

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5281405号  
(P5281405)

(45) 発行日 平成25年9月4日(2013.9.4)

(24) 登録日 平成25年5月31日(2013.5.31)

(51) Int. Cl. F I  
**G06F 17/21 (2006.01)** G O 6 F 17/21 5 5 0 A  
**G06F 17/30 (2006.01)** G O 6 F 17/30 2 2 0 Z

請求項の数 18 (全 25 頁)

(21) 出願番号	特願2008-533768 (P2008-533768)	(73) 特許権者	502208397
(86) (22) 出願日	平成18年9月29日 (2006. 9. 29)		グーグル インコーポレイテッド
(65) 公表番号	特表2009-510637 (P2009-510637A)		アメリカ合衆国 カリフォルニア州 94
(43) 公表日	平成21年3月12日 (2009. 3. 12)		043 マウンテン ビュー アンフィシ
(86) 国際出願番号	PCT/US2006/038552		アター パークウェイ 1600
(87) 国際公開番号	W02007/041545	(74) 代理人	100079108
(87) 国際公開日	平成19年4月12日 (2007. 4. 12)		弁理士 稲葉 良幸
審査請求日	平成21年6月19日 (2009. 6. 19)	(74) 代理人	100093861
(31) 優先権主張番号	11/241, 701		弁理士 大賀 眞司
(32) 優先日	平成17年9月30日 (2005. 9. 30)	(74) 代理人	100109346
(33) 優先権主張国	米国 (US)		弁理士 大貫 敏史

最終頁に続く

(54) 【発明の名称】 表示のための高品質レビューの選択

(57) 【特許請求の範囲】

【請求項1】

コンピュータにより実施される、レビューを処理する方法であって、  
 クライアントからの検索要求に回答するステップを含み、前記回答するステップが、  
 前記クライアントからの前記検索要求に従って1つ以上のレビューソースからレビュー  
 のセットを識別するステップであって、前記レビューのセットが、前記検索要求と一致  
 する1つ以上のエンティティの2つ以上のレビューを含み、前記2つ以上のレビューのそ  
 れぞれは、前記エンティティに対するレーティングスコアに関連付けられている、ステッ  
 プと、

前記識別されたレビューのセットに対する全体的レーティングスコアを生成するステ  
 ップであって、前記全体的レーティングスコアは、前記レビューのセットに含まれる各レ  
 ビューに関連付けられた前記レーティングスコアに少なくとも基づいて決定される、ステ  
 ップと、

複数の所定のレーティング範囲からレーティング範囲を識別するステップであって、  
 前記識別されたレーティング範囲が、前記識別されたレビューのセットに対する前記全体的  
 レーティングスコアに対応する、ステップと、

前記識別されたレビューのセットのサブセットを選択するステップであって、前記サブ  
 セットは、前記識別されたレーティング範囲に前記レーティングスコアが含まれるレビ  
 ユーを少なくとも含む、ステップと、

前記クライアントでの表示のために前記検索要求に対する回答を生成し、前記クライ

10

20

アントに送信するステップであって、前記応答が、前記選択されたサブセットからの内容を含む、ステップとによって行われる、方法。

【請求項 2】

前記識別されたレーティング範囲が、前記複数の所定のレーティング範囲のうちの 1 つを含み、前記複数の所定のレーティング範囲のうちの前記 1 つが、前記識別されたレビューのセットに対する前記全体的レーティングスコアを含む、請求項 1 に記載の方法。

【請求項 3】

全体的レーティングスコアを生成するステップが、  
前記レビューソースからの 1 つ以上の集合レーティングスコアを識別するステップと、  
前記集合レーティングスコアおよび各レビューソースに含まれる前記レビューのセットのそれぞれのレビューの数に基づいて全体的レーティングスコアを決定するステップとを含む、請求項 1 または 2 に記載の方法。

10

【請求項 4】

応答を生成するステップが、  
前記選択されたサブセットにおける複数のレビューから抜粋を生成するステップを含む、請求項 1 ~ 3 のいずれか 1 項に記載の方法。

【請求項 5】

抜粋を生成するステップが、  
レビューを 1 つ以上のパーティションに分割するステップと、  
所定の基準に基づいて前記パーティションのサブセットを選択するステップと、  
前記パーティションの前記選択されたサブセットからの内容を含む抜粋を生成するステップとを含む、請求項 4 に記載の方法。

20

【請求項 6】

前記全体的レーティングスコアが、前記識別されたレビューのセットの個別のレーティングスコアの加重平均を含む、請求項 1 ~ 5 のいずれか 1 項に記載の方法。

【請求項 7】

レビューを処理するためのサーバシステムであって、  
1 つ以上のプロセッサと、  
クライアントからの検索要求に応答するために、前記 1 つ以上のプロセッサによって実行される 1 つ以上のプログラムを格納するメモリとを含む、

30

前記 1 つ以上のプログラムが、  
前記クライアントからの前記検索要求に従って 1 つ以上のレビューソースからレビューのセットを識別する命令であって、前記レビューのセットが、前記検索要求と一致する 1 つ以上のエンティティの 2 つ以上のレビューを含み、前記 2 つ以上のレビューのそれぞれは、前記エンティティに対するレーティングスコアに関連付けられている、命令と、

前記識別されたレビューのセットに対する全体的レーティングスコアを生成する命令であって、前記全体的レーティングスコアは、前記レビューのセットに含まれる各レビューに関連付けられた前記レーティングスコアに少なくとも基づいて決定される、命令と、

複数の所定のレーティング範囲からレーティング範囲を識別する命令であって、前記識別されたレーティング範囲が、前記識別されたレビューのセットに対する前記全体的レーティングスコアに対応する、命令と、

40

前記識別されたレビューのセットのサブセットを選択する命令であって、前記サブセットは、前記識別されたレーティング範囲に前記レーティングスコアが含まれるレビューを少なくとも含む、命令と、

前記選択されたサブセットから内容を抽出する命令であって、前記抽出された内容が、前記選択されたサブセットにおける複数のレビューの抜粋を含む、命令と、

前記クライアントでの表示のために前記検索要求に対する応答を生成し、前記クライアントに送信する命令であって、前記応答が、前記選択されたサブセットからの前記抽出された内容を含む、命令

50

とを含む、サーバシステム。

【請求項 8】

前記識別されたレーティング範囲が、前記複数の所定のレーティング範囲のうちの 1 つを含み、前記複数の所定のレーティング範囲のうちの前記 1 つが、前記識別されたレビューのセットに対する前記全体的レーティングスコアを含む、請求項 7 に記載のシステム。

【請求項 9】

前記全体的レーティングスコアを生成する命令が、  
前記レビューソースからの 1 つ以上の集合レーティングスコアを識別する命令と、  
前記集合レーティングスコアおよび各レビューソースに含まれる前記レビューのセットのそれぞれのレビューの数に基づいて全体的レーティングスコアを決定する命令と  
を含む、請求項 7 または 8 に記載のシステム。

10

【請求項 10】

前記抜粋を生成する命令が、前記選択されたサブセットにおける複数のレビューから抜粋を生成する命令を含む、請求項 7 ~ 9 のいずれか 1 項に記載のシステム。

【請求項 11】

前記抜粋を生成する命令が、  
レビューを 1 つ以上のパーティションに分割する命令と、  
所定の基準に基づいて前記パーティションのサブセットを選択する命令と、  
前記パーティションの前記選択されたサブセットからの内容を含む抜粋を生成する命令と  
を含む、請求項 10 に記載のシステム。

20

【請求項 12】

前記全体的レーティングスコアが、前記識別されたレビューのセットの個別のレーティングスコアの加重平均を含む、請求項 7 ~ 11 のいずれか 1 項に記載のシステム。

【請求項 13】

クライアントからの検索要求に回答するためにサーバシステムによる実行のための 1 つ以上のコンピュータプログラムを格納するコンピュータ可読記憶媒体であって、前記 1 つ以上のコンピュータプログラムが、

前記クライアントからの前記検索要求に従って 1 つ以上のレビューソースからレビューのセットを識別する命令であって、前記レビューのセットが、前記検索要求と一致する 1 つ以上のエンティティの 2 つ以上のレビューを含み、前記 2 つ以上のレビューのそれぞれは、前記エンティティに対するレーティングスコアに関連付けられている、命令と、

30

前記識別されたレビューのセットに対する全体的レーティングスコアを生成する命令であって、前記全体的レーティングスコアは、前記レビューのセットに含まれる各レビューに関連付けられた前記レーティングスコアに少なくとも基づいて決定される、命令と、

複数の所定のレーティング範囲からレーティング範囲を識別する命令であって、前記識別されたレーティング範囲が、前記識別されたレビューのセットに対する前記全体的レーティングスコアに対応する、命令と、

前記識別されたレビューのセットのサブセットを選択する命令であって、前記サブセットは、前記識別されたレーティング範囲に前記レーティングスコアが含まれるレビューを少なくとも含む、命令と、

40

前記クライアントでの表示のために前記検索要求に対する応答を生成し、前記クライアントに送信する命令であって、前記応答は、前記選択されたサブセットからの内容を含む、コンピュータ可読記憶媒体。

【請求項 14】

前記識別されたレーティング範囲が、前記複数の所定のレーティング範囲のうちの 1 つを含み、前記複数の所定のレーティング範囲のうちの前記 1 つが、前記識別されたレビューのセットに対する前記全体的レーティングスコアを含む、請求項 13 に記載のコンピュータ可読記憶媒体。

