

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3622602号
(P3622602)

(45) 発行日 平成17年2月23日(2005.2.23)

(24) 登録日 平成16年12月3日(2004.12.3)

(51) Int. Cl.⁷

G06F 17/30

F I

G06F 17/30 340A

G06F 17/30 170A

G06F 17/30 350C

請求項の数 7 (全 10 頁)

<p>(21) 出願番号 特願平11-325389 (22) 出願日 平成11年11月16日(1999.11.16) (65) 公開番号 特開2001-142899(P2001-142899A) (43) 公開日 平成13年5月25日(2001.5.25) 審査請求日 平成13年11月21日(2001.11.21)</p>	<p>(73) 特許権者 000004226 日本電信電話株式会社 東京都千代田区大手町二丁目3番1号 (74) 代理人 100070150 弁理士 伊東 忠彦 (72) 発明者 長谷川 隆明 東京都千代田区大手町二丁目3番1号 日 本電信電話株式会社内 審査官 辻本 泰隆</p>
--	---

最終頁に続く

(54) 【発明の名称】 話題別関心度計算方法及び装置及び話題別関心度計算プログラムを格納した記憶媒体

(57) 【特許請求の範囲】

【請求項1】

話題データベース、文書送受信手段、主題表現抽出手段、話題獲得手段、データベース更新手段、関心度決定手段及び電子化文書格納部とを備えた話題別関心度計算装置において、電子化文書に含まれる話題に対するユーザの関心度を計算する話題別関心度計算方法であって、

前記文書送受信手段が、前記電子化文書を送受信し、該電子化文書を前記電子化文書格納部に格納する文書送受信ステップと、

前記主題表現抽出手段が、前記電子化文書格納部に格納されている前記電子化文書に対して主題を表現するパターンでパターンマッチングを行い、主題表現を抽出する主題表現抽出ステップと、

前記話題獲得手段が、前記主題表現に対して形態素解析を行い、該形態素解析により品詞分割された品詞情報に基づいて名詞句の集合を話題として抽出する話題獲得ステップと、前記データベース更新手段が、前記話題に対し、前記話題データベースにタプルが存在しない場合に、該話題データベースに話題名、受信回数、送信回数、返信率からなるタプルを作成し、前記話題に対する受信回数と送信回数を0にセットし、

前記電子化文書を受信した際には、該電子化文書に含まれる話題に対する受信回数をインクリメントし、

前記電子化文書を新規に送信した際には、該電子化文書に含まれる話題に対する受信回数と送信回数をそれぞれインクリメントし、

10

20

受信した前記電子化文書に対して返信する際には、前記電子化文書に含まれる前記話題に対する送信回数をインクリメントするデータベース更新ステップと、前記関心度決定手段が、話題の関心度として、前記話題データベースの話題名における受信回数に対する送信回数の割合である返信率を計算することにより、該話題の関心度として決定する関心度決定ステップと、
を行なうことを特徴とする話題別関心度計算方法。

【請求項 2】

前記主題表現抽出ステップにおいて、前記主題表現抽出手段が、前記電子化文書に対し、サブジェクトの文字列と、本文に含まれる主題を表す特徴を持つ文字列を主題表現として抽出するステップを行なう請求項 1 記載の話題関心度計算方法。

10

【請求項 3】

前記話題獲得ステップにおいて、前記話題獲得手段が、前記主題表現に対し、形態素解析を行い、名詞、複合名詞、未定義語からなる名詞句を前記話題として抽出するステップと、同一の主題表現に複数の名詞句が含まれる場合は、それらの名詞句の集合も話題とするステップと、を行う請求項 1 記載の話題別関心度計算方法。

【請求項 4】

電子化文書に含まれる話題に対するユーザの関心度を計算する話題別関心度計算装置であって、

20

話題を格納する話題データベースと、
 電子化文書を格納する電子化文書格納部と、
 前記電子化文書を送受信し、該電子化文書を前記電子化文書格納部に格納する文書送受信手段と、

前記電子化文書格納部に格納されている前記電子化文書に対して主題を表現するパターンでパターンマッチングを行い、主題表現を抽出する主題表現抽出手段と、

前記主題表現に対して形態素解析を行い、該形態素解析により品詞分割された品詞情報に基づいて名詞句の集合を話題として抽出する話題獲得手段と、

前記話題に対し、前記話題データベースにタプルが存在しない場合に、該話題データベースに話題名、受信回数、送信回数、返信率からなるタプルを作成し、前記話題に対する受信回数と送信回数を 0 にセットし、

