

明 細 書

発明の名称： 含意判定装置、方法、およびプログラム

技術分野

[0001] 本発明は、与えられた仮説を与えられた判定対象文が含意するか否かを判定する含意判定装置、含意判定方法および含意判定プログラムに関する。

背景技術

[0002] 含意判定とは、与えられた仮説Hを与えられた判定対象文Tが含意するかどうかを判定すること（タスク）である。判定対象文Tが仮説Hを含意するとは、判定対象文Tを読んだ人間が仮説Hを事実として推論できることを言う。含意判定を行う対象として、例えば、「K首相がアメリカを訪れた」という仮説Hと、「K首相がホワイトハウスを訪問した」という判定対象文Tが与えられたとする。この場合、判定対象文Tを読んだ人間は仮説Hを事実として推論できるため、判定対象文Tが仮説Hを含意すると判定される。

[0003] 含意判定の代表的な方式が非特許文献1に記載されている。非特許文献1に記載された方式では、仮説Hに含まれる単語または係り受けの組が文Tにおいて共通する割合（以下、共通割合と記す。）を算出する。算出した共通割合が基準とする類似度閾値（基準値）以上の場合、仮説Hを文Tが含意すると判定される。また、非特許文献1には、含意判定の方式として、共通割合を算出する際に用いられる単語をソーラスの範囲に拡張した方式、文の否定形やモダリティを判断する方式についても記載されている。

[0004] また、特許文献1には、話題が類似する関連文書を検索する文書検索装置が記載されている。特許文献1に記載された文書検索装置は、各文書に含まれる話題の類似性を基に関連文書か否かを判断する。また、特許文献1に記載された文書検索装置では、各関連文書が発表または報道された日付に基づいて、初期関連文書とするか、後続関連文書とするかが判断される。

[0005] なお、特許文献2には、対話を円滑に行うコミュニケーション支援装置が

記載されている。特許文献2に記載された装置は、対象とする発話をそれ以前の発話と比較し、以前の発話には含まれるが、対象とする発話には含まれない要素を新情報と判定する。具体的には、特許文献2に記載された装置は、第1言語の話者が発話した第1言語文の言語構造、および第2言語文の言語構造を比較して、第2言語構造文で新たに発話された内容を表す新情報を抽出する。抽出された新情報は、優先して対話の相手に出力される。

先行技術文献

特許文献

[0006] 特許文献1：特開2004-145753号公報（段落0044）

特許文献2：特開2008-83993号公報（段落0050、0053）

非特許文献

[0007] 非特許文献1：Adrian Iftene, "Textual Entailment", October 2009., P H.D.Thesis, TR 09-02, [online], [平成22年10月4日検索]、インターネット〈<http://profs.info.uaic.ro/~tr/tr09-02.pdf>〉

発明の概要

発明が解決しようとする課題

[0008] 以下、含意判定を行い、判定結果を出力する装置のことを含意判定装置と記す。また、正解（具体的には、仮説とその仮説を含意する文の組）の集合が存在するときに、含意判定装置が正解であると判定して出力した結果のうち、その出力結果に含まれる本来の正解の割合を「適合率」と記す。また、正解の集合のうち含意判定装置が正解であると判定して出力した正解の割合を「再現率」と記す。すなわち、「適合率」とは含意判定装置の出力の正確性を図る指標であり、「再現率」とは含意判定装置の出力の網羅性を図る指標である。

[0009] 非特許文献1に記載されている代表的な含意判定の方式では、含意判定の性能が低いという問題がある。誤判定を防ぐ（すなわち、適合率の高める）ためには、一般的に、基準値を高く設定する必要がある。しかし、基準値を

高く設定した場合、判定漏れが大量に発生し、再現率が低下してしまうことになる。文の表現には多様性があるため、仮説を含意する判定対象文であっても共通割合が高いとは限らないためである。

[0010] また、特許文献 1 に記載された文書検索装置では、関連文書か否かは、各文書の話題同士の類似度で判断されるが、類似度の計算方法としては、一般的に知られた方法が用いられるにすぎない。また、関連文書を日付順に並べ替えたとしても、文書の前後関係が明確になるにすぎず、関連文書間の類似度が変わるわけではない。そのため、特許文献 1 に記載された装置を用いても、含意判定の性能を向上させることは困難である。

[0011] そこで、本発明は、含意判定の性能を向上させることができる含意判定装置、含意判定方法および含意判定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

[0012] 本発明による含意判定装置は、与えられた仮説を含意する文である仮説含意文で、その仮説が初めて明らかにされた事実を示す新事実であるか否かを、その仮説含意文に記載された特定の表現をもとに判定する新事実性判定手段と、与えられた仮説を、その仮説を含むか否かが判定される文である判定対象文が含意するか否かを判定する含意判定手段とを備え、含意判定手段が、与えられた仮説が仮説含意文で新事実であると判定された場合に、判定対象文の発生日時とその仮説を含意する仮説含意文の発生日時とを比較して、与えられた仮説を判定対象文が含意するか否かを判定することを特徴とする。

[0013] 本発明による含意判定方法は、与えられた仮説を含意する文である仮説含意文で、その仮説が初めて明らかにされた事実を示す新事実であるか否かを、その仮説含意文に記載された特定の表現をもとに判定し、与えられた仮説が仮説含意文で新事実であると判定された場合に、与えられた仮説を含むか否かが判定される文である判定対象文の発生日時とその仮説を含意する仮説含意文の発生日時とを比較して、与えられた仮説を判定対象文が含意するか否かを判定することを特徴とする。

[0014] 本発明による含意判定プログラムは、コンピュータに、与えられた仮説を
含意する文である仮説含意文で、その仮説が初めて明らかにされた事実を示
す新事実であるか否かを、その仮説含意文に記載された特定の表現をもとに
判定する新事実性判定処理、および、与えられた仮説を、その仮説を含むか
否かが判定される文である判定対象文が含意するか否かを判定する含意判定
処理を実行させ、含意判定処理で、与えられた仮説が仮説含意文で新事実で
あると判定された場合に、判定対象文の発生日時とその仮説を含意する仮説
含意文の発生日時とを比較させ、与えられた仮説を判定対象文が含意するか
否かを判定させることを特徴とする。