【請求項 15】

50

前記全体的レーティングスコアを生成する命令が、  
 前記レビューソースからの1つ以上の集合レーティングスコアを識別する命令と、  
 前記集合レーティングスコアおよび各レビューソースに含まれる前記レビューのセット  
 のそれぞれのレビューの数に基づいて全体的レーティングスコアを決定する命令と  
 を含む、請求項13または14に記載のコンピュータ可読記憶媒体。

【請求項16】

前記応答を生成する命令が、前記選択されたサブセットにおける複数のレビューから抜  
 粋を生成する命令を含む、請求項13～15のいずれか1項に記載のコンピュータ可読記  
 憶媒体。

【請求項17】

前記抜粋を生成する命令が、  
 レビューを1つ以上のパーティションに分割する命令と、  
 所定の基準に基づいて前記パーティションのサブセットを選択する命令と、  
 前記パーティションの前記選択されたサブセットからの内容を含む前記抜粋を生成する  
 命令と  
 を含む、請求項16に記載のコンピュータ可読記憶媒体。

【請求項18】

前記全体的レーティングスコアが、前記識別されたレビューのセットの個別のレーティ  
 ングスコアの加重平均を含む、請求項13～17のいずれか1項に記載のコンピュータ可  
 読記憶媒体。

【発明の詳細な説明】

【関連出願の相互参照】

【0001】

関連出願

本出願は、以下の出願に関連し、またこれらの出願のそれぞれは、参照によって本明細  
 書に援用される。

【0002】

2005年9月30日出願の米国特許出願第11/241,698号「表示用の代表的  
 レビューの選択 (Selecting Representative Reviews  
 for Display)」。

【0003】

2005年9月30日出願の米国特許出願第11/241,702号「レビュー抜粋で  
 表示するための、識別されたレビュー内における高品質テキストの選択 (Selecti  
 ng High Quality Text Within Identified R  
 eviews for Display in Review Snippets)」。

【0004】

2005年9月30日出願の米国特許出願第11/241,694号「類似レビューの  
 クラスターの識別および多数のクラスターからの代表的レビューの表示 (Identifi  
 ng Clusters of Similar Reviews and Displ  
 aying Representative Reviews from Multip  
 le Clusters)」。

【0005】

2005年9月30日出願の米国特許出願第11/241,693号「評判管理のため  
 のシステムおよび方法 (Systems and Methods for Reput  
 ation Management)」。

【技術分野】

【0006】

技術分野

開示される実施形態は、一般に、検索エンジンに関する。特に、開示される実施形態は  
 、プレゼンテーションのためにレビューおよびレビューの内容を選択するための方法およ

10

20

30

40

50

びシステムに関する。

【背景技術】

【0007】

背景

多くのインターネットユーザは、製品またはサービス入手する前に、それらを調査する。多くのインターネットユーザはまた、製品またはサービスの提供者を調査してから、その提供者と取引する。現在、多くのユーザが従うアプローチは、製品、サービスおよび/またはそれらの提供者のレーティングおよびレビューを提供するウェブサイトを利用することである。たとえば、www.pricegrabber.com、www.bizrate.com、およびwww.resellerratings.comなどのウェブサイトが、製品ならびにその提供者のレーティングおよびレビューを提供する。

10

【0008】

製品、サービスまたは提供者に対するレビューおよびレーティングの全体的視野を得るために、ユーザは、レビューおよびレーティングを提供する多くのウェブサイトアクセスして、それらのウェブサイトによって提供される多くのレーティングおよびレビューを読んでもよい。しかしながら、このプロセスは、かなり時間を浪費し、厄介である。ユーザは、様々なウェブサイトのレビューおよびレーティングを厳選することに時間を費やすのを避けるために、レーティングおよびレビューの簡単な要約で満足することがある。

【発明の開示】

【発明が解決しようとする課題】

20

【0009】

したがって、ユーザが、(たとえば、購入、リース、レンタルまたは他の類似の取引によって)入手することに関心がある製品およびサービスに関して、ならびに取引することに関心がある製品およびサービスの提供者に関して、調査をより効率的に実施できるようにすることは、非常に望ましいであろう。

【課題を解決するための手段】

【0010】

発明の概要

本発明のいくつかの実施形態において、レビューの処理方法には、複数のレビューを識別するステップ、少なくとも所定の品質基準に基づいて複数のレビューのサブセットを選択するステップ、および選択されたサブセットの内容を含む応答を生成するステップが含まれる。

30

【0011】

類似の参照数字は、図面全体を通して対応する部分を指す。

【発明を実施するための最良の形態】

【0012】

実施形態の説明

(製品、サービスまたはこれらの提供者などの)対象に関して調査を行なうユーザは、いくつかのウェブサイトにはわたって多数のレビューおよびレーティングを読むことに時間を費やすことを望まず、その対象のレビューおよびレーティングの要約で満足することがある。要約は、対象に対するレビューのサンプルを含んでもよい。しかしながら、単にサンプルに含むためにランダムにレビューを選択することは、ユーザにはあまり役に立たない。開示される実施形態は、レビューサンプルに含むためのレビューを、所定の非ランダム基準に基づいて選択し、またレビュー抜粋で用いるテキストをレビューから選択する。

40

【0013】

図1は、本発明のいくつかの実施形態によるネットワークを示す。ネットワーク100には、1つ以上のクライアント102、1つ以上のドキュメントホスト104、およびレビューエンジン106が含まれる。ネットワーク100にはまた、これらの構成要素を結合するネットワーク108が含まれる。

【0014】

50

ドキュメントホスト104は、ドキュメントを格納し、ドキュメントへのアクセスを提供する。ドキュメントは、テキスト、グラフィックス、マルチメディアコンテンツ等の任意の組み合わせを含む任意の機械可読データであってもよい。いくつかの実施形態において、ドキュメントは、テキスト、グラフィックス、およびハイパーテキストマークアップ言語（HTML）で書かれた他の可能な情報形態すなわちウェブページの組み合わせであってもよい。ドキュメントには、他のドキュメントへの1つ以上のハイパーリンクを含んでもよい。ドキュメントホスト102に格納されたドキュメントは、ユニフォームリソースロケータ（URL）もしくはウェブアドレス、または識別および/もしくは位置特定のための任意の他の適切な形態によって、位置特定および/または識別してもよい。ドキュメントホスト104はまた、ユーザによってこれらのホストに投稿されたレビューを格納し、ウェブページなどのドキュメントを介してレビューへのアクセスを提供する。

10

**【0015】**

クライアント102には、ユーザがウェブページなどのドキュメントにアクセスできるようにするクライアントアプリケーションが含まれる。いくつかの実施形態において、クライアントアプリケーションには、ウェブブラウザが含まれる。ウェブブラウザの例には、ファイヤフォックス、インターネットエクスプローラおよびオペラが含まれる。いくつかの実施形態において、ユーザはまた、クライアント102を介して、ドキュメントホスト104またはレビューエンジン106にレビューを投稿することができる。

**【0016】**

レビューには、対象または対象のクラスに関する内容（たとえばコメント、評価、意見等）が含まれる。いくつかの実施形態において、内容はテキストである。他の実施形態において、内容にはまた、オーディオ、ビデオ、またはテキスト、オーディオおよびビデオの任意の組み合わせを含んでもよい。

20

**【0017】**

レビューの対象は、特定のエンティティまたは物体であって、それらに対してレビューの内容がコメント、評価、意見等を提供する。いくつかの実施形態において、レビューの対象は、対象のタイプによって分類してもよい。対象タイプの例には、製品、サービス、製品の提供者、サービスの提供者などが含まれる。レビューは、対象のクラスに向けられてもよい。対象のクラスには、共通の特性、特徴または機能を共有する複数の特定のエンティティまたは物体が含まれる。たとえば、特定の製品ラインは、レビューの対象となる可能性がある対象のクラスになり得る。別の例として、特定のブランドを有する全ての製品は、レビューの対象となる可能性がある対象のクラスになり得る。

30

**【0018】**

レーティングは、レビューに関連付けされてもよく、またレビューと共に格納してもよい。レーティング（または「レーティングスコア」）は、レビューの対象（または対象のクラス）のスコアを所定の尺度に基づいて表現する。レーティングのフォーマットは、数値または数値にマップすることができる任意の非数値フォーマットであってもよい。たとえば、非数値の同意または不同意のレーティングは、それぞれ、2進値の1または0にマップされてもよい。レーティング形態の例には、記号または記述的フォーマット（肯定的/否定的、同意/不同意など）および数値フォーマット（1~3、1~5、1~10、1~100など）が含まれる。いくつかの実施形態において、レーティングに加えて、レビューはまた、対象の特定の側面のためのサブレーティングと関連付けられてもよい。サブレーティングは、対象の特定の側面のためのスコアであってもよい。

40

**【0019】**

レビューエンジン106には、レビューサーバ110、レビューリポジトリ112、レビューコレクタ114、およびドキュメントリポジトリ116が含まれる。レビューサーバ110は、クライアント102へ伝送するためのレビューおよび/またはレビュー抜粋を含む応答を生成する。レビューサーバ110はまた、レビューエンジン106にレビューおよびレーティングを投稿するためのインタフェースを、クライアント102のユーザに提供する。

50

## 【0020】

レビューコレクタ114は、ドキュメントからレビューを収集する。レビューコレクタ114は、ドキュメントを解析し、ドキュメントからレビュー、レーティングおよび他の関連情報（レビュー作成者、レビューの日付、レビューの対象等）を抽出する。抽出されたレビューは、格納のためにレビューリポジトリ112へ伝送される。レビューコレクタ114がレビューを抽出する元となるドキュメントは、ドキュメントホスト104および/またはドキュメントリポジトリ116に格納してもよい。