30

前記電子化文書を受信した際には、該電子化文書に含まれる話題に対する受信回数をインクリメントし、

前記電子化文書を新規に送信した際には、該電子化文書に含まれる話題に対する受信回数と送信回数をそれぞれインクリメントし、

受信した前記電子化文書に対して返信する際には、前記電子化文書に含まれる前記話題に対する送信回数をインクリメントするデータベース更新手段と、

話題の関心度として、前記話題データベースの話題名における受信回数に対する送信回数の割合である返信率を計算することにより、該話題の関心度として決定する関心度決定手段と、

40

を有することを特徴とする話題別関心度計算装置。

【請求項 5】

前記主題表現抽出手段は、
 前記電子化文書に対し、サブジェクトの文字列と、本文に含まれる主題を表す特徴を持つ文字列を主題表現として抽出する手段を含む請求項 4 記載の話題関心度計算装置。

【請求項 6】

前記話題獲得手段は、
 前記主題表現に対し、形態素解析を行い、名詞、複合名詞、未定義語からなる名詞句を前記話題として抽出する手段と、

同一の主題表現に複数の名詞句が含まれる場合は、それらの名詞句の集合も話題とする手

50

段と、を含む請求項4記載の話題別関心度計算装置。

【請求項7】

電子化文書に含まれる話題に対するユーザの関心度を計算する処理を実行させるための話題別関心度計算プログラムを格納した記憶媒体であって、

前記請求項1乃至3記載の話題別関心度計算方法を実現するための処理をコンピュータに実行させるプログラムを格納したことを特徴とする話題別関心度計算プログラムを格納した記憶媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、話題別関心度計算方法及び装置及び話題別関心度計算プログラムを格納した記憶媒体に係り、特に、利用者の電子化文書に対する行動履歴に基づいて、電子化文書に含まれる話題への利用者の関心度を計算するための話題別関心度計算方法及び装置及び話題別関心度計算プログラムを格納した記憶媒体に関する。

【0002】

【従来の技術】

従来の電子メール等の電子化文書に含まれる話題の関心度を決定するために必要となる話題を獲得する方法は、サブジェクトに含まれる単語を話題とする方法が提案されている。一方、電子化文書の本文を対象とした方法では、電子化文書の本文全文に対して形態素解析を行い、文書に存在するほぼ全ての名詞の頻度を用いた特徴ベクトルを作成する方法が提案されている。

【0003】

【発明が解決しようとする課題】

しかしながら、上記従来の電子化文書のサブジェクトからの話題の獲得方法では、文書の本文の内容を考慮していないこと、また、本文全文に対して形態素解析を行うことは文書に含まれる話題とは関係のない単語が混在すること、ユーザの話題の関心度の動的な変化に対応しにくいという問題がある。

【0004】

本発明は、上記の点に鑑みなされたもので、電子化文書に含まれる話題を効率よく獲得し、少ない計算量で利用者の話題に対する関心度を計算し、ユーザの関心度の動的な変化に対応するため、利用者の電子化文書に対する行動履歴に基づいて、電子化文書の本文に存在する主題表現文字列に含まれる話題への利用者の関心度を計算することが可能な話題別関心度計算方法及び装置及び話題別関心度計算プログラムを格納した記憶媒体を提供することを目的とする。

【0005】

【課題を解決するための手段】

図1は、本発明の原理を説明するための図である。

本発明(請求項1)は、話題データベース、文書送受信手段、主題表現抽出手段、話題獲得手段、データベース更新手段、関心度決定手段及び電子化文書格納部とを備えた話題別関心度計算装置において、電子化文書に含まれる話題に対するユーザの関心度を計算する

話題別関心度計算方法であって、文書送受信手段が、電子化文書を送受信し、該電子化文書を電子化文書格納部に格納する文書送受信ステップ(ステップ1)と、

主題表現抽出手段が、電子化文書格納部に格納されている電子化文書に対して主題を表現するパターンでパターンマッチングを行い、主題表現を抽出する主題表現抽出ステップ(ステップ2)と、

話題獲得手段が、主題表現に対して形態素解析を行い、該形態素解析により品詞分割された品詞情報に基づいて名詞句の集合を話題として抽出する話題獲得ステップ(ステップ3)と、

