

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2017-54434
(P2017-54434A)

(43) 公開日 平成29年3月16日(2017.3.16)

(51) Int.Cl.	F I	テーマコード(参考)
G06F 17/28 (2006.01)	G06F 17/28 681	5B091
G06F 17/27 (2006.01)	G06F 17/27 695	5B109
G06F 17/24 (2006.01)	G06F 17/24	

審査請求 未請求 請求項の数 7 O L (全 14 頁)

(21) 出願番号	特願2015-179839 (P2015-179839)	(71) 出願人	000004226 日本電信電話株式会社 東京都千代田区大手町一丁目5番1号
(22) 出願日	平成27年9月11日(2015.9.11)	(71) 出願人	504132272 国立大学法人京都大学 京都府京都市左京区吉田本町36番地1
		(74) 代理人	110001519 特許業務法人太陽国際特許事務所
		(72) 発明者	東中 竜一郎 東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内
		(72) 発明者	松尾 義博 東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内

最終頁に続く

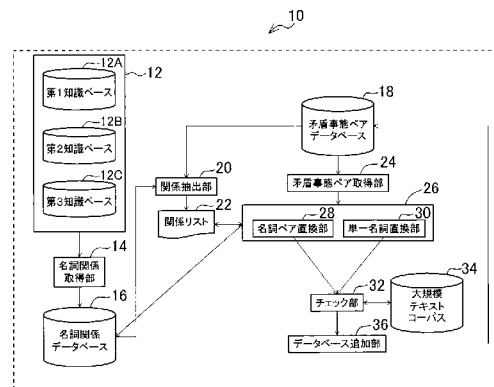
(54) 【発明の名称】 矛盾作成装置、方法、及びプログラム

(57) 【要約】

【課題】新しい矛盾事態ペアを作成できる。

【解決手段】関係抽出部20が、矛盾事態ペアデータベース18と、名詞関係データベース16とを照合することで、矛盾に関する名詞ペアの関係を抽出し、抽出した名詞ペアの関係を格納した関係リストを作成する。名詞ペア置換部28が、矛盾事態ペアデータベース18から得られる矛盾事態ペアについて、矛盾事態ペアに含まれる名詞ペアの関係が、関係リスト22に含まれる場合、名詞ペアの関係と同じ関係を持つ名詞ペアを名詞関係データベース16から取得し、取得した名詞ペアによって前記矛盾事態ペアの名詞ペアを置き換えることで矛盾事態ペアを作成する。単一名詞置換部30が、名詞関係データベース16から、取得した矛盾事態ペアが持つ共有名詞の同位語を抽出し、取得した同位語により矛盾事態ペアの共有名詞を置き換えることで、矛盾事態ペアを作成する。

【選択図】図1



【特許請求の範囲】**【請求項 1】**

矛盾する事態を表す文のペアである矛盾事態ペアを格納した矛盾事態ペアデータベースと、名詞ペア及び前記名詞ペアの関係を格納した名詞関係データベースとを照合することで、矛盾に関する名詞ペアの関係を抽出し、抽出した名詞ペアの関係を格納した関係リストを作成する関係抽出部と、

前記矛盾事態ペアデータベースから得られる前記矛盾事態ペアについて、前記矛盾事態ペアに含まれる名詞ペアの関係が、前記関係リストに含まれる場合、前記名詞ペアの関係と同じ関係を持つ名詞ペアを前記名詞関係データベースから取得し、前記取得した名詞ペアによって前記矛盾事態ペアの名詞ペアを置き換えることで矛盾事態ペアを作成する名詞ペア置換部と、

を含む矛盾作成装置。

【請求項 2】

矛盾する事態を表す文のペアである矛盾事態ペアを格納した矛盾事態ペアデータベースから、前件と後件とで共通する共有名詞を持つ矛盾事態ペアを取得する矛盾事態ペア取得部と、

名詞ペア及び前記名詞ペアの関係を格納した名詞関係データベースから、前記矛盾事態ペア取得部によって取得した矛盾事態ペアが持つ共有名詞の同位語を抽出し、前記抽出した同位語により前記矛盾事態ペアの共有名詞を置き換えることで、矛盾事態ペアを作成する単一名詞置換部と、

を含む矛盾作成装置。

【請求項 3】

前記作成された矛盾事態ペアに含まれる前記事態を表す文の各々について、大規模テキストコーパスにおける前記事態を表す文の頻度が閾値以上であるかをチェックして、前記作成された矛盾事態ペアに含まれる前記事態を表す文の各々の頻度が閾値以上であれば、前記矛盾事態ペアが適切であると判断するチェック部を更に含む請求項 1 又は 2 記載の矛盾作成装置。

【請求項 4】

関係抽出部及び名詞ペア置換部を含む矛盾作成装置における矛盾作成方法であって、

前記関係抽出部が、矛盾する事態を表す文のペアである矛盾事態ペアを格納した矛盾事態ペアデータベースと、名詞ペア及び前記名詞ペアの関係を格納した名詞関係データベースとを照合することで、矛盾に関する名詞ペアの関係を抽出し、抽出した名詞ペアの関係を格納した関係リストを作成し、

