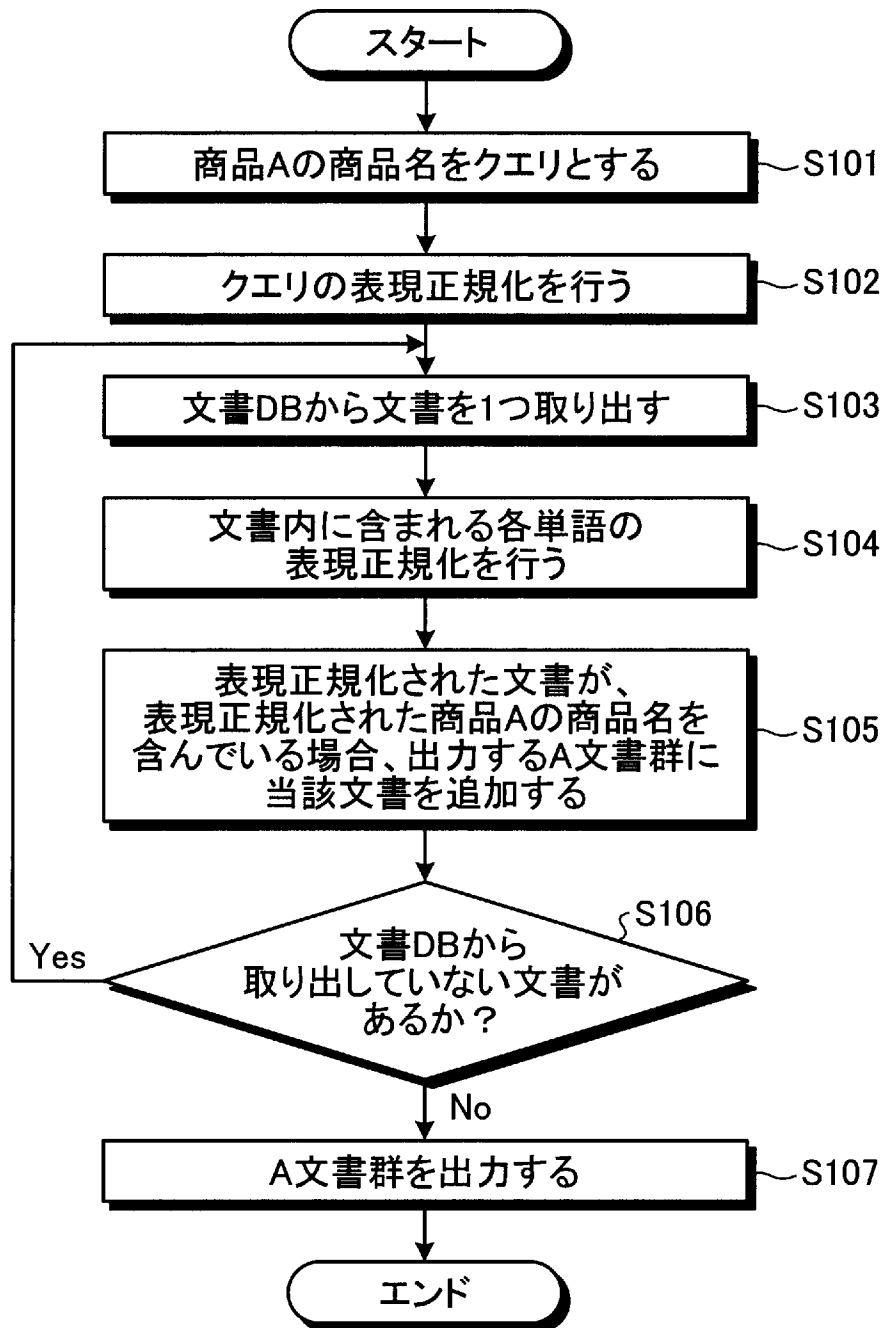
前記統合スコアに基づいて所定の基準に従って選択された 1 以上の重要単語、または、該重要単語を含む前記第 3 文書群中の 1 以上のテキストの少なくとも一方を、前記第 1 商品と前記第 2 商品との組み合

わせ効果に関する情報を含む推薦理由として提示する機能と、を実現させるためのプログラム。

[図1]