データベース更新手段が、話題に対し、話題データベースにタプルが存在しない場合に、

10

20

30

40

50

該話題データベースに話題名、受信回数、送信回数、返信率からなるタプルを作成し(ステップ4)、話題に対する受信回数と送信回数を0にセットし、
 電子化文書を受信した際には、該電子化文書に含まれる話題に対する受信回数をインクリメントし、
 電子化文書を新規に送信した際には、該電子化文書に含まれる話題に対する受信回数と送信回数をそれぞれインクリメントし、
 受信した電子化文書に対して返信する際には、電子化文書に含まれる話題に対する送信回数をインクリメントするデータベース更新ステップ(ステップ5)と、
 関心度決定手段が、話題の関心度として、話題データベースの話題名における受信回数に対する送信回数の割合である返信率を計算することにより、該話題の関心度として決定する関心度決定ステップ(ステップ6)と、を行なう。

10

【0006】

本発明(請求項2)は、主題表現抽出ステップ(ステップ2)において、
 主題表現抽出手段が、電子化文書に対し、サブジェクトの文字列と、本文に含まれる主題を表す特徴を持つ文字列を主題表現として抽出するステップを行なう。
 本発明(請求項3)は、話題獲得ステップ(ステップ3)において、
 話題獲得手段が、主題表現に対し、形態素解析を行い、名詞、複合名詞、未定義語からなる名詞句を話題として抽出するステップと、
 同一の主題表現に複数の名詞句が含まれる場合は、それらの名詞句の集合も話題とするステップと、を行う。

20

【0008】

図2は、本発明の原理構成図である。
 本発明(請求項4)は、電子化文書に含まれる話題に対するユーザの関心度を計算する話題別関心度計算装置であって、
 話題を格納する話題データベース6と、
 電子化文書を格納する電子化文書格納部と、
 電子化文書を送受信し、該電子化文書を電子化文書格納部に格納する文書送受信手段1と、
 電子化文書格納部に格納されている電子化文書に対して主題を表現するパターンでパターンマッチングを行い、主題表現を抽出する主題表現抽出手段2と、
 主題表現に対して形態素解析を行い、該形態素解析により品詞分割された品詞情報に基づいて名詞句の集合を話題として抽出する話題獲得手段3と、
 話題に対し、話題データベース6にタプルが存在しない場合に、該話題データベース6に話題名、受信回数、送信回数、返信率からなるタプルを作成し、話題に対する受信回数と送信回数を0にセットし、
 電子化文書を受信した際には、該電子化文書に含まれる話題に対する受信回数をインクリメントし、
 電子化文書を新規に送信した際には、該電子化文書に含まれる話題に対する受信回数と送信回数をそれぞれインクリメントし、
 受信した電子化文書に対して返信する際には、電子化文書に含まれる話題に対する送信回数をインクリメントする更新するデータベース更新手段4と、
 話題の関心度として、話題データベース6の話題名における受信回数に対する送信回数の割合である返信率を計算することにより、該話題の関心度として決定する関心度決定手段5と、を有する。

30

40

【0009】

本発明(請求項5)は、主題表現抽出手段2において、
 電子化文書に対し、サブジェクトの文字列と、本文に含まれる主題を表す特徴を持つ文字列を主題表現として抽出する手段を含む。
 本発明(請求項6)は、話題獲得手段3において、
 主題表現に対し、形態素解析を行い、名詞、複合名詞、未定義語からなる名詞句を話題と

50

して抽出する手段と、
同一の主題表現に複数の名詞句が含まれる場合は、それらの名詞句の集合も話題とする手段と_を含む。

【0011】

本発明（請求項7）は、電子化文書に含まれる話題に対するユーザの関心度を計算する処理を実行させるための話題別関心度計算プログラムを格納した記憶媒体であって、
請求項1乃至3記載の話題別関心度計算方法を実現するための処理をコンピュータに実行させるプログラムを格納した記憶媒体である。

【0014】

上記のように、本発明では、まず、電子化文書の主題を表す言語の特徴を手がかりに主題表現を抽出し、抽出された主題表現に対して形態素解析を行って名詞句を取り出し、利用者の電子化文書に対する行動履歴を考慮することにより、名詞句に対する利用者の関心度を計算する。ここで、行動履歴とは、話題における受信回数と送信回数であり、当該受信回数に対する送信回数（受信された話題に対する相手への利用者による返信回数）の割合を関心度とする。

