(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
2012 年 9 月 13 日(13.09.2012)

WIPO | PCT

(10) 国際公開番号
WO 2012/121011 A1

(51) 国際特許分類:
G06F 17/30 (2006.01)

(21) 国際出願番号: PCT/JP2012/054211

(22) 国際出願日: 2012 年 2 月 22 日(22.02.2012)

(25) 国際出願の言語: 日本語

(26) 国際公開の言語: 日本語

(30) 優先権データ:
特願 2011-048124 2011 年 3 月 4 日(04.03.2011) JP

(71) 出願人 (米国を除く全ての指定国について): 楽天株式会社(Rakuten, Inc.) [JP/JP]; 〒1400002 東京都品川区東品川四丁目１２番３号 Tokyo (JP).

(72) 発明者; および
(75) 発明者/出願人 (米国についてのみ): 萩原 正人(HAGIWARA Masato) [JP/JP]; 〒1400002 東京都品川区東品川四丁目１２番３号 楽天株式会社内 Tokyo (JP).

(74) 代理人: 木村 満(KIMURA Mitsuru); 〒1010054 東京都千代田区神田錦町二丁目７番地 協販ビル２階 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, 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, IS, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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.

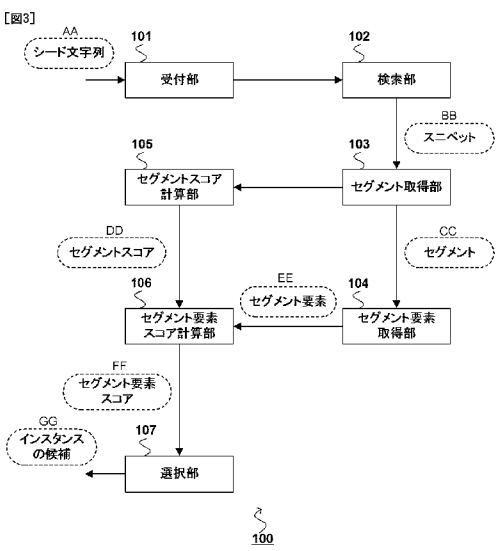
(84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR),

[続葉有]

(54) **Title**: SET-EXPANSION DEVICE, SET-EXPANSION METHOD, PROGRAM, AND NON-TRANSITORY STORAGE MEDIUM

(54) 発明の名称 : 集合拡張処理装置、集合拡張処理方法、プログラム、及び、非一時的な記録媒体

[図3]



101... INPUT UNIT
102... SEARCH UNIT
103... SEGMENT-ACQUISITION UNIT
104... SEGMENT-ELEMENT ACQUISITION UNIT
105... SEGMENT-SCORE CALCULATION UNIT
106... SEGMENT-ELEMENT-SCORE CALCULATION UNIT
107... SELECTION UNIT
AA... SEED STRING
BB... SNIPPETS
CC... SEGMENTS
DD... SEGMENT SCORES
EE... SEGMENT ELEMENTS
FF... SEGMENT-ELEMENT SCORES
GG... CANDIDATE INSTANCE

(57) **Abstract**: A seed string is inputted into an input unit (101). A search unit (102) acquires document snippets containing said seed string. A segment-acquisition unit (103) acquires segments by segmenting said snippets at a segment-delimiter string. A segment-element acquisition unit (104) acquires segment elements by segmenting the segments at a segment-element delimiter string. A segment-score calculation unit (105) uses the standard deviation of the lengths of the segment elements to calculate a score for each segment. A segment-element-score calculation unit (106) uses the segment scores and distances between the positions of the seed string and the positions of the segment elements to calculate a score for each segment element. On the basis of said segment-element scores, a selection unit (107) selects one of the segment elements as a candidate instance in an expanded set for the seed string.

(57) 要約: 受付部（１０１）がシード文字列を受け付ける。検索部（１０２）がシード文字列を含む文書のスニペットを得る。セグメント取得部（１０３）が当該スニペットをセグメント区切文字列で区切ってセグメントを得る。セグメント要素取得部（１０４）がセグメントをセグメント要素区切文字列で区切ってセグメント要素を得る。セグメントスコア計算部（１０５）がセグメントのセグメントスコアをセグメント要素の長さの標準偏差から計算する。セグメント要素スコア計算部（１０６）がセグメント要素のセグメント要素スコアをシード文字列の位置とセグメント要素の位置との距離とセグメントスコアから計算する。選択部（１０７）がセグメント要素スコアに基づいてセグメント要素からいずれかをシード文字列の拡張集合に含まれるインスタンスの候補として選択する。

WO 2012/121011 A1

OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:

— 国際調査報告（条約第 21 条(3)）

明　細　書

発明の名称 ：

集合拡張処理装置、集合拡張処理方法、プログラム、及び、非一時的な記録媒体

技術分野

[0001] 本発明は、集合拡張処理装置、集合拡張処理方法、プログラム、及び、非一時的な（non-transitory）記録媒体に関し、特に、意味的に同一のカテゴリに含まれる語の獲得に関するものである。

背景技術

[0002] ネットショッピングにおいて、ショッピングサイトで取り扱われる商品はカテゴリ分けされて、ユーザに提示される。例えば、特許文献１には、商品を掲載するページにおいて、商品のカテゴリ“家電商品”、“書籍”、“コンピュータ”等を表示する情報送受信システムが開示されている。ユーザは、これらのカテゴリの中から購入を希望する商品のカテゴリを選択することにより、容易に商品を絞り込むことができる。

[0003] 一方、人名、地名、あるいは、商品名などの固有表現を体系的に構築・維持するには膨大なコストがかかる。そのため、固有表現の意味的関係性を計算機により自動的に獲得する自動獲得手法が盛んに研究されている。例えば、非特許文献１には、分かち書き文から意味的語彙カテゴリを抽出するアルゴリズム（「ｇ－Ｅｓｐｒｅｓｓｏアルゴリズム」という）が開示されている。また、非特許文献２には、非分かち書き文から意味的語彙カテゴリを抽出するアルゴリズム（「ｇ－Ｍｏｎａｋａアルゴリズム」という）が開示されている。

先行技術文献

特許文献

[0004] 特許文献1：特開２００９－４８２２６号公報

非特許文献

[0005] 非特許文献1：Mamoru Komachi, Taku Kudo, Masahi Shimbo, and Yuji Matsumoto, "Graph-based analysis of semantic drift in espresso-like bootstrapping algorithms." In Proc. of the EMNLP 2008, pp. 1011-1020, 2008.
非特許文献2：萩原正人、小川泰弘、外山勝彦、「グラフカーネルに基づく非分かち書き文からの意味的語彙カテゴリの抽出」、言語処理学会第１５回年次大会講演論文集、ｐｐ．６９７－７００、２００９年

発明の概要

発明が解決しようとする課題

[0006] 上記のようなショッピングサイトにおいて、日々新たな商品が登場するため、手動では商品のカテゴリの登録作業が追いつかなく、多くのユーザが検索する商品であっても、その商品が属するカテゴリが設けられていない場合がある。しかしながら、店舗側にとっては、新たな商品が登場する度に登録すべきカテゴリを調査するのは負担が大きく、登録すべきカテゴリの候補を自動的に選択してもらいたいという要望があった。

[0007] 本発明は、上記のような課題を解決するもので、意味的に同一のカテゴリに属する語の候補を選択するのに好適な集合拡張処理装置、集合拡張処理方法、プログラム、及び、非一時的な記録媒体を提供することを目的とする。

課題を解決するための手段

[0008] 本発明の第１の観点に係る集合拡張処理装置は、

シード文字列を受け付ける受付部、

前記受け付けられたシード文字列を含む文書を検索して、当該検索された文書のスニペットを得る検索部、

前記得られたスニペットを所定のセグメント区切文字列で区切ることにより、前記受け付けられたシード文字列の前後に出現する文字列と、当該シード文字列とを出現順に並べた文字列からなるセグメントを得るセグメント取得部、

前記得られたセグメントのそれぞれを、所定のセグメント要素区切文字列で区切ることにより、セグメント要素を得るセグメント要素取得部、

前記得られたセグメントのそれぞれのセグメントスコアを、当該セグメントに出現するセグメント要素のそれぞれの長さの分散もしくは標準偏差に基づいて計算するセグメントスコア計算部、

前記得られたセグメントのそれぞれに含まれるセグメント要素のそれぞれのセグメント要素スコアを、当該セグメントにおいて前記受け付けられたシード文字列が出現する位置と当該セグメントにおいて当該セグメント要素が出現する位置との距離、ならびに、当該セグメントについて計算されたセグメントスコアに基づいて計算するセグメント要素スコア計算部、

前記得られたセグメント要素のそれぞれについて計算されたセグメント要素スコアに基づいて、当該セグメント要素からいずれかを、前記受け付けられたシード文字列を含む集合を拡張した拡張集合に含まれるインスタンスの候補として選択する選択部、

を備えることを特徴とする。

[0009] また、上記観点に係る集合拡張処理装置において、

前記インスタンスの候補を用いて検索することにより得られたスニペットから、前記抽出されたインスタンスの候補を含むｎグラムの接続グラフを生成し、当該接続グラフにおける前記受け付けられたシード文字列の前後の文脈と当該インスタンスの候補の前後の文脈とに基づいて当該シード文字列と当該インスタンスの候補との類似度を計算し、当該類似度に基づいて、当該インスタンスの候補から、当該シード文字列を含む集合を拡張した拡張集合に含めるべきインスタンスを抽出する抽出部

をさらに備えることを特徴とする。

[0010] また、上記観点に係る集合拡張処理装置において、

前記得られたセグメントのそれぞれについて、当該セグメントに出現するセグメント要素のそれぞれの長さの標準偏差が所定の閾値を超える場合、前記セグメントスコアならびに前記セグメント要素スコアは、当該セグメントに含まれるセグメント要素が前記インスタンスの候補として前記選択部により選択されることがないような値となる

ことを特徴とする。

[0011] また、上記観点に係る集合拡張処理装置において、

前記得られたセグメントのそれぞれに出現するセグメント要素のそれぞれのセグメント要素スコアは、当該セグメントにおいて前記受け付けられたシード文字列が出現する位置と当該セグメントにおいて当該セグメント要素が出現する位置との最短距離に対して指数的に減衰する

ことを特徴とする。

[0012] 本発明の第２の観点に係る集合拡張処理方法は、

受付部と、検索部と、セグメント取得部と、セグメント要素取得部と、セグメントスコア計算部と、セグメント要素スコア計算部と、選択部と、を備える集合拡張処理装置が実行する集合拡張処理方法であって、

前記受付部が、シード文字列を受け付ける受付工程、

前記検索部が、前記受け付けられたシード文字列を含む文書を検索して、当該検索された文書のスニペットを得る検索工程、

前記セグメント取得部が、前記得られたスニペットを所定のセグメント区切文字列で区切ることにより、前記受け付けられたシード文字列の前後に出現する文字列と、当該シード文字列とを出現順に並べた文字列からなるセグメントを得るセグメント取得工程、

前記セグメント要素取得部が、前記得られたセグメントのそれぞれを、所定のセグメント要素区切文字列で区切ることにより、セグメント要素を得るセグメント要素取得工程、

前記セグメントスコア計算部が、前記得られたセグメントのそれぞれのセグメントスコアを、当該セグメントに出現するセグメント要素のそれぞれの長さの分散もしくは標準偏差に基づいて計算するセグメントスコア計算工程、

前記セグメント要素スコア計算部が、前記得られたセグメントのそれぞれに含まれるセグメント要素のそれぞれのセグメント要素スコアを、当該セグメントにおいて前記受け付けられたシード文字列が出現する位置と当該セグ

メントにおいて当該セグメント要素が出現する位置との距離、ならびに、当該セグメントについて計算されたセグメントスコアに基づいて計算するセグメント要素スコア計算工程、

前記選択部が、前記得られたセグメント要素のそれぞれについて計算されたセグメント要素スコアに基づいて、当該セグメント要素からいずれかを、前記受け付けられたシード文字列を含む集合を拡張した拡張集合に含まれるインスタンスの候補として選択する選択工程、

を備えることを特徴とする。

[0013] 本発明の第３の観点に係るプログラムは、

コンピュータを、

シード文字列を受け付ける受付部、

前記受け付けられたシード文字列を含む文書を検索して、当該検索された文書のスニペットを得る検索部、

前記得られたスニペットを所定のセグメント区切文字列で区切ることにより、前記受け付けられたシード文字列の前後に出現する文字列と、当該シード文字列とを出現順に並べた文字列からなるセグメントを得るセグメント取得部、

前記得られたセグメントのそれぞれを、所定のセグメント要素区切文字列で区切ることにより、セグメント要素を得るセグメント要素取得部、

前記得られたセグメントのそれぞれのセグメントスコアを、当該セグメントに出現するセグメント要素のそれぞれの長さの分散もしくは標準偏差に基づいて計算するセグメントスコア計算部、

前記得られたセグメントのそれぞれに含まれるセグメント要素のそれぞれのセグメント要素スコアを、当該セグメントにおいて前記受け付けられたシード文字列が出現する位置と当該セグメントにおいて当該セグメント要素が出現する位置との距離、ならびに、当該セグメントについて計算されたセグメントスコアに基づいて計算するセグメント要素スコア計算部、

前記得られたセグメント要素のそれぞれについて計算されたセグメント要

素スコアに基づいて、当該セグメント要素からいずれかを、前記受け付けられたシード文字列を含む集合を拡張した拡張集合に含まれるインスタンスの候補として選択する選択部、

として機能させることを特徴とする。

[0014] 本発明の第４の観点に係る非一時的なコンピュータ読み取り可能な記録媒体は、

コンピュータを、

シード文字列を受け付ける受付部、

前記受け付けられたシード文字列を含む文書を検索して、当該検索された文書のスニペットを得る検索部、

前記得られたスニペットを所定のセグメント区切文字列で区切ることにより、前記受け付けられたシード文字列の前後に出現する文字列と、当該シード文字列とを出現順に並べた文字列からなるセグメントを得るセグメント取得部、