発明の効果

[0015] 本発明によれば、含意判定の性能を向上させることができる。

図面の簡単な説明

[0016] [図1]本発明の第1の実施形態における含意判定装置の例を示すブロック図で
ある。

[図2]第1の実施形態における含意判定装置の動作を示すフローチャートであ
る。

[図3]記憶装置が記憶する情報の例を示す説明図である。

[図4]記憶装置が記憶する情報の例を示す説明図である。

[図5]本発明の第2の実施形態における含意判定装置の例を示すブロック図で
ある。

[図6]第2の実施形態における含意判定装置の動作を示すフローチャートであ
る。

[図7]記憶装置が記憶する情報の例を示す説明図である。

[図8]本発明の第3の実施形態における含意判定装置の例を示すブロック図で
ある。

[図9]第3の実施形態における含意判定装置の動作を示すフローチャートであ
る。

[図10]記憶装置が記憶する情報の例を示す説明図である。

[図11]記憶装置が記憶する情報の例を示す説明図である。

[図12]記憶装置が記憶する情報の例を示す説明図である。

[図13]記憶装置が記憶する情報の例を示す説明図である。

[図14]本発明による含意判定装置の最小構成の例を示すブロック図である。

発明を実施するための形態

[0017] 以下、本発明の実施形態を図面を参照して説明する。

[0018] 実施形態 1 .

図 1 は、本発明の第 1 の実施形態における含意判定装置の例を示すブロック図である。本実施形態における含意判定装置は、プログラム制御により動作するデータ処理装置 1 と、情報を記憶する記憶装置 2 とを備えている。

[0019] データ処理装置 1 は、新事実性判定手段 10 と、含意判定手段 11 とを含む。

[0020] 新事実性判定手段 10 は、与えられた仮説を含意する文（以下、仮説含意文と記す。）において、その仮説が初めて明らかにされた事実（新事実）であるか否かをその仮説含意文に記載された特定の表現をもとに判定する。なお、新事実の判定方法については後述する。

[0021] 含意判定手段 11 は、与えられた仮説を、仮説を含むか否かが判定される文（以下、判定対象文と記す。）が含意するか否かを判定する。そして、含意判定手段 11 は、仮説含意文において、与えられた仮説が新事実であると判定された場合に、判定対象文の発生日時が、その仮説含意文の発生日時よりも古い場合、与えられた仮説を判定対象文が含意しないと判定する。ここで、発生日時とは、文書（文）が作成された日時や、発行された日時など、その文書（文）が発生した日時のことを意味する。以下、判定対象文の発生日時、仮説含意文の発生日時を、それぞれ判定対象文の日時、仮説含意文の日時と記すこともある。

[0022] また、含意判定手段 11 は、判定対象文が仮説を含意する度合いを示す含意度（以下、仮説と判定対象文との含意度と記すこともある。）を用いて、判定対象文が仮説を含意するか否かを判定してもよい。このとき、含意判定

手段 1 1 は、例えば、仮説と判定対象文とで記載が類似する度合いを含意度として用いてもよい。具体的には、判定対象文の発生日時が、与えられた仮説を含意する仮説含意文の発生日時よりも古い場合、含意判定手段 1 1 は、与えられた仮説と判定対象文との間の含意度にペナルティを付与してもよい。また、ペナルティ付与後の含意度が予め定められた基準値を下回る場合、含意判定手段 1 1 は、その判定対象文が仮説を含意しないと判定してもよい。含意度は、例えば、仮説および判定対象文に記載された単語や係り受け組などが共通する度合いなどによって算出される。なお、含意度の算出方法については後述する。

[0023] 記憶装置 2 は、判定対象文書記憶部 2 0 と、仮説記憶部 2 1 と、仮説含意文書記憶部 2 2 と、新事実性判定結果記憶部 2 3 と、含意度記憶部 2 4 と、含意判定結果記憶部 2 5 とを含む。判定対象文書記憶部 2 0 と、仮説記憶部 2 1 と、仮説含意文書記憶部 2 2 と、新事実性判定結果記憶部 2 3 と、含意度記憶部 2 4 と、含意判定結果記憶部 2 5 とは、それぞれ磁気ディスク等により実現される。

[0024] 判定対象文書記憶部 2 0 は、判定対象文を含む文書集合を記憶する。具体的には、判定対象文書記憶部 2 0 は、判定対象文と、その判定対象文の日時とを対応付けて記憶する。判定対象文は、ユーザ等により予め判定対象文書記憶部 2 0 に記憶される。

[0025] 仮説記憶部 2 1 は、判定対象とする仮説を記憶する。なお、仮説記憶部 2 1 が記憶する仮説は、1 つであってもよく、複数であってもよい。仮説は、ユーザ等により予め仮説記憶部 2 1 に記憶される。

[0026] 仮説含意文書記憶部 2 2 は、仮説を含意する文（すなわち、仮説含意文）を含む文書集合を格納する。具体的には、仮説含意文書記憶部 2 2 は、仮説含意文と、その仮説含意文の日時とを対応付けて記憶する。仮説含意文書記憶部 2 2 には、特定の仮説を含意することが既知の文（仮説含意文）が予め記憶される。なお、この仮説含意文は、ユーザ等により予め仮説含意文書記憶部 2 2 に記憶される。また、各仮説に対して、少なくとも 1 つ以上の仮説

含意文を含む文書が対応する。

[0027] 新事実性判定結果記憶部 23 は、新事実性判定手段 10 が判定した結果を記憶する。具体的には、新事実性判定結果記憶部 23 は、与えられた仮説が仮説含意文において初めて明らかにされた事実であるか否かを判定した結果を記憶する。