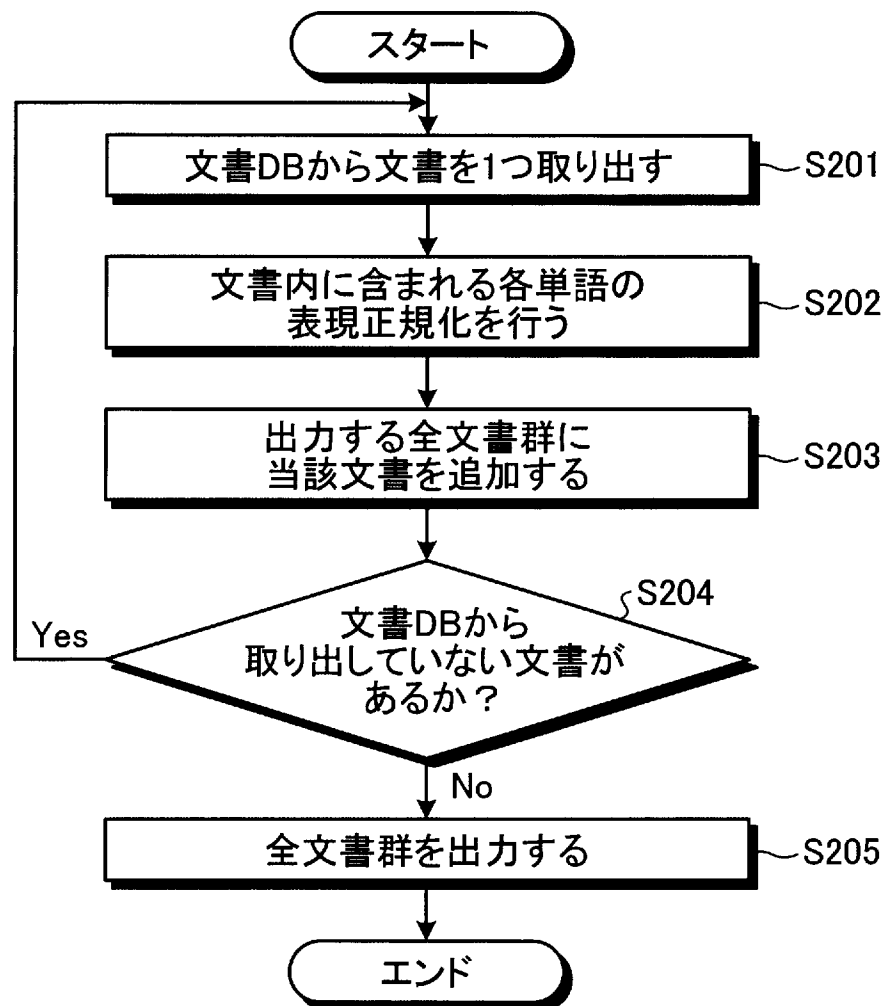
[図2]



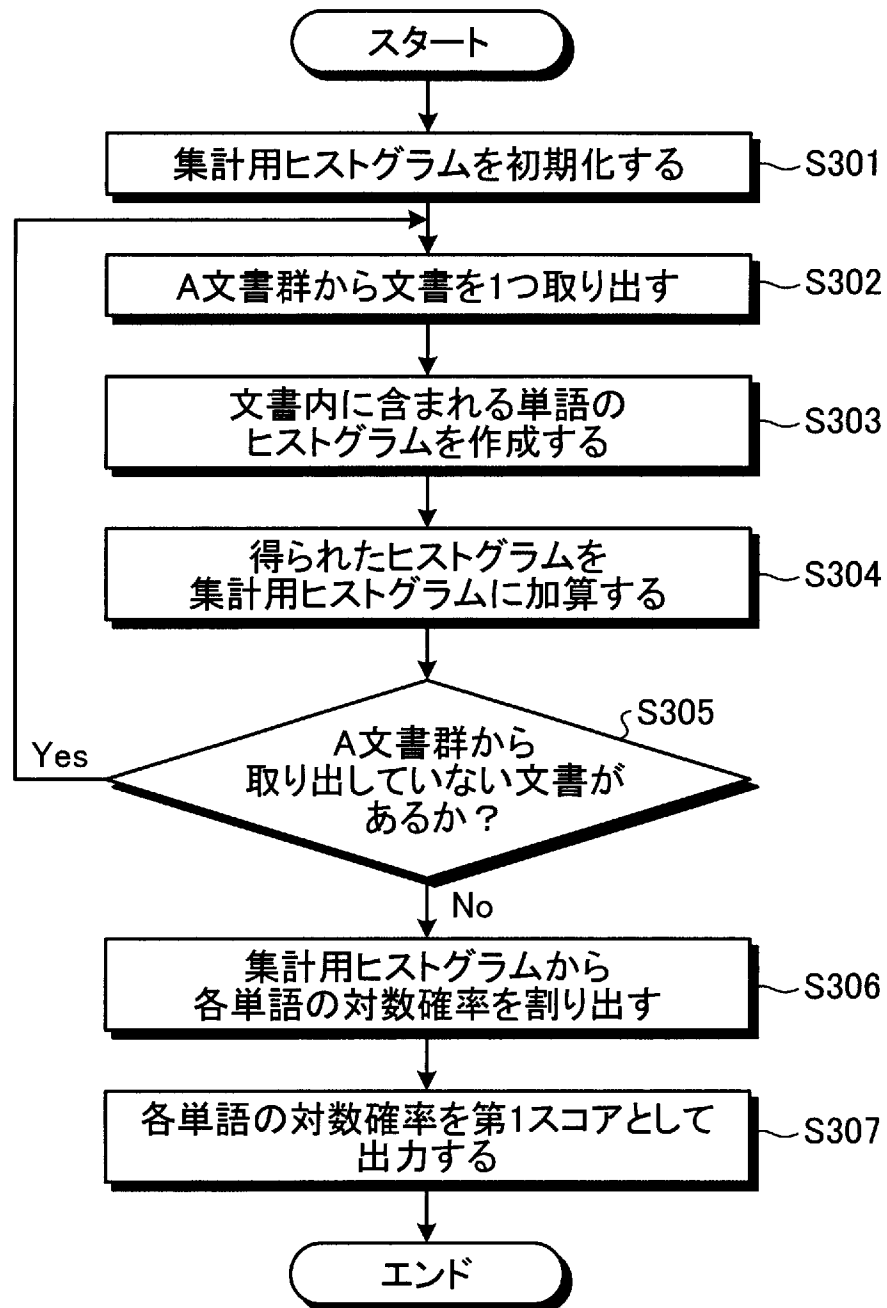
[図3]