前記名詞ペア置換部が、前記矛盾事態ペアデータベースから得られる前記矛盾事態ペアについて、前記矛盾事態ペアに含まれる名詞ペアの関係が、前記関係リストに含まれる場合、前記名詞ペアの関係と同じ関係を持つ名詞ペアを前記名詞関係データベースから取得し、前記取得した名詞ペアによって前記矛盾事態ペアの名詞ペアを置き換えることで矛盾事態ペアを作成する

矛盾作成方法。

【請求項 5】

矛盾事態ペア取得部及び単一名詞置換部を含む矛盾作成装置における矛盾作成方法であって、

前記矛盾事態ペア取得部が、矛盾する事態を表す文のペアである矛盾事態ペアを格納した矛盾事態ペアデータベースから、前件と後件とで共通する共有名詞を持つ矛盾事態ペアを取得し、

前記単一名詞置換部が、名詞ペア及び前記名詞ペアの関係を格納した名詞関係データベースから、前記矛盾事態ペア取得部によって取得した矛盾事態ペアが持つ共有名詞の同位語を抽出し、前記抽出した同位語により前記矛盾事態ペアの共有名詞を置き換えることで、矛盾事態ペアを作成する

矛盾作成方法。

10

20

30

40

50

【請求項 6】

チェック部が、前記作成された矛盾事態ペアに含まれる前記事態を表す文の各々について、大規模テキストコーパスにおける前記事態を表す文の頻度が閾値以上であるかをチェックして、前記作成された矛盾事態ペアに含まれる前記事態を表す文の各々の頻度が閾値以上であれば、前記矛盾事態ペアが適切であると判断することを更に含む請求項 4 又は 5 記載の矛盾作成方法。

【請求項 7】

コンピュータを、請求項 1 ~ 請求項 3 の何れか 1 項記載の矛盾作成装置を構成する各部として機能させるためのプログラム。

【発明の詳細な説明】

10

【技術分野】**【0001】**

本発明は、矛盾作成装置、方法、及びプログラムに係り、特に、新たな矛盾事態ペアを作成するための矛盾作成装置、方法、及びプログラムに関する。

【背景技術】**【0002】**

自然言語処理における言語生成の分野では一貫した内容を持つ文章を生成することが重要である。しかし、コンピュータの生成する文章が一貫した内容を持つかを判定することは常識的な知識が必要となり簡単ではない。

【0003】

20

たとえば、一貫性を欠く場合として、生成された文章に矛盾が含まれる場合がある。たとえば、「パリに行ったことがないが、エッフェル塔に登った」という文は矛盾を含んでいる。「パリに行ったことがない」とことと「エッフェル塔に登ったこと」は矛盾するからである。

【0004】

このような矛盾を含む文章の生成を避けるためには、矛盾の知識を用いてチェックする必要がある。

【0005】

たとえば、「パリに行ったことがない」とことと「エッフェル塔に登ったこと」が矛盾しているという知識をコンピュータが持っていれば、先の例のような文は生成されることはない。

30

【0006】

このような矛盾知識を取得する方法論として、コーパスから自動獲得する方法（非特許文献 1 参照）やクラウドソーシング（インターネット上のユーザにデータ作成を依頼する手法）を用いる方法などがある（非特許文献 2 参照）。

【先行技術文献】**【非特許文献】****【0007】**

【非特許文献 1】 Chikara Hashimoto, Kentaro Torisawa, Stijn De Saeger, Jong-Hoon Oh, and Jun'ichi Kazama. Excitatory or inhibitory: A new semantic orientation extracts contradiction and causality from the web. In Proceedings of EMNLP2012, pp. 619-630, 2012.

40

【非特許文献 2】 高畠悠, 森田一, 河原大輔, 黒橋禎夫, 東中竜一郎, 松尾義博. クラウドソーシングを活用した事態間矛盾の分析と分類, 言語処理学会第 21 回年次大会, pp.305-308, 2015.

【発明の概要】**【発明が解決しようとする課題】****【0008】**

コーパスから自動的に矛盾を抽出するとしても、クラウドソーシングを用いるとしても、獲得できる矛盾知識の量には限界がある。人間が備える常識的知識（矛盾知識を含む）

50

は膨大と考えられ、矛盾知識の量を増やすことは重要な課題である。

【0009】

本発明は上記事情を鑑みてなされたものであり、新しい矛盾事態ペアを作成できる矛盾作成装置、方法、及びプログラム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】

【0010】

上記目的を達成するために、第1の発明に係る矛盾作成装置は、矛盾する事態を表す文のペアである矛盾事態ペアを格納した矛盾事態ペアデータベースと、名詞ペア及び前記名詞ペアの関係を格納した名詞関係データベースとを照合することで、矛盾に関する名詞ペアの関係を抽出し、抽出した名詞ペアの関係を格納した関係リストを作成する関係抽出部と、前記矛盾事態ペアデータベースから得られる前記矛盾事態ペアについて、前記矛盾事態ペアに含まれる名詞ペアの関係が、前記関係リストに含まれる場合、前記名詞ペアの関係と同じ関係を持つ名詞ペアを前記名詞関係データベースから取得し、前記取得した名詞ペアによって前記矛盾事態ペアの名詞ペアを置き換えることで矛盾事態ペアを作成する名詞ペア置換部と、を含んで構成されている。