## 【0021】

ドキュメントリポジトリ116は、ドキュメントホスト104に格納されたドキュメントの少なくともサブセットのコピーの記憶装置（a store）である。ドキュメントリポジトリ116に格納されるドキュメントは、ドキュメントホスト104から収集して、レビューエンジン106によってリポジトリ116に格納してもよい。いくつかの実施形態において、ドキュメントリポジトリ116は、レビューエンジン106がアクセス可能な検索エンジン（図示せず）に位置してもよく、検索エンジンは、ドキュメントホスト104からドキュメントを収集し、それらをドキュメントリポジトリ116に格納する働きをする。

10

## 【0022】

レビューエンジン106に格納されるレビューは、クライアント102のユーザによって書かれ、ドキュメントホスト104またはレビューエンジン106に投稿される。ドキュメントホスト104に投稿されたレビューは、ドキュメントホスト104に格納されたドキュメントまたはドキュメントリポジトリ116に格納されたドキュメントのコピーから抽出してもよい。レビューはまた、ユーザがレビューエンジン106に投稿してもよい。ドキュメントから抽出されたレビュー、およびレビューエンジン106に投稿されたレビューの両方は、格納のためにレビューリポジトリ112へ伝送される。

20

## 【0023】

ドキュメントホスト104またはレビューエンジン106は、ユーザがそれらにレビューを投稿する機能を提供してもよい。たとえば、ドキュメントホスト104またはレビューエンジン106は、ユーザがレビューおよびレーティングで埋めて、次に投稿できるオンラインフォームを提供してもよい。レビューは、投稿および保存の後で、ウェブページなどのドキュメントを通して他のユーザがアクセスしてもよい。

30

## 【0024】

レビューのソースは、レビューを投稿されたエンティティである。ソースは、レビューを投稿されたドキュメントホスト104の位置および/または識別子によって識別してもよい。いくつかの実施形態において、レビューのソースは、レビューを投稿されたドキュメントホスト104のドメインによって識別してもよい。たとえば、レビューが、「www.xyz.com」というドメイン下のドキュメントホストに投稿された場合には、抽出されたレビューのソースは「xyz.com」であってもよい。ユーザによってレビューエンジン106に投稿されたレビューの場合には、レビューエンジン106をソースと見なしてもよい。

## 【0025】

レビューリポジトリ112は、レビューおよび関連するレーティングを格納する。レビューリポジトリ112はまた、各レビューの対象または対象のクラス、および対象タイプ（すなわち、対象または対象のクラスが、製品、製品提供者等かどうか）を格納する。レビューリポジトリ112はまた、各レビューのソース、作成者および日付を格納してもよい。いくつかの実施形態において、レビューおよびレーティングを、レビューリポジトリ112において、レビューおよびレーティング自体の1つ以上の評価と関連付けてもよい。レビューおよびレーティングの評価は、レビューならびにレーティングの有用性および/または信頼性を評価してもよい。たとえば、レビューおよびレーティングの評価には、有用/有用でないというレーティングを含んでもよい。別の例として、レビューおよびレーティングは、その作成者の評判の程度に基づいたメトリック値に関連付けてもよい。評

40

50

判をベースにしたメトリック値の例が、2005年9月30日出願の米国特許出願11/241,693号明細書「評判管理のためのシステムおよび方法 (Systems and Methods for Reputation Management)」に開示され、この出願の開示が、参照により本明細書に援用される。

【0026】

レビューエンジン106の構成要素のそれぞれが多数のコンピュータに分散されてもよいことを理解されたい。たとえば、レビューリポジトリ112は、どのレビューがM台のサーバのそれぞれに格納されるかを決定するために用いられる「モジュロM」関数などのマッピング関数を備えた、M台のサーバに配置してもよい。同様に、レビューサーバ110は、多数のサーバに分配されてもよく、レビューコレクタ114およびドキュメントリポジトリ116は、それぞれ、多数のコンピュータに分配されてもよい。しかしながら、説明の便宜上、レビューエンジン106の構成要素が、単一のコンピュータに実現されているかのように、これらの構成要素について説明する。

10

【0027】

図2は、本発明のいくつかの実施形態による、レビュー要約の要求を受信し、要求に回答するプロセスの流れ図である。上記のように、レビューエンジン106は、ユーザによってレビューエンジン106に投稿されたレビューと同様に、ドキュメントホスト104に投稿されたレビューを収集および格納する。ユーザは、クライアント102を通して、製品、サービスまたは提供者といった対象のレビュー情報をレビューエンジンに要求してもよい。たとえば、ユーザは、クライアント102に表示されたウェブページにおいて、

20

【0028】

クライアント102を介して、ユーザは、対象または対象のクラスのレビュー要約をレビューエンジン106に要求してもよい。レビューエンジン106は、対象のレビュー要約に対する要求をクライアント102から受信する(202)。レビューリポジトリ112に格納された、対象のレビューが、識別される(204)。識別されたレビューのサブセットが、選択される(206)。選択されたサブセットの内容を含む応答が生成される(208)。応答は、クライアント102に伝送される(210)。クライアント102は、応答を受信すると、ユーザに提示するために、ウェブブラウザなどのクライアントアプリケーションで応答を表示する。

30

【0029】

生成された応答は、ユーザへの提供および表示のためにクライアント102に伝送されるドキュメントである。応答ドキュメントには、対象のレビュー要約を含んでもよい。レビュー要約には、対象の全体的レーティングなどの情報を含んでもよいが、そのさらなる詳細は、図3に関連して下記で説明する。レビュー要約にはまた、利用可能な場合には、レビューソースによって与えられた対象の集合レーティングを含んでもよい。レビューソースによって対象に与えられる集合レーティングは、そのソースに投稿された対象のレビューに関連付けられたレーティングに基づいて、レビューソースによって決定されたレーティングである。集合レーティングがどのように決定されるかは、レビューソースごとに

40

【0030】

レビュー要約にはまた、レビューサンプルが含まれる。いくつかの実施形態において、レビューサンプルには、前記選択されたレビューのうちの少なくともいくつかの完全な内容を含んでもよい。テキストベースのレビューにとって、レビューの完全な内容は、レビ

50



ューの全テキストである。ビデオベースのレビューにとって、レビューの完全な内容は、レビューの完全なビデオクリップである。他のいくつかの実施形態において、レビューサンプルには、前記選択されたレビューのうちの少なくともいくつかにおける抜粋を含んでもよいが、そのさらなる詳細は、図6に関連して下記で説明する。しかしながら、いくつかの実施形態において、レビューサンプルには、いくつかの選択されたレビューの完全な内容、および他の選択されたレビューの抜粋の両方を含んでもよいことを理解されたい。レビューサンプルにはまた、完全な内容または抜粋がレビューサンプルに含まれているレビューソースへの1つ以上のリンクを含んでもよい。

【0031】

図3は本発明のいくつかの実施形態による、代表的なレビューを選択するためのプロセスの流れ図である。対象のレビュー要約に対する要求をユーザから受信すると、レビューエンジン106は、対象のレビューサンプルに含めるための多くのレビューを選択して、サンプルにおけるレビューが対象の全体的レーティングの代表であるようにすることができる。

【0032】

特定の対象のレビューおよび該レビューのソースが識別される(302)。レビューは、特定の対象に関連する全てのレビューに対してレビューリポジトリ112を探索することによって、レビューリポジトリ112から識別してもよい。前記識別されたレビューが、特定の対象のためのレビューのコーパス(corpus)を形成する。利用可能な場合には、対象の集合レーティングが、各識別されたソースから識別される(304)。各識別されたレビューソースについて、それぞれのソースにある、コーパスのレビュー数が識別される(306)。これは、単に、コーパスにおけるどれくらいの数のレビューが各ソースに含まれるかの計算である。

【0033】

全体的レーティングスコアが、対象に対して決定される(308)。全体的レーティングスコアは、レビューソースによって与えられた対象の集合レーティングの数学的な組み合わせであってもよい。いくつかの実施形態において、全体的レーティングスコアは、集合レーティングの加重平均である。重み付けは、各ソースに含まれる、コーパスにおけるレビューの数に基づいている。したがって、より多くのレビューがコーパスにあるソースからの集合レーティングは、加重平均では有利になる。全体的レーティングを計算する例示的な式は、

【数1】

$$OR = \frac{\sum_{i=1}^S r_i \log n_i}{\sum_{i=1}^S \log n_i}$$

であり、式中、ORは全体的レーティングであり、Sは、コーパスにおける少なくとも1つのレビュー(すなわち、対象の少なくとも1つのレビュー)および対象の集合レーティングを有するレビューソース数であり、 $r_i$ は、ソースiからの集合レーティングであり、 $n_i$ は、ソースiにある、コーパスのレビュー数である。レビューソースが、それぞれ、その集合レーティングのために異なる尺度および/または形式を用いる場合には、集合

10

20

30

40

50

レーティングは、最初に、全体的レーティングのために用いられる尺度／形式と同じ尺度および形式に変換および／または正規化される。いくつかの実施形態において、全体的レーティングは、1～5の数値レーティング尺度に基づいており、したがって、集合レーティングは、この尺度に変換および／または正規化される。しかしながら、全体的レーティングのために代替レーティング尺度を用いてもよいことを理解されたい。いくつかの実施形態において、上記の式に示すように、集合レーティングは、各レビューソースにある、コーパスのレビュー数の対数によって重み付けされる。対数は、底2、底10または底eなどの任意の適切な底であってもよい。他のいくつかの実施形態において、次の式に示すように、集合レーティングは、各レビューソースにある、コーパスのレビュー数によって重み付けしてもよい。