PC	パソコン	パーソナル コンピュータ	
ポテト	じゃがいも	じゃが	馬鈴薯
スマートフォン	スマホ	スマフォ	
⋮	⋮	⋮	⋮

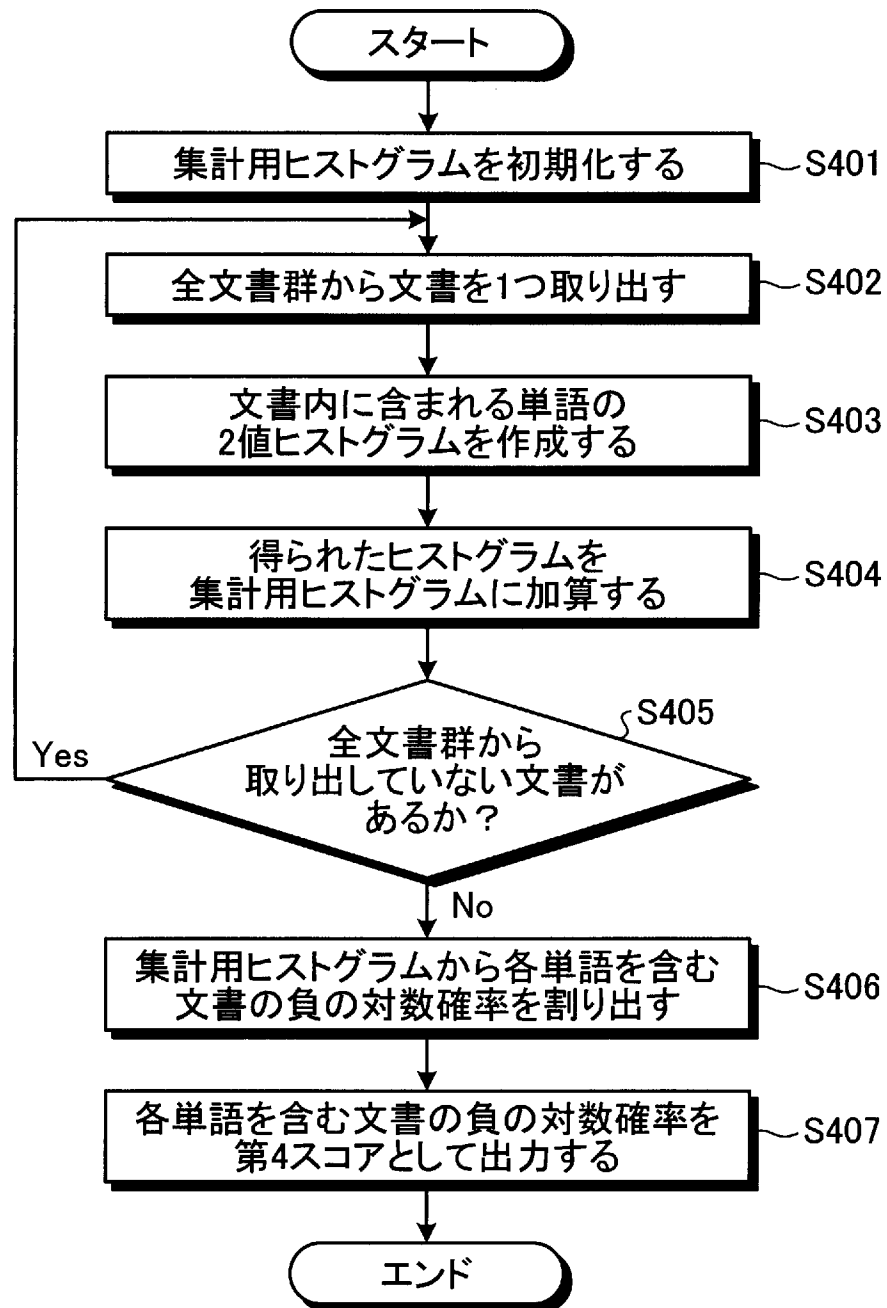
[図4]



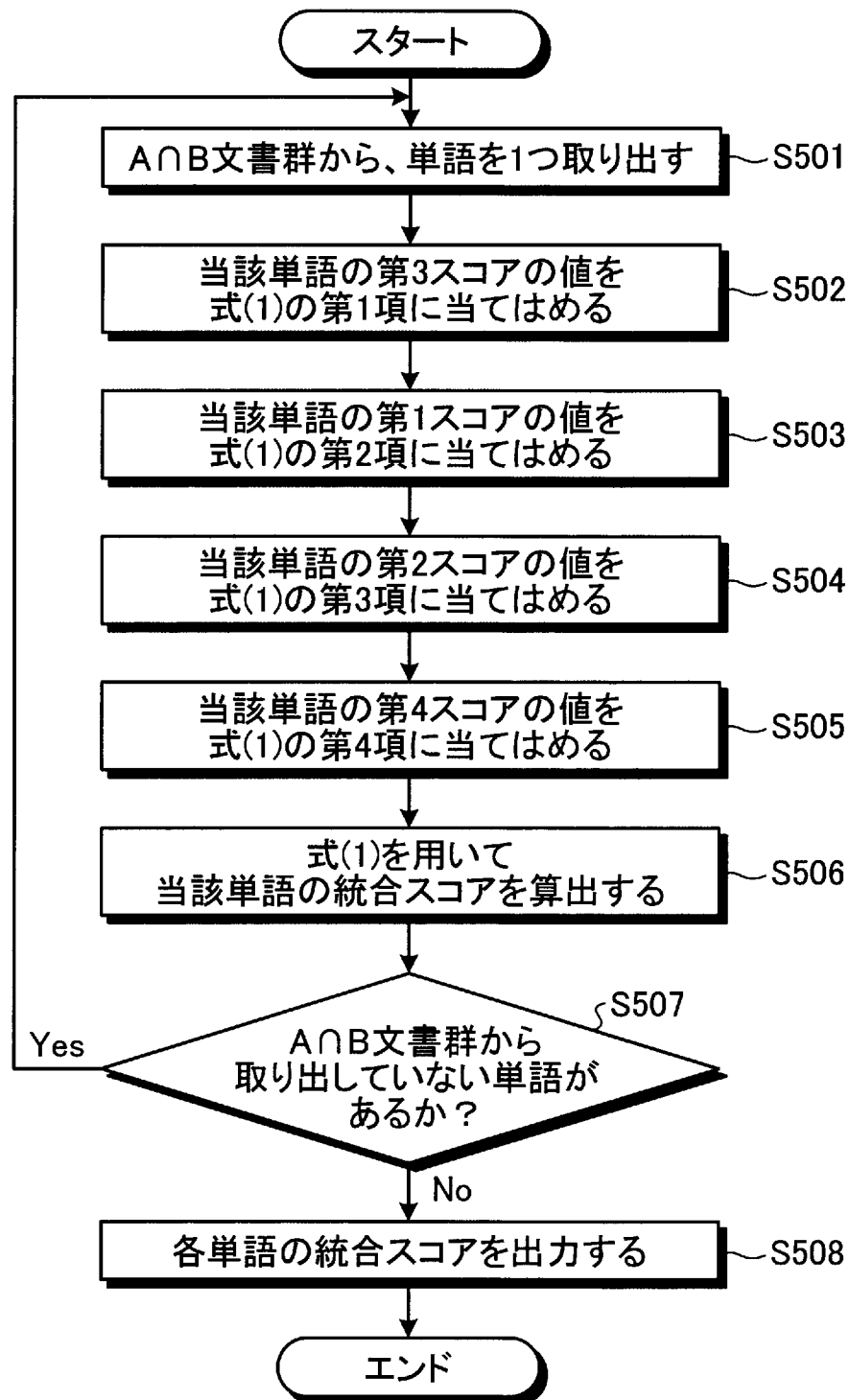
[図5]