10

【0011】

第2の発明に係る矛盾作成方法は、関係抽出部及び名詞ペア置換部を含む矛盾作成装置における矛盾作成方法であって、前記関係抽出部が、矛盾する事態を表す文のペアである矛盾事態ペアを格納した矛盾事態ペアデータベースと、名詞ペア及び前記名詞ペアの関係を格納した名詞関係データベースとを照合することで、矛盾に関する名詞ペアの関係を抽出し、抽出した名詞ペアの関係を格納した関係リストを作成し、前記名詞ペア置換部が、前記矛盾事態ペアデータベースから得られる前記矛盾事態ペアについて、前記矛盾事態ペアに含まれる名詞ペアの関係が、前記関係リストに含まれる場合、前記名詞ペアの関係と同じ関係を持つ名詞ペアを前記名詞関係データベースから取得し、前記取得した名詞ペアによって前記矛盾事態ペアの名詞ペアを置き換えることで矛盾事態ペアを作成する。

20

【0012】

第1の発明及び第2の発明によれば、矛盾に関する名詞ペアの関係を抽出して関係リストを作成し、データベースから得られた矛盾事態ペアに含まれる名詞ペアの関係が、関係リストに含まれる場合、名詞ペアの関係と同じ関係を持つ名詞ペアを名詞関係データベースから取得し、取得した名詞ペアによって矛盾事態ペアの名詞ペアを置き換えることで、新しい矛盾事態ペアを作成できる。

30

【0013】

第3の発明に係る矛盾作成装置は、矛盾する事態を表す文のペアである矛盾事態ペアを格納した矛盾事態ペアデータベースから、前件と後件とで共通する共有名詞を持つ矛盾事態ペアを取得する矛盾事態ペア取得部と、名詞ペア及び前記名詞ペアの関係を格納した名詞関係データベースから、前記矛盾事態ペア取得部によって取得した矛盾事態ペアが持つ共有名詞の同位語を抽出し、前記抽出した同位語により前記矛盾事態ペアの共有名詞を置き換えることで、矛盾事態ペアを作成する単一名詞置換部と、を含んで構成されている。

【0014】

第4の発明に係る矛盾作成方法は、矛盾事態ペア取得部及び単一名詞置換部を含む矛盾作成装置における矛盾作成方法であって、前記矛盾事態ペア取得部が、矛盾する事態を表す文のペアである矛盾事態ペアを格納した矛盾事態ペアデータベースから、前件と後件とで共通する共有名詞を持つ矛盾事態ペアを取得し、前記単一名詞置換部が、名詞ペア及び前記名詞ペアの関係を格納した名詞関係データベースから、前記矛盾事態ペア取得部によって取得した矛盾事態ペアが持つ共有名詞の同位語を抽出し、前記抽出した同位語により前記矛盾事態ペアの共有名詞を置き換えることで、矛盾事態ペアを作成する。

40

【0015】

第3の発明及び第4の発明によれば、矛盾事態ペアデータベースから、共有名詞を持つ矛盾事態ペアを取得し、名詞関係データベースから、取得した矛盾事態ペアが持つ共有名詞の同位語を抽出し、取得した同位語により矛盾事態ペアの共有名詞を置き換えることで

50

、新しい矛盾事態ペアを作成できる。

【0016】

また、第5の発明に係るプログラムは、コンピュータを、上記の矛盾作成装置を構成する各部として機能させるためのプログラムである。

【発明の効果】

【0017】

以上説明したように、本発明の矛盾作成装置、方法、及びプログラムによれば、新しい矛盾事態ペアを作成できる、という効果が得られる。

【図面の簡単な説明】

【0018】

【図1】本発明の実施の形態に係る矛盾作成装置の構成を示すブロック図である。

【図2】本発明の実施の形態における矛盾作成処理ルーチンの内容を示すフローチャートである。

【図3】実験結果を示す図である。

【発明を実施するための形態】

【0019】

以下、図面を参照して本発明の実施の形態を詳細に説明する。本発明の実施の形態では、矛盾事態ペアデータベースに含まれる矛盾事態ペアを拡充する場合を例に説明する。

【0020】

< 本実施の形態の概要 >

矛盾事態ペアデータベースが存在するときに、データベース中の矛盾事態ペアを、知識ベースから得られる名詞関係データベースを用いて拡張することによって、新しい矛盾事態ペアを獲得する。

【0021】

ここで、矛盾事態ペアデータベースは矛盾事態ペアのリストである。矛盾事態ペアとは矛盾する二つの事態を表す文であり、それぞれの事態を表す文は述語項構造で表現されるとする。述語項構造は、一つの述語と一つ以上の項（名詞句）からなるとする。矛盾事態ペアの例を次に示す。

【0022】

< フランスへ行かない，エッフェル塔に行く >

< 旅行の予定を立てない，海外ツアーに申し込む >

【0023】

名詞関係データベースとは、2つの名詞と、当該2つの名詞間の関係を示したもので、既存の知識ベースから構築される。

【0024】

既存の知識ベースとは、たとえば、Freebase（登録商標）(<https://www.freebase.com/>)である。Freebase（登録商標）では、二つの物事とその関係を三つ組みで表したデータが大量に格納されている。そして、Freebase（登録商標）には名詞とその関係として以下のようなものが含まれており、それらを取得して名詞関係データベースとする。