[0028] 含意度記憶部 24 は、仮説と判定対象文との含意度を記憶する。

[0029] 含意判定結果記憶部 25 は、含意判定手段 11 が判定した結果を記憶する。具体的には、含意判定結果記憶部 25 は、判定対象文が与えられた仮説を含意するか否かを判定した結果を記憶する。

[0030] 新事実性判定手段 10 と、含意判定手段 11 とは、プログラム（含意判定プログラム）に従って動作するコンピュータの CPU によって実現される。例えば、プログラムは、データ処理装置 1 の記憶部（図示せず）に記憶され、CPU は、そのプログラムを読み込み、プログラムに従って、新事実性判定手段 10 及び含意判定手段 11 として動作してもよい。また、新事実性判定手段 10 と、含意判定手段 11 とは、それぞれが専用のハードウェアで実現されていてもよい。

[0031] 次に、本実施形態の含意判定装置の動作を説明する。図 2 は、第 1 の実施形態における含意判定装置の動作を示すフローチャートである。また、図 3 及び図 4 は、記憶装置 2 が記憶する情報の例を示す説明図である。以下の説明では、判定対象文書記憶部 20、仮説記憶部 21 および仮説含意文書記憶部 22 が、図 3 に例示するデータを記憶しているものとする。

[0032] 仮説記憶部 21 は、図 3（a）に例示するように、仮説 ID 項目、仮説項目、文書 ID 項目および文 ID 項目を記憶する。仮説記憶部 21 は、仮説 ID 項目として仮説の識別子（以下、仮説 ID）を記憶し、仮説項目として仮説の内容を記憶する。また、仮説記憶部 21 は、文書 ID 項目として仮説を含意する文書の識別子を記憶し、文 ID 項目として仮説を含意する文の識別子を記憶する。以下の説明では、文の識別子を文 ID と記し、文書の識別子を文書 ID と記す。

[0033] 仮説含意文書記憶部 22 は、図 3 (b) に例示するように、文書 I D 項目、文 I D 項目、日時項目および文内容項目を記憶する。仮説含意文書記憶部 22 は、文書 I D 項目として仮説を含意する文を含む文書の文書 I D を記憶し、文 I D 項目として仮説を含意する文の文 I D を記憶する。また、仮説含意文書記憶部 22 は、日時項目として文書の発生日時を記憶し、文内容項目として仮説含意文の内容を記憶する。仮説と、その仮説を含意する文との対応関係は、仮説記憶部 21 及び仮説含意文書記憶部 22 が記憶する文書 I D と文 I D とで判断される。図 3 (a) および図 3 (b) に示す例では、仮説記憶部 21 に記憶された「仮説 I D = 1」で特定される仮説は、仮説含意文書記憶部 22 に記憶された「文書 I D = 1」かつ「文 I D = 1」で特定される文が含まれることを示す。

[0034] 判定対象文書記憶部 20 は、図 3 (c) に例示するように、文書 I D 項目、文 I D 項目、日時項目および文内容項目を記憶する。判定対象文書記憶部 20 は、文書 I D 項目として、判定対象文を含む文書の文書 I D を記憶し、文 I D 項目として判定対象文の文 I D を記憶する。また、判定対象文書記憶部 20 は、日時項目として文書の発行日時を記憶し、文内容項目として判定対象文の内容を記憶する。以下、本実施形態では、判定対象文書記憶部 20、仮説記憶部 21 および仮説含意文書記憶部 22 が、図 3 に示す情報をそれぞれ記憶している場合を例に説明する。

[0035] まず、新事実性判定手段 10 は、仮説が仮説含意文において初めて明らかにされた事実（すなわち、新事実）であるか否かを、その仮説含意文に含まれる特定の表現を用いて判定する（図 2 におけるステップ S1）。具体的には、新事実性判定手段 10 は、仮説含意文書記憶部 22 に記憶された仮説含意文に対し、以下の（１）～（５）に例示する表現を手がかりに、新事実である、または、新事実ではない、と判定する。

[0036] なお、以下の説明において、（１），（２），（３）の説明に用いる仮説 H を「Q Q R R 容疑者が逮捕された峠道は若いころに車で通り詰めた場所である」とし、（４），（５）の説明に用いる仮説 H を「Q Q R R は逮捕され

た」とする。なお、QQは姓を表し、RRは名を表すものとする。

[0037] また、以下の説明において、(1)～(4)の説明に用いる仮説含意文T_—Hを『M本社工場（Z市XX区、W県YY町）で乗用車が社員12人をはね死傷させた事件から29日で1週間。殺人未遂容疑で逮捕されたQQRR容疑者（42）が、自ら110番して逮捕された峠道について、「若いころに車で通い詰めた思い出の場所」と供述していることが捜査本部への取材で分かった。』とし、(5)の説明に用いる仮説含意文T_—Hを「QQRR容疑者が逮捕されてから29日で1週間になる。」とする。

[0038] (1) 仮説含意文T_—Hに「情報の発信元を示す表現が存在する」ならば、T_—Hにおいて仮説Hを新事実と判定する。

[0039] 上記例の場合、仮説含意文T_—Hの「捜査本部への取材で」という表現が、情報の発信元を示す表現であると言える。したがって、新事実性判定手段10は、この仮説含意文T_—Hにおいて仮説Hを新事実と判定する。他にも、「関係筋への取材で」という表現や、「発表した」などの表現を、情報の発信元を示す表現とすることが出来る。新事実性判定手段10は、このような表現を含んでいる場合に、仮説含意文T_—Hにおいて仮説Hを新事実と判定してもよい。さらに、仮説含意文が英文の場合、「according to」、「said」、「told」、「reported」、「announced」などの表現が仮説含意文に含まれている場合、情報の発信元を示す表現とすることが出来る。

[0040] このように、新事実性判定手段10は、情報の発信元を示す所定の文字列が含まれている仮説含意文T_—Hが含意する仮説Hを新事実と判定する。なお、判定に用いられる所定の文字列については、情報の発信元を示すと考えられる表現を予め定めておけばよい。