10

【数2】

$$OR = \frac{\sum_{i=1}^S r_i n_i}{\sum_{i=1}^S n_i}$$

20

【0034】

全体的レーティングを決定する際に、全体的レーティングが収まるレーティング範囲が識別される(310)。レーティング尺度は、2つ以上のレーティング範囲に分割してもよい。たとえば、1～5の尺度は、3つの範囲に分割してもよい。3.66以上5以下のレーティングは、対象に関する経験が全体として肯定的だったことを示してもよい。1以上2.33以下のレーティングは、対象に関する経験が全体として否定的だったことを示してもよい。2.34以上3.65以下のレーティングは、対象に関する経験が全体的にいろいろ混在していたことを示してもよい。別の例として、同じ1～5の尺度は、4つの範囲に分割してもよい。4.1以上5以下のレーティングは、非常によいレーティングを示してもよい。3.1以上4以下のレーティングは、よいレーティングを意味してもよい。2.1以上3以下のレーティングは、中程度のレーティングを意味してもよい。1以上2以下のレーティングは、悪いレーティングを意味してもよい。上記のレーティング範囲の例はあくまで例示であり、レーティング尺度を分割する代替方法を用いてもよいことを理解されたい。しかしながら、説明の便宜上、レーティング尺度が、3つの範囲すなわち高／肯定的範囲、低／否定的範囲および中間／混在範囲に分割されるかのように、図3に示すプロセスを説明する。

30

【0035】

全体的レーティングが低範囲(310-低)に入る場合には、低範囲のレーティングに関連する、コーパスのレビューが選択される(312)。レビューは、ソースごとに選択するか、または全体としてのコーパスから選択してもよい。レビューがソースごとに選択される場合には、低範囲におけるレーティングに関連する、第1の所定の数のレビューまで、各ソースから選択してもよい。レビューが全体としてのコーパスから選択される場合には、第2の所定の数のレビューまで、レビューソースに関係なく、コーパスから選択してもよい。

40

【0036】

全体的レーティングが中間範囲(310-中間)に入る場合には、高範囲のレーティングに関連する、コーパスのレビュー、および低範囲のレーティングに関連する、コーパスのレビューが、選択される(314)。換言すれば、選択されたレビューの中には、高範

50

囲のレーティングに関連するレビューおよび低範囲のレーティングに関連するレビューがある。代替実施形態では、中間範囲のレーティングに関連する、コーパスのレビューが、選択される。上記のように、レビューは、ソースごとか、または全体としてのコーパスから選択してもよい。

【0037】

全体的レーティングが高範囲(310 - 高)に入る場合には、高範囲のレーティングに関連する、コーパスのレビューが、選択される(316)。上記のように、レビューは、ソースごとか、または全体としてのレビューセットから選択してもよい。

【0038】

いくつかの実施形態では、追加選択基準を含んでもよい。たとえば、追加基準は、選択されるレビューが、冒瀆または性的に露骨な内容などの不快な内容を有さないということであってもよい。別の例として、追加基準は、選択されるレビューが、所定の閾値を超える、評判に基づいたメトリック値を有さなければならないということであってもよい。より一般的には、全体的レーティングが収まるレーティング範囲のレーティングに関連するレビューであって、かつゼロ以上の他の所定の基準を満たすレビューを選択してもよい。

10

【0039】

前記選択されたレビューの内容を含む応答が生成される(318)。前記生成された応答は、ユーザへの提供および表示のためにクライアント102に伝送されるドキュメントである。応答ドキュメントには、対象のレビュー要約が含まれる。レビュー要約には、対象の全体的レーティング、および任意に、レビューソースによって与えられた対象の集合レーティングなどの情報を含んでもよい。レビュー要約にはまた、レビューサンプルが含まれるが、このサンプルには、上記のように、前記選択されたレビューのうちの少なくともいくつか、またはそれらの抜粋が含まれる。

20

【0040】

図4は、本発明のいくつかの実施形態による、高品質レビューを選択するためのプロセスの流れ図である。対象のレビュー要約の要求をユーザから受信すると、レビューエンジン106は、レビューが高品質な内容を含むように、対象のレビューサンプルに含めるための多くのレビューを選択することができる。

【0041】

特定の対象のレビューおよび該レビューのソースが、識別される(402)。レビューは、特定の対象に関連する全てのレビューに対してレビューリポジトリ112を探索することによって、レビューリポジトリ112から識別してもよい。前記識別されたレビューは、対象のためのレビューのコーパスを形成する。いくつかの実施形態において、最初に識別されたレビューは、402かまたはプロセスのより後の段階でフィルタリングされ、不快な内容を含むあらゆるレビューを除去するようにする。

30

【0042】

各識別されたレビューのために、品質スコアが決定される(404)。品質スコアは、レビューの内容の品質の尺度である。品質スコアによって、レビューの品質に関して、レビューを互いに比較する基礎が提供される。品質スコアは、1つ以上の所定の要因に基づいてもよい。いくつかの実施形態において、所定の要因には、レビューの長さ、レビューにおける文の長さ、レビューの単語に関連する値、およびレビューの文法的品質(grammatical quality)が含まれる。各要因に基づき、レビューに対してサブスコアを決定してもよく、またサブスコアは、レビューの品質スコアを決定するために組み合わせられてもよい。しかしながら、追加および/または代替要因を含んでもよいことを理解されたい。

40

【0043】

レビューの文法的品質に関して、適切な文法および大文字使用(たとえば、実際には、全てを大文字としたものでない文、レビューを用いる)を有するレビューが有利である。したがって、「適切な」文法および大文字使用を伴ったレビューは、この要因に対して、より高いサブスコアを得る。貧弱な文法および不適切な大文字使用を伴ったレビューは、あまり読み易くない傾向がある。さらに、全てが大文字のレビューは、無作法であると見

50

なされることが多い。いくつかの実施形態において、レビューにおける文の検出は、レビューにおけるピリオドなどの文の区切り文字の検出に基づいてもよい。いくつかの実施形態において、レビューは、主語 - 動詞の一致、無終止文または不完全な部分がない (absence of run-on sentences or fragments) など、文法的品質の追加的な特徴にしたがって評価してもよい。いくつかの実施形態において、レビューの文法および大文字使用の評価は、文法チェッカの助けを借りて実行してもよいが、文法チェッカは、当該技術分野において周知であり、さらに説明する必要はない。

【0044】

レビューの長さに関して、長すぎず、短すぎないレビューが、有利である。短いレビュー（たとえば数語）は、情報価値がない傾向があり、長いレビュー（たとえば多くの段落）は、より短いレビューほど読み易くない傾向がある。いくつかの実施形態において、レビュー長さは、単語数に基づいてもよい。他のいくつかの実施形態において、レビュー長さは、文字数または文の数に基づいてもよい。レビュー長さのサブスコアは、そのレビュー長さと所定の「最適な」レビュー長さとの間の差に基づいてもよい。

【0045】

いくつかの実施形態において、レビューにおける文の長さもまた、考慮してよい。レビューエンジンは、極端に長いまたは短い文よりも「適度な」長さの文を好んでもよい。いくつかの実施形態において、レビューの文の長さサブスコアは、そのレビューにおける文の長さとの間の「最適な」文の長さとの間における差の平均に基づいてもよい。

【0046】

レビューにおける単語に関連する値に関して、高い値の単語を備えたレビューは、低い値の単語を備えたレビューより有利である。いくつかの実施形態において、単語値 (the word values) は、単語に関連する逆文献頻度 (the inverse document frequency) (IDF) 値に基づいている。高いIDF値を備えた単語は、一般に、より「価値がある」と見なされる。単語のIDFは、テキストの1セットのテキスト数に基づいており、その単語の少なくとも1回の出現を含むテキストのセットのテキスト数で割られたものである。レビューエンジン106は、レビューリポジトリ112のレビュー全体にわたってIDF値を決定し、それらの値を1つ以上の表に格納してもよい。いくつかの実施形態において、IDF値の表が、各タイプのレビューのために生成される。たとえば、IDF値の表が、全ての製品レビューのために生成される。表が、全ての製品提供者レビューのために生成される、等である。すなわち、製品レビューのためにIDF値の表を決定するために用いられるテキストのセットは、レビューリポジトリ112における全ての製品レビューである。製品提供者レビューのためにIDF値の表を決定するために用いられるテキストセットは、レビューリポジトリ112における全ての製品提供者レビューである、等である。各対象タイプは、それ自身のIDF値の表を有する。なぜなら、ある対象タイプのレビューでは価値がある単語が、別の対象タイプのレビューでは、同じほどには価値がない可能性があるからである。

【0047】

任意の識別されたレビューに関して、レビューにおける各個別の単語の頻度が、決定され、その単語のためのIDFで乗算される。レビューのための単語値サブスコアは、

【数3】

$$WV_R = \sum_{w \in R} f_{w,R} \log IDF_w$$

10

20

30

40

50

であり、式中、 $WV_R$  は、レビュー R のための単語値サブスコアであり、 $f_{w, R}$  は、レビュー R における個別の単語  $w$  の出現回数（用語頻度または「TF」）であり、 $\log IDF_w$  は、単語  $w$  の IDF 値の対数である。単語  $w$  の IDF 値は、レビューの対象タイプに適した IDF 値の表から取られる。たとえば、レビュー R の対象が製品である場合には、 $IDF_w$  値は、製品レビューのための IDF 値の表から取られる。