【0025】

(エッフェル塔, contained-by, フランス)

(箱根, contained-by, 神奈川県)

【0026】

ここで、contained-by は「含まれる」という関係であり、エッフェル塔はフランスに含まれるという意味である。また二つ目の例では、箱根が神奈川県に含まれるという意味である。このような2つの名詞とその関係を表す名詞関係データベースを用いて矛盾事態ペアを新しく作成することができる。たとえば、< フランスへ行かない，エッフェル塔に行く > という矛盾事態ペアには、それぞれ項としてフランスとエッフェル塔が含まれている。そして、これらは名詞関係データベースによるとcontained-by関係にある。よって、同じくcontained-by 関係にある箱根と神奈川県によって置き換えられると仮定し、

10

20

30

40

50

< 神奈川県へ行かない, 箱根に行く > という新しい矛盾事態ペアを作成することができる。なお、ここでは、contained-by 関係のみに着目した例を示したが、名詞関係データベースに含まれる他の関係を用いてもよい。

【0027】

新しく作られた矛盾事態ペアに含まれる内容は、自動的に作成したものであるため、文法や意味的な誤りなどのノイズが含まれている可能性がある。そこで、大規模テキストコーパスに、矛盾事態ペアに含まれる内容が閾値以上の回数存在するかどうかによって、内容としての適切性をチェックする。

【0028】

< システム構成 >

本発明の実施の形態に係る矛盾作成装置10は、CPU (Central Processing Unit) と、RAM (Random Access Memory) と、後述する矛盾作成処理ルーチンを実行するための矛盾作成プログラムを記憶したROM (Read Only Memory) とを備えたコンピュータで構成されている。CPUが矛盾作成プログラムを、内部記憶装置であるROMから読み込んで実行することにより、コンピュータが矛盾作成装置10として機能する。

【0029】

図1に示すように、矛盾作成装置10は、知識ベース群12、名詞関係取得部14、名詞関係データベース16、矛盾事態ペアデータベース18、関係抽出部20、矛盾事態ペア取得部24、矛盾事態ペア作成部26、チェック部32、大規模テキストコーパス34、及びデータベース追加部36を備えている。

【0030】

知識ベース群12は、第1知識ベース12A、第2知識ベース12B、及び第3知識ベース12Cを備えている。

【0031】

本実施の形態では、第1知識ベース12Aとして、Freebase(登録商標)を用い、第2知識ベース12Bとして、Wikipedia(登録商標)を用い、第3知識ベース12Cとして、国語辞典を用いる。なお、これらの知識ベース以外の知識ベースも名詞とそれらの関係が得られるものであれば何でもよい。たとえば、NTT(登録商標)の日本語語彙大系を用いてもよい。

【0032】

名詞関係取得部14は、第1知識ベース12A~第3知識ベース12Cから、名詞ペアと当該名詞ペアの関係を取得し、名詞関係データベース16に格納する。

【0033】

Freebase(登録商標)には、基本的に2つの名詞とその関係が三つ組みとして格納されているため、これらの三つ組みをすべて抽出すればよい。Wikipedia(登録商標)、国語辞典については、辞書の見出しとその定義文をもとに、上位下位関係にある名詞を取得することが可能である。たとえば、下記の非特許文献3や非特許文献4の手法を使えばよい。例えば、国語辞典に、見出し「三輪車」の定義文として「乗り物の一つ」とあれば、三輪車の上位語が乗り物であることが分かり、「三輪車」と「乗り物」が上位下位関係であるという名詞関係が取得できる。

【0034】

[非特許文献3]: 隅田飛鳥, 吉永直樹, 鳥澤健太郎, Wikipediaの記事構造からの上位下位関係抽出, 自然言語処理, vol.16(3), pp.3-24, 2008.

【0035】

[非特許文献4]: 大福泰樹, 河原大輔, 黒橋禎夫, 大規模コーパスと国語辞典の統合的利用によるソーラスの自動構築, 言語処理学会第10回年次大会, 2004.

【0036】

Freebase(登録商標)やWikipedia(登録商標)、国語辞典から抽出された名詞関係データベース16に含まれるデータの例は以下の通りである。

【0037】

10

20

30

40

50

(エッフェル塔, contained-by, フランス)
 (箱根, contained-by, 神奈川県)
 (ゴルフ, sports-equipment, ゴルフクラブ)
 (野球, sports-equipment, バット)
 (海外ツアー, is-a, 旅行)
 (国内ツアー, is-a, 旅行)
 (TOEIC テスト, is-a, テスト)
 (テニス, is-a, 球技)
 (ゴルフ, is-a, 球技)

【0038】

最初の4つはFreebase(登録商標)によるものである。contained-byは先に示した通り、「～に含まれる」という関係を表し、sports-equipmentはスポーツとそのスポーツにて使われる道具の関係を表す。たとえば、「ゴルフ」と「ゴルフクラブ」はsports-equipmentの関係にある。以降のis-aは上位下位関係(左側が下位語, 右側が上位語)を表す。たとえば、海外ツアーは旅行の下位語である。

【0039】

また、同じ上位語に対してis-a関係をもつ下位語集合(下位語間の関係を「同位語」と呼ぶ)についても、名詞関係データベース16に登録しておく。たとえば、「旅行」に対して「海外ツアー」「国内ツアー」など、「球技」に対して「テニス」「ゴルフ」などの同位語が得られる。同位語については、以下の様に名詞関係データベース16に格納されている。