10

【0015】

例えば、電子メールに含まれる話題に対する利用者の関心度を計算する場合に、電子メールを受信し、受信したメールから予め用意しておいた主題を表す表現にマッチする文字列が抜き出される。抜き出された文字列に対して形態素解析を行うことによって、名詞や未定義を中心とした名詞句が話題として取り出される。取り出された話題を対象に、利用者の電子メールに対する送受信履歴に基づいて、話題データベースの新規作成と更新が行われることによって、話題に対する利用者の関心度が得られる。

20

【0016】

これにより、ユーザの動的な関心度の変化に対応可能となる。ここで、動的な関心度とは、固定的に話題のリストを持つのではなく、今まで現れていない新しい話題に対してもそれを話題として登録することを可能とするものである。

【0017】

【発明の実施の形態】

図3は、本発明の話題別関心度計算装置の構成を示す。

同図に示す話題別関心度計算装置は、文書送受信部1、主題表現抽出部2、話題獲得部3、データベース更新部4、関心度決定部5、話題データベース6、及び電子化文書格納部7から構成される。

30

【0018】

文書受信部1は、電子メール等の電子化文書を送受信し、ディスク等の電子化文書格納部7に格納する。

主題表現抽出部2は、電子化文書格納部7に格納されている電子メールに対し、形態素解析を行い、名詞、複合名詞、未定義語からなる名詞句を話題として抽出し、同一の主題表現に複数の名詞句が含まれる場合は、それらの名詞句の集合も話題とする。

【0019】

データベース更新部4は、話題に対し、話題データベース6にタプル（データベースの項目の一覧）が存在しない場合に、話題データベース6に話題名、受信回数、送信回数、返信率からなるタプルを作成し、話題に対する受信回数と送信回数を0にセットし（初期化）、電子化文書を受信した時には受信回数をインクリメントし、電子化文書を新規に送信した時には話題に対する受信回数と送信回数をインクリメントし、受信した電子化文書に対して返信する時には電子化文書に含まれる話題に対する送信回数をインクリメントする。

40

【0020】

関心度決定部5は、話題の関心度として、話題データベース6の話題名における受信回数に対する送信回数の割合である返信率を計算することによって、話題の関心度として決定する。

50

次に、上記の構成における動作を説明する。

図4は、本発明の話題別関心度計算装置の動作のフローチャートである。

【0021】

ステップ101) まず、文書送受信部1において、電子メールが届いているかを調べ、届いている場合にはステップ102に移行する。

ステップ102) 届いている場合には、文書送受信部1において、電子メールを電子化文書格納部7に格納する。

ステップ103) 主題表現抽出部2は、電子化文書格納部7に格納された電子メールから、主題を表現する言語の特徴を手がかりに主題表現を抽出する。

【0022】

ステップ104) 話題獲得部3において、抽出された主題表現を対象として形態素解析を行い、名詞句を話題として獲得する。

ステップ105) データベース更新部5において、話題データベース6に話題名のタプルが存在しなければ、話題名、受信回数、送信回数、返信率からなるタプルを作成し、受信回数と送信回数を0にセットする。

【0023】

ステップ106) 電子化文書格納部7に格納された電子メールの種類を調べ、他者から受信したものである場合にはステップ107に移行し、新規送信の場合にはステップ108に移行し、返信である場合にはステップ109に移行する。

ステップ107) 他者から受信した電子メールの場合、電子メールに含まれる話題の受信回数をインクリメントする。

【0024】

ステップ108) 新規に利用者が送信した電子メールの場合、電子メールに含まれる話題の受信回数と送信回数をそれぞれインクリメントする。

ステップ109) 既に受信していた電子メールへの返信の場合は、受信していた電子メールに含まれる話題の送信回数をインクリメントする。

上記のようにして求められた話題データベース6から、話題の受信回数に対する送信回数の割合を返信率として求めることにより、利用者の話題に対する関心度を決定する。

【0025】

【実施例】

以下、図面と共に本発明の実施例を説明する。

最初に、図4のフローチャートのステップ103における主題表現の抽出について説明する。

主題表現は、電子メールのサブジェクトだけでなく、予め主題となりそうな文字列をパターンとして用意しておき、本文にパターンがマッチする文字列を主題表現として抽出する。このとき、まず、本文全文を文単位で区切り、パターンがマッチするかどうかを各文に対して調べ、マッチした文に対し、マッチした文字列のうち、読点より前の文字列を除いた部分を主題表現として抽出する。

【0026】

図5は、本発明の一実施例の主題表現のパターン例を示し、図6は、本発明の一実施例の電子メールの例を示す。

例えば、図6の例において、格納されている当該電子メールでは、サブジェクトの「refrigerator」の他に、「新しい冷蔵庫」、「冷蔵庫に保存してある個人のものに名前」、「名前のないもの」が主題表現として抽出される。