【0048】

他のいくつかの実施形態において、単語値は、レビュー文脈において価値があると考えられる単語の所定の辞書に基づいている。別個の辞書を異なる対象タイプのために定義してもよい。なぜなら、異なる単語が、異なる対象タイプに関連するレビューで用いる価値があり得るからである。たとえば、対象が製品である場合のレビュー用の価値のある単語の辞書があってもよく、対象が提供者である場合のレビュー用の価値のある単語の別の辞書があってもよい。これらの実施形態において、単語値サブスコアは、所定の辞書中の単語のうちいくつか、それぞれのレビューに含まれているかの計算に基づいてもよい。

10

【0049】

レビューエンジン 106 は、各識別されたレビューを各所定の要因に基づいて評価し、その評価に基づいて各要因のサブスコアを決定する。要因のそれぞれに対するサブスコアは、下記の例示的な式を用いて、品質スコアに組み合わせてもよい。

【数 4】

$$Q = \sum_{j=1}^F q_j \text{重さ}_j$$

20

式中、 $Q$  は、レビューのための品質スコアであり、 $F$  は、品質スコアに關与する要因の数であり、 $q_j$  は、要因  $j$  のためのサブスコアであり、 $\text{重さ}_j$  は、要因  $j$  のための重み付けである。いくつかの実施形態において、重み付けは、全て 1 に等しいが、この場合には、品質スコア  $Q$  は、要因のためのスコアの合計である。他のいくつかの実施形態において、重み付けは、各要因に対して別々に定義してもよい。一般に、重み付けは、品質スコアに対する各要因の重要性、および要因がレビューの品質に肯定的または否定的に寄与するかどうかに基づいて定義してもよい。

30

【0050】

いくつかの実施形態において、レビューの古さ（age）を、レビューの品質スコアの要因と見なしてもよい。一般に、より新しいレビューが有利である。なぜなら、それらは、より遠く離れた過去の経験より重要な、レビュー対象に対する最近の経験をより反映しているからである。品質スコアを増加させるボーナス点を、レビューの古さに基づいたレビューの品質スコアに適用してもよい。たとえば、1日の古さのレビューは、（加算または乗算のいずれかによって）その品質スコアを増加してもよく、一方で1年の古さのレビューは、ボーナスを得ない。

40

【0051】

品質スコアに基づいて、レビューが選択される（406）。最高の品質スコアを備えたレビューが選択される。レビューは、ソースごとにかまたは全体としてのコーパスから選択してもよい。レビューがソースごとに選択される場合には、各ソースに対して多くの最高得点レビューが選択される。たとえば、ソースごとに10の最高得点レビューを選択してもよい。いくつかの実施形態において、選択は、品質スコアによってレビューをソートすることによって実行され、レビューは、所望のレビュー数が選択されるまで、最高得点レビューから取られる。

【0052】

50

いくつかの実施形態において、所定の内容基準はまた、レビューを選択するための追加基準であってもよい。所定の基準を満たす内容に関して、基準は、冒読および性的に露骨な内容など、ユーザを怒らせる可能性があるレビュー内容を備えたレビューの優先度を下げる (disfavor) ように定義してもよい。かかる単語および句は、対象の理解にはほとんどまたは全く寄与せず、レビューを読んでいるユーザを不快にする可能性がある。所定の基準を満たす内容に関するレビューの評価は、無礼または不快な内容に共通に関連する内容の辞書を定義し、かつレビューにおける内容を辞書と照合することによって実行してもよい。冒読または性的に露骨な言葉などの不快な内容を有するレビューは、選択のための考慮から除外される。所定の内容基準を満たす内容に関するレビュー内容の評価は、スコア決定中 (404) か、またはレビュー選択時 (406) に行ってもよい。評価がいつ実行されるかは、設計上の選択の問題である。

10

## 【0053】

いくつかの実施形態において、レーティングスコア基準は、レビュー選択のための追加基準であってもよい。たとえば、上記したような、代表的なレビューを選択するプロセスを目下のプロセスと組み合わせて、対象の全体的レーティングを代表する高品質レビューを選択するようにしてもよい。したがって、全体的レーティングが収まるレーティング範囲のレーティングに関連するレビューであって、高品質スコアを有するレビューが、選択可能である。

## 【0054】

上記の追加基準があくまで例示であること、ならびに上記の基準および他の基準の任意の組み合わせを、レビュー選択のために追加的に考慮してもよいことを理解されたい。より一般的には、レビューエンジンは、ゼロ以上の他の所定の基準を満たす (品質スコアの観点における) 最高得点レビューを選択してもよい。

20

## 【0055】

前記選択されたレビューを含む応答が生成される (408)。前記生成された応答は、ユーザへ提供および表示するためにクライアント102に伝送されるドキュメントである。応答ドキュメントには、対象のレビュー要約が含まれる。レビュー要約には、対象の全体的レーティング、および任意に、レビューソースによって与えられた対象の集合レーティングなどの情報を含んでもよい。レビュー要約にはまたレビューサンプルが含まれるが、このサンプルには、図2に関連して上記したように、前記選択されたレビューの内容が含まれる。

30

## 【0056】

図5は、本発明のいくつかの実施形態による、レビューをクラスタリングするための、およびクラスタからレビューを選択するための、プロセスの流れ図である。特定の対象のレビューが識別される (502)。レビューは、特定の対象に関連する全てのレビューに対してレビューリポジトリ112を検索することによって、レビューリポジトリ112から識別してもよい。前記識別されたレビューは、対象のレビューコーパスを形成する。

## 【0057】

レビューの単語値ベクトルが生成される (504)。単語値ベクトルには用語頻度 - レビューにおける単語のための逆文献頻度値が含まれる。用語頻度 - 逆文献頻度 (「TF-IDF」または「TFIDF」としてもまた知られている) は、ドキュメントにおける、またはこれらの実施形態の場合にはレビューにおける、単語の重要性を評価するための技術である。レビューに関連する単語の値は、その単語がレビューに現われる回数と共に増加するが、しかしそれは、その単語を含むレビューのコーパスにおけるレビュー数によって相殺される。識別されたレビューのコーパスの任意のレビューのために、単語値のベクトルを生成してもよい。たとえば、レビューRは、重み付けベクトル、

40

$$R = [v_1 \quad v_2 \quad v_3 \quad \dots \quad v_n]$$

を有してもよく、式中、 $v_1 \sim v_n$  は、レビューのコーパスにおける別個の単語全ての、レビューTに関連する単語値である。いくつかの実施形態において、単語およびその関連する変化形が、一緒にカウントされる。たとえば、ある動詞の動詞時制は、単にスペルが

50

異なっているだけの可能性があるので、別個の単語としてではなく、同じ動詞の出現としてカウントしてもよい。

【0058】

レビューRに関連する単語wの値は、例示的な式、

$$v_{w, R} = f_{w, R} \log IDF_w$$

によって決定してもよく、式中、 $v_{w, R}$  は、レビューRに関連する単語wの値であり、 $f_{w, R}$  は、レビューR内における単語wの出現回数（用語頻度）であり、 $\log IDF_w$  は、上記のように、単語wのためのIDF値の対数である。レビューRが単語wを有さない（ $f_{w, R} = 0$ ）場合には、単語値 $v_{w, R}$  は0である。単語値 $v_{w, R}$  は、決して負になり得ない。なぜなら、 $f_{w, R} \geq 0$ （出現回数は決して負ではない）および $\log IDF_w \geq 0$ であるからである。

10

【0059】

コーパスにおける各レビューのための単語値ベクトルが生成されると、コーパスにおけるレビューは、単語値ベクトルに基づいてクラスタに組織される（506）。単語値ベクトルは、ベクトル空間に埋め込まれるが、そこでは、各単語値ベクトルは、そのベクトル空間における「ポイント」である。「ポイント」は、クラスタリングアルゴリズムを用いて、1つ以上のクラスタにグループ化してもよい。1つの例示的なクラスタリングアルゴリズムは、K平均クラスタリングアルゴリズムである。K平均クラスタリングアルゴリズムは、当該技術分野において周知である。しかしながら、開示される実施形態の理解を容易にするために、K平均アルゴリズムを下記で説明する。

20

【0060】

次の擬似コードが、K平均アルゴリズムの基礎的なステップを示す。

k個のクラスタに関するk個の重心をランダムに生成する。

各ベクトルをk個のクラスタのうちの1つに割り当てる。

終了条件が満たされるまで以下を繰り返す：クラスタ重心を決定し直す。各ベクトルを1つのクラスタに割り当てる。

【0061】

K平均アルゴリズムにおいて、任意の数kが事前に定義される。いくつかの実施形態において、kは、2～16の値であり、一方で他のいくつかの実施形態では、kは、2～50の値である。単語値ベクトルのベクトル空間におけるKランダムベクトルが、生成される。k個のランダムベクトルは、ベクトル空間のための初期重心である。各初期重心は、クラスタの「中心」を表わす。換言すれば、k個の初期クラスタおよびそれらの中心が、任意に定義される。各単語値ベクトルは、それぞれの単語値ベクトルと各重心との間の類似度（距離）に基づいて、kクラスタのうちの1つに割り当てられる。単語値ベクトルは、それが最も類似した（最短距離の）重心に割り当てられる。