【0040】

(海外ツアー, 同位語, 国内ツアー)
 (テニス, 同位語, ゴルフ)

【0041】

矛盾事態ペアデータベース18は、矛盾する事態を表す文のペアである矛盾事態ペアを複数記憶している。

【0042】

矛盾事態ペアデータベース18を、人手で作成してもよいし、非特許文献1や非特許文献2の手法で構築してもよい。

【0043】

関係抽出部20は、矛盾事態ペアデータベース18と名詞関係データベース16を照合し、矛盾事態ペアデータベース18において閾値以上の回数出現する、矛盾に関する名詞ペアの関係を抽出し、抽出した名詞ペアの関係を格納した関係リスト22を作成する。

【0044】

具体的には、まず、矛盾事態ペアデータベース18に記憶されている、それぞれの矛盾事態(述語項構造)ペア中の項部分から名詞ペアを抽出する。ここで、名詞ペアのそれぞれの名詞は、異なる述語項構造から抽出することとし、名詞は複合名詞(2つ以上の名詞の連続)も含める。

【0045】

次に、これらの名詞のペアが名詞関係データベース16に存在するかどうかをチェックし、存在するならば、どういう関係で存在するかを検索し、集計する。

【0046】

たとえば、前述の矛盾事態ペアデータベース18からは、(フランス, エッフェル塔)、(旅行, 海外ツアー)、(予定, 海外ツアー)という名詞ペアが抽出され、そのうち(フランス, エッフェル塔)、(旅行, 海外ツアー)の二つが名詞関係データベース16に存在することが分かる。そのため、contained-by, is-aの関係がそれぞれ1回ずつカウントアップされる。このように矛盾事態ペアデータベース18に含まれるすべての矛盾事態ペアにおける名詞ペアについて、名詞関係データベース16を照合し、どの関係が何回照合できたかを数え上げる。

10

20

30

40

50

【 0 0 4 7 】

なお、照合においては、表記の揺らぎに対処するため、形態素解析器JUMAN の出力する代表表記を用いて行う。もちろん、揺らぎ情報を含むような辞書があればそれを用いる。

【 0 0 4 8 】

最後に、見つかった名詞ペアの関係の出現回数が、それぞれ所定の閾値（たとえば、10）を超えているかどうかを確認し、超えていれば、その名詞ペアの関係を矛盾作成に有効な関係として関係リスト22に追加する。ここでの処理は、どのような関係を持つ名詞が矛盾事態ペアに現われやすいかを調べ上げる処理と言える。

【 0 0 4 9 】

矛盾事態ペア取得部24は、予め用意された矛盾事態ペアデータベース18から、前件と後件とで項部分に共通する共有名詞を持つ矛盾事態ペアを一つ以上取り出し、後述する単一名詞置換部30へ送る。

10

【 0 0 5 0 】

また、矛盾事態ペア取得部24は、予め用意された矛盾事態ペアデータベース18から、共有名詞をもたない矛盾事態ペアを一つ以上取り出し、後述する名詞ペア置換部28へ送る。

【 0 0 5 1 】

ここで取り出された矛盾事態ペアは、一つずつ、後段の名詞ペア置換部28及び単一名詞置換部30で処理される。

【 0 0 5 2 】

以下は、矛盾事態ペアデータベース18に含まれる、共有名詞をもたない矛盾事態ペアの例である。

20

【 0 0 5 3 】

< フランスへ行かない，エッフェル塔に行く >
< 旅行の予定を立てない，海外ツアーに申し込む >

【 0 0 5 4 】

矛盾事態ペア作成部26は、名詞ペア置換部28及び単一名詞置換部30を備えている。

【 0 0 5 5 】

名詞ペア置換部28は、矛盾事態ペアデータベース18から得られる、共有名詞をもたない矛盾事態ペアの各々について、当該矛盾事態ペアに含まれる名詞ペアの関係が、関係リスト22に含まれる場合、名詞ペアの関係と同じ関係を持つ名詞ペアを名詞関係データベース16から取得し、取得した名詞ペアによって当該矛盾事態ペアの名詞ペアを置き換えることで矛盾事態ペアを作成する。

30

【 0 0 5 6 】

たとえば、< フランスへ行かない，エッフェル塔に行く > が処理対象だとすると、まず、ここから名詞ペアを抽出する。名詞ペアは(フランス，エッフェル塔)である。そして、名詞関係データベース16を参照し、フランスとエッフェル塔の間の関係を探す。そうすると、contained-by の関係であることが分かる。この関係は関係抽出部20が得た関係リスト22に含まれているため、同じ関係をもつ別の名詞ペアを名詞関係データベース16から取得して、当該矛盾事態ペアの名詞ペアと置換し、新しい矛盾事態ペアを生成する。たとえば、この例からは次の新しい矛盾事態ペアとして以下が生成される。

40

【 0 0 5 7 】

< 神奈川県へ行かない，箱根に行く >

【 0 0 5 8 】

ここでは、神奈川県と箱根の関係がcontained-by であることから、フランスを神奈川県、エッフェル塔を箱根で置換している。このように置き換えることで新しい矛盾事態ペアを作成できる。