【0027】

次に、図4のフローチャートのステップ103における話題の獲得について説明する。

ステップ103で抽出された主題表現に対して形態素解析を行う。このうち、品詞が普通名詞とサ変名詞と未定義語の単語または、単語の連続を話題とする。図6の例で説明すると、「refrigerator」、「新しい冷蔵庫」、「冷蔵庫に保存してある個人のものに名前」、「名前のないもの」に対し、それぞれ形態素解析を行う。形態素解析の結

10

20

30

40

50

果は図7の通りである。

【0028】

上記の処理により、品詞が普通名詞、サ変名詞、未定義語である、「refrigerator」、「冷蔵庫」、「保存」、「個人」、「名前」を話題として抽出する。

また、上記の実施例は、図3の構成に基づいて説明しているが、同図に示す構成要素をプログラムとして構築し、話題別関心度計算装置として利用されるコンピュータに接続されるディスク装置や、フロッピーディスク、CD-ROM等の可搬記憶媒体に格納しておき、本発明を実施する際にインストールすることにより、容易に本発明を実現できる。

【0029】

なお、本発明は、上記の実施例に限定されることなく、特許請求の範囲内において、種々 10
変更・応用が可能である。

【0030】

【発明の効果】

上述のように、本発明によれば、電子化文書の中の主題表現から話題を獲得し、電子化文書に対する利用者の行動履歴を利用することにより、電子化文書に含まれる話題に対する利用者の関心を効率よく計算することができる。

また、話題に対するランク付けやそれに応じたユーザへの情報提供の手がかりとすることができる。

【0031】

さらに、送信という行為をクリックという行為に置き換えることにより、ホームページで 20
情報を配信するサイトなどにおけるトレンド解析が可能となる。

【図面の簡単な説明】

【図1】本発明の原理を説明するための図である。

【図2】本発明の原理構成図である。

【図3】本発明の話題別関心度計算装置の構成図である。

【図4】本発明の話題関心度計算装置の動作のフローチャートである。

【図5】本発明の一実施例の主題表現のパターン例である。

【図6】本発明の一実施例の電子メールの例である。

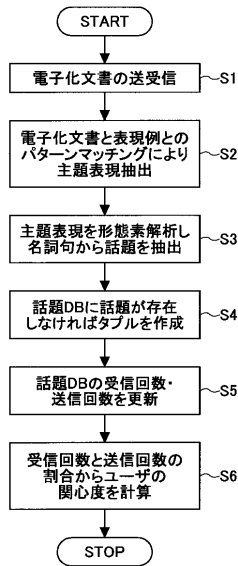
【図7】本発明の一実施例の形態素解析結果である。

【符号の説明】

- 1 文書送受信手段、文書送受信部
- 2 主題表現抽出手段、主題表現抽出部
- 3 話題獲得手段、話題獲得部
- 4 データベース更新手段、データベース更新部
- 5 関心度決定手段、関心度決定部
- 6 話題データベース
- 7 電子化文書格納部