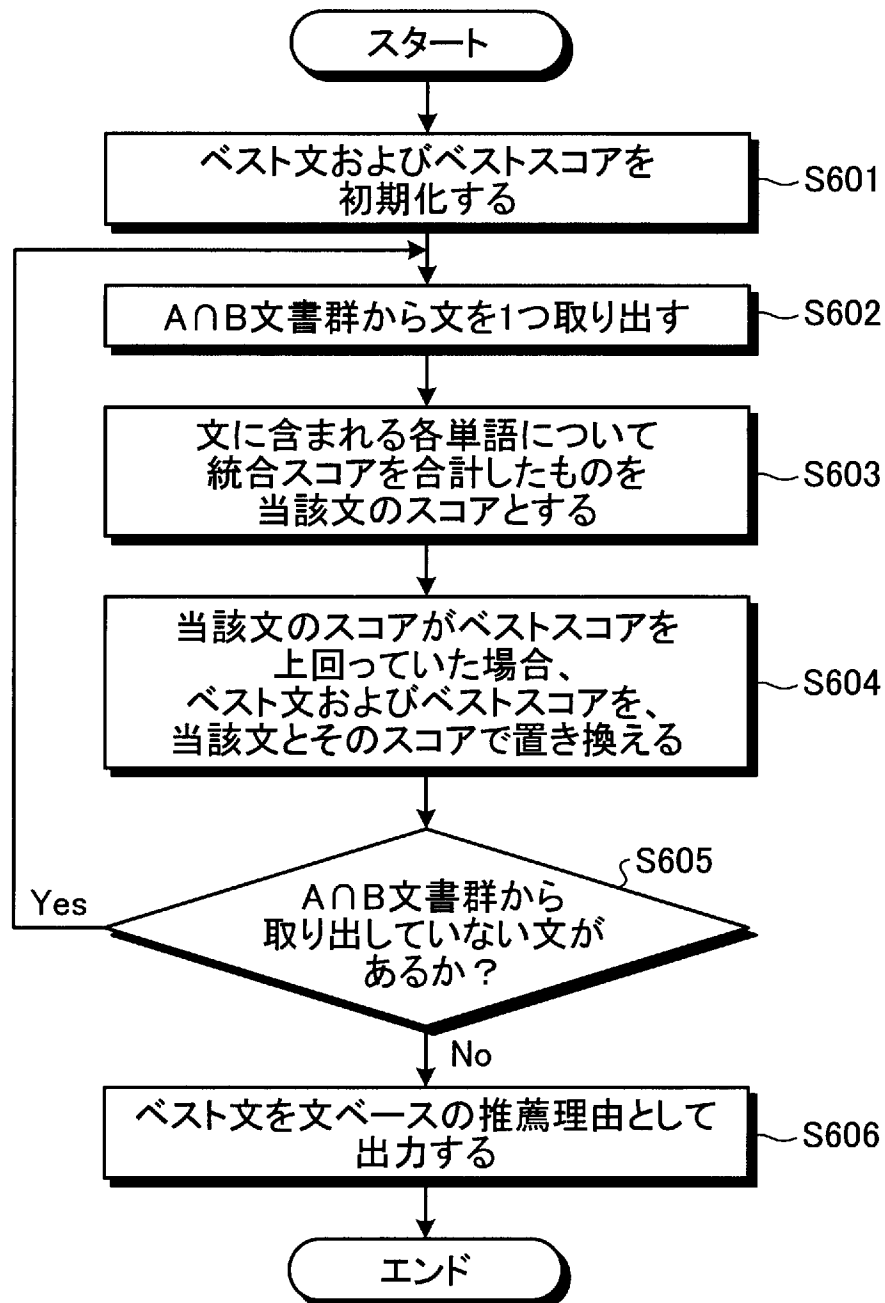
[図6]