【 0 0 5 9 】

新しく作成された矛盾事態ペアはチェック部32に送られる。

50

【0060】

単一名詞置換部30は、矛盾事態ペアデータベース18から得られる、共有名詞をもつ矛盾事態ペアの各々について、名詞関係データベース16から、当該矛盾事態ペアが持つ共有名詞の同位語を抽出し、取得した同位語により当該矛盾事態ペアの共有名詞を置き換えることで、矛盾事態ペアを作成する。

【0061】

たとえば、以下の矛盾事態ペアがあったとする。

【0062】

< 動物園の場所を調べない、動物園に行く >
< テニスに興味がない、テニスをしたい >

10

【0063】

ここから、単一名詞置換部30は以下の新たな矛盾事態ペアを作成する。

【0064】

< 植物園の場所を調べない、植物園に行く >
< ゴルフに興味がない、ゴルフをしたい >

【0065】

ここで、「動物園」と「植物園」は同位語であり、「テニス」と「ゴルフ」は、どちらも国語辞典における「球技」を上位語に持つ同位語である。

【0066】

新しく作成された矛盾事態ペアはチェック部32に送られる。

20

【0067】

チェック部32は、新しく作成された矛盾事態ペアの各々に対し、当該矛盾事態ペアに含まれる事態を表す文の各々について、大規模テキストコーパス34における当該事態を表す文の頻度が閾値以上であるかをチェックして、当該矛盾事態ペアに含まれる事態を表す文の各々の頻度が閾値以上であれば、矛盾事態ペアが適切であると判断し、データベース追加部36に送られる。

【0068】

具体的には、新しく作成された矛盾事態ペアの前件（左側の文）、後件（右側の文）それぞれについて独立に大規模テキストコーパス34における出現回数をチェックし、どちらも閾値以上（たとえば、10回以上）の回数出現しているなら、この矛盾事態ペアが適切であると判断し、新しい矛盾事態ペアとして採用する。出現のチェックは、事態を表す述語項構造表現を一般的なWeb検索エンジン、もしくは、大規模なテキストから作成した検索インデックスでフレーズ検索することによって行う。検索は、表記の揺らぎを吸収するため、述語項構造表現をJUMAN代表表記に正規化して行ってもよい。

30

【0069】

なお、獲得したい矛盾事態ペアの量と精度との兼ね合いに基づいて、チェック部32の処理ステップはスキップしてもよい。その場合は、矛盾事態ペア作成部26によって新しく作成された矛盾事態ペアはそのままデータベース追加部36に送られる。

【0070】

データベース追加部36は、チェック部32から送られてきた矛盾事態ペアを矛盾事態ペアデータベース18に追加する。

40

【0071】

< 矛盾作成装置の作用 >

次に、本実施の形態に係る矛盾作成装置10の作用について説明する。矛盾作成装置10では、図2に示す矛盾作成処理ルーチンが実行される。

【0072】

まず、ステップS100において、第1知識ベース12A～第3知識ベース12Cから、名詞ペアと当該名詞ペアの関係との組み合わせを各々取得して、名詞関係データベース16に記憶する。

【0073】

50

そして、ステップ S 1 0 2 では、矛盾事態ペアデータベース 1 8 と名詞関係データベース 1 6 を照合し、矛盾事態ペアデータベース 1 8 において閾値以上の回数出現する、矛盾に関する名詞ペアの関係を抽出し、抽出した名詞ペアの関係を格納した関係リスト 2 2 を作成する。

【 0 0 7 4 】

ステップ S 1 0 4 では、矛盾事態ペアデータベース 1 8 から、共有名詞をもたない矛盾事態ペアを取得する。ステップ S 1 0 6 では、上記ステップ S 1 0 4 で取得した、共有名詞をもたない矛盾事態ペアについて、当該矛盾事態ペアに含まれる名詞ペアの関係が、関係リスト 2 2 に含まれる場合、名詞ペアの関係と同じ関係を持つ名詞ペアを名詞関係データベース 1 6 から取得する。

10

【 0 0 7 5 】

そして、ステップ S 1 0 8 において、上記ステップ S 1 0 6 で取得した名詞ペアによって、上記ステップ S 1 0 4 で取得した矛盾事態ペアの名詞ペアを置き換えることで矛盾事態ペアを作成する。

【 0 0 7 6 】

ステップ S 1 1 0 では、上記ステップ S 1 0 8 で作成した矛盾事態ペアに含まれる事態を表す文の各々について、大規模テキストコーパス 3 4 における当該事態を表す文の頻度が閾値以上であるかをチェックして、当該矛盾事態ペアに含まれる事態を表す文の各々の頻度が閾値以上であるか否かを判定する。当該矛盾事態ペアに含まれる事態を表す文の何れの頻度も閾値以上である場合には、作成した矛盾事態ペアが適切であると判断し、ステップ S 1 1 2 へ移行する。

20

【 0 0 7 7 】

一方、当該矛盾事態ペアに含まれる事態を表す文の少なくとも一方の頻度が閾値未満である場合には、作成した矛盾事態ペアが適切でない判断し、ステップ S 1 1 4 へ移行する。