[0041] (2) 仮説含意文T_—Hを含む文書の「仮説Hとヘッドライン（タイトル）との類似度が高い（あらかじめ設定した閾値以上）」ならば、T_—Hにおいて仮説Hを新事実と判定する。

[0042] 仮説Hとヘッドラインとの間の類似度は、後述する含意度の計算方法と同

様の方法で計算される。例えば、上記の例における仮説含意文 T__H のヘッドライン（見出し）が『Q Q R R 容疑者、逮捕の峠道「思い出の場所」 元” 走り屋” に何が』であるとする。仮説 H を形態素解析した結果得られた単語のうち、自立語として、「Q Q」、「R R」、「容疑者」、「逮捕」、「峠道」、「若い」、「車」、「通い詰める」、「場所」の 9 つの単語 t が選択される。同様に、仮説含意文 T__H から単語 t を選択する。仮説 H に含まれる 9 つの単語のうち、仮説含意文 T__H にも含まれる単語は、「Q Q」、「R R」、「容疑者」、「逮捕」、「峠道」、「場所」の 6 つである。したがって、単語 t の重要度（以下、weight (t)）を全て 1 とした場合、H とヘッドラインの類似度は 0.67 (= 6 / 9) と算出される。ここで、仮説と判定する際の基準値（以下、仮説判定基準値と記す。）を 0.5 と定めた場合、本例の類似度は仮説判定基準値よりも大きいため、仮説 H は、T__H において新事実と判定される。

[0043] なお、仮説判定基準値は、ユーザ等により予め定められる。また、各仮説含意文のヘッドラインもユーザ等により予め作成され、例えば、仮説含意文書記憶部 22 に別項目として記憶される。また、仮説 H とヘッドラインとの類似度は、上述する方法に限定されない。このように、新事実性判定手段 10 は、仮説含意文 T__H のヘッドラインと仮説 H との類似度が仮説判定基準値よりも大きい場合、その仮説含意文 T__H が含意する仮説 H を新事実と判定する。

[0044] (3) 仮説含意文 T__H に「新事実を表す直接的な表現が存在する」ならば、仮説 H は T__H において新事実と判定する。

[0045] 上記例の場合、仮説含意文 T__H の「分かった」という表現が、新事実を示す直接的な表現に該当する。したがって、新事実性判定手段 10 は、この仮説含意文 T__H において仮説 H を新事実と判定する。他にも、「判明した」などの表現を、新事実を示す直接的な表現とすることが出来る。このように、新事実性判定手段 10 は、新事実を示す直接的な所定の表現が含まれている仮説含意文 T__H が含意する仮説 H を新事実と判定する。なお、判定に

用いられる所定の表現については、新事実を示すと考えられる表現を予め定めておけばよい。

[0046] (4) 仮説含意文 T__H に「過去を表す時間表現が存在する」ならば、T__H において仮説 H を新事実ではないと判定する。

[0047] 上記例の場合、仮説含意文 T__H の「日で〇〇週間」という表現が、過去を表す時間表現であると言える。したがって、新事実性判定手段 10 は、この仮説含意文 T__H において仮説 H を新事実ではないと判定する。

[0048] 例えば、過去を表す表現のパターンを事前にいくつか用意しておき、仮説含意文 T__H が、それらのパターンのいずれかと合致する場合に、新事実性判定手段 10 は、仮説含意文 T__H に過去を表す時間表現が存在すると判定してもよい。また、この場合、正規表現で記述されたパターンを用いてもよい。正規表現で記述されたパターンとして、例えば、「¥ d + (週間 | 年)」という表現が該当する。ここで、¥ d は数値、+ は一回以上の繰り返し、(週間 | 年) は週間と年のいずれかを表す。上記例の場合、仮説含意文 T__H には、このパターンと合致する「1 週間」という表現が含まれるので、新事実性判定手段 10 は、仮説 H が仮説含意文 T__H において新事実ではないと判定する。

[0049] なお、過去を表す所定の時間表現が含まれているか否かを判定する方法は、上記方法に限定されない。このように、新事実性判定手段 10 は、過去を表す所定の時間表現が含まれている仮説含意文 T__H が含意する仮説 H を新事実ではないと判定する。

[0050] (5) 仮説含意文 T__H に「仮説が挿入句として記載されている」ならば、T__H において仮説 H を新事実ではないと判定する。

[0051] ここで、「記載されている」とは、仮説 H の記載と仮説含意文 T__H に含まれる記載とが、完全に一致する場合に限られず、記載内容が極めて近い（記載内容が同等である）場合も、ここで言う「記載されている」に該当する。

[0052] 上記例の場合、仮説 H 「Q Q R R は逮捕された」が、仮説含意文 T__H の

「・・・殺人未遂容疑で逮捕されたＱＱＲＲ容疑者（４２）・・・」の部分に挿入句として記載されている。したがって、新事実性判定手段１０は、この仮説含意文Ｔ__Ｈにおいて仮説Ｈを新事実ではないと判定する。

[0053] 新事実性判定手段１０は、仮説Ｈと一致する句が仮説含意文Ｔ__Ｈに記載されている場合に、挿入句と判定してもよい。ただし、新事実性判定手段１０が挿入句を判定する方法は、内容が一致するか否かを判定する方法に限定されない。新事実性判定手段１０は、仮説Ｈが仮説含意文Ｔ__Ｈの挿入句であるか否かを、例えば、仮説Ｈに含まれる体言と用言が、仮説含意文Ｔ__Ｈにおいて連体修飾関係で接続しているか否かで判定してもよい。新事実性判定手段１０は、上記例の仮説Ｈ「ＱＱＲＲは逮捕された」に対して、体言と用言とを入れ替えた句である「逮捕されたＱＱＲＲ」を挿入句としてもよい。このように、仮説Ｈの記載内容と同等の記載内容を仮説含意文Ｔ__Ｈが含む場合、新事実性判定手段１０は、その仮説Ｈを新事実と判定する。