前記得られたセグメントのそれぞれを、所定のセグメント要素区切文字列で区切ることにより、セグメント要素を得るセグメント要素取得部、

前記得られたセグメントのそれぞれのセグメントスコアを、当該セグメントに出現するセグメント要素のそれぞれの長さの分散もしくは標準偏差に基づいて計算するセグメントスコア計算部、

前記得られたセグメントのそれぞれに含まれるセグメント要素のそれぞれのセグメント要素スコアを、当該セグメントにおいて前記受け付けられたシード文字列が出現する位置と当該セグメントにおいて当該セグメント要素が出現する位置との距離、ならびに、当該セグメントについて計算されたセグメントスコアに基づいて計算するセグメント要素スコア計算部、

前記得られたセグメント要素のそれぞれについて計算されたセグメント要素スコアに基づいて、当該セグメント要素からいずれかを、前記受け付けられたシード文字列を含む集合を拡張した拡張集合に含まれるインスタンスの候補として選択する選択部、

として機能させることを特徴とするプログラムを記録する。

[0015] 上記プログラムは、プログラムが実行されるコンピュータとは独立して、コンピュータ通信網を介して配布・販売することができる。また、上記記録媒体は、コンピュータとは独立して配布・販売することができる。

発明の効果

[0016] 本発明によれば、意味的に同一のカテゴリに属する語の候補を選択するのに好適な集合拡張処理装置、集合拡張処理方法、プログラム、及び、非一時的な記録媒体を提供することができる。

図面の簡単な説明

[0017] [図1]本発明の実施形態に係る集合拡張処理装置と、ショッピングサーバとの関係を示す図である。

[図2]本発明の実施形態に係る集合拡張処理装置が実現される典型的な情報処理装置の概要構成を示す図である。

[図3]実施形態１の集合拡張処理装置の概要構成を説明するための図である。

[図4]検索された文書を説明するための図である。

[図5]セグメントを説明するための図である。

[図6]セグメント要素を説明するための図である。

[図7]セグメントスコア及びセグメント要素スコアを説明するための図である。

[図8]選択されたインスタンスの候補を説明するための図である。

[図9]実施形態１に係る集合拡張処理装置の各部が行う集合拡張処理を説明するためのフローチャート図である。

[図10]実施形態２の集合拡張処理装置の概要構成を説明するための図である。

[図11]接続グラフを説明するための図である。

[図12]抽出されたインスタンスを説明するための図である。

[図13]実施形態２に係る集合拡張処理装置の各部が行う集合拡張処理を説明するためのフローチャート図である。

発明を実施するための形態

[0018] 本発明の実施形態に係る集合拡張処理装置１００は、図１に示すように、ショッピングサーバ２００に接続される。ショッピングサーバ２００はインターネット３００に接続される。インターネット３００には、ユーザが操作する複数の端末装置４０１、４０２～４０ｎが接続されている。ショッピングサーバ２００は、インターネット３００を介し、端末装置４０１～４０ｎに、ショッピングサーバ２００に登録されている商品の情報を提示し、端末装置４０１～４０ｎから商品の注文を受け付ける。一般的に、ショッピングサーバ２００に登録されている商品は、商品の種類に基づいてカテゴリ分けされて、端末装置４０１～４０ｎのユーザに提示される。集合拡張処理装置１００は、ショッピングサーバ２００で扱う商品について集合拡張処理を行い、商品のカテゴリの候補を提示するものである。

[0019] ここで、「集合拡張」とは、少数の正解セットをシードとして与え、シードと意味的に同一のカテゴリに属する語の集合を獲得するタスクをいう。例えば、キッチン用品の“中華鍋”、“圧力鍋”をシードとした場合、意味的に同一のカテゴリに属する語とは“土鍋”、“雪平鍋”、及び“タジン鍋”等である。すなわち、集合拡張処理装置１００は、“中華鍋”、“圧力鍋”がシードとして与えられると、それらと同一のカテゴリ“鍋”に属する語として、“土鍋”、“雪平鍋”や“タジン鍋”等を獲得する。

[0020] 以下、本発明の実施形態に係る集合拡張処理装置１００が実現される典型的な情報処理装置５００について説明する。

[0021] （１．情報処理装置の概要構成）

情報処理装置５００は、図２に示すように、ＣＰＵ（Central Processing Unit）５０１と、ＲＯＭ（Read only Memory）５０２と、ＲＡＭ（Random Access Memory）５０３と、ＮＩＣ（Network Interface Card）５０４と、画像処理部５０５と、音声処理部５０６と、ＤＶＤ－ＲＯＭ（Digital Versatile Disc ROM）ドライブ５０７と、インターフェース５０８と、外部メモリ５０９と、コントローラ５１０と、モニタ５１１と、スピーカ５１２と、を備え

る。

[0022] ＣＰＵ　５０１は、情報処理装置５００全体の動作を制御し、各構成要素と接続され制御信号やデータをやりとりする。

[0023] ＲＯＭ　５０２には、電源投入直後に実行されるＩＰＬ（Initial Program Loader）が記録され、これが実行されることにより、所定のプログラムをＲＡＭ　５０３に読み出してＣＰＵ　５０１による当該プログラムの実行が開始される。また、ＲＯＭ　５０２には、情報処理装置５００全体の動作制御に必要なオペレーティングシステムのプログラムや各種のデータが記録される。

[0024] ＲＡＭ　５０３は、データやプログラムを一時的に記憶するためのもので、ＤＶＤ－ＲＯＭから読み出したプログラムやデータ、その他、通信に必要なデータ等が保持される。

[0025] ＮＩＣ　５０４は、情報処理装置５００をインターネット３００等のコンピュータ通信網に接続するためのものであり、ＬＡＮ（Local Area Network）を構成する際に用いられる１０ＢＡＳＥ－Ｔ／１００ＢＡＳＥ－Ｔ規格にしたがうものや、電話回線を用いてインターネットに接続するためのアナログモデム、ＩＳＤＮ（Integrated Services Digital Network）モデム、ＡＤＳＬ（Asymmetric Digital Subscriber Line）モデム、ケーブルテレビジョン回線を用いてインターネットに接続するためのケーブルモデム等と、これらとＣＰＵ　５０１との仲立ちを行うインターフェース（図示せず）により構成される。

[0026] 画像処理部５０５は、ＤＶＤ－ＲＯＭ等から読み出されたデータをＣＰＵ　５０１や画像処理部５０５が備える画像演算プロセッサ（図示せず）によって加工処理した後、これを画像処理部５０５が備えるフレームメモリ（図示せず）に記録する。フレームメモリに記録された画像情報は、所定の同期タイミングでビデオ信号に変換され、モニタ５１１に出力される。これにより、各種のページ表示が可能となる。

[0027] 音声処理部５０６は、ＤＶＤ－ＲＯＭ等から読み出した音声データをアナ

ログ音声信号に変換し、これに接続されたスピーカ５１２から出力させる。また、ＣＰＵ　５０１の制御の下、情報処理装置５００が行う処理の進行の中で発生させるべき音を生成し、これに対応した音声をスピーカ５１２から出力させる。

[0028] ＤＶＤ－ＲＯＭドライブ５０７に装着されるＤＶＤ－ＲＯＭには、例えば、実施形態に係る集合拡張処理装置１００を実現するためのプログラムが記憶される。ＣＰＵ　５０１の制御によって、ＤＶＤ－ＲＯＭドライブ５０７は、これに装着されたＤＶＤ－ＲＯＭに対する読み出し処理を行って、必要なプログラムやデータを読み出し、これらはＲＡＭ　５０３等に一時的に記憶される。

[0029] インターフェース５０８には、外部メモリ５０９、コントローラ５１０、モニタ５１１、及びスピーカ５１２が、着脱可能に接続される。

[0030] 外部メモリ５０９には、ユーザの個人情報に関するデータなどが書き換え可能に記憶される。

[0031] コントローラ５１０は、情報処理装置５００の各種の設定時などに行われる操作入力を受け付ける。情報処理装置５００のユーザは、コントローラ５１０を介して指示入力を行うことにより、これらのデータを適宜外部メモリ５０９に記録することができる。

[0032] モニタ５１１は、画像処理部５０５により出力されたデータを情報処理装置５００のユーザに提示する。

[0033] スピーカ５１２は、音声処理部５０６により出力された音声データを情報処理装置５００のユーザに提示する。

[0034] この他、情報処理装置５００は、ハードディスク等の大容量外部記憶装置を用いて、ＲＯＭ　５０２、ＲＡＭ　５０３、外部メモリ５０９、ＤＶＤ－ＲＯＭドライブ５０７に装着されるＤＶＤ－ＲＯＭ等と同じ機能を果たすように構成してもよい。

[0035] 以下、上記情報処理装置５００において実現される実施形態に係る集合拡張処理装置１００の概要構成について、図１乃至１３を参照して説明する。

情報処理装置５００の電源を投入することにより、実施形態に係る集合拡張処理装置１００として機能させるプログラムが実行され、実施形態に係る集合拡張処理装置１００が実現される。

[0036] （２．実施形態１の集合拡張処理装置の概要構成）

実施形態１の集合拡張処理装置１００は、シード文字列を含む集合を拡張した拡張集合に含まれるインスタンスの候補を選択するものである。

[0037] 本実施形態に係る集合拡張処理装置１００は、図３に示すように、受付部１０１と、検索部１０２と、セグメント取得部１０３と、セグメント要素取得部１０４と、セグメントスコア計算部１０５と、セグメント要素スコア計算部１０６と、選択部１０７と、から構成される。

[0038] 以下、集合拡張処理装置１００が、キッチン商品の鍋のカテゴリに属する語として適当な語（インスタンス）の候補の提示を行う場合を例に説明する。

[0039] 受付部１０１は、シード文字列を受け付ける。シード文字列とは、例えば、“鍋”のカテゴリに属する語の集合に含まれる正解の語（“中華鍋”や“圧力鍋”等）である。例えば、図４に示すように、ユーザがＷＥＢページの検索エンジンの検索欄６０１に、全てのシード文字列をスペース区切りで連結させたものをクエリとして入力し、検索ボタン６０２を押圧する。この場合、受付部１０１は、検索欄６０１に入力された“中華鍋”及び“圧力鍋”をシード文字列として受け付ける。なお、検索エンジンの種類は任意である。

[0040] 本実施形態では、ＣＰＵ　５０１及びコントローラ５１０が協働して、受付部１０１として機能する。

[0041] 検索部１０２は、受け付けられたシード文字列を含む文書を検索し、スニペットを得る。ここで、スニペットとは、例えば、ＷＥＢページの検索エンジンを使用した際に、検索結果として表示されるクエリを含むテキスト部分である。検索部１０２は、ＷＥＢページの検索エンジンに対して、全てのシード文字列をスペース区切りで連結させたものをクエリとして入力し、検索

結果の、例えば、上位３００件のスニペットのリストを得る。例えば、検索部１０２は、“中華鍋　圧力鍋”をクエリとして検索エンジンを用いてＷＥＢページの検索を行い、与えられたシード文字列“中華鍋”及び“圧力鍋”を含む図４のスニペット１、２、３～３００（図示せず）を得る。なお、検索部１０２は、上記のように外部装置を利用して文書を得ることに限らず、内部に検索機能を備えるようにしてもよい。例えば、検索部１０２は、Ｗｅｂ検索ＡＰＩを使用してスニペットを得ることとしてもよい。

[0042] 本実施形態では、検索部１０２は、ＣＰＵ　５０１及びＮＩＣ　５０４が協働して、検索部１０２として機能する。

[0043] セグメント取得部１０３は、得られたスニペットを所定のセグメント区切文字列で区切ることにより、シード文字列の前後に出現する文字列と、シード文字列とを出現順に並べた文字列からなるセグメントを得る。スニペットは、検索語が含まれるページにおいて当該検索語がどのように用いられているかがユーザにとって一目で分かるように、所定の区切文字列で区切られているのが一般的である。例えば、所定のセグメント区切文字列を“・・・”とする。例えば、セグメント取得部１０３は、得られたスニペット１、２、３～３００をＵｎｉｃｏｄｅのＮＦＫＣにより正規化して、小文字に統一し、セグメント区切文字列“・・・”によって複数の文字列に分割する。そして、セグメント取得部１０３は、分割された文字列のうち重複している文字列は除外し、残りの文字列をセグメントとして得る。得られたスニペットを小文字に統一することにより、例えば、型番等の文字列が大文字・小文字で統一されていない場合に対応することができる。図５に、セグメント取得部１０３がスニペット１から得たセグメント１－１～１－３を示す。

[0044] なお、セグメント区切文字列は、“・・・”の文字列に限らない。検索部１０２が使用する検索エンジン又はＷｅｂ検索ＡＰＩが提示するスニペットが、例えば、“－－－”や“＃＃”の文字列で区切られている場合、セグメント区切文字列を“－－－”や“＃＃”の文字列とする。また、セグメントを得る手法は、セグメント区切文字列を用いてセグメントを得る手法に限ら

ない。セグメントは、使用する検索エンジン又はＷｅｂ検索ＡＰＩが提示するスニペットに応じて、適宜取得される。例えば、１つのスニペットが、“・・・”等の記号により区切られずに提示される場合は、当該スニペットを１つのセグメントとする。また、予め、スニペット内のセグメントに相当する部分が箇条書き等で提示される場合は、箇条書きの１行に該当する部分を１つのセグメントとする。