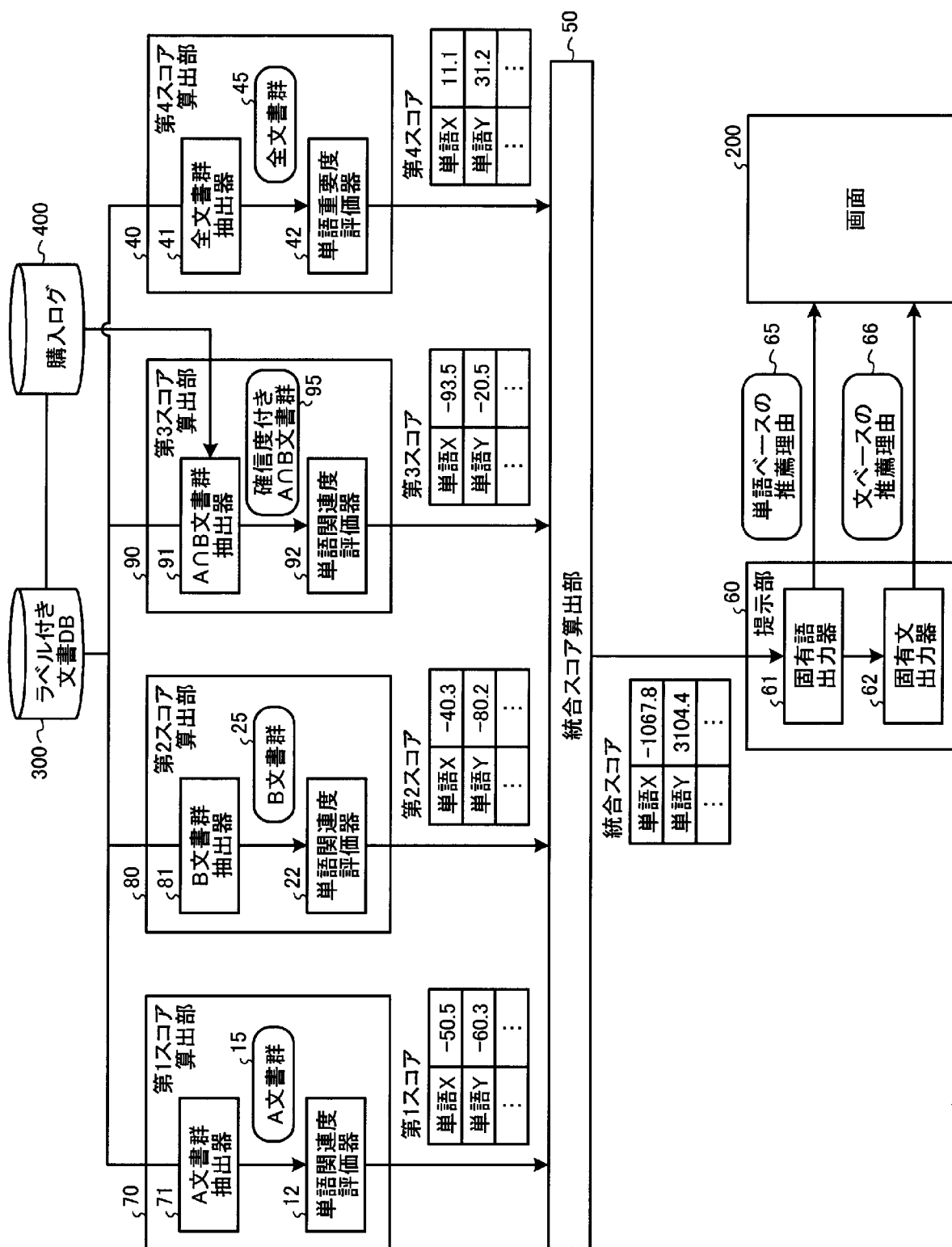
[図7]



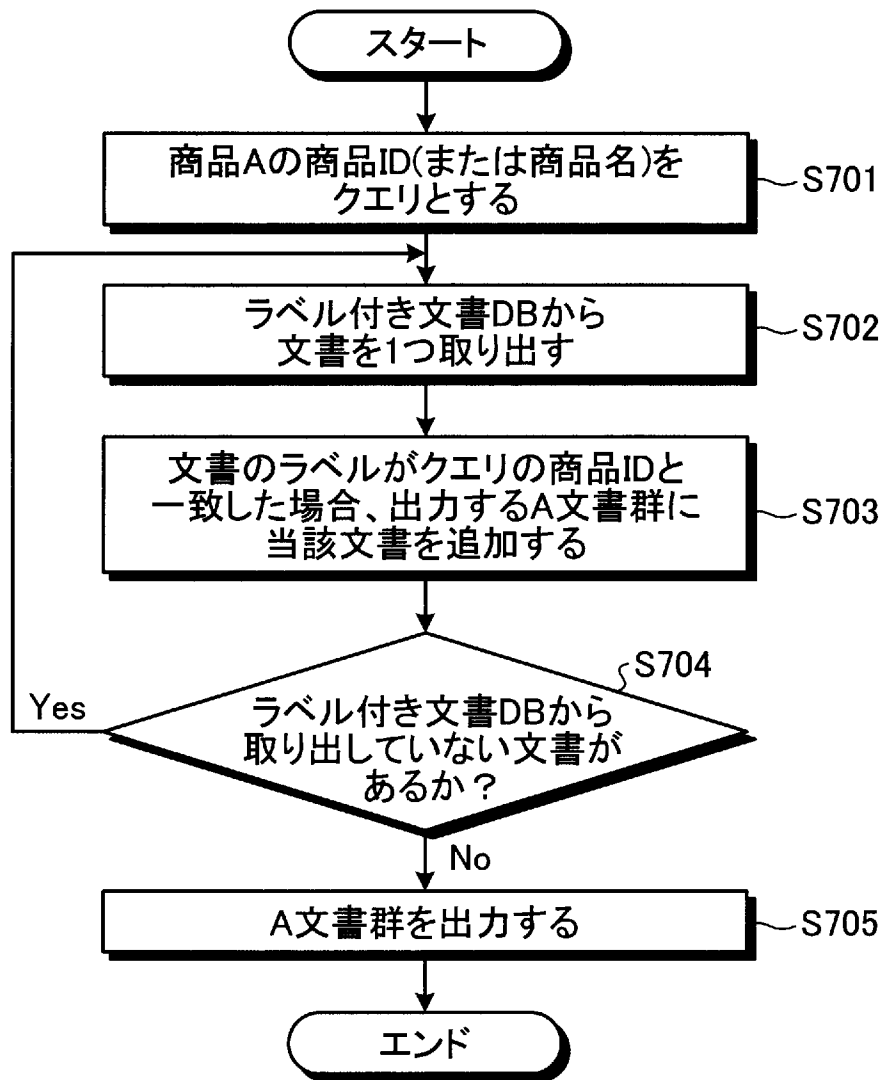
[図8]