[0054] 以上のように、新事実性判定手段１０は、仮説含意文Ｔ__Ｈが予め定められた表現を含む場合、その表現に応じて、その表現を含む仮説含意文Ｔ__Ｈにおいて仮説Ｈが新事実か否かを判定してもよい。

[0055] なお、新事実か否かを判定する手がかりが複数存在する場合、これらの手がかりに予め優先順位を設けておけばよい。上記（１）～（５）の手がかりが存在する場合、例えば、優先順位を（５）、（３）、（１）、（４）、（２）の順に設け、新事実性判定手段１０は、この順に判定処理を行ってもよい。また、どの手がかりにも一致しない場合、新事実性判定手段１０は、仮説含意文Ｔ__Ｈにおいて仮説Ｈが新事実ではないと判定してもよい。

[0056] 新事実性判定手段１０は、判定した結果を新事実性判定結果記憶部２３に記憶させる。図４（ａ）は、新事実性判定結果記憶部２３が記憶する判定結果の例を示す説明図である。図４（ａ）に示す例は、図３に例示する状態に対して、新事実性判定手段１０が判定処理を実行した結果である。図４（ａ）に示す例は、仮説記憶部２１に格納された仮説ＩＤ＝１に対する仮説含意文が仮説含意文書記憶部２２の文書ＩＤ＝１、文ＩＤ＝１で特定される文で

あることを示す。

[0057] 具体的には、仮説含意文書記憶部 22 の文書 I D = 1、文 I D = 1 で特定される仮説含意文が (1) 及び (3) で示す手がかりの条件に一致するため、新事実性判定手段 10 は、仮説 H を新事実と判定する。そのため、新事実性判定手段 10 は、新事実性判定結果記憶部 23 の判定結果項目に、新事実であることを示す情報を設定する。図 4 (a) に示す例では、文書 I D = 1、文 I D = 1 で特定される仮説含意文において仮説 I D = 1 で特定される仮説が新事実 (判定結果 = 1) であることを示している。

[0058] 次に、含意判定手段 11 は、仮説含意文で仮説が新事実であると判定された場合であって、判定対象文の発生日時が仮説含意文の発生日時よりも古い場合、その仮説を判定対象文が含意しないと判定する。なお、この場合、含意判定手段 11 は、その仮説とその判定対象文間の含意度にペナルティを付与してもよい (図 2 におけるステップ S 2)。

[0059] まず、ステップ S 2 において、含意判定手段 11 が、仮説を判定対象文が含意しないと判定する動作を説明する。

[0060] 図 4 (c) は、含意判定結果記憶部 25 が記憶する判定結果の例を示す説明図である。図 4 (c) に例示する判定結果は、図 4 (a) に例示する判定結果をもとに、含意判定手段 11 が判定処理を行った結果を示す。仮説記憶部 21 に格納された仮説 I D = 1 で特定される仮説は、新事実性判定結果記憶部 23 を参照すると、文書 I D = 1、文 I D = 1 で特定される仮説含意文において新事実と判定されている。また、仮説含意文書記憶部 22 を参照すると、文書 I D = 1、文 I D = 1 で特定される仮説含意文の発生日時は「2010/04/01」である。

[0061] 一方、判定対象文書記憶部 20 に記憶された文書 I D = 10、文 I D = 1 で特定される判定対象文の日時は「2010/03/01」であるため、判定対象文の日時は仮説含意文の日時よりも古いことが分かる。したがって、含意判定手段 11 は、文書 I D = 10、文 I D = 1 で特定される判定対象文が仮説 I D = 1 で特定される仮説を含意しないと判定し、含意判定結果記憶

部 2 5 の判定結果に 0 を格納する。ここで、判定結果が 0 とは、判定対象文が仮説を含まないことを意味する。

[0062] 次に、ステップ S 2 において、含意判定手段 1 1 が、仮説と判定対象文間の含意度にペナルティを付与する動作を説明する。なお、含意度が用いられる場合、仮説と判定対象文との含意度は予め算出され、含意度記憶部 2 4 に記憶される。図 4 (b) は、含意度記憶部 2 4 が記憶する含意度の例を示す説明図である。図 4 (b) に示す例では、仮説記憶部 2 1 に格納された仮説 1 D = 1 に対する仮説含意文が仮説含意文書記憶部 2 2 の文書 1 D = 1、文 1 D = 1 で特定される文であり、その仮説と仮説含意文との含意度が 0.5 6 であることを示す。なお、含意度の算出方法については後述する。

[0063] 以下、仮説と判定対象文との含意度にペナルティを付与する方法を説明する。まず 1 つ目の方法は、含意度を一定の値または一定の割合だけ低減させる方法である。この場合、含意判定手段 1 1 は、低減後の含意度が基準値未満であればその仮説を判定対象文が含意しないと判定する。2 つ目の方法は、基準値を一定の値または一定の割合だけ増加させる方法である。この場合、含意判定手段 1 1 は、含意度が増加した基準値未満であればその仮説を判定対象文が含意しないと判定する。また、これらの方法は、同時に実行されてもよい。すなわち、含意度にペナルティを付与する第一の方法として、含意度の値を低減させる方法が挙げられる。また、第二の方法として、予め定められた基準値の値を増加させる方法が挙げられる。さらに、第三の方法として、含意度の値を低減させるとともに基準値の値を増加させる方法が挙げられる。

[0064] 図 4 (a) ~ 図 4 (c) を用いて、2 つ目の方法を用いてペナルティを付与した場合の判定処理を説明する。仮説記憶部 2 1 に格納された仮説 1 D = 1 で特定される仮説は、図 4 (a) に例示する新事実性判定結果記憶部 2 3 を参照すると、文書 1 D = 1、文 1 D = 1 で特定される仮説含意文において新事実 (判定結果 = 1) と判定されている。また、仮説含意文書記憶部 2 2 を参照すると、文書 1 D = 1、文 1 D = 1 で特定される仮説含意文の日は「