[0045] 本実施形態では、ＣＰＵ　５０１がセグメント取得部１０３として機能する。

[0046] セグメント要素取得部１０４は、得られたセグメントのそれぞれを、所定のセグメント要素区切文字列で区切ることによりセグメント要素を得る。例えば、所定のセグメント要素区切文字列とは、句読点や記号（“、”、“，”、“。”、“！”、“［”、“］”等）であり、これらのセグメント要素区切文字列によりセグメントを区切り、セグメント要素を得る。例えば、セグメント要素取得部１０４は、図５のセグメント１－１、１－２、１－３をセグメント要素区切文字列で区切ると、図６のセグメント要素群１－１Ｐ（セグメント要素P_i（ｉ＝１～５））、１－２Ｐ（セグメント要素P_i（ｉ＝１～１２））、１－３Ｐ（セグメント要素P_i（ｉ＝１～５））を得る。

[0047] 本実施形態では、ＣＰＵ　５０１がセグメント要素取得部１０４として機能する。

[0048] セグメントスコア計算部１０５は、得られたセグメントのそれぞれのセグメントスコアを、当該セグメントに出現するセグメント要素のそれぞれの長さの分散もしくは標準偏差に基づいて計算する。ここで、得られたセグメントのそれぞれについて、当該セグメントに出現するセグメント要素のそれぞれの長さの標準偏差が所定の閾値を超える場合、セグメントスコアならびに後述のセグメント要素スコアは、当該セグメントに含まれるセグメント要素がインスタンスの候補として選択部１０７により選択されることがないような値となるとする。本実施形態では、セグメント要素の長さを、Ｕｎｉｃｏｄｅの文字数で定義するが、これに限られない。例えば、セグメント要素の

長さとして、その他の文字コードにおけるバイト数を採用することも可能である。

[0049] 例えば、図５に示すように、セグメント１－１、１－３は、通常の文を含んでいるが、セグメント１－２は、通常の文を含んでいない。そして、セグメント１－１、１－３に含まれるセグメント要素の長さのばらつきは、セグメント１－２に含まれるセグメント要素の長さのばらつきよりも大きい。すなわち、通常の文を含んでいるセグメントは、一般的に、通常の文を含んでいないセグメントよりも、セグメントに含まれる各セグメント要素の長さが揃っていないという傾向がある。そして、通常の文を含むセグメントには、シード文字列と同じ意味範囲に属するインスタンスが含まれていないことが多いので、インスタンスの候補を得るセグメントとして適当ではない。したがって、以下では、セグメント要素の長さの標準偏差が所定の閾値を越えるセグメントは、インスタンスの候補を得るセグメントから除外することとする。

[0050] 本実施形態では、所定の閾値を５．００とする。また、セグメントスコア計算部１０５は、セグメント要素の長さの標準偏差が５．００未満の場合は、標準偏差の値そのものを、セグメントスコアとし、標準偏差が５．００以上の場合は、セグメントスコアを５．００とする。

[0051] 図７に、セグメントスコア計算部１０５が計算したセグメントスコアを示す。図７のテーブルには、シード文字列をクエリとして得た「スニペット７０１ａ」と、スニペット７０１ａに含まれる「セグメント７０２ａ」と、セグメント７０２ａに含まれる「セグメント要素７０３ａ」と、セグメント要素７０３ａの「長さ７０４ａ」と、長さ７０４ａの「標準偏差７０５ａ」と、標準偏差７０５ａに基づいて計算される「セグメントスコア７０６ａ」と、後述するセグメント要素スコア計算部１０６により計算される「セグメント要素スコア７０７ａ」と、が対応づけて記載されている。

[0052] 例えば、セグメントスコア計算部１０５は、図７の７０４ａに示すように、セグメント１－１に含まれるセグメント要素P_i（ｉ＝１～５）、セグメン

ト１－２に含まれるセグメント要素P_i（ｉ＝１～１２）、及び、セグメント１－３に含まれるセグメント要素P_i（ｉ＝１～５）の長さを求める。そして、セグメントスコア計算部１０５は、図７の７０５ａに示すように、セグメント１－１に含まれるセグメント要素P_i（ｉ＝１～５）の長さの標準偏差を“５．８９”、セグメント１－２に含まれるセグメント要素P_i（ｉ＝１～１２）の長さの標準偏差を“１．３４”、セグメント１－３に含まれるセグメント要素P_i（ｉ＝１～５）の長さの標準偏差を“５．２７”と求める。したがって、セグメントスコア計算部１０５は、図７の７０６ａに示すように、セグメント１－１のセグメントスコアを“５．００”、セグメント１－２のセグメントスコアを“１．３４”、セグメント１－３のセグメントスコアを“５．００”と求める。

[0053] 本実施形態では、ＣＰＵ　５０１がセグメントスコア計算部１０５として機能する。

[0054] セグメント要素スコア計算部１０６は、得られたセグメントのそれぞれに含まれるセグメント要素のそれぞれのセグメント要素スコアを、当該セグメントにおいて受け付けられたシード文字列が出現する位置と当該セグメントにおいて当該セグメント要素が出現する位置との距離、ならびに、当該セグメントについて計算されたセグメントスコアに基づいて計算する。

[0055] 例えば、前述のように、セグメント要素のそれぞれの長さの標準偏差が所定の閾値を超える場合、セグメント要素スコアを、セグメント要素がインスタンスの候補として選択部１０７により選択されないような値になるとする。例えば、セグメント要素スコア計算部１０６は、セグメントスコアが“５．００”の場合は、セグメント要素スコアを“０”とする。一方、セグメントスコアが“５．００”未満の場合は、セグメント要素スコア計算部１０６は、セグメントにおいて受け付けられたシード文字列が出現する位置と当該セグメントにおいて当該セグメント要素が出現する位置との距離に基づいてセグメント要素スコアを計算する。ここで、セグメントにおいてシード文字列が出現する位置s_j（ｊ：シード文字列の数）、及びセグメントにおいてセ

グメント要素が出現する位置p_iとは、図６に示すように、セグメントにおいて出現順にセグメント要素を並べた時のセグメント内での出現順位であり、距離とは位置s_jと位置p_iとの出現順位の差である。すなわち、シード文字列が“中華鍋”及び“圧力鍋”とすると、セグメント１－２においてシード文字列“圧力鍋”（P_4）が出現する位置s_1は“４”番目であり、シード文字列“中華鍋”（P_8）が出現する位置s_2は“８”番目である。また、セグメント１－２においてセグメント要素“親子鍋”（P_5）が出現する位置p_5は“５”番目であり、シード文字列“中華鍋”（P_8）とセグメント要素“親子鍋”（P_5）との距離は３となる。

[0056] そして、セグメント要素スコア計算部１０６は、セグメント要素スコアS_iを、セグメントにおいてシード文字列が出現する位置s_jと、セグメントにおいてセグメント要素が出現する位置p_iとから、以下の式（数１）に基づいて計算する。この式（数１）によれば、最も近いシード文字列との距離に従い指数的に減衰するスコアが、各セグメント要素のセグメント要素スコアとされる。本実施形態では$\alpha = 0.8$とする。計算結果を図７のセグメント要素スコア７０７ａに示す。

[0057] [数1]

$$S_i = \max_j \exp(-\alpha \left| p_i - s_j \right|)$$

[0058] 上記においては、最も近いシード文字列との距離に従い指数的に減衰するスコアを求めることとしたが、スコアの求め方には様々な変形が可能である。例えば、シード文字列が複数存在する場合に、各シード文字列とセグメント要素との距離をそれぞれ求め、求めた距離の平均値に従い線形的に減衰するスコアを各セグメント要素のセグメント要素スコアとしてもよい。

[0059] 以上、セグメント内にシード文字列が出現する場合の一例を記載したが、シード文字列の類似語が出現する場合も同様に計算できる。具体的には、“中華鍋”及び“圧力鍋”をシード文字列とした場合に、検索部ではシード文

字列に加えてシード文字列の類似語で検索を行うと、“中華なべ”や“圧力なべ”といったシード文字列の類似語が含まれるスニペットが得られる。このような場合には、セグメント要素スコア計算部１０６において、公知の漢字かな文字変換プログラム等を用いることで、シード文字列の類似語をシード文字列として同様に取り扱うことができる。このように、シード文字列の類似語がセグメント内に出現した場合であっても、数１に従ってセグメント要素スコアS_iを計算できる。

[0060] 本実施形態では、ＣＰＵ　５０１がセグメント要素スコア計算部１０６として機能する。

[0061] 選択部１０７は、得られたセグメント要素のそれぞれについて計算されたセグメント要素スコアに基づいて、当該セグメント要素からいずれかを、受け付けられたシード文字列を含む集合を拡張した拡張集合に含まれるインスタンスの候補として選択する。ここで、拡張集合とは、集合拡張処理を施した後に得られる集合であり、シード文字列と意味的に同一のカテゴリに含まれる語の集合である。例えば、選択部１０７は、セグメント要素スコアの値が“０．１０”未満のセグメント要素をインスタンスの候補から除外し、残りのセグメント要素をインスタンスの候補として選択する。すなわち、選択部１０７は、セグメント１－１、１－３から得たセグメント要素のセグメント要素スコアがすべて“０”なので（図７）、セグメント１－１、１－３から得たセグメント要素を候補から除外する。そして、選択部１０７は、図８に示すように、セグメント１－２から得たセグメント要素のうち、セグメント要素スコアが“０．１０”未満の“パスタマシーン”、“その他”、及び、“さらに価格が”のセグメント要素を除外し、残りのセグメント要素を、“中華鍋”及び“圧力鍋”と意味的に同一のカテゴリに含まれるインスタンスの候補として選択する。なお、本実施形態では、１つのスニペットを例に、インスタンスの候補を選択する手法について説明したが、実際には多数のスニペットからセグメント要素を得てセグメント要素スコアを求め、インスタンスの候補を選択する。この場合、同じセグメント要素に、異なるスニペ

ットからそれぞれセグメント要素スコアが求められることがある。特に、シード文字列と意味的に同じカテゴリに含まれるセグメント要素は、複数のスニペットに含まれることが多いと考えられるので、複数のセグメント要素スコアが得られる可能性が高い。したがって、複数のセグメント要素スコアが得られた場合、それらの和や最大値等を当該セグメント要素のセグメント要素スコアの値とする。このように処理することにより、より適当なインスタンスの候補を選択することができる。

[0062] 本実施形態では、ＣＰＵ　５０１が選択部１０７として機能する。

[0063] （３．実施形態１の集合拡張処理装置の動作）

次に、本実施形態の集合拡張処理装置１００の各部が行う動作について図９のフローチャートを用いて説明する。集合拡張処理装置１００に電源が入れられ、所定の操作が行われると、ＣＰＵ　５０１は図９のフローチャートに示す集合拡張処理を開始する。

[0064] まず、受付部１０１は、シード文字列を受け付ける（ステップＳ１０１）。例えば、受付部１０１は、図４に示すように、ＷＥＢページの検索エンジンの検索欄６０１にクエリとして入力された“中華鍋”及び“圧力鍋”を、シード文字列として受け付ける。

[0065] 次に、検索部１０２は、受け付けられたシード文字列を含む文書を検索し、スニペットを得る（ステップＳ１０２）。例えば、検索部１０２は、シード文字列“中華鍋”及び“圧力鍋”をクエリとして検索し、図４に示すように、検索結果の上位３００件のスニペット１、２、３～３００を得る。なお、検索部１０２が得るスニペットの数は、任意であるが、およそ１００件以上のスニペットを得ることにより、より適当なインスタンスの候補を選択することができる。

[0066] 次に、セグメント取得部１０３は、検索部１０２が得たスニペットを、セグメント区切文字列で区切ることによりセグメントを得る（ステップＳ１０３）。例えば、セグメント取得部１０３は、スニペット１、２、３～３００をセグメント区切文字列“・・・”で区切り、セグメントを得る。例えば、

セグメント取得部１０３は、スニペット１から、図５に示すように、セグメント１－１～１－３を得る。

[0067] セグメントが得られると（ステップＳ１０３）、セグメント要素取得部１０４は、当該セグメントを所定のセグメント要素区切文字列で区切ることによりセグメント要素を得る（ステップＳ１０４）。例えば、セグメント１－１～１－３を、セグメント要素区切文字列（“、”、“，”、“。”、“！”、“［”、“］”等）で区切り、図６のセグメント要素（セグメント要素群１－１Ｐ、１－２Ｐ、１－３Ｐ）を得る。

[0068] セグメント要素が得られると（ステップＳ１０４）、セグメントスコア計算部１０５は、当該セグメントのそれぞれのセグメントスコアをセグメントが含むセグメント要素の長さの標準偏差に基づいて計算する（ステップＳ１０５）。例えば、セグメントスコア計算部１０５は、セグメント要素の長さの標準偏差が５．００未満の場合は、標準偏差の値そのものをセグメントスコアとし、セグメント要素の長さの標準偏差が５．００以上の場合は、セグメントスコアを５．００とする。すなわち、セグメントスコア計算部１０５は、標準偏差が“５．８９”のセグメント１－１のセグメントスコアを“５．００”、標準偏差が“１．３４”のセグメント１－２のセグメントスコアを“１．３４”、標準偏差が“５．２７”のセグメント１－３のセグメントスコアを“５．００”と求める。

[0069] 次に、セグメント要素スコア計算部１０６は、セグメント要素のセグメント要素スコアを、セグメントにおいて受け付けられたシード文字列が出現する位置と当該セグメントにおいて当該セグメント要素が出現する位置との距離、ならびに、当該セグメントについて計算されたセグメントスコアに基づいて計算する（ステップＳ１０６）。例えば、セグメント要素スコア計算部１０６は、セグメントスコアが“５．００”の場合は、セグメント要素スコアを“０”とし、セグメントスコアが“５．００”未満の場合は、セグメントにおいてシード文字列が出現する位置とセグメント要素が出現する位置との距離を用いた式（数１）に基づいて、セグメント要素スコア７０７ａ（図