30

【0062】

いくつかの実施形態において、単語値ベクトルと重心との間の類似度（距離）は、コサイン類似度（「コサイン距離」としてもまた知られている）である。すなわち、

【数5】

40

$$\cos \theta = \frac{X \cdot Y}{\|X\| \times \|Y\|}$$

であり、式中、 $X \cdot Y$  は、ベクトルXおよびYのドット積であり、 $\|X\| \times \|Y\|$  は、ベクトルXの長さ×ベクトルYの長さであり、 $\cos$  は、コサイン類似度である。ベクトル

50

ルXおよびYが正確に同じである場合には、コサイン類似度の値は、1である。これらの実施形態においてコサイン類似度の値域は、0以上1以下である（コサイン類似度は、決して負になり得ない。なぜなら、単語値が決して負になり得ないからである）。したがって、1により近いコサイン類似度を備えたレビューは、より類似しており（より短い距離）、一方で0に近いコサイン類似度を備えたレビューは、それほど類似していない（より長い距離）。他のいくつかの実施形態では、距離または類似度を決定する代替方法を用いてもよい。

【0063】

いくつかの実施形態において、多くの所定の規準レビューを、初期重心として用いてもよい。規準レビューは、対象の特定の側面に関してコメントするレビューの典型として働く所定のレビューのセットである。規準レビューのセットは、レビューのコーパスの対象が何であるかに依存して、異なってもよい。たとえば、対象が製品である規準レビューのセット（使い易さおよび性能などの側面の規準レビューを含んでいる可能性がある）は、対象が製品提供者である規準レビューのセット（顧客サービスおよび発送の適時性などの側面の規準レビューを含んでいる可能性がある）と異なってもよい。

10

【0064】

単語値ベクトルがk個のクラスタに割り当てられた後、k個のクラスタの重心が新たに決定される。すなわち、重心は、各クラスタのために再決定される。クラスタの重心は、クラスタ（初期重心を含まない。初期重心は、初期クラスタ割り当てだけに関連する）における単語値ベクトルの「平均」を取ることで決定してもよい。重心Cを決定する式は、

20

【数6】

$$C = \frac{\sum_{i=1}^{CS} V_i}{CS}$$

30

であり、式中、CSは、クラスタのサイズ（クラスタにおける単語値ベクトルの数）であり、 $V_i$ は、クラスタにおける単語値ベクトルの正規化された（単位長さのベクトルに変換された）ベクトルである。

【0065】

新しい重心が決定されると、単語ベクトル値は、今度は新しい重心に対する類似度に基づいて、クラスタに再割り当てされる。単語ベクトルは、それが最も類似した重心に割り当てられる。各単語ベクトルがクラスタに再割り当てされた後で、重心の再決定および単語ベクトルの再割り当ての繰り返しは、反復される。終了条件が満たされるまで、繰り返しは反復される。いくつかの実施形態において、終了条件は、収束基準が満たされたときである。収束基準は、繰り返しの完了後に、異なるクラスタに再割り当てされる単語ベクトルがないということであってもよい。他のいくつかの実施形態において、終了条件は、所定数の繰り返しが実行されたということである。

40

【0066】

階層クラスタリング、ファジィc平均アルゴリズムおよびその他などの、クラスタリングの代替方法を用いてもよいことを理解されたい。

【0067】

レビューをクラスタにグループ化すると、レビュークラスタのサイズが識別される（508）。これは、単に、各クラスタにおける（重心を含まない、単語値ベクトルによって

50



表わされた)レビュー数である。

【0068】

レビューが、各クラスタから選択される(510)。いくつかの実施形態において、レビューは、クラスタサイズに比例して各クラスタから選択される。所定の合計数のレビューが、レビューのコーパスのサンプルとして働くために、レビューコーパスから選択される。サンプルにおけるレビューは、クラスタのサイズに比例してクラスタから選択される。サンプルは、より小さなクラスタより、より大きなクラスタから、より多くのレビューが選択されることになる。いくつかの実施形態において、極端に小さなクラスタ(たとえば、所定のレビュー数未満、またはコーパスにおける合計レビュー数の所定の割合未満)は、レビュー選択から除外してもよい。そのクラスタからのレビューは、サンプルに含めるためには選択されない。クラスタが除外された場合には、サンプルにおけるレビュー数が所定の合計数に達するように、1つ以上のレビューを他のクラスタから選択してもよい。

10

【0069】

いくつかの実施形態において、レビューは、追加的な所定の基準に基づいてクラスタから選択してもよい。たとえば、図4に関連して上記したように、レビューは、レビューの品質に基づいてクラスタから選択してもよい。高品質のレビューは、一般に、より情報価値があり、低品質のレビューより読み易い。したがって、たとえば、クラスタから10のレビューが選択されることになる場合には、追加品質基準を用いて、そのクラスタの10の最高品質のレビューを選択してもよい。別の例として、図3に関連して上記した選択プロセスのように、レビューは、レビューに関連するレーティングに基づいてクラスタから選択してもよい。より一般的には、クラスタが、クラスタサイズに比例する数のレビューをレビューサンプルに与える限り、そのクラスタからのレビューは、ゼロ以上の所定の基準に基づいて選択してもよい。

20

【0070】

前記選択されたレビューを含む応答が生成される(512)。前記生成された応答は、ユーザへの提供および表示のためにクライアント102に伝送されるドキュメントである。応答ドキュメントには、対象のレビュー要約が含まれる。レビュー要約には、対象の全体的レーティング、および任意にレビューソースによって与えられた対象の集合レーティングなどの情報を含んでもよい。レビュー要約にはまた、レビューサンプルが含まれるが、このサンプルには、図2に関連して上記したように、前記選択されたレビューの内容が含まれる。

30

【0071】

レビューをクラスタリングし、かつクラスタからレビューを選択することによって、レビューの話題の中心を代表するレビューサンプルが選択される。クラスタリングは、レビューエンジンが、対象の特定の側面に焦点を当てたレビューを識別するのを支援する。レビューが焦点を当てた側面によってレビューを(クラスタに)分類し、かつレビューサンプルに含めるためにクラスタからレビューを選択することによって、ユーザは、レビューサンプルを示されるとすぐに、対象のどの側面が特に注目に値するか、またはその対象を経験した他のユーザにとって、対象のどの側面が特に重要だったかを、よりよく理解することができる。

40

【0072】

図6は、本発明のいくつかの実施形態による、レビュー内の高品質の内容から抜粋を生成するためのプロセスの流れ図である。時間を節約するために、ユーザは、レビューの完全な内容よりレビューにおける部分だけを読むのを好む場合がある。レビューエンジンは、レビュー抜粋としてレビューサンプルに含めるために、レビュー内の特定の内容を選択してもよい。

【0073】

レビューが識別される(602)。前記識別されたレビューは、パーティションに分割される(604)。いくつかの実施形態において、パーティションは、レビューの文であ

50

る。すなわち、レビューの各文が、レビューのパーティションである。レビューの文は、ピリオドなどの文区切り文字に基づいて識別してもよい。レビューが1つの文だけを有する場合など、レビューが1つのパーティションだけを有することもあり得る。説明の便宜上、図5のプロセスは、レビューのパーティションがレビューの文であるかのように、下記で説明される。しかしながら、レビューを分割する代替方法（所定の整数Zの単語のパーティションなど）を用いてもよいことを理解されたい。

【0074】

レビューの各文のために品質スコアが決定される（606）。図4に関連して上記したように、レビュー文の品質スコアは、レビューの品質スコアに類似している。文品質スコアによって、文の品質に関して、1レビューにおける文の相対的な順序付けのための基礎が提供される。品質スコアは、1つ以上の要因に基づいてもよい。要因のそれぞれに基づいて、サブスコアを決定してもよい。サブスコアは、上記で図3に関連して説明した式に類似した加重和の式を用いて、文の品質スコアへと組み合わせてもよい。いくつかの実施形態において、所定の要因には、文の長さ、文の単語に関連する値、およびレビュー内の文の位置が含まれる。

10

【0075】

レビューの文の長さに関連して、長すぎず短すぎない文（すなわち「適度な長さ」の文）は、有利である。極端に短い文は、あまり情報を含んでいない可能性があり、極端に長い文は、読み難い可能性がある。いくつかの実施形態において、文の長さに基づいたサブスコアは、所定の「最適な」文の長さからの、レビューにおける文の偏差に基づいてもよい。文の長さは、単語数または文字数に基づいてもよい。

20

【0076】

文の単語に関連する値に関して、高い値の単語を備えた文は、低い値の単語を備えた文より有利である。いくつかの実施形態において、単語値は、図4に関連して上記したように、レビューをスコアリングする際に用いられる単語値の要因に類似した、単語に関連する逆文献頻度（IDF）値に基づいている。1文について、該文における各別個の単語の頻度が決定され、その単語用のIDFにより乗算される。レビューのための単語値サブスコアは、

【数7】

30

$$WV_P = \sum_{w \in P} f_{w,P} \log IDF_w$$

であり、式中、 $WV_P$  は、文Pのための単語値サブスコアであり、 $f_{w,P}$  は、文Pにおける単語wの出現回数であり、 $\log IDF_w$  は、単語wのためのIDF値の対数である。

40

【0077】

他のいくつかの実施形態において、単語値は、レビュー文脈において価値があると考えられる単語の所定の辞書に基づいている。別個の辞書を異なる対象タイプのために定義してもよい。なぜなら、異なる単語が、異なる対象タイプに関連するレビューで用いる価値がある場合があるからである。たとえば、対象が製品である場合のレビューに対して価値のある単語の辞書があってもよく、対象が提供者である場合のレビューに対して価値のある単語の別の辞書があってもよい。これらの実施形態において、単語値サブスコアは、所定の辞書の単語のうちいくつか、それぞれの文に含まれているかの計算に基づいてもよい。