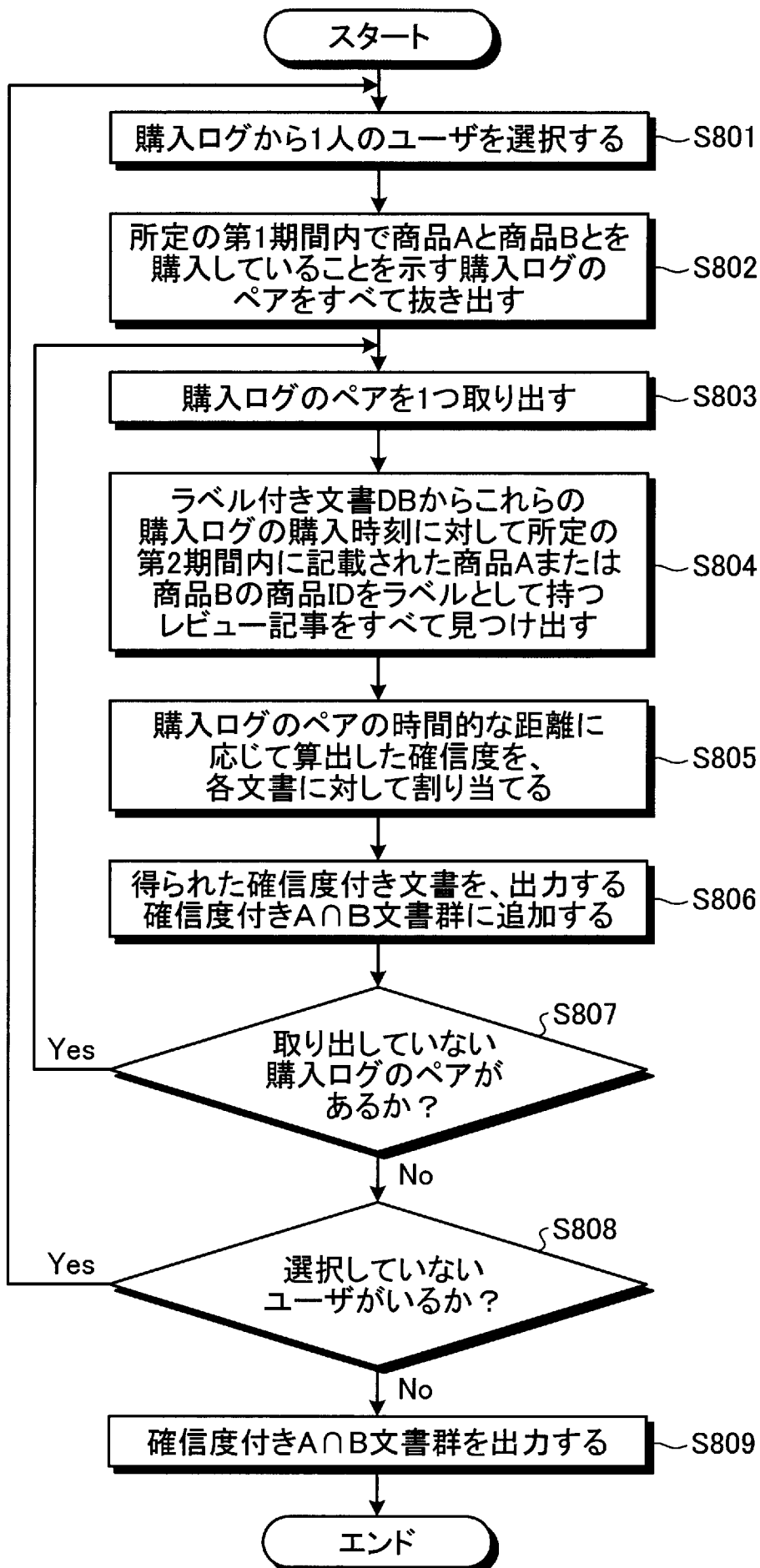
[図9]