７）を計算する。

[0070] そして、選択部１０７は、得られたセグメント要素についてのセグメント要素スコアに基づいて、シード文字列と意味的に同一のカテゴリに属するインスタンスの候補を選択する（ステップＳ１０７）。例えば、選択部１０７は、図８に示すように、セグメント要素スコアの値が“０．１０”以上のセグメント要素をインスタンスの候補として選択する。

[0071] 本実施形態によれば、“親子鍋”や“タジン鍋”は、シード文字列の“中華鍋”や“圧力鍋”と同じ“鍋”のカテゴリに含まれる用語であるので、意味的に同一のカテゴリに属する語の候補を選択することができる。

[0072] （４．実施形態２の集合拡張処理装置の概要構成）

実施形態２の集合拡張処理装置１００は、拡張集合に含まれるインスタンスの候補について、文脈に基づいてフィルタをかけることにより、意味的に無関係な語を排除するものである。

[0073] 本実施形態に係る集合拡張処理装置１００は、図１０に示すように、受付部１０１と、検索部１０２と、セグメント取得部１０３と、セグメント要素取得部１０４と、セグメントスコア計算部１０５と、セグメント要素スコア計算部１０６と、選択部１０７と、抽出部１０８と、から構成される。本実施形態の受付部１０１、検索部１０２、セグメント取得部１０３、セグメント要素取得部１０４、セグメントスコア計算部１０５、セグメント要素スコア計算部１０６、及び、選択部１０７は、実施形態１と同様の機能を有する。以下、異なる機能を有する抽出部１０８について説明する。

[0074] まず、インスタンスの候補は、シード文字列の前後の文脈とインスタンスの候補の前後の文脈とが類似するほど、シード文字列と意味的に類似していると考えられる。そこで、実施形態２の集合拡張処理装置１００は、シード文字列の前後の文脈とインスタンスの候補の前後の文脈とに基づいてシード文字列とインスタンスの候補との類似度を求め、当該類似度に基づき、インスタンスの候補の中からインスタンスを抽出する。これにより、意味的に無関係な語を排除することができる。以下、集合拡張装置１００は、ｇ－Ｍｏ

ｎａｋａアルゴリズムに基づいて計算した類似度から、インスタンスの候補をランク付けし、所定の値以上の類似度を有するものをインスタンスとして抽出する。なお、類似度を求める手法はｇ－Ｍｏｎａｋａアルゴリズムに限らない。例えば、ｇ－Ｅｓｐｒｅｓｓｏアルゴリズムを用いてもよい。

[0075] 抽出部１０８は、インスタンスの候補を用いて検索することにより得られたスニペットから、抽出されたインスタンスの候補を含むｎグラムの接続グラフを生成する。そして、抽出部１０８は、当該接続グラフにおける、受け付けられたシード文字列の前後の文脈とインスタンスの候補の前後の文脈とに基づいて当該シード文字列と当該インスタンスとの類似度を計算し、当該類似度に基づいて、当該インスタンスの候補から、当該シード文字列を含む集合を拡張した拡張集合に含めるべきインスタンスを抽出する。以下、ｇ－Ｍｏｎａｋａアルゴリズムに基づく類似度の計算手法を詳細に説明する。

[0076] 抽出部１０８は、選択部１０７が選択したインスタンスの候補のそれぞれを、ＷＥＢページの検索エンジンに対してクエリとして入力し、検索結果の上位３００件のスニペットのリストを得る。そして、抽出部１０８は、得られたスニペットに対して、ＵｎｉｃｏｄｅのＮＦＫＣにより正規化して、小文字に統一し、重複を取り除く。また、日本語の割合が極端に少ない、記号が多いなど、スニペットとして適当でないものを除外する。

[0077] 次に、抽出部１０８は、残ったスニペットの集合に含まれるすべての文字ｎグラムについて、接続行列Ｍ（ｕ，ｖ）を構築する。接続行列Ｍ（ｕ，ｖ）は、式（数２）で表される。

[0078] [数2]

$$M(u,v)=\frac{pmi(u,v)}{\max pmi},\qquad pmi(u,v)=\log\frac{|u,v|}{|u,*||*,v|}$$

[0079] ここで、｜ｕ，ｖ｜は、ｎグラムｕの後にｎグラムｖが続く頻度であり、｜ｕ，＊｜、｜＊，ｖ｜はそれぞれ、ｎグラムｕ、ｎグラムｖそのものの出現頻度である。本実施形態では、｜ｕ，ｖ｜、｜ｕ，＊｜、｜＊，ｖ｜は、

それら自体をクエリとして検索した場合の検索結果数であり、ｐｍｉ（ｕ，ｖ）は、それらの検索結果数の自然対数をとったものを用いている。

[0080] 次に、抽出部１０８は、全てのｎグラムの集合Ｖを節点集合とし、Ｍを接続行列として表現される有向重み付きグラフ（以下、「接続グラフ」という）G_Mを生成する。生成した接続グラフG_Mの例を図１１に示す。このグラフにおいて、ｎグラムｕ及びｎグラムｖのそれぞれの右側文脈及び左側文脈が類似しているほど、それらの意味は類似しているとみなすことができる。

[0081] ここで、まず、ｎグラムｕの右側文脈とｎグラムｖの右側文脈とが類似しているか否かは、引用解析手法の書誌結合の概念に対応付けて考えることができる。書誌結合とは、文献ｘ、ｙが同じ文献を引用することをいう。すなわち、書誌結合は、ｎグラムｕとｎグラムｖが同じｎグラムに接続しているか否かということに対応付けて考えることができる。一方、ｎグラムｕの左側文脈とｎグラムｖの左側文脈とが類似しているか否かは、引用解析手法の共引用の概念に対応付けて考えることができる。共引用とは、文献ｘ、ｙが同じ文献により引用されることをいう。すなわち、ｎグラムｕとｎグラムｖが同じｎグラムから接続されているか否かということに対応付けて考えることができる。

[0082] したがって、ｎグラムｕ及びｎグラムｖの右側文脈及び左側文脈が類似しているか否かを示す類似度行列A_R、A_Lを、書誌結合行列及び共引用行列にそれぞれ対応させて求めることとする。右側文脈の類似度行列A_R、及び、左側文脈の類似度行列A_Lは、接続行列Ｍを用いて、式（数３）で表すことができる。

[0083] [数3]

$$A_R = \frac{1}{|V|^2} MM^T, \qquad A_L = \frac{1}{|V|^2} M^T M$$

[0084] 抽出部１０８は、全てのｎグラムについて右側文脈の類似度行列A_R、及び、左側文脈の類似度行列A_Lを求める。

[0085] また、ｎグラムｕとｎグラムｖとが意味的に類似しているとみなすためには、右側文脈及び左側文脈の両者が類似している必要がある（以下、「両側近接制約」という）。そこで、抽出部１０８は、式（数４）に示すように、要素毎の重み付き一般化平均によって、ｎグラムｕとｎグラムｖとの類似度を示す類似度行列Ａを求める。ここで、ｍは、この両側近接制約の強さを調節するパラメータであり、本実施形態では、ｍ＝０．１とする。

[0086] [数4]

$$A(i,j)=\sqrt[m]{\frac{1}{2}(A_R(i,j)^m+A_L(i,j)^m)}$$

[0087] そして、抽出部１０８は、この類似度行列Ａを用いてラプラシアンカーネルR_β（Ａ）を、数５、数６の式から求める。

[0088] [数5]

$$\tilde{R}_\beta(A)=\sum_{n=0}^{\infty}\beta^n(-\tilde{L})$$

[0089] [数6]

$$\tilde{L}=I-D^{-\frac{1}{2}}AD^{-\frac{1}{2}},\qquad D(i,j)=\sum_j A(i,j)$$

[0090] R_β（Ａ）の（ｉ，ｊ）要素が、ｎグラムｉとｎグラムｊとの類似度に対応する。そこで、抽出部１０８は、シードベクトルv_0（シード文字列に対応する要素が１、それ以外が０となっているようなベクトル）を用いて、R_β（Ａ）v_0計算し、計算された値を類似度とする。βの値は、任意であり、例えば、１．０－２である。

[0091] 例えば、図１１の接続グラフG_Mにおいて、“中華鍋”は“の”に接続し、“調理はさみ”、“タジン鍋”は両者とも“の”に接続している。また、“

中華鍋”に接続する“の”は、“タジン鍋”に接続しているが、“調理はさみ”には接続していない。このような場合において、“調理はさみ”の“中華鍋”に対する類似度R_{β}（Ａ）v_0は、“タジン鍋”の“中華鍋”に対する類似度R_{β}（Ａ）v_0よりも小さい値となる。

[0092] 抽出部１０８は、例えば、計算した類似度が所定の値を超えたものを、インスタンスとして抽出する。例えば、図１２に示すように類似度が求められ、所定の値を“０．１０”とすると、抽出部１０８は、“圧力鍋”、“中華鍋”、“親子鍋”、“タジン鍋”、“伊賀焼”を、インスタンスとして抽出する。

[0093] 本実施形態では、ＣＰＵ　５０１が抽出部１０８として機能する。

[0094]（５．実施形態２の集合拡張処理装置の動作）

次に、本実施形態の集合拡張処理装置１００の各部が行う動作について図１３のフローチャートを用いて説明する。集合拡張処理装置１００に電源が入れられ、所定の操作が行われると、ＣＰＵ　５０１は図１３のフローチャートに示す集合拡張処理を開始する。なお、図１３のフローチャートにおいて、図９のフローチャートと同じステップ番号が付されているステップは、図９のフローチャートにおける処理と同様の処理を行う。したがって、これらの説明は省略する。

[0095] 選択部１０７によりインスタンスの候補が選択されると（ステップＳ１０７）、抽出部１０８は、インスタンスの候補を用いて検索エンジンで検索することによりスニペットを取得する（ステップＳ２０８）。例えば、抽出部１０８は、インスタンスの候補をクエリとしてＷＥＢページの検索エンジンに入力し、検索結果の上位３００件のスニペットのリストを得る。

[0096] 次に、抽出部１０８は得られたスニペットからインスタンスの候補を含むｎグラムの接続グラフを生成する（ステップＳ２０９）。例えば、抽出部１０８は、３００件のスニペットから、不適当なものを除外し、残ったスニペットの集合に含まれるすべての文字のｎグラムについて、接続行列Ｍを求める。そして、図１１に示すように、すべてのｎグラムの集合Ｖを節点集合と

し、Ｍ（数２）を接続行列として表現される接続グラフG_Mを生成する。

[0097] 抽出部１０８は、接続グラフにおける、シード文字列の前後の文脈と、インスタンスの候補の前後の文脈とに基づいて、シード文字列とインスタンスの候補との類似度を計算する（ステップＳ２１０）。例えば、抽出部１０８は、式（数３）に基づいて、右側文脈の類似度行列A_R、及び、左側文脈の類似度行列A_Lを求め、式（数４）に示すように、要素毎に重み付き一般化平均を行った類似度行列Ａを求める。さらに、式（数５、６）に基づいて、類似度行列Ａを用いたラプラシアンカーネルR_β（Ａ）を求め、シードベクトルv_0を乗じることにより、シード文字列に対するインスタンスの候補の類似度を求める。

[0098] 抽出部１０８は、類似度に基づいてインスタンスを抽出する（ステップＳ２１１）。例えば、抽出部１０８は、計算した類似度が“０．１０”を超えたものを、図１２に示すように、インスタンスとして抽出する。また、あるいは、抽出部１０８は、類似度の高いものから所定の個数だけ抽出することとしてもよい。例えば、インスタンスの候補が図１２に示すように９個有る場合、所定の個数を４個とすると、抽出部１０８は、類似度において上位４個の“圧力鍋”、“中華鍋”、“親子鍋”、及び、“タジン鍋”をインスタンスとして抽出する。

[0099] 本実施形態によれば、意味的に無関係な語を排除することができ、意味的に同一のカテゴリに含まれるとみなすのにより適当な用語を抽出することができる。

[0100] なお、実施形態１、２では、集合拡張処理装置１００は、ショッピングサイトの商品のカテゴリ生成に適用する例を示したが、これに限らない。例えば、固有表現獲得や辞書構築等に応用可能である。

[0101] 本発明は、２０１１年３月４日に出願された日本国特許出願２０１１－０４８１２４号に基づく。本明細書中に日本国特許出願２０１１－０４８１２４号の明細書、特許請求の範囲、図面全体を参照として取り込むものとする。

産業上の利用可能性

[0102] 本発明によれば、意味的に同一のカテゴリに属する語の候補を選択するのに好適な集合拡張処理装置、集合拡張処理方法、プログラム、及び、非一時的な記録媒体を提供することができる。

符号の説明

[0103] １００　集合拡張処理装置
１０１　受付部
１０２　検索部
１０３　セグメント取得部
１０４　セグメント要素取得部
１０５　セグメントスコア計算部
１０６　セグメント要素スコア計算部
１０７　選択部
１０８　抽出部
２００　ショッピングサーバ
３００　インターネット
４０１、４０２～４０ｎ　端末装置
５００　情報処理装置
５０１　ＣＰＵ
５０２　ＲＯＭ
５０３　ＲＡＭ
５０４　ＮＩＣ
５０５　画像処理部
５０６　音声処理部
５０７　ＤＶＤ－ＲＯＭドライブ
５０８　インターフェース
５０９　外部メモリ
５１０　コントローラ

５１１　モニタ

５１２　スピーカ

６０１　検索欄

６０２　検索ボタン

請求の範囲

[請求項1] シード文字列を受け付ける受付部、

前記受け付けられたシード文字列を含む文書を検索して、当該検索された文書のスニペットを得る検索部、

前記得られたスニペットを所定のセグメント区切文字列で区切ることにより、前記受け付けられたシード文字列の前後に出現する文字列と、当該シード文字列とを出現順に並べた文字列からなるセグメントを得るセグメント取得部、

前記得られたセグメントのそれぞれを、所定のセグメント要素区切文字列で区切ることにより、セグメント要素を得るセグメント要素取得部、

前記得られたセグメントのそれぞれのセグメントスコアを、当該セグメントに出現するセグメント要素のそれぞれの長さの分散もしくは標準偏差に基づいて計算するセグメントスコア計算部、