【 0 0 7 8 】

ステップ S 1 1 2 では、上記ステップ S 1 0 8 で作成した矛盾事態ペアを、矛盾事態ペアデータベース 1 8 に追加する。

【 0 0 7 9 】

ステップ S 1 1 2 では、矛盾事態ペアデータベース 1 8 に格納されている、共有名詞をもたない矛盾事態ペアの全てについて、上記ステップ S 1 0 4 ~ S 1 1 2 の処理を実行したか否かを判定し、上記ステップ S 1 0 4 ~ S 1 1 2 の処理を実行していない、共有名詞をもたない矛盾事態ペアが存在する場合には、上記ステップ S 1 0 4 へ戻り、当該矛盾事態ペアを取得する。

30

【 0 0 8 0 】

一方、矛盾事態ペアデータベース 1 8 に格納されている、共有名詞をもたない矛盾事態ペアの全てについて、上記ステップ S 1 0 4 ~ S 1 1 2 の処理を実行した場合には、ステップ S 1 1 6 へ進む。

【 0 0 8 1 】

ステップ S 1 1 6 では、矛盾事態ペアデータベース 1 8 から、共有名詞をもつ矛盾事態ペアを取得する。ステップ S 1 1 8 では、名詞関係データベース 1 6 から、上記ステップ S 1 1 6 で取得した矛盾事態ペアが持つ共有名詞の同位語を取得する。

40

【 0 0 8 2 】

ステップ S 1 2 0 では、上記ステップ S 1 1 8 で取得した同位語により、上記ステップ S 1 1 6 で取得した矛盾事態ペアの共有名詞を置き換えることで、矛盾事態ペアを作成する。

【 0 0 8 3 】

ステップ S 1 2 2 では、上記ステップ S 1 2 0 で作成した矛盾事態ペアに含まれる事態を表す文の各々について、大規模テキストコーパス 3 4 における当該事態を表す文の頻度が閾値以上であるかをチェックして、当該矛盾事態ペアに含まれる事態を表す文の各々の

50

頻度が閾値以上であるか否かを判定する。当該矛盾事態ペアに含まれる事態を表す文の何れの頻度も閾値以上である場合には、作成した矛盾事態ペアが適切であると判断し、ステップS 1 2 4へ移行する。

【0084】

一方、当該矛盾事態ペアに含まれる事態を表す文の少なくとも一方の頻度が閾値未満である場合には、作成した矛盾事態ペアが適切でないと判断し、ステップS 1 2 6へ移行する。

【0085】

ステップS 1 2 4では、上記ステップS 1 2 0で作成した矛盾事態ペアを、矛盾事態ペアデータベース18に追加する。

10

【0086】

ステップS 1 2 6では、矛盾事態ペアデータベース18に格納されている、共有名詞をもつ矛盾事態ペアの全てについて、上記ステップS 1 1 6～S 1 2 4の処理を実行したか否かを判定し、上記ステップS 1 1 6～S 1 2 4の処理を実行していない、共有名詞をもつ矛盾事態ペアが存在する場合には、上記ステップS 1 1 6へ戻り、当該矛盾事態ペアを取得する。

【0087】

一方、矛盾事態ペアデータベース18に格納されている、共有名詞をもつ矛盾事態ペアの全てについて、上記ステップS 1 1 6～S 1 2 4の処理を実行した場合には、矛盾作成処理ルーチンを終了する。

20

【0088】

<矛盾事態ペアデータベースの拡張実験>

本発明の実施の形態に基づく手法を実行した結果、獲得できた矛盾事態ペアの一例は図3に示す通りである。図3のいずれもが、チェック部32を通過したものであり、それらは、すべて正しい矛盾事態ペアであった。

【0089】

また、チェック部32により除外された例として、<市が好きでない、政令指定都市にハマる>がある。これは、<寿司が好きでない、回転寿司にハマる>から、(政令指定都市, is-a, 市)の関係によって生成されるが、「政令指定都市にハマる」は大規模テキストコーパス34において出現回数が少なかったため、チェック部32によって除外され、正しく獲得されなかった。

30

【0090】

以上説明したように、本実施の形態に係る矛盾作成装置によれば、矛盾に関する名詞ペアの関係を抽出して関係リストを作成し、矛盾事態ペアデータベースから得られた矛盾事態ペアに含まれる名詞ペアの関係が、関係リストに含まれる場合、名詞ペアの関係と同じ関係を持つ名詞ペアを名詞関係データベースから取得し、取得した名詞ペアによって矛盾事態ペアの名詞ペアを置き換えることで、新しい矛盾事態ペアを作成できる。

【0091】

また、矛盾事態ペアデータベースから、共有名詞を持つ矛盾事態ペアを取得し、名詞関係データベースから、取得した矛盾事態ペアが持つ共有名詞の同位語を抽出し、取得した同位語により矛盾事態ペアの共有名詞を置き換えることで、新しい矛盾事態ペアを作成できる。

40

【0092】

また、矛盾事態ペアデータベースが拡張されることにより、言語生成システムにおいて、矛盾しない発話の生成が実現され、ユーザにとって理解しやすい文章がコンピュータによって生成されるようになる。これにより、人間の知的活動を促進することができる。