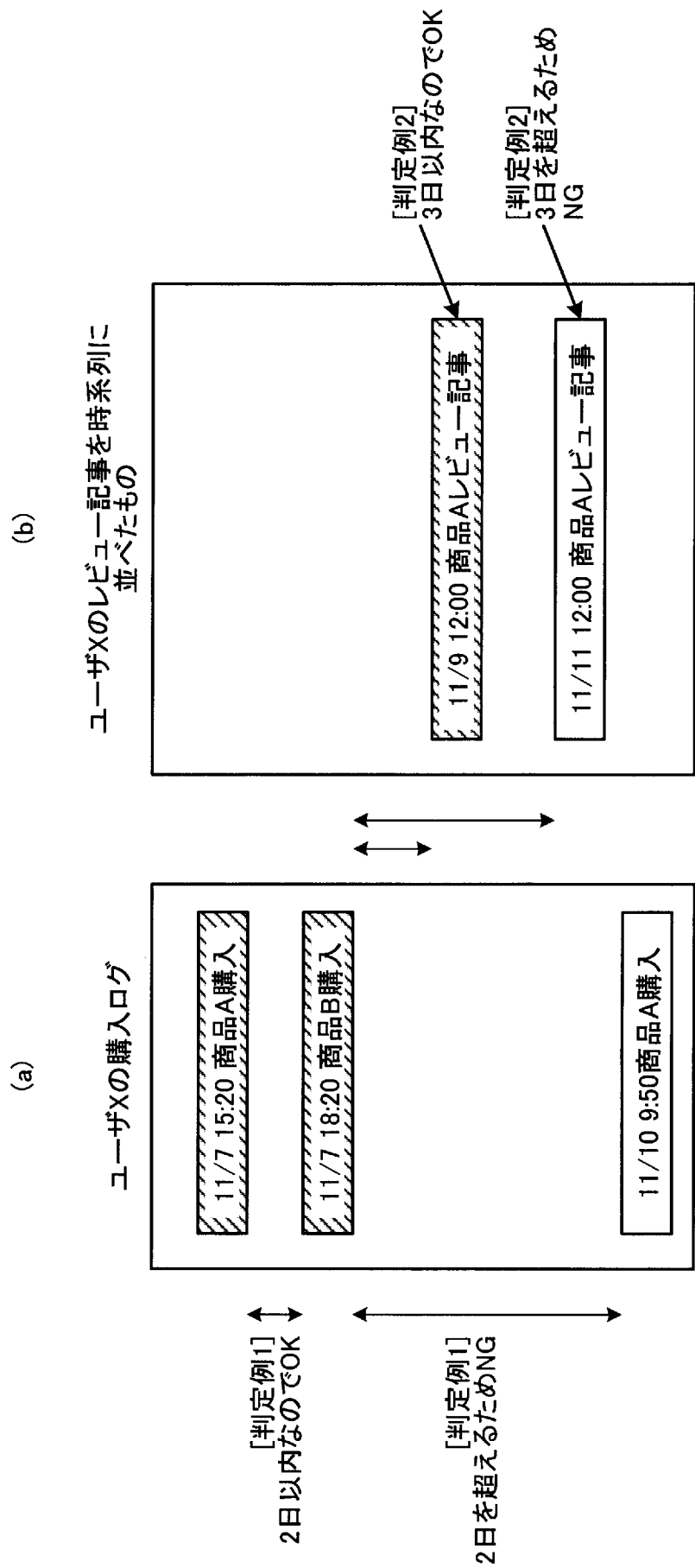
[図10]



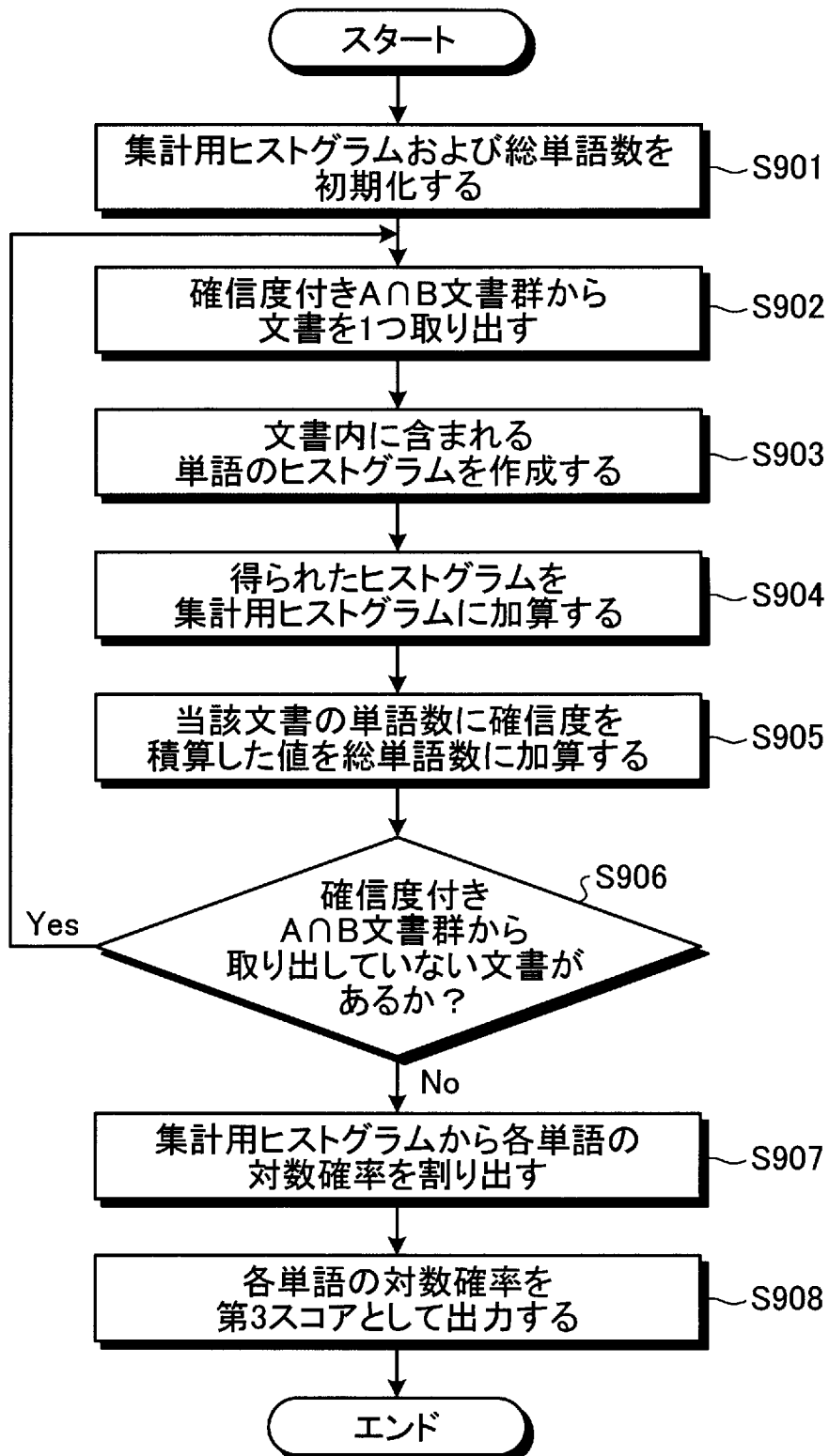
[図11]