前記得られたセグメントのそれぞれに含まれるセグメント要素のそれぞれのセグメント要素スコアを、当該セグメントにおいて前記受け付けられたシード文字列が出現する位置と当該セグメントにおいて当該セグメント要素が出現する位置との距離、ならびに、当該セグメントについて計算されたセグメントスコアに基づいて計算するセグメント要素スコア計算部、

前記得られたセグメント要素のそれぞれについて計算されたセグメント要素スコアに基づいて、当該セグメント要素からいずれかを、前記受け付けられたシード文字列を含む集合を拡張した拡張集合に含まれるインスタンスの候補として選択する選択部、

を備えることを特徴とする集合拡張処理装置。

[請求項2] 請求項１に記載の集合拡張処理装置であって、

前記インスタンスの候補を用いて検索することにより得られたスニペットから、前記抽出されたインスタンスの候補を含むｎグラムの接

続グラフを生成し、当該接続グラフにおける前記受け付けられたシード文字列の前後の文脈と当該インスタンスの候補の前後の文脈とに基づいて当該シード文字列と当該インスタンスの候補との類似度を計算し、当該類似度に基づいて、当該インスタンスの候補から、当該シード文字列を含む集合を拡張した拡張集合に含めるべきインスタンスを抽出する抽出部

をさらに備えることを特徴とする集合拡張処理装置。

[請求項3] 請求項１又は２に記載の集合拡張処理装置であって、

前記得られたセグメントのそれぞれについて、当該セグメントに出現するセグメント要素のそれぞれの長さの標準偏差が所定の閾値を超える場合、前記セグメントスコアならびに前記セグメント要素スコアは、当該セグメントに含まれるセグメント要素が前記インスタンスの候補として前記選択部により選択されることがないような値となる

ことを特徴とする集合拡張処理装置。

[請求項4] 請求項１に記載の集合拡張処理装置であって、

前記得られたセグメントのそれぞれに出現するセグメント要素のそれぞれのセグメント要素スコアは、当該セグメントにおいて前記受け付けられたシード文字列が出現する位置と当該セグメントにおいて当該セグメント要素が出現する位置との最短距離に対して指数的に減衰する

ことを特徴とする集合拡張処理装置。

[請求項5] 受付部と、検索部と、セグメント取得部と、セグメント要素取得部と、セグメントスコア計算部と、セグメント要素スコア計算部と、選択部と、を備える集合拡張処理装置が実行する集合拡張処理方法であって、

前記受付部が、シード文字列を受け付ける受付工程、

前記検索部が、前記受け付けられたシード文字列を含む文書を検索して、当該検索された文書のスニペットを得る検索工程、

前記セグメント取得部が、前記得られたスニペットを所定のセグメント区切文字列で区切ることにより、前記受け付けられたシード文字列の前後に出現する文字列と、当該シード文字列とを出現順に並べた文字列からなるセグメントを得るセグメント取得工程、

前記セグメント要素取得部が、前記得られたセグメントのそれぞれを、所定のセグメント要素区切文字列で区切ることにより、セグメント要素を得るセグメント要素取得工程、

前記セグメントスコア計算部が、前記得られたセグメントのそれぞれのセグメントスコアを、当該セグメントに出現するセグメント要素のそれぞれの長さの分散もしくは標準偏差に基づいて計算するセグメントスコア計算工程、

前記セグメント要素スコア計算部が、前記得られたセグメントのそれぞれに含まれるセグメント要素のそれぞれのセグメント要素スコアを、当該セグメントにおいて前記受け付けられたシード文字列が出現する位置と当該セグメントにおいて当該セグメント要素が出現する位置との距離、ならびに、当該セグメントについて計算されたセグメントスコアに基づいて計算するセグメント要素スコア計算工程、

前記選択部が、前記得られたセグメント要素のそれぞれについて計算されたセグメント要素スコアに基づいて、当該セグメント要素からいずれかを、前記受け付けられたシード文字列を含む集合を拡張した拡張集合に含まれるインスタンスの候補として選択する選択工程、

を備えることを特徴とする集合拡張処理方法。

［請求項6］ コンピュータを、

シード文字列を受け付ける受付部、

前記受け付けられたシード文字列を含む文書を検索して、当該検索された文書のスニペットを得る検索部、

前記得られたスニペットを所定のセグメント区切文字列で区切ることにより、前記受け付けられたシード文字列の前後に出現する文字列

と、当該シード文字列とを出現順に並べた文字列からなるセグメントを得るセグメント取得部、

前記得られたセグメントのそれぞれを、所定のセグメント要素区切文字列で区切ることにより、セグメント要素を得るセグメント要素取得部、

前記得られたセグメントのそれぞれのセグメントスコアを、当該セグメントに出現するセグメント要素のそれぞれの長さの分散もしくは標準偏差に基づいて計算するセグメントスコア計算部、

前記得られたセグメントのそれぞれに含まれるセグメント要素のそれぞれのセグメント要素スコアを、当該セグメントにおいて前記受け付けられたシード文字列が出現する位置と当該セグメントにおいて当該セグメント要素が出現する位置との距離、ならびに、当該セグメントについて計算されたセグメントスコアに基づいて計算するセグメント要素スコア計算部、

前記得られたセグメント要素のそれぞれについて計算されたセグメント要素スコアに基づいて、当該セグメント要素からいずれかを、前記受け付けられたシード文字列を含む集合を拡張した拡張集合に含まれるインスタンスの候補として選択する選択部、

として機能させることを特徴とするプログラム。

[請求項7] コンピュータを、

シード文字列を受け付ける受付部、

前記受け付けられたシード文字列を含む文書を検索して、当該検索された文書のスニペットを得る検索部、

前記得られたスニペットを所定のセグメント区切文字列で区切ることにより、前記受け付けられたシード文字列の前後に出現する文字列と、当該シード文字列とを出現順に並べた文字列からなるセグメントを得るセグメント取得部、

前記得られたセグメントのそれぞれを、所定のセグメント要素区切

文字列で区切ることにより、セグメント要素を得るセグメント要素取得部、

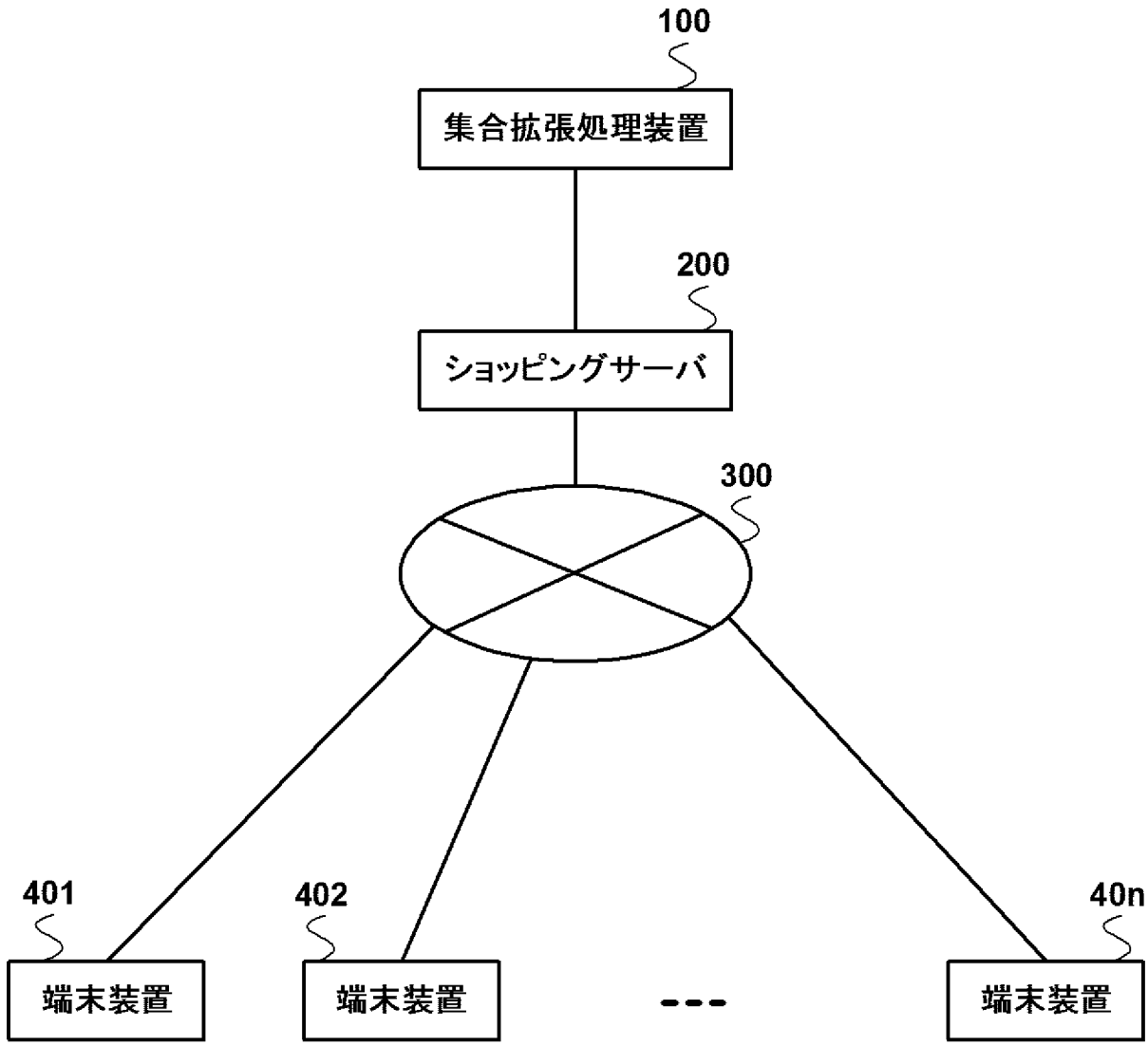
前記得られたセグメントのそれぞれのセグメントスコアを、当該セグメントに出現するセグメント要素のそれぞれの長さの分散もしくは標準偏差に基づいて計算するセグメントスコア計算部、

前記得られたセグメントのそれぞれに含まれるセグメント要素のそれぞれのセグメント要素スコアを、当該セグメントにおいて前記受け付けられたシード文字列が出現する位置と当該セグメントにおいて当該セグメント要素が出現する位置との距離、ならびに、当該セグメントについて計算されたセグメントスコアに基づいて計算するセグメント要素スコア計算部、

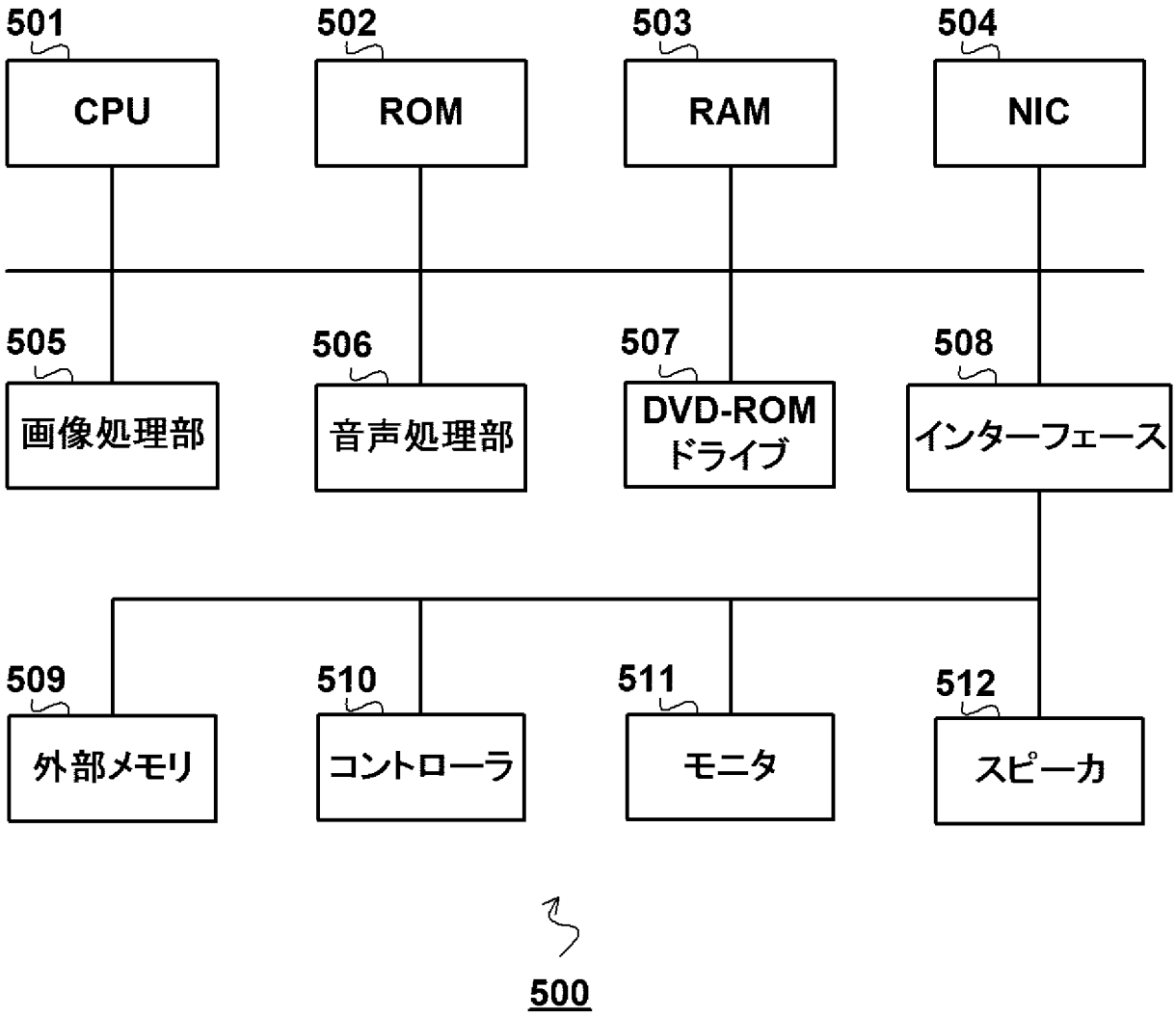
前記得られたセグメント要素のそれぞれについて計算されたセグメント要素スコアに基づいて、当該セグメント要素からいずれかを、前記受け付けられたシード文字列を含む集合を拡張した拡張集合に含まれるインスタンスの候補として選択する選択部、