【0093】

なお、本発明は、上記実施の形態に限定されるものではなく、この発明の要旨を逸脱しない範囲内で様々な変形や応用が可能である。

【0094】

50

例えば、本願明細書中において、プログラムが予めインストールされている実施の形態として説明したが、外部の記憶装置や記録媒体等に格納されたプログラムを随時読み込んで、またインターネットを介してダウンロードして実行するようにしてもよい。また、当該プログラムを、コンピュータ読み取り可能な記録媒体に格納して提供することも可能である。

【符号の説明】

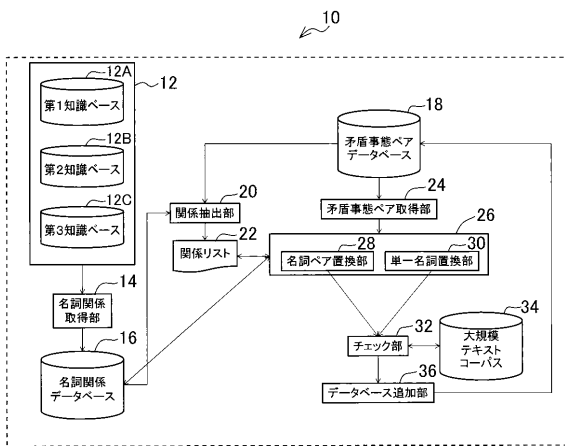
【0095】

- 10 矛盾作成装置
- 12 知識ベース群
- 14 名詞関係取得部
- 16 名詞関係データベース
- 18 矛盾事態ペアデータベース
- 20 関係抽出部
- 22 関係リスト
- 24 矛盾事態ペア取得部
- 26 矛盾事態ペア作成部
- 28 名詞ペア置換部
- 30 単一名詞置換部
- 32 チェック部
- 34 大規模テキストコーパス
- 36 データベース追加部

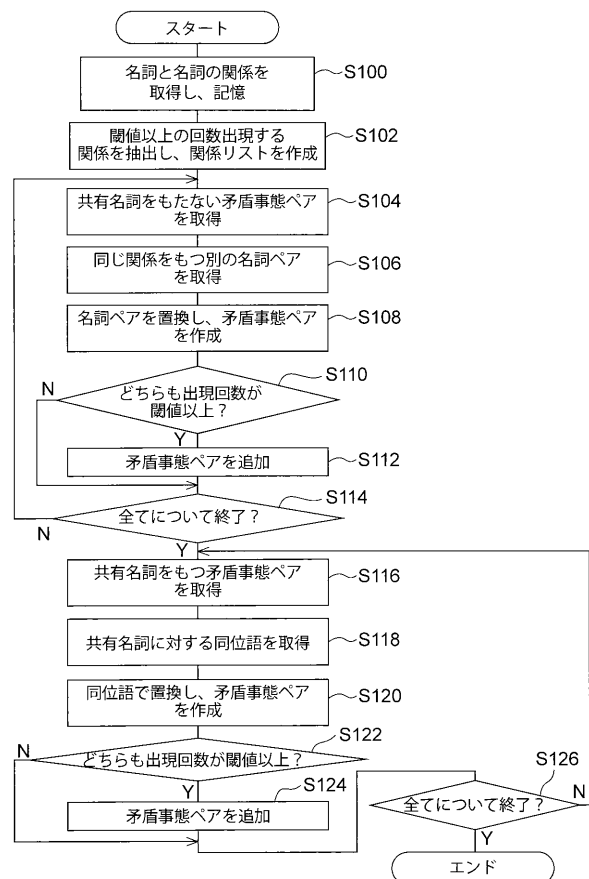
10

20

【図1】



【図2】



【 図 3 】

矛盾事態ペアデータベース中の矛盾事態ペア	本発明によって作成された矛盾事態ペア
(ゴルフをしない、ゴルフクラブを買う)	(野球をしない、バットを買う)
(ゴルフをしない、ゴルフクラブを買う)	(サイクリングをしない、自転車を買う)
(シューズを持っていない、スポーツシューズを持っている)	(装置を持っていない、電動助走装置を持っている)
(シューズを持っていない、スポーツシューズを持っている)	(博士業を持っていない、 박사 服を持っている)
(寿司が好きでない、回転寿司にハマる)	(芸能事務所が好きでない、大人計画にハマる)
(寿司が好きでない、回転寿司にハマる)	(面白いコンビが好きでない、エレキコミックにハマる)
(スーパーに行く時間がない、近くのスーパーに行く)	(コンビニに行く時間がない、近くのコンビニに行く)
(スーパーに行く時間がない、近くのスーパーに行く)	(銀行に行く時間がない、近くの銀行に行く)
(野球の試合に興味がない、野球の試合を見に行く)	(バレーの試合に興味がない、バレーの試合を見に行く)
(野球の試合に興味がない、野球の試合を見に行く)	(サッカーの試合に興味がない、サッカーの試合を見に行く)

フロントページの続き

- (72)発明者 黒橋 禎夫
京都府京都市左京区吉田本町3番地1 国立大学法人京都大学内
- (72)発明者 河原 大輔
京都府京都市左京区吉田本町3番地1 国立大学法人京都大学内
- (72)発明者 高畠 悠
京都府京都市左京区吉田本町3番地1 国立大学法人京都大学内
- Fターム(参考) 5B091 AA15 AB17 CA12 CA21 CC04
5B109 QB14