「2010/04/01」である。一方、判定対象文書記憶部20に格納された文書ID=10、文ID=1で特定される判定対象文の日時は「2010/03/01」である。

[0065] さらに、図4(b)に例示する含意度記憶部24を参照すると、仮説ID=1で特定される仮説と、文書ID=10、文ID=1で特定される判定対象文の間の含意度は0.56である。基準値を0.5、増加させる一定の値を0.2とした場合、基準値は、0.7となる。仮説ID=1で特定される仮説と文書ID=10、文ID=1で特定される判定対象文との間の含意度は0.56であり、基準値0.5に0.2を足した0.7よりも小さい。そのため、含意判定手段11は、その仮説を判定対象文が含意しないと判定し、含意判定結果記憶部25に判定結果0を格納する。

[0066] ここで、含意度の計算方法を説明する。含意度は、仮説と判定対象文の双方に含まれる単語または係り受け組の共通する割合として計算できる。具体的には、含意度は、下記の式で計算される。

[0067]
$$\text{Sim}(H, T) = \sum_{t \in H \cap T} \text{weight}(t) / \sum_{t \in H} \text{weight}(t)$$

[0068] ここで、Hは仮説を表し、Tは判定対象文を表す。また $t \in H$ は、Hに含まれる単語を表し、 $t \in H \cap T$ は、HとTに共通する単語を表し、 $\text{weight}(t)$ は単語tの重要度を表す。

[0069] 仮説および判定対象文から、形態素解析により単語tを抽出すればよい。具体的には、形態素解析ソフトウェアを用いて仮説および判定対象文から単語tを抽出してもよい。形態素解析ソフトウェアを用いた場合、文が単語に分割され、それぞれの単語に品詞が付与される。この際、全ての品詞を採用するのではなく、名詞、動詞、形容詞、形容動詞等の自立語のみを単語tとして採用することが可能である。

[0070] 例えば、仮説Hが「Aさんを東京で逮捕した」であれば、形態素解析によって、「Aさん」、「東京」、「逮捕」が抽出され、これらの単語が仮説Hに含まれる単語tの集合になる。なお、形態素解析ソフトウェアとして、例

えば、C h a S e n が利用可能である。なお、C h a S e n については、以下の参考文献 1 に紹介されている。

[0071] <参考文献 1>” ChaSen -- 形態素解析器”、2007年、Nara Institute of Science and Technology、[online]、[平成 22 年 10 月 18 日検索]、インターネット <<http://chasen-legacy.sourceforge.jp/>>

[0072] 最も単純な方法として、単語 t の重要度 $w e i g h t (t)$ に、全て 1 を設定する方法がある。この場合、 $S i m (H, T)$ は、仮説 H に含まれる単語と判定対象文 T に含まれる単語が共通する割合を意味する。また、別の方法として、単語 t の出現文書数を用いる方法もある。この場合、単語 t の出現文書数を $d f (t)$ 、全文書数を $|D|$ としたとき、単語 t の重要度は、 $w e i g h t (t) = \log (|D| / d f (t))$ で計算される。なお、 $\log$ は、底を 2 または 10 とする対数である。

[0073] ここで、図 3 に例示する仮説記憶部 21 に記憶された仮説 $ID = 1$ で特定される仮説 H と、判定対象文書記憶部 20 に記憶された文書 $ID = 10$ 、文 $ID = 1$ で特定される判定対象文 T の間の含意度を計算する方法を説明する。仮説 H を形態素解析して抽出された単語のうち自立語を選択すると、仮説 H に含まれる単語 t の集合は、「QQ」、「RR」、「容疑者」、「逮捕」、「峠道」、「若い」、「車」、「通い詰める」、「場所」になる。同様の方法で、判定対象文 T から単語 t が抽出される。仮説 H に含まれる 9 つの単語のうち、判定対象文 T にも含まれる単語は「峠道」、「若い」、「車」、「通い詰める」、「場所」の 5 つである。ここで、 $w e i g h t (t)$ を全て 1 とした場合、類似度は $5 / 9 = 0.56$ と計算される。なお、このように図 3 に例示する状態に対して計算された含意度が、図 4 (b) に例示する含意度記憶部 24 に記憶されている

[0074] なお、上述した含意度の計算方法は、仮説と判定対象文間の含意度を計算する一例であり、上記の方式に限定されない。例えば、 $S i m (H, T)$ として、 $C o s i n e$ 類似度や、 $J a c c a r d$ 係数を用いても良い。

[0075] また、上記説明では、 $S i m (H, T)$ を算出する際、 t として単語を用

いる方法を説明したが、単語を用いる代わりに、単語の係り受け組を用いてもよい。係り受け組とは、係り受け関係にある2つの単語の組を表す。例えば、仮説Hが、「Aさんを東京で逮捕した」である場合、係り受け組として、「Aさんー逮捕」および「東京ー逮捕」が仮説Hに含まれるtの集合になる。

[0076] また、含意度の計算に分類器を用いても良い。分類器とは、データを自動的に2つのカテゴリに分類する方式であり、分類器を実装したソフトウェアも広く知られている。分類器を実装したソフトウェアを使用する際、利用者は、(i) 2つのカテゴリに事前分類したデータの単語ベクトル、および、(ii) 未分類のデータの単語ベクトル、の2つを事前に用意する。特徴ベクトルとは、データに含まれる単語をベクトルの次元とし、各次元の値がそのデータにおける単語の有無(0/1)または単語の重要度を表すベクトルである。

[0077] 分類器を実装したソフトウェアは、学習処理および分類処理の2つの部分から構成されている。まず学習処理では、事前分類したデータの単語ベクトルを入力データとして分類器が作成される。この際、分類器では、通常、どのような単語が含まれる場合に文書が一方のカテゴリに所属する可能性が高いかを表す分類基準が決定される。次に、分類処理では、学習処理で作成された分類器を用いて、未分類のデータが2つのカテゴリのいずれかに分類される。