として機能させることを特徴とするプログラムを記録した非一時的なコンピュータ読み取り可能な記録媒体。

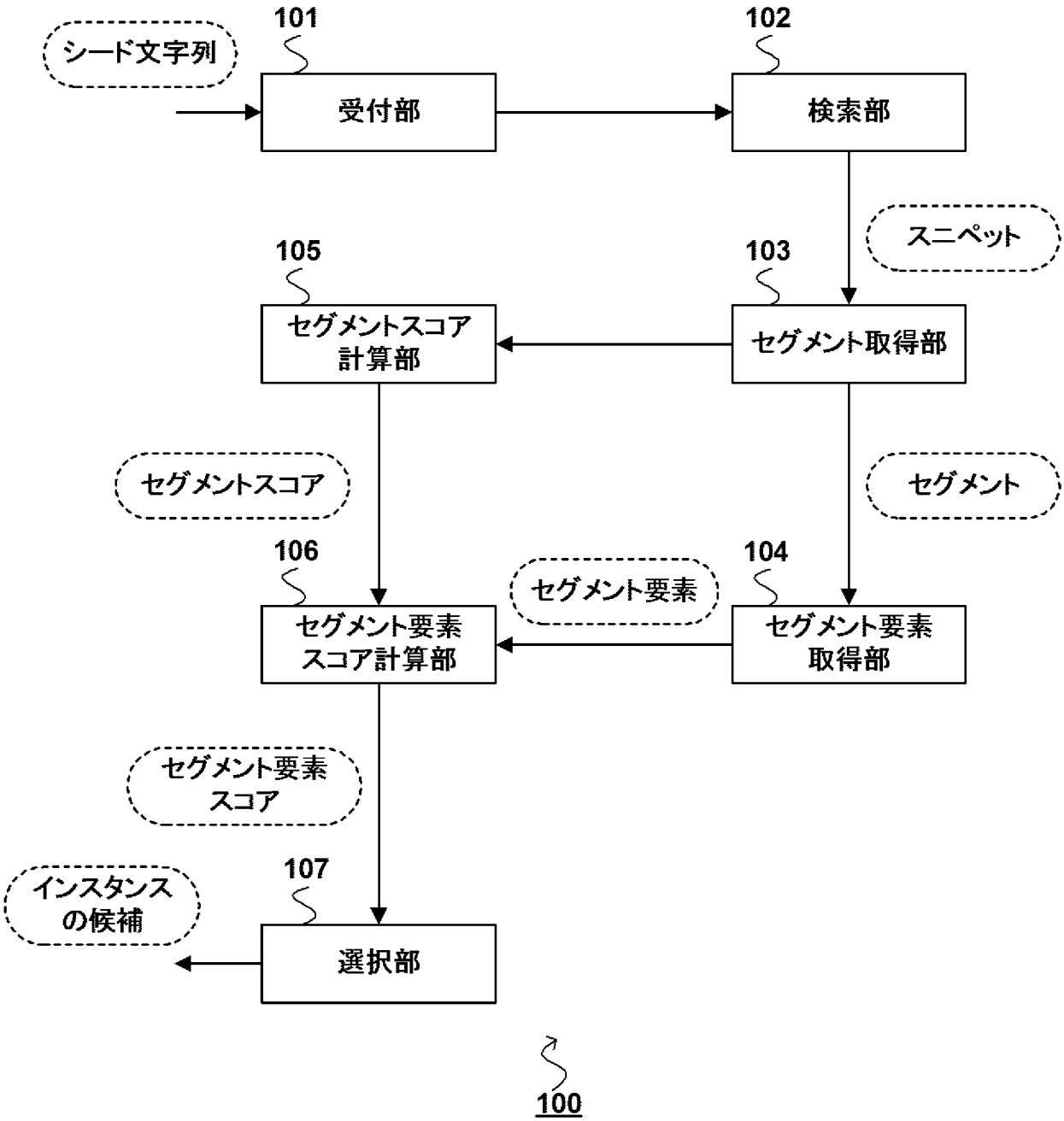
[図1]



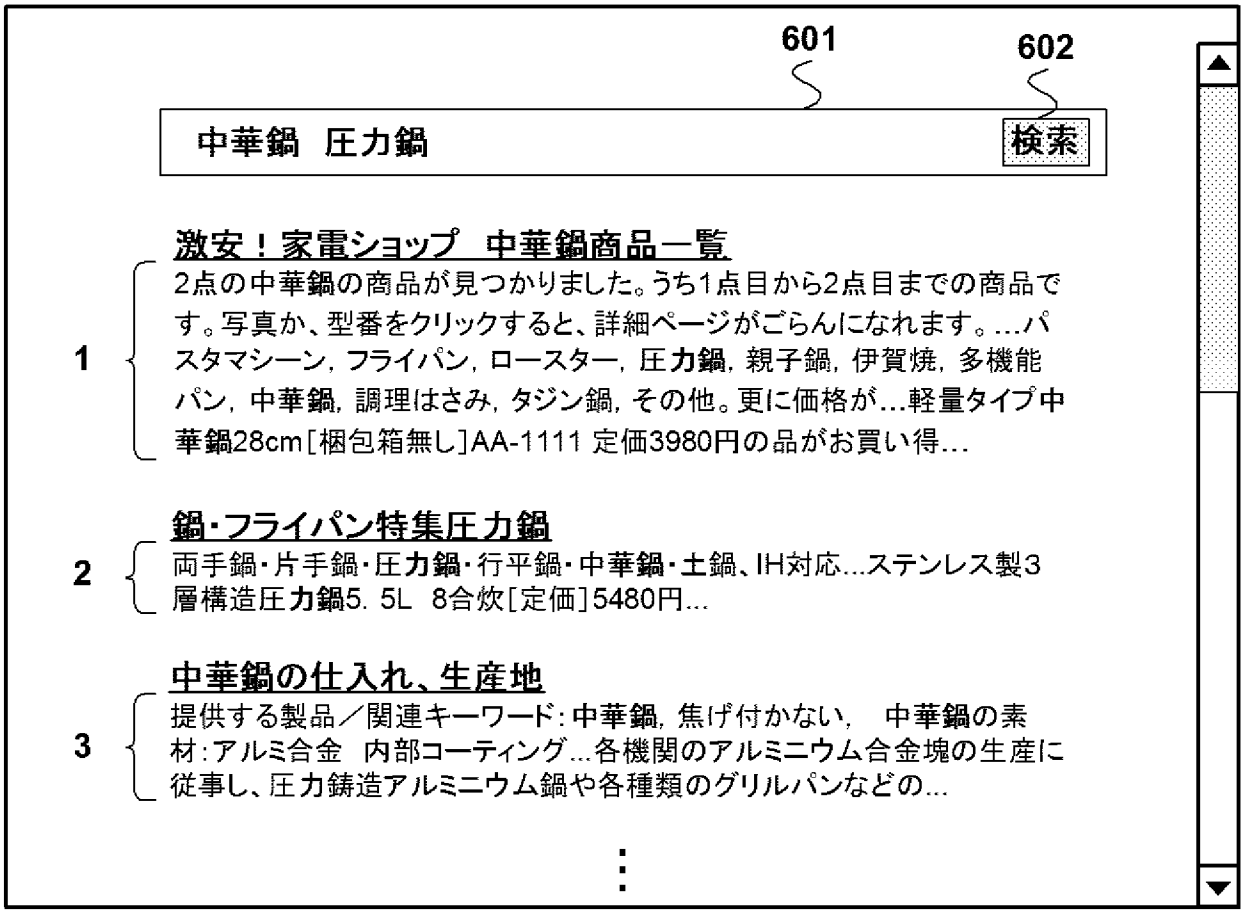
[図2]


501
CPU
502
ROM
503
RAM
504
NIC
505
画像処理部
506
音声処理部
507
DVD-ROM
ドライブ
508
インターフェース
509
外部メモリ
510
コントローラ
511
モニタ
512
スピーカ
500

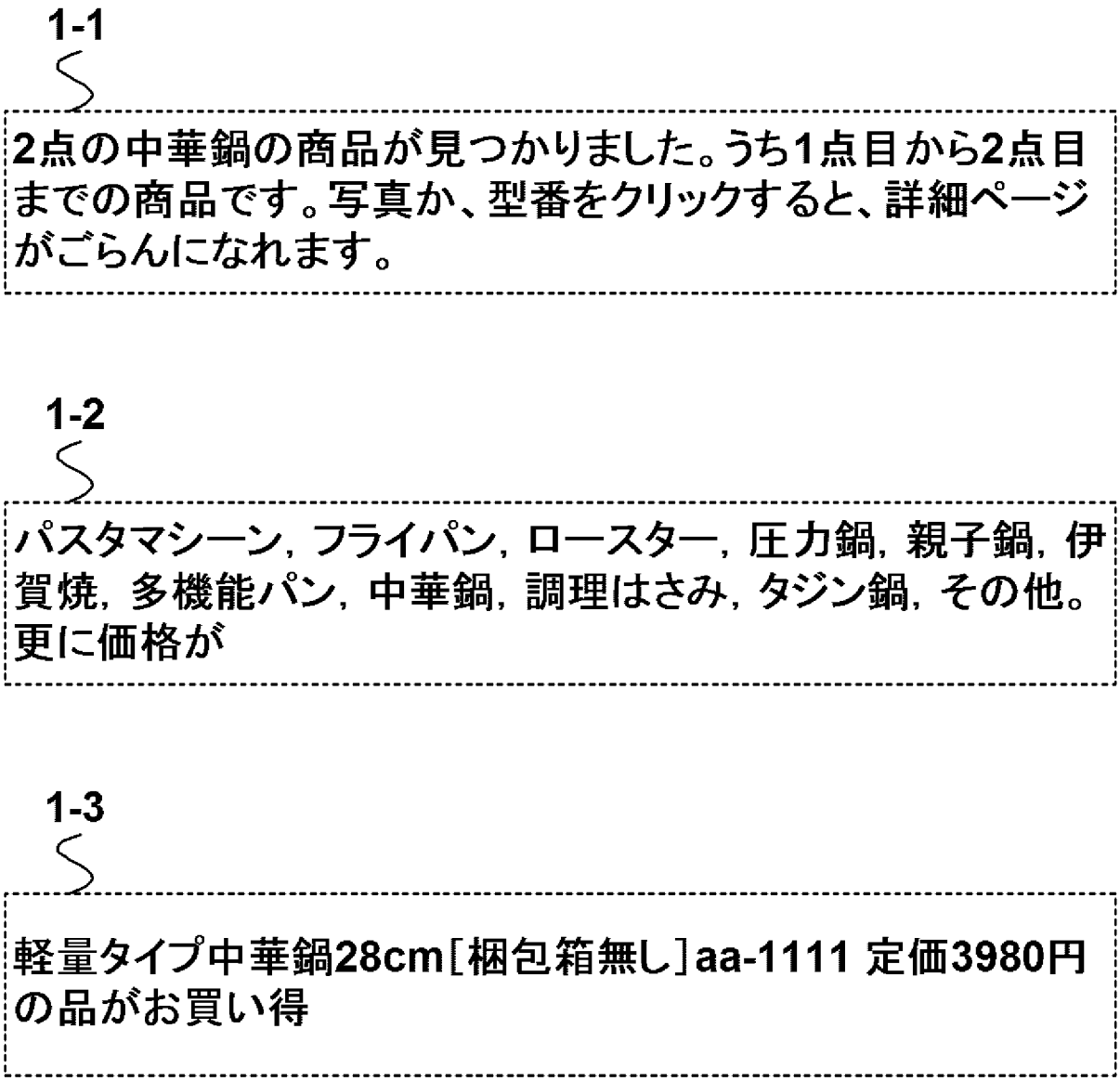
[図3]


シード文字列
101
受付部
102
検索部
スニペット
105
セグメントスコア計算部
103
セグメント取得部
セグメントスコア
セグメント
106
セグメント要素スコア計算部
セグメント要素
104
セグメント要素取得部
セグメント要素スコア
107
インスタンスの候補
選択部
100