【0078】

50

レビュー内の文の位置に関して、いくつかの実施形態では、レビューエンジンは、レビューの始めに出現する文を優先してもよい。したがって、位置に基づくサブスコアは、レビューの文の数に対して正規化された、レビューの文の位置に基づいてもよい。たとえば、10の文を備えたレビューの4番目の文について、その文のための位置サブスコアは、 $4 / 10 = 0.4$ であってよい。

【0079】

文のサブスコアを決定すると、サブスコアは、図4に関連して上記した式と類似の式を用いて、文の品質スコアへと数学的に組み合わせてもよい。

【0080】

レビュー文の組み合わせが識別される(608)。各組み合わせには、所定の長さ基準を満たす、レビューの1つ以上の連続する文が含まれる。いくつかの実施形態において、長さ基準は、組み合わせの長さが、所定の最大抜粋長さ(単語数または文字数に基づいてもよい)に等しいか、または組み合わせにおける最後の文の一部だけ最大抜粋長さを超えるということである。組み合わせを識別するための例示的なアルゴリズムを、下記に擬似コードで示す。

```

For each sentence i in the review :
    integer j = i
    combination i = sentence j
    while (length(combination i) < max_snippet_length)
        combination i = combination i + sentence(++j)

```

上記の擬似コードに示すように、組み合わせは、レビューにおける一文として出発し、組み合わせの長さを最大抜粋長さ以上にする最初の文を含むまで、続く文が組み合わせに追加される。したがって、組み合わせは、レビューにおけるできるだけ多くの連続する文の連結であるが、これは、組み合わせの長さが、最大抜粋長さと、組み合わせに追加された場合に、組み合わせの長さを最大抜粋長さ以上にするおそらく1つの追加文との、合計を超えないように行われる。

【0081】

他のいくつかの実施形態において、追加される文のどれくらいが最大抜粋長さ内にあるか、すなわち、追加文を収容するために、組み合わせにどれくらいの「スペース」が残っているかもまた考慮するように、アルゴリズムを洗練してもよい。たとえば、組み合わせが、1つまたは2つの単語だけ最大抜粋長さに達しない場合には、追加文を組み合わせに追加しないことが、より価値があり得る。

【0082】

最高の組み合わせ品質スコアを備えた組み合わせが選択される(610)。いくつかの実施形態において、組み合わせのための組み合わせ品質スコアは、組み合わせ内における文の品質スコアの単純な合計である。他のいくつかの実施形態において、組み合わせ品質スコアは、組み合わせ内における文の品質スコアの加重和、単純平均または加重平均であってよい。

【0083】

前記選択された組み合わせを用いて、抜粋が生成される(612)。抜粋には、最大抜粋長さまで、前記選択された組み合わせが含まれる。組み合わせが最大抜粋長さを超えた場合には、組み合わせ長さが最大抜粋長さに等しくなるまで、内容が、組み合わせの最後から切り捨てられる。いくつかの実施形態において、最大抜粋長さまで切り捨てた後に、組み合わせにおける最後の文の小部分(たとえば1つまたは2つの単語)だけが残る場合には、組み合わせは、最大抜粋長さより短くなるように切り捨ててもよい。換言すれば、最大抜粋長さまで組み合わせを切り捨てた後に、その文の少数の単語だけが残る場合には、組み合わせの最後の文を削除することによって切り捨てること、より価値があり得る。

。

10

20

30

40

50

## 【 0 0 8 4 】

前記抜粋を含む応答が生成される(614)。前記生成された応答は、ユーザへの提供および表示のためにクライアント102に伝送されるドキュメントである。応答ドキュメントには、対象のレビュー要約が含まれる。レビュー要約には、対象の全体的レーティング、および任意にレビューソースによって与えられた対象の集合レーティングなどの情報を含んでもよい。レビュー要約にはまた、レビューサンプルが含まれるが、このサンプルには、図2に関連して上記したように、前記選択されたレビューの内容が含まれる。

## 【 0 0 8 5 】

レビューエンジン106は、クライアント102へ伝送するために、そのレビューリポジトリからレビューを選択し、前記選択されたレビューの内容(完全なレビューおよび/または抜粋など)を含む応答を生成する。図3、4および5は、サンプル用のレビューを選択するための3つのプロセスを示す。図6は、レビューの抜粋を生成するためのプロセスを示すが、このレビューは、図3、4および/または5のプロセスで選択されたレビューであってもよい。上記のプロセスを組み合わせてもよいことを理解されたい。たとえば、レビューエンジン106は、全体的スコアが収まるレーティング範囲に対応し、かつ高い品質スコアを有する多くのレビューを選択してもよい。別の例として、レビューエンジン106は、対象のレビューをクラスタリングし、全体的スコアが収まるレーティング範囲に対応しかつ高い品質スコアを有する多くのレビューを、クラスタサイズに比例して各クラスタから選択してもよい。これらの選択されたレビューの抜粋が生成され、前記抜粋を含む応答が生成される。より一般的には、レビューを1つ以上の所定の基準に基づいて選択してもよく、これらのレビューの抜粋を生成して、クライアント102に送信される応答に含めてもよい。

## 【 0 0 8 6 】

図7は、本発明のいくつかの実施形態による、レビュー処理システム700を示すブロック図である。システム700には、典型的には、1つ以上の処理装置(CPU)702、1つ以上のネットワークまたは他の通信インタフェース710、メモリ712、およびこれらの構成要素を相互接続するための1つ以上の通信バス714が含まれる。任意に、システム700には、表示装置706およびキーボード/マウス708を含むユーザインタフェース704を含んでもよい。メモリ712には、DRAM、SRAM、DDR-RAMまたは他のランダムアクセス固体メモリ装置などの高速ランダムアクセスメモリが含まれ、また1つ以上の磁気ディスク記憶装置、光ディスク記憶装置、フラッシュメモリ装置、または他の不揮発性ソリッドステート記憶装置などの不揮発性メモリを含んでもよい。任意に、メモリ712には、CPU702から遠隔に位置する1つ以上の記憶装置を含んでもよい。いくつかの実施形態において、メモリ712は、下記のプログラム、モジュールおよびデータ構造、またはそれらのサブセットを格納する。すなわち、

- ・ 様々な基本システムサービスを扱うための、およびハードウェア依存タスクを実行するための手順を含むオペレーティングシステム716と、
- ・ インターネット、他のワイドエリアネットワーク、ローカルエリアネットワーク、メトロポリタンエリアネットワークなどの1つ以上の通信ネットワークインタフェース710(有線または無線)を介して、レビュー処理システム700を他のコンピュータに接続するために用いられるネットワーク通信モジュール718と、
- ・ レビュー記憶システムとインタフェースするレビュー記憶インタフェース720と、
- ・ レビューソースを識別するソース識別モジュール722と、
- ・ レビューソースからのレビューおよび関連するレーティングを識別するレビュー識別モジュール724と、
- ・ 対象の全体的レーティングを決定し、かつその全体的レーティングがどのレーティング範囲に収まるかを決定する全体的レーティングモジュール726と、
- ・ レビューの品質スコアを決定するレビュー品質スコアリングモジュール728と、
- ・ レビューをクラスタに組織するレビュークラスタリングモジュール730と、

10

20

30

40

50

- ・ レビューをパーティションに分割し、パーティションの品質スコアを決定し、パーティションの組み合わせを識別し、かつ最高組み合わせ品質スコアを備えた組み合わせを選択するレビューパーティションモジュール732と、

- ・ 1つ以上の所定の基準に基づいてレビューを選択するレビュー選択モジュール734と、

- ・ 不快な内容など、所定の内容基準を満たす内容に対して、レビューおよびレビューパーティションを評価する内容フィルタ736と、

- ・ レビューおよび/またはレビュー抜粋を含む応答を生成する応答生成モジュール738と、

である。

#### 【0087】

システム700にはまた、レビュー記憶システム740が含まれる。レビュー記憶システム740は、レビューおよび関連するレーティングを格納する。レビュー記憶システム740には、レビューの抜粋を生成する抜粋生成器742が含まれる。いくつかの実施形態において、抜粋生成器742は、レビュー記憶システム740ではなく、メモリ712に位置してもよい。

#### 【0088】

上記で識別した要素のそれぞれは、前に言及したメモリ装置の1つ以上に格納してもよく、また上記の機能を実行するための命令セットに対応する。上記で識別したモジュールまたはプログラム(すなわち命令セット)は、別個のソフトウェアプログラム、手順またはモジュールとして実行する必要はなく、したがって、これらのモジュールの様々なサブセットは、様々な実施形態において、組み合わせるか、さもなければ再編してもよい。いくつかの実施形態において、メモリ712は、上で識別したモジュールおよびデータ構造のサブセットを格納してもよい。さらに、メモリ712は、上記しなかった追加モジュールおよびデータ構造を格納してもよい。

#### 【0089】

図7は「レビュー処理システム」を示すが、図7は、本明細書で説明する実施形態の構造的概略図としてよりも、サーバのセットに存在し得る様々な特徴の機能的説明として意図されている。実際には、当業者が理解されるように、別々に示したアイテムは結合させることができ、いくつかのアイテムは分離することができる。たとえば、図7に別々に示したいくつかのアイテムは、単一のサーバに実現してもよく、また単一のアイテムは、1つ以上のサーバで実現してもよい。レビュー処理システムを実現するために用いられるサーバの実際の数、およびそれらの中で機能がどのように割り当てられるかは、実装ごとに変化し、また平均使用期間中に加えてピーク使用期間中にシステムが処理しなければならないデータトラフィック量に部分的に依存してもよい。