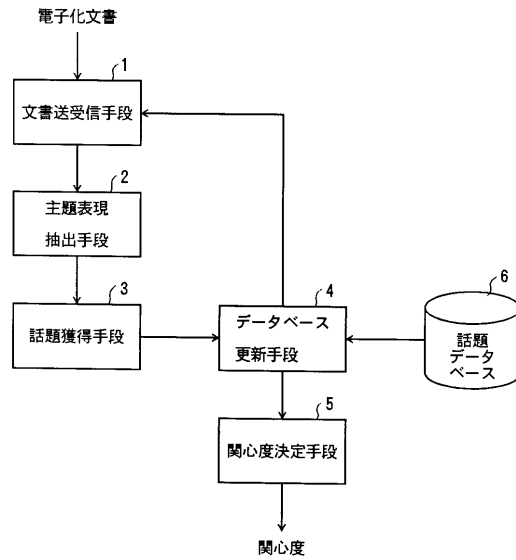
【 図 1 】

本発明の原理を説明するための図



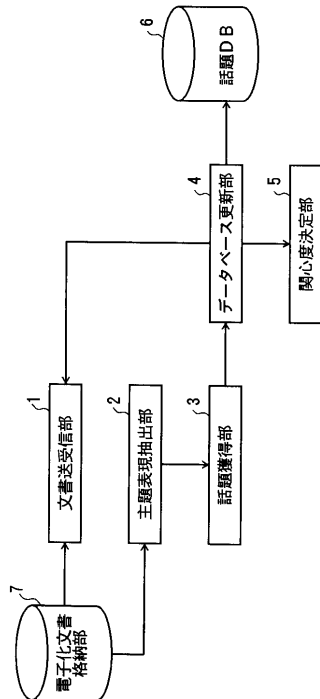
【 図 2 】

本発明の原理構成図



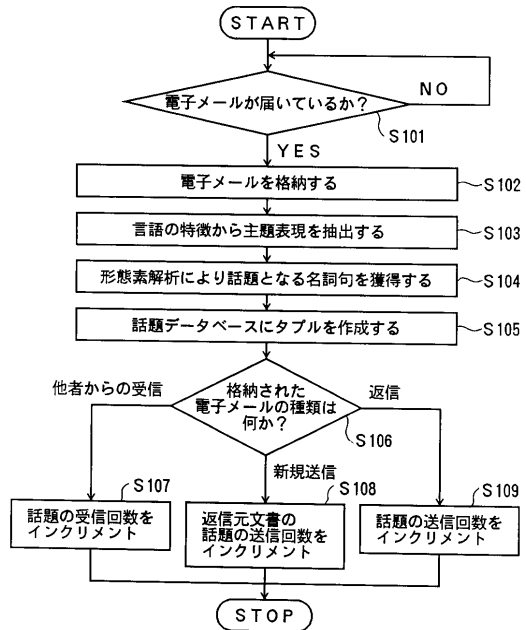
【 図 3 】

本発明の話題別関心度計算装置の構成図



【 図 4 】

本発明の話題別関心度計算装置の動作のフローチャート



【 図 5 】

本発明の一実施例の主題表現のパターン例

正規表現	表現例
<Topic> (について の件です)	明日の打ち合わせの件ですが
<Topic> (を が は) .*ます	来週打ち合わせを実施します
.*は<Topic> (の日) です	本日は献血の日です

<Topic>と.*は任意の文字列である。
 <Topic>に主題を格納する。ただし、読点より前の部分は除く。

【 図 6 】

本発明の一実施例の電子メールの例

```
Subject: refrigerator
To: tigi@ntt.co.jp
From: Chinatsu TAHAKASHI <chinatsu@ntt.co.jp>
Date:Thu, 30 Jul 1998 14:01:37 +0900
Message-id:<4.0.1-J.19980730135519.00df7730@ntt.co.jp>
```

高橋 (千) です。

現在使用中の冷蔵庫が故障のため
 急ぎよ明日、新しい冷蔵庫がはいります。

明日のまでに、
 冷蔵庫に保存してある個人のものに名前を
 つけてくださるようお願いいたします。

新しい冷蔵庫に移し替える際に、
 名前のないものは処分させていただきます。

ご協力お願いします。

 Chinatsu TAKAHASHI (chinatsu@ntt.co.jp)

【 図 7 】

本発明の一実施例の形態素解析結果

単語	読み	基本型	品詞	活用の種類	活用形
refrigerator	refrigerator	refrigerator	品詞 未定義語		活用形
新しい	あたらしい	新しい	形容詞	イ形容詞イ段	基本形
冷蔵庫	れいぞうこ	冷蔵庫	普通名詞		
に	に	に	普通名詞 格助詞		
保存	ほぞん	保存	サ変動詞	サ変動詞	タ系連用テ形
して	して	する	動詞	動詞性接尾辞ある型	基本形
ある	ある	ある	動詞性接尾辞		
個人	こじん	個人	普通名詞		
の	の	の	名詞接続助詞		
もの	もの	もの	形式名詞		
に	に	に	格助詞		
名前	なまえ	名前	普通名詞		
名前	なまえ	名前	普通名詞		
の	の	の	名詞接続助詞		
ない	ない	ない	形容詞	イ形容詞アウオ段	基本形
もの	もの	もの	形式名詞		

フロントページの続き

(56)参考文献 特開平10-312347(JP,A)

朝倉 敬喜,個人の興味に基づくインターネット情報フィルタ,NEC技報,日本,株式会社
NECクリエイティブ,1996年7月30日,第49巻 第7号,17~23

朝倉 敬喜,エージェントによる情報フィルタリング,情報処理学会研究報告,日本,社団法人
情報処理学会,1995年3月10日,第95巻 第30号,49~55

(58)調査した分野(Int.Cl.⁷,DB名)

G06F 17/30