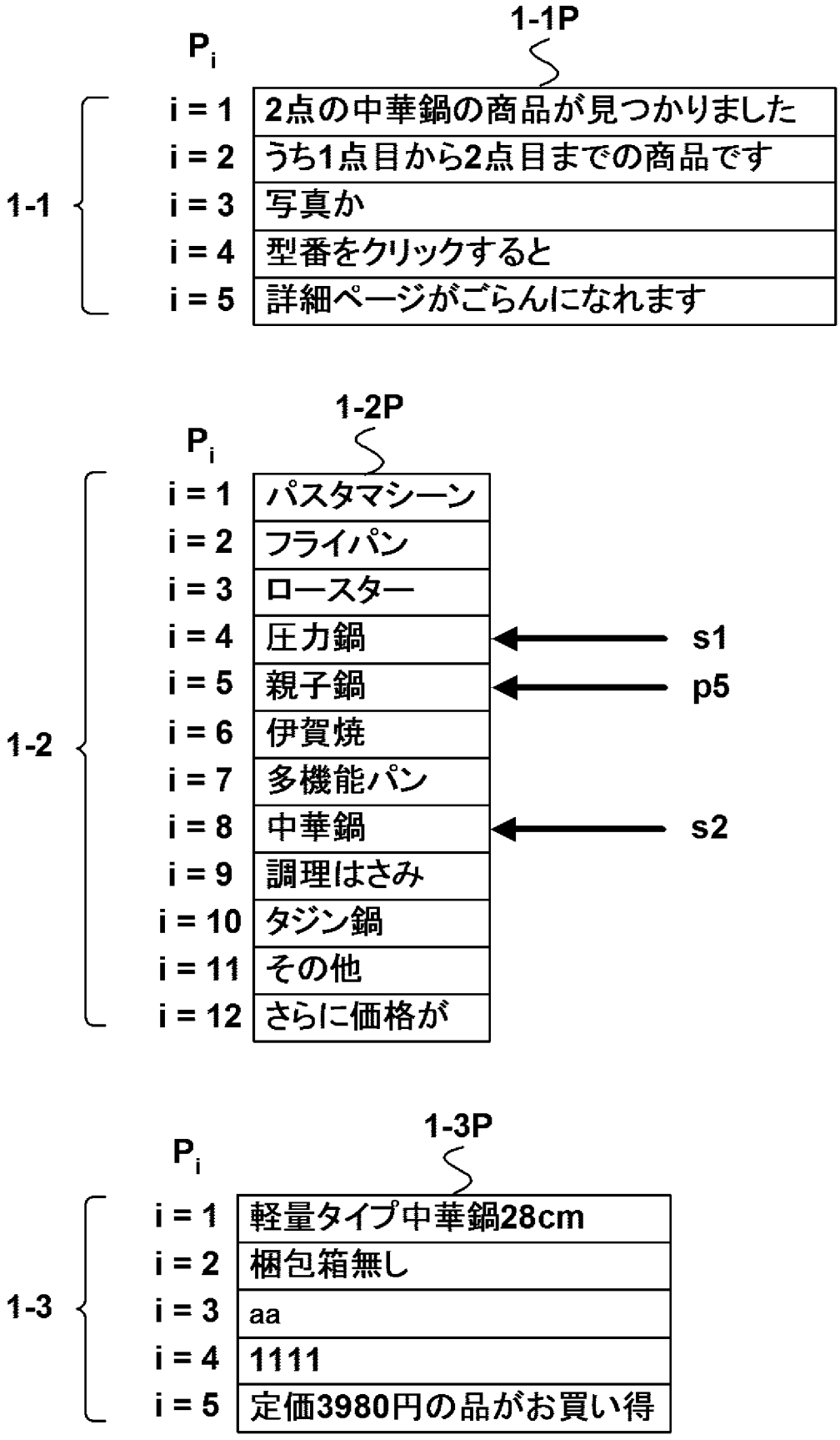
[図4]


601
602
中華鍋 圧力鍋
検索
激安！家電ショップ 中華鍋商品一覧
2点の中華鍋の商品が見つかりました。うち1点目から2点目までの商品です。写真か、型番をクリックすると、詳細ページがごらんになれます。...パスタマシーン, フライパン, ロースター, 圧力鍋, 親子鍋, 伊賀焼, 多機能パン, 中華鍋, 調理はさみ, タジン鍋, その他。更に価格が...軽量タイプ中華鍋28cm[梱包箱無し]AA-1111 定価3980円の品がお買い得...
1
鍋・フライパン特集圧力鍋
両手鍋・片手鍋・圧力鍋・行平鍋・中華鍋・土鍋、IH対応...ステンレス製3層構造圧力鍋5. 5L 8合炊[定価]5480円...
2
中華鍋の仕入れ、生産地
提供する製品／関連キーワード:中華鍋, 焦げ付かない, 中華鍋の素材:アルミ合金 内部コーティング...各機関のアルミニウム合金塊の生産に従事し、圧力鋳造アルミニウム鍋や各種類のグリルパンなどの...
3

[図5]


1-1
2点の中華鍋の商品が見つかりました。うち1点目から2点目までの商品です。写真か、型番をクリックすると、詳細ページがごらんになれます。
1-2
パスタマシーン，フライパン，ロースター，圧力鍋，親子鍋，伊賀焼，多機能パン，中華鍋，調理はさみ，タジン鍋，その他。更に価格が
1-3
軽量タイプ中華鍋28cm[梱包箱無し]aa-1111 定価3980円の品がお買い得

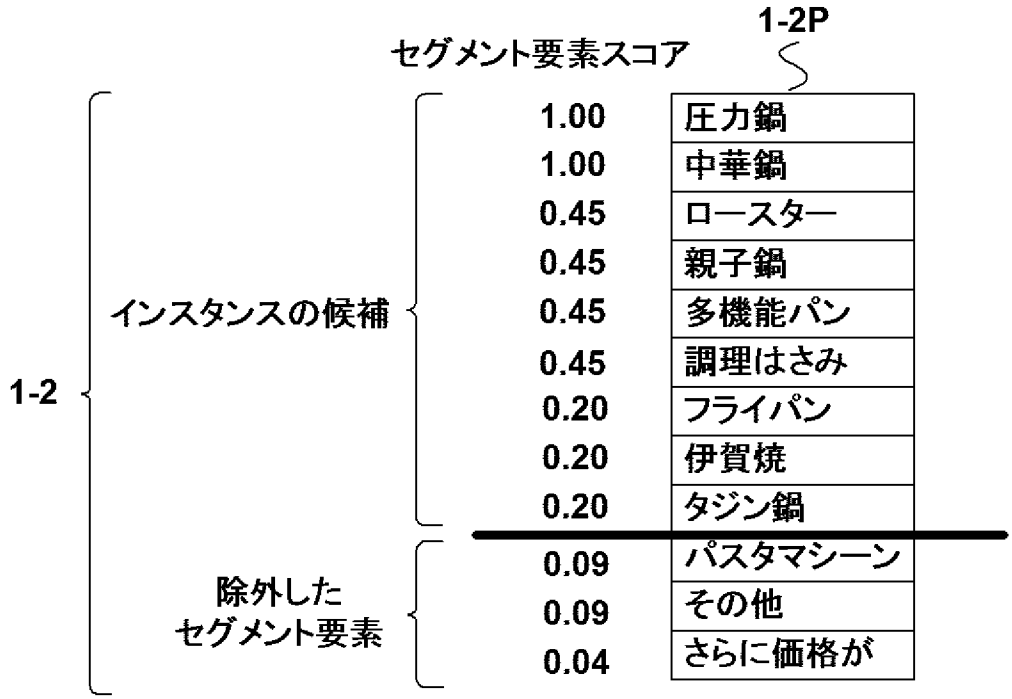
[図6]


1-1P
Pi
1-1
i = 1
2点の中華鍋の商品が見つかりました
i = 2
うち1点目から2点目までの商品です
i = 3
写真か
i = 4
型番をクリックすると
i = 5
詳細ページがごらんになれます
1-2P
Pi
1-2
i = 1
パスタマシーン
i = 2
フライパン
i = 3
ロースター
i = 4
圧力鍋
s1
i = 5
親子鍋
p5
i = 6
伊賀焼
i = 7
多機能パン
i = 8
中華鍋
s2
i = 9
調理はさみ
i = 10
タジン鍋
i = 11
その他
i = 12
さらに価格が
1-3P
Pi
1-3
i = 1
軽量タイプ中華鍋28cm
i = 2
梱包箱無し
i = 3
aa
i = 4
1111
i = 5
定価3980円の品がお買い得

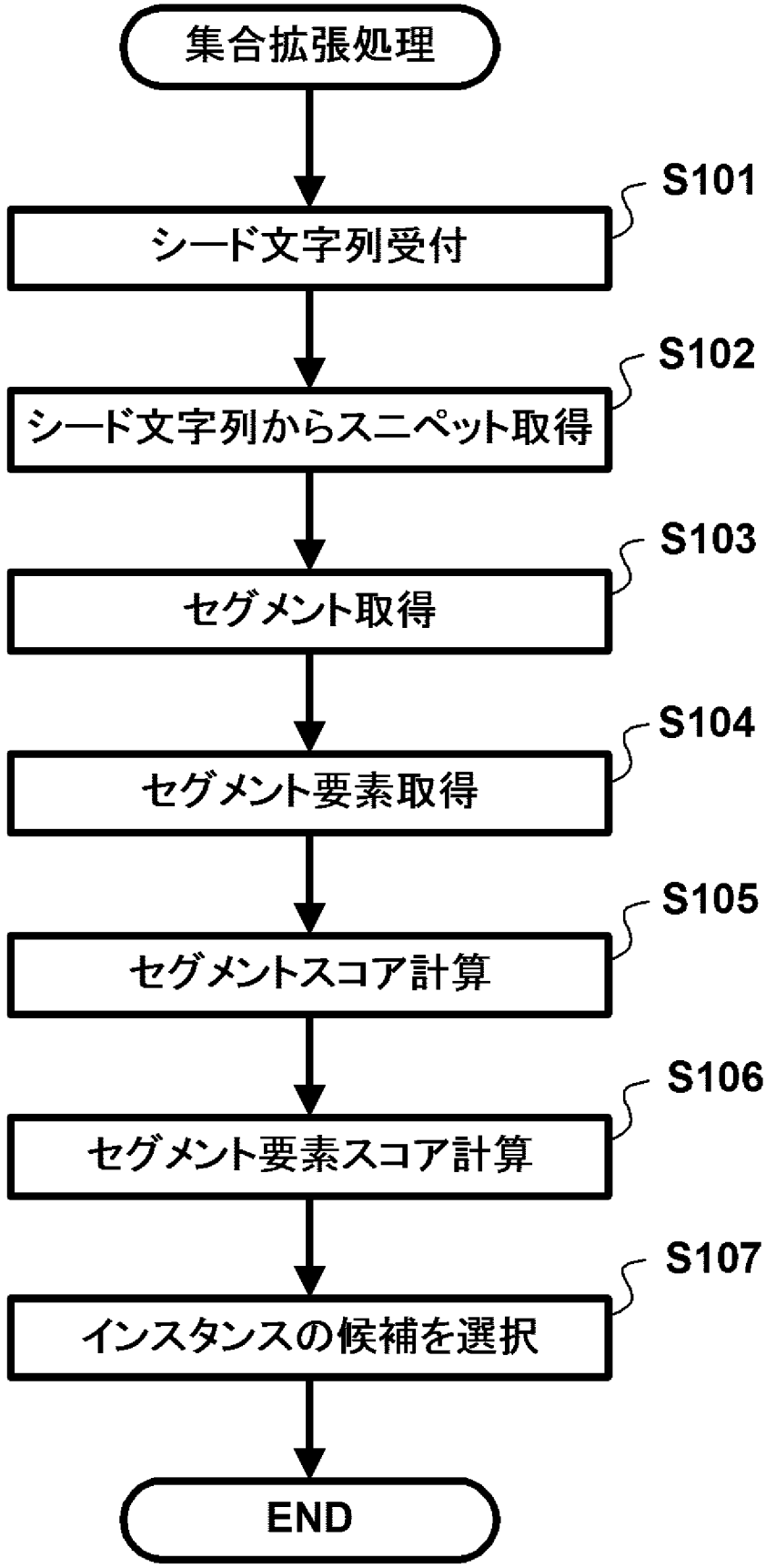
[図7]

701a	702a	703a	704a	705a	706a	707a
スニペット	セグメント	セグメント要素	長さ	標準偏差	セグメントスコア	セグメント要素スコア
1	1-1	P1	17	5.89	5.00	0
		P2	17			0
		P3	3			0
		P4	7			0
		P5	14			0
	1-2	P1	7	1.34	1.34	0.09
		P2	5			0.20
		P3	5			0.45
		P4	3			1.00
		P5	3			0.45
		P6	3			0.20
		P7	5			0.45
		P8	3			1.00
		P9	5			0.45
		P10	3			0.20
		P11	3			0.09
		P12	5			0.04
	1-3	P1	12	5.27	5.00	0
		:	:			:
		P5	14			0
2	:	:	:	:	:	:
3	:	:	:	:	:	:
:	:	:	:	:	:	:
300	:	:	:	:	:	:

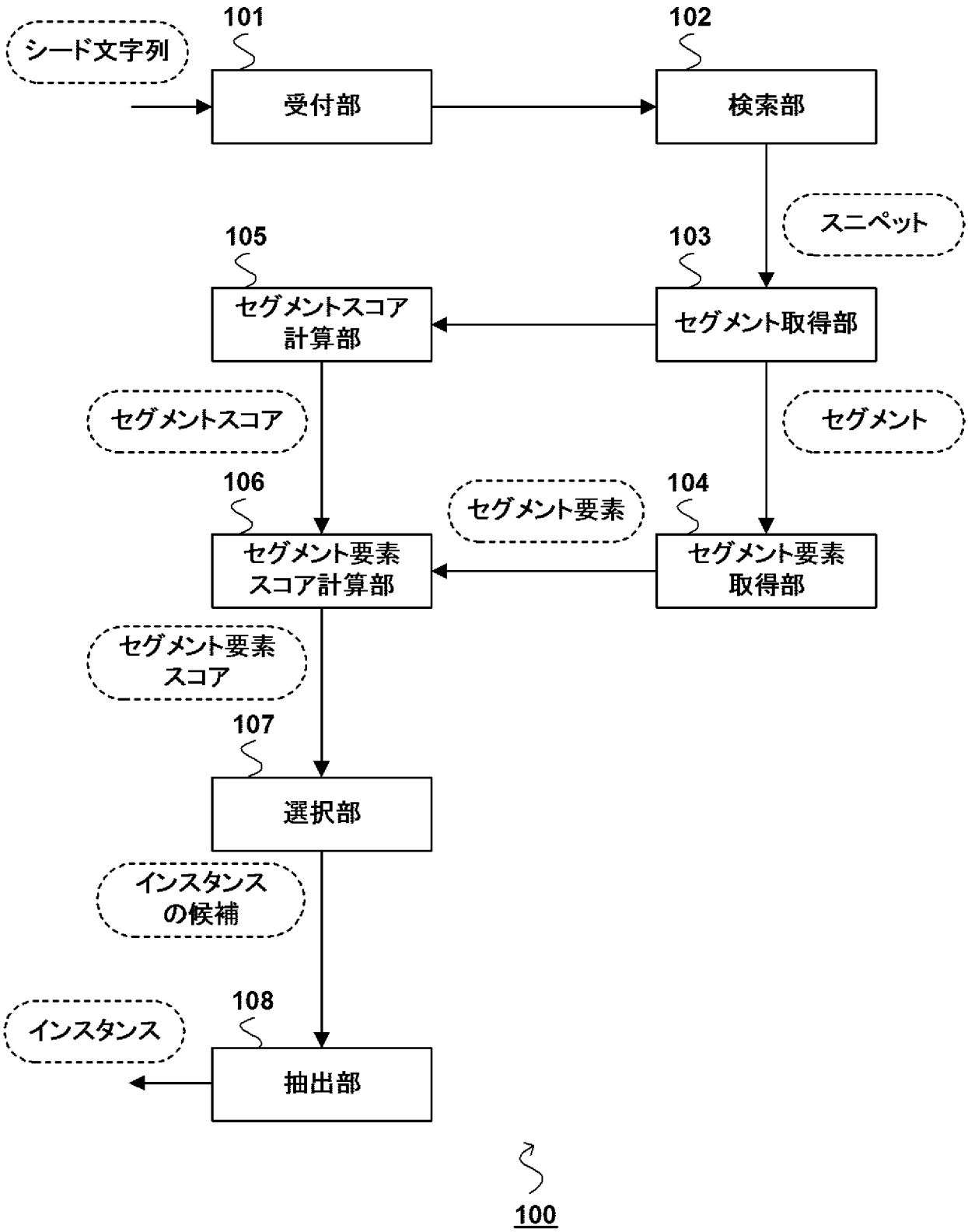
[図8]


1-2P
セグメント要素スコア
1-2
インスタンスの候補
除外した
セグメント要素
1.00 圧力鍋
1.00 中華鍋
0.45 ロースター
0.45 親子鍋
0.45 多機能パン
0.45 調理はさみ
0.20 フライパン
0.20 伊賀焼
0.20 タジン鍋
0.09 パスタマシーン
0.09 その他
0.04 さらに価格が

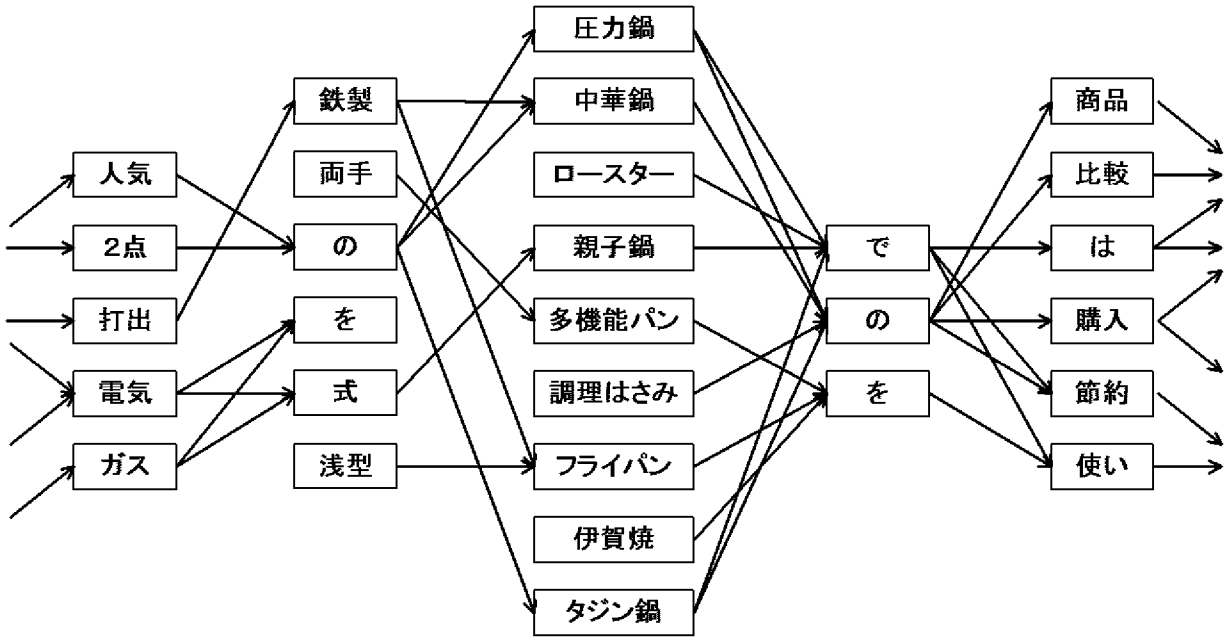
[図9]


集合拡張処理
S101
シード文字列受付
S102
シード文字列からスニペット取得
S103
セグメント取得
S104
セグメント要素取得
S105
セグメントスコア計算
S106
セグメント要素スコア計算
S107
インスタンスの候補を選択
END

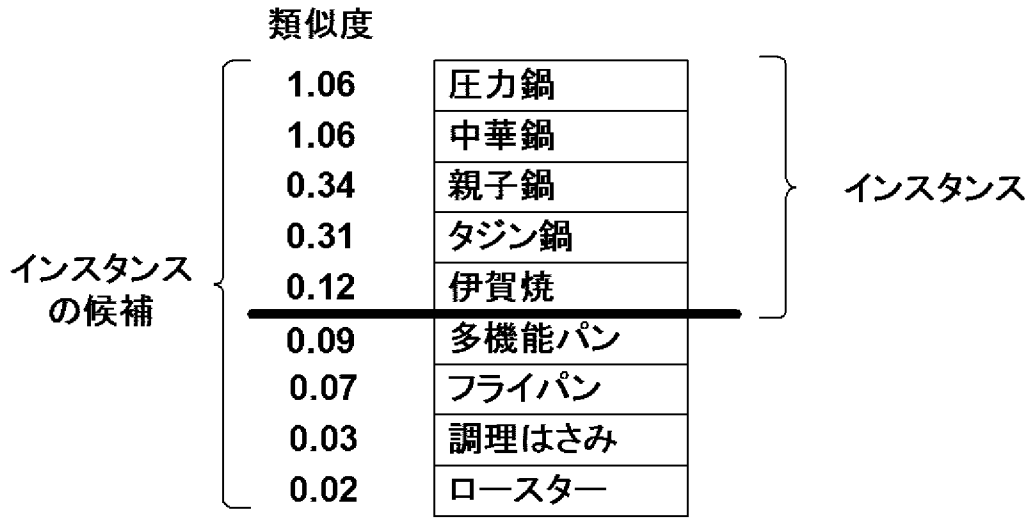
[図10]


シード文字列
101
受付部
102
検索部
スニペット
105
セグメントスコア計算部
103
セグメント取得部
セグメントスコア
セグメント
106
セグメント要素スコア計算部
セグメント要素
104
セグメント要素取得部
セグメント要素スコア
107
選択部
インスタンスの候補
108
インスタンス
抽出部
100

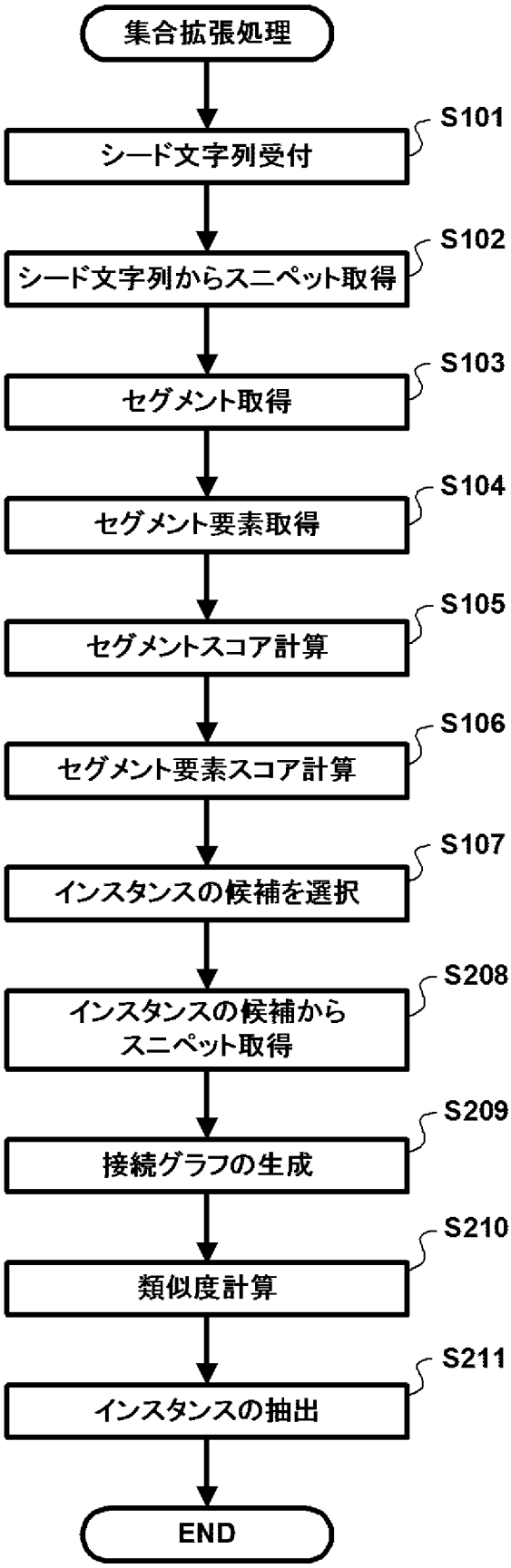
[図11]


人気
2点
打出
電気
ガス
鉄製
両手
の
を
式
浅型
圧力鍋
中華鍋
ロースター
親子鍋
多機能パン
調理はさみ
フライパン
伊賀焼
タジン鍋
で
の
を
商品
比較
は
購入
節約
使い

[図12]


類似度
1.06 圧力鍋
1.06 中華鍋
0.34 親子鍋
0.31 タジン鍋
0.12 伊賀焼
0.09 多機能パン
0.07 フライパン
0.03 調理はさみ
0.02 ロースター
インスタンス
インスタンス
の候補

[図13]


集合拡張処理
S101
シード文字列受付
S102
シード文字列からスニペット取得
S103
セグメント取得
S104
セグメント要素取得
S105
セグメントスコア計算
S106
セグメント要素スコア計算
S107
インスタンスの候補を選択
S208
インスタンスの候補から
スニペット取得
S209
接続グラフの生成
S210
類似度計算
S211
インスタンスの抽出
END

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2012/054211

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	1996-2012
Kokai Jitsuyo Shinan Koho	1971-2012	Toroku Jitsuyo Shinan Koho	1994-2012

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 2010-055164 A (Nippon Telegraph and Telephone Corp.), 11 March 2010 (11.03.2010), entire text; all drawings (Family: none)	1-7
A	JP 2010-198269 A (Yahoo Japan Corp.), 09 September 2010 (09.09.2010), entire text; all drawings (Family: none)	1-7
A	JP 2009-110231 A (Nippon Telegraph and Telephone Corp.), 21 May 2009 (21.05.2009), entire text; all drawings (Family: none)	1-7

☒ 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	Date of mailing of the international search report
10 May, 2012 (10.05.12)	22 May, 2012 (22.05.12)
Name and mailing address of the ISA/ Japanese Patent Office Facsimile No.	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2012/054211

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2010-123036 A (Nippon Telegraph and Telephone Corp.), 03 June 2010 (03.06.2010), entire text; all drawings (Family: none)	1-7
A	JP 4-293161 A (Hitachi, Ltd.), 16 October 1992 (16.10.1992), entire text; all drawings & US 5757983 A	1-7
A	Shin'ya MURATA et al., "Ranking search results based on information needs in conjunction with click log analysis", Database Society of Japan Ronbunshi, 27 March 2009 (27.03.2009), vol.7, no.4, pages 37 to 42	1-7

国際調査報告

国際出願番号　ＰＣＴ／ＪＰ２０１２／０５４２１１

Ａ．　発明の属する分野の分類（国際特許分類（ＩＰＣ））

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（ＩＰＣ））

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	１９２２－１９９６年
日本国公開実用新案公報	１９７１－２０１２年
日本国実用新案登録公報	１９９６－２０１２年
日本国登録実用新案公報	１９９４－２０１２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 2010-055164 A (日本電信電話株式会社) 2010.03.11, 全文, 全図 (ファミリーなし)	1-7
A	JP 2010-198269 A (ヤフー株式会社) 2010.09.09, 全文, 全図 (ファミリーなし)	1-7

☑　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	国際調査報告の発送日
１０．０５．２０１２	２２．０５．２０１２

国際調査機関の名称及びあて先
日本国特許庁（ＩＳＡ／ＪＰ）
郵便番号１００－８９１５
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）　５Ｍ　４７７４
齊藤　貴孝
電話番号　０３－３５８１－１１０１　内線　３５９９

Ｃ（続き）．　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 2009-110231 A (日本電信電話株式会社) 2009.05.21, 全文, 全図 (ファミリーなし)	1-7
A	JP 2010-123036 A (日本電信電話株式会社) 2010.06.03, 全文, 全図 (ファミリーなし)	1-7
A	JP 4-293161 A (株式会社日立製作所) 1992.10.16, 全文, 全図 & US 5757983 A	1-7
A	村田　眞哉、外３名，クリックログ解析による情報要求ベースの検索結果ランキング，日本データベース学会論文誌，2009.03.27，第７巻，第４号，ｐ．３７－４２	1-7