#### 【0090】

上記の説明が、その適用において、純粋なテキストの、すなわち文字列からなるレビューに限定されないことを理解されたい。この説明は、オーディオ、ビデオまたは他のメディア形態を含むレビューに適合させることができる。たとえば、オーディオ(オーディオだけのレビューまたはオーディオトラックを備えたビデオレビューなど)を含むレビューに関して、オーディオは、音声テキスト変換を用いてテキストに変換してもよいが、この変換は、当該技術分野において周知である。変換されたテキストは、上記の選択および抜粋生成プロセス用の「レビュー」として用いてもよい。オーディオまたはビデオレビューの抜粋は、レビューの変換されたテキストに基づいて抜粋のために選択された単語を備えた音声有する、オーディオまたはビデオにおける部分ということになる。レビュー品質が、オーディオ/ビデオレビューを選択するための基準である場合には、文法的な品質要因は、媒体に適合させてもよい。たとえば、レビューの内容が、テキストではなく口頭である場合には、大文字使用は、それほど関係なく、したがって無視することができる。

#### 【0091】

説明目的のために、前述の記載は、特定の実施形態に関連して説明した。しかしながら

10

20

30

40

50

、上記の実例的な説明は、包括的であるようにも、開示された形態に本発明を厳密に限定するようにも意図されていない。上記の教示を考慮すれば、多くの修正および変更が可能である。実施形態は、本発明の原理およびその実例的な用途を最もよく説明し、それによって、当業者が、本発明および考えられた特定の使用方法に適した様々な修正を含む様々な実施形態を最もよく利用できるようにするために、選択され説明された。

【図面の簡単な説明】

【0092】

図面の簡単な説明

【図1】本発明のいくつかの実施形態に基づくネットワークを示す。

【図2】本発明のいくつかの実施形態に基づく、レビュー要約の要求を受信し、それに  
10 応答するプロセスの流れ図である。

【図3】本発明のいくつかの実施形態に基づく、代表的なレビューを選択するためのプロセスの流れ図である。

【図4】本発明のいくつかの実施形態に基づく、高品質レビューを選択するためのプロセスの流れ図である。

【図5】本発明のいくつかの実施形態に基づく、レビューをクラスタリングし、クラスタからレビューを選択するためのプロセスの流れ図である。

【図6】本発明のいくつかの実施形態に基づく、レビュー内の高品質な内容から抜粋を生成するためのプロセスの流れ図である。

【図7】本発明のいくつかの実施形態に基づく、レビューを処理するためのシステムを示  
20 す。

【図1】

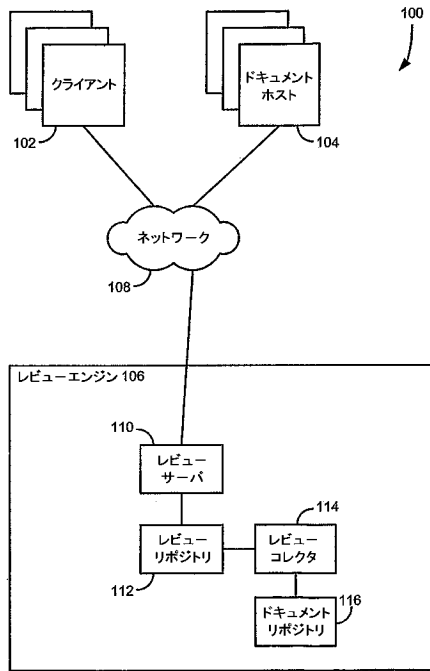


図 1

【図2】

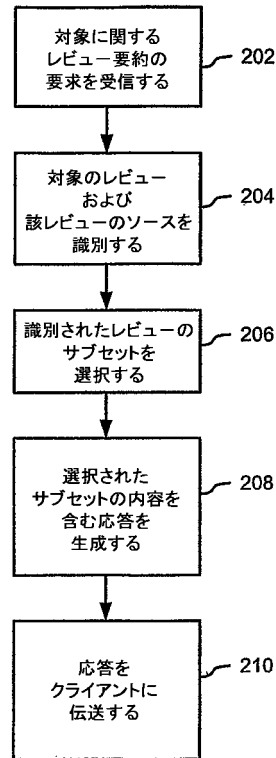


図 2

【 図 3 】

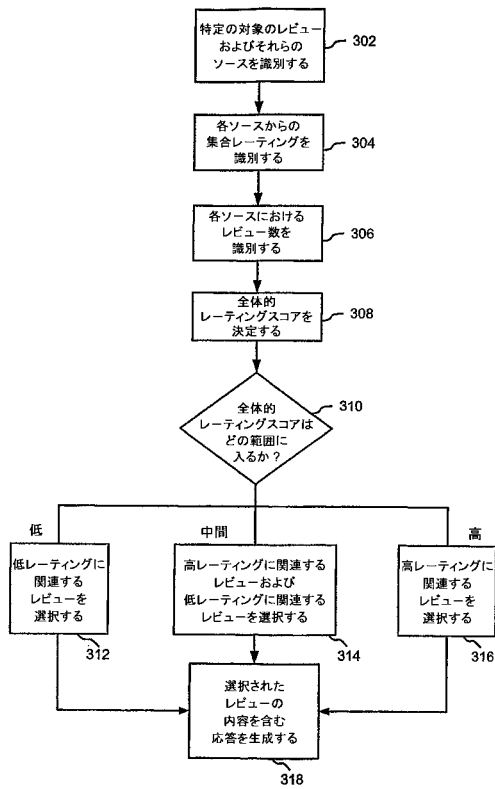


図 3

【 図 4 】

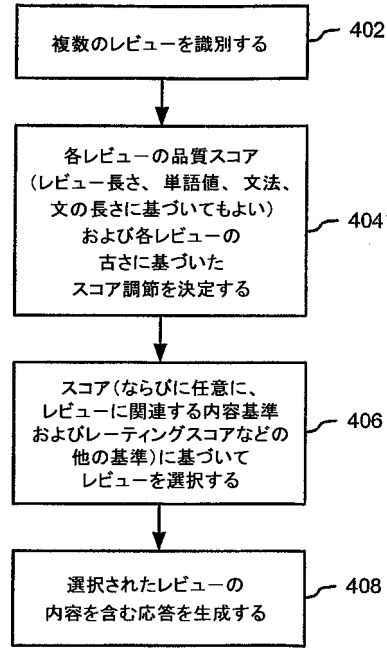


図 4

【 図 5 】

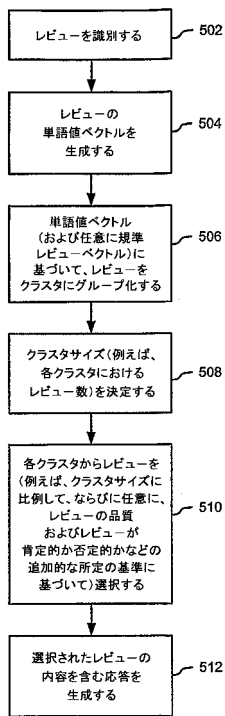


図 5

【 図 6 】

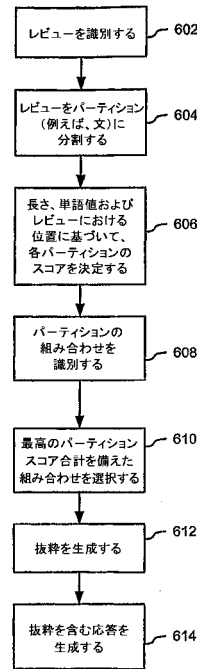


図 6

【図7】

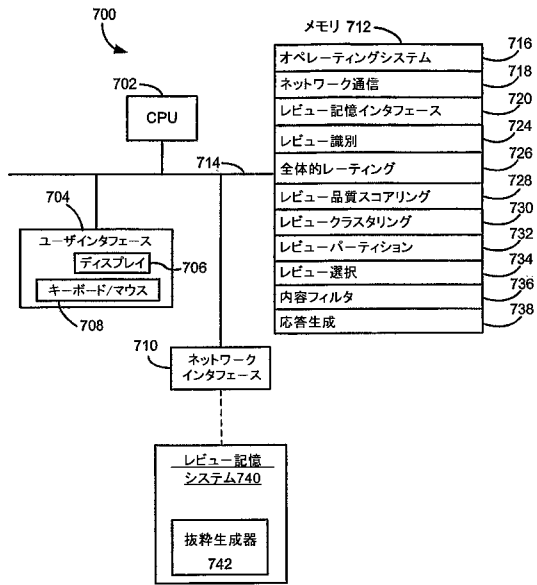


図7



---

フロントページの続き

- (72)発明者 デイブ, クシャル, ビー.  
アメリカ合衆国, ニューヨーク州 10011, ニューヨーク, ウェスト 17番 ストリート  
248, アpartment 201
- (72)発明者 ヒルトン, ジェレミー, エー.  
アメリカ合衆国, ペンシルベニア州 18042, イーストン, パーク ストリート 610

審査官 成瀬 博之

- (56)参考文献 国際公開第2004/025516(WO, A1)  
立石健二 他2名, インターネットからの評判情報検索, 人工知能学会誌, 日本, 社団法人人工  
知能学会, 2004年 5月 1日, 第19巻第3号, 317-323頁

- (58)調査した分野(Int.Cl., DB名)  
G06F 17/20 - 17/30