[0078] この分類器を実装したソフトウェアを本発明に適用するには、分類対象のデータを仮説と判定対象文の組とし、2つのカテゴリを「仮説を判定対象文が含意するか、含意しないか」とすればよい。つまり、(i) 2つのカテゴリに事前分類した仮説と判定対象文の組の単語ベクトルと、(ii) 未分類の仮説と判定対象文の組の単語ベクトルとを用意すれば、処理内容は上記と同様である。

[0079] なお、仮説と判定対象文それぞれの単語ベクトルの和を、仮説と判定対象文の組の単語ベクトルとすればよい。分類器を用いた結果、データが「含意

する」とするカテゴリに分類された場合、そのデータの判定結果を1とし、データが「含意しない」とするカテゴリに分類された場合、そのデータの判定結果を0として、それぞれの判定結果を含意度記憶部24に記憶させる。

[0080] なお、分類器を実装したソフトウェアの例として、Support Vector Machine (SVM) を実装したSVM-Lightや、決定木を実装したC4.5などが知られている。なお、SVM-Lightについては、以下の参考文献2に紹介されている。また、C4.5については、以下の参考文献3に示すWebページからダウンロード可能である。

[0081] <参考文献2> Thorsten Joachims, "Support Vector Machine"、2008年8月14日、[online]、[平成22年10月18日検索]、インターネット〈<http://svmlight.joachims.org/>〉

<参考文献3> "Ross Quinlan"、[online]、[平成22年10月18日検索]、インターネット〈<http://www.rulequest.com/Personal/>〉

[0082] 以上のように、本実施形態によれば、新事実性判定手段10が、仮説含意文T__Hにおいて、与えられた仮説Hが初めて明らかにされた事実（新事実）であるか否かを、その仮説含意文T__Hに記載された特定の表現をもとに判定する。そして、含意判定手段11が、その仮説含意文T__Hにおいて、与えられた仮説Hが新事実であると判定された場合に、判定対象文Tの発生日時とその仮説含意文T__Hの発生日時とを比較して、与えられた仮説Hを判定対象文Tが含意するか否かを判定する。

[0083] 具体的には、判定対象文Tの発生日時が仮説含意文T__Hの発生日時よりも古い場合、含意判定手段11が、与えられた仮説Hを判定対象文T__Hが含意しないと判定する。このような構成により、含意判定の性能を向上させることができる。具体的には、再現率を低下させずに適合率を向上させることができる。

[0084] すなわち、仮説Hが含意判定文T__Hにおいて新事実であるかを新事実性判定手段10が事前に判定する。その仮説が新事実と判定された場合に、判定対象文Tの日付が含意判定文T__Hの日付よりも古ければ、仮説Hを判定

対象文 T が含意しないと含意判定手段 11 が判定する。そのため、含意判定の性能を向上させることができる。

[0085] これは、仮説 H が含意判定文 T__H において新事実ならば、含意判定文 T__H よりも古い判定対象文 T が仮説 H を含意することはないからである。よって、一般的な方式のように、基準値を一律に高く設定しなくても（すなわち、再現率を低下させずに）、仮説 H を含意しない判定対象文 T を正確に判定できる（すなわち、適合率を向上できる）。

[0086] 例えば、非特許文献 1 に記載された一般的な方式では、含意度（すなわち、含意度記憶部 24 に記憶された含意度）のみを用いて判定処理を行う。この場合、図 4（b）に例示する含意度記憶部 24 の内容を参照すると、仮説 ID = 1 で特定される仮説と、文書 ID = 10、文 ID = 1 で特定される判定対象文間の類似度は 0.56 である。ここで、基準値が 0.5 の場合、非特許文献 1 に記載された一般的な方式では、その判定対象文が仮説を含意すると判定される。しかし、この判定対象文は仮説を含意しないため誤りである。一方、本実施形態によれば、含意判定手段 11 は、新事実性判定手段 10 による判定結果を用いて判定を行うため、この判定対象文が仮説を含意しないと判定できる。

[0087] 実施形態 2.

図 5 は、本発明の第 2 の実施形態における含意判定装置の例を示すブロック図である。なお、第 1 の実施形態と同様の構成については、図 1 と同一の符号を付し、説明を省略する。本実施形態における含意判定装置は、プログラム制御により動作するデータ処理装置 1 と、情報を記憶する記憶装置 2 とを備えている。

[0088] データ処理装置 1 は、新事実性判定手段 10 と、含意判定手段 11 と、新規仮説判定手段 12 とを含む。このうち、新事実性判定手段 10 及び含意判定手段 11 の内容については、第 1 の実施形態と同様のため、説明を省略する。

[0089] 新規仮説判定手段 12 は、いずれの判定対象文にも含意されない仮説を特

定する。具体的には、新規仮説判定手段１２は、含意判定手段１１の判定結果をもとに、判定対象文書集合のうち、含意する判定対象文が一つも存在しない仮説を、いずれの判定対象文にも含意されない仮説と判定する。新規仮説判定手段１２は、例えば、含意判定手段１１が含意判定結果記憶部２５に記憶させた仮説のうち、判定結果に「１」が存在しない仮説を、いずれの判定対象文にも含意されない仮説と特定してもよい。

[0090] 記憶装置２は、判定対象文書記憶部２０と、仮説記憶部２１と、仮説含意文書記憶部２２と、新事実性判定結果記憶部２３と、含意度記憶部２４と、含意判定結果記憶部２５と、新規仮説記憶部２６とを含む。このうち、判定対象文書記憶部２０、仮説記憶部２１、仮説含意文書記憶部２２、新事実性判定結果記憶部２３、含意度記憶部２４、および、含意判定結果記憶部２５の内容については、第１の実施形態と同様のため、説明を省略する。

[0091] 新規仮説記憶部２６は、新規仮説判定手段１２でいずれの判定対象文にも含意されないと判定された仮説を記憶する。新規仮説記憶部２６は、仮説の内容を記憶してもよく、仮説ＩＤを記憶してもよい。新規仮説記憶部２６は、例えば、磁気ディスク等により実現される。