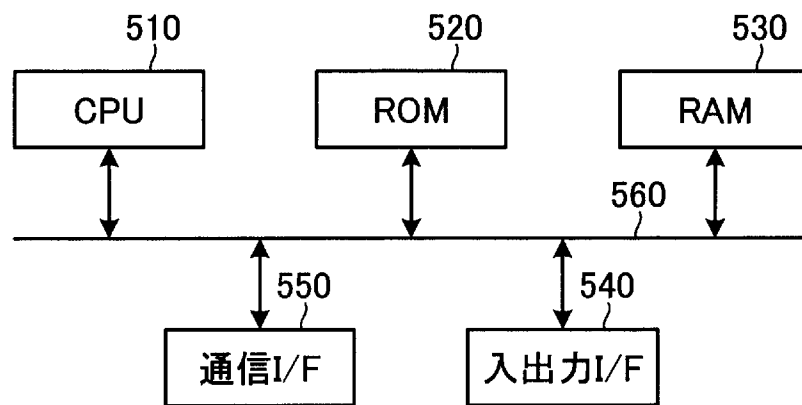
[図12]



[図13]



[図14]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2015/063532

A. CLASSIFICATION OF SUBJECT MATTER

G06F17/30(2006.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F17/30

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2015
Kokai Jitsuyo Shinan Koho	1971-2015	Toroku Jitsuyo Shinan Koho	1994-2015

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2008-225659 A (Sony Corp.), 25 September 2008 (25.09.2008), entire text; all drawings (Family: none)	1-8
A	JP 2009-064187 A (Sony Corp.), 26 March 2009 (26.03.2009), entire text; all drawings & US 8332208 B2 & US 2009/0063132 A1 & CN 101382946 A	1-8
A	JP 2010-113557 A (Nippon Telegraph and Telephone Corp.), 20 May 2010 (20.05.2010), entire text; all drawings (Family: none)	1-8

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance
 "E" earlier application or patent but published on or after the international filing date
 "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
 "O" document referring to an oral disclosure, use, exhibition or other means
 "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
 "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
 "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
 "&" document member of the same patent family

Date of the actual completion of the international search
29 May 2015 (29.05.15)

Date of mailing of the international search report
09 June 2015 (09.06.15)

Name and mailing address of the ISA/
Japan Patent Office
3-4-3, Kasumigaseki, Chiyoda-ku,
Tokyo 100-8915, Japan

Authorized officer

Telephone No.

A. 発明の属する分野の分類（国際特許分類（I P C））

Int.Cl. G06F17/30(2006.01)i

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（I P C））

Int.Cl. G06F17/30

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1 9 2 2 - 1 9 9 6 年
日本国公開実用新案公報	1 9 7 1 - 2 0 1 5 年
日本国実用新案登録公報	1 9 9 6 - 2 0 1 5 年
日本国登録実用新案公報	1 9 9 4 - 2 0 1 5 年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 2008-225659 A（ソニー株式会社）2008.09.25, 全文, 全図（ファミリーなし）	1-8
A	JP 2009-064187 A（ソニー株式会社）2009.03.26, 全文, 全図 & US 8332208 B2 & US 2009/0063132 A1 & CN 101382946 A	1-8
A	JP 2010-113557 A（日本電信電話株式会社）2010.05.20, 全文, 全図（ファミリーなし）	1-8

☐ C 欄の続きにも文献が列挙されている。☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの
「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）
「O」口頭による開示、使用、展示等に言及する文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「&」同一パテントファミリー文献

国際調査を完了した日

2 9 . 0 5 . 2 0 1 5

国際調査報告の発送日

0 9 . 0 6 . 2 0 1 5

国際調査機関の名称及びあて先

日本国特許庁（I S A / J P）
郵便番号 1 0 0 - 8 9 1 5
東京都千代田区霞が関三丁目4番3号

特許庁審査官（権限のある職員）

田中 秀樹

5 M

3 2 4 6

電話番号 0 3 - 3 5 8 1 - 1 1 0 1 内線 3 5 9 9