[0092] 新事実性判定手段１０と、含意判定手段１１と、新規仮説判定手段１２とは、プログラム（含意判定プログラム）に従って動作するコンピュータのＣＰＵによって実現される。また、新事実性判定手段１０と、含意判定手段１１と、新規仮説判定手段１２とは、それぞれが専用のハードウェアで実現されていてもよい。

[0093] 次に、本実施形態の含意判定装置の動作を説明する。図６は、第２の実施形態における含意判定装置の動作を示すフローチャートである。また、図７は、記憶装置２が記憶する情報の例を示す説明図である。なお、仮説が新事実か否かを新事実性判定手段１０が判定し、含意判定手段１１が、その仮説を判定対象文が含意するか否かを判定するまでの処理は、図２におけるステップＳ１～ステップＳ２の処理と同様であるため、説明を省略する。

[0094] 新規仮説判定手段１２は、含意判定手段１１の判定結果をもとに、含意す

る文が一つも存在しない仮説を判定対象文書集合に記載されていないと判定する（図6におけるステップS3）。

[0095] 図7は、新規仮説記憶部26の内容の例を示す説明図である。図4（c）に例示する含意判定結果記憶部25には、仮説ID=1で特定される仮説が記憶されているが、判定結果=1のデータは存在しない。そのため、仮説ID=1で特定される仮説を含意する判定対象文は存在しない。したがって、新規仮説判定手段12は、いずれの判定対象文にも仮説ID=1で特定される仮説が含まれない仮説と判定し、図7に例示するように、新規仮説記憶部26に仮説ID=1を記憶させる。

[0096] 以上のように、本実施形態によれば、第1の実施形態における効果に加え、仮説が判定対象文書集合に記載されているかを判別できる。その理由は、新規仮説判定手段12が、含意判定手段11の結果を参照して、含意する判定対象文が一つも存在しない仮説を判定対象文書集合に記載されていないと判定するためである。すなわち、新規仮説判定手段12が、含意判定手段11によって含意する判定対象文が存在すると判定されなかった仮説が存在する場合、その仮説をいずれの判定対象文にも含意されない仮説と判定するためである。

[0097] また、本実施形態における含意判定装置を用いることにより、Update Summarization Task（UST）における要約の候補文の選択が可能になる。USTとは、文書集合Aの記載内容を、それ以前に記述された文書集合Bの内容が既知であることを条件として要約するタスクである。文書集合Aに含まれるそれぞれの文を仮説とし、文書集合Bの文を判定対象文とすれば、本実施形態による含意判定装置が、文書集合Bに記載されていない文書集合Aに含まれる文を判定できる。このようにして判定された文を要約の候補文とすればよい。

[0098] 実施形態3.

図8は、本発明の第3の実施形態における含意判定装置の例を示すブロック図である。なお、第1の実施形態と同様の構成については、図1と同一の

符号を付し、説明を省略する。本実施形態における含意判定装置は、プログラム制御により動作するデータ処理装置１と、情報を記憶する記憶装置２とを備えている。

[0099] 記憶装置２は、判定対象文書記憶部２０と、仮説記憶部２１と、仮説含意文書記憶部２２と、新事実性判定結果記憶部２３と、含意度記憶部２４と、含意判定結果記憶部２５を有する。記憶装置２の内容は、第１の実施形態と同様のため、説明を省略する。

[0100] データ処理装置１は、新事実性判定手段１０と、含意判定手段１１と、仮説含意文追加手段１３とを含む。このうち、新事実性判定手段１０及び含意判定手段１１の内容については、第１の実施形態と同様のため、説明を省略する。

[0101] 仮説含意文追加手段１３は、含意判定手段１１の判定結果をもとに、仮説を含意する判定対象文を仮説含意文として追加し、その内容を新事実性判定手段１０に通知する。具体的には、含意判定手段１１が与えられた仮説を判定対象文が含意すると判定した場合、仮説含意文追加手段１３は、その判定対象文を仮説含意文として追加する。このとき、仮説含意文追加手段１３は、仮説を含意すると判定した判定対象文の内容を仮説記憶部２１および仮説含意文書記憶部２２に記憶させてもよい。

[0102] 新事実性判定手段１０と、含意判定手段１１と、仮説含意文追加手段１３とは、プログラム（含意判定プログラム）に従って動作するコンピュータのＣＰＵによって実現される。また、新事実性判定手段１０と、含意判定手段１１と、仮説含意文追加手段１３とは、それぞれが専用のハードウェアで実現されていてもよい。

[0103] 次に、本実施形態の含意判定装置の動作を説明する。図９は、第３の実施形態における含意判定装置の動作を示すフローチャートである。また、図１０～図１３は、記憶装置２が記憶する情報の例を示す説明図である。なお、仮説が新事実か否かを新事実性判定手段１０が判定し、含意判定手段１１が、その仮説を判定対象文が含意するか否かを判定するまでの処理は、図２に

おけるステップS 1～ステップS 2の処理と同様である。ただし、本動作の説明では、初期状態として図10に例示する内容が、仮説記憶部21（図10（a）参照。）、仮説含意文書記憶部22（図10（b）参照。）、判定対象文書記憶部20（図10（c）参照。）に記憶されているものとする。具体的には、仮説記憶部21が仮説ID=2で特定される仮説を記憶している点、および、判定対象文書記憶部20が文書ID=11、文ID=1で特定される文、および、文書ID=12、文ID=1で特定される文を記憶している点において、第1の実施形態と異なる。

[0104] ステップS 2までの処理の結果、図11に例示する内容が、新事実性判定結果記憶部23（図11（a）参照。）、含意度記憶部24（図11（b）参照。）に記憶される。図11（a）に例示するように、仮説ID=2で特定される仮説は、文書ID=1、文ID=1で特定される仮説含意文において新事実ではない（判定結果=0）と判定されている。これは、文書ID=1、文ID=1で特定される仮説含意文が、第1の実施形態で示す手がかり（5）の条件に一致するためである。

[0105] また、含意度記憶部24には、仮説ID=2で特定される仮説と、判定対象文書記憶部20に格納された2つの判定対象文（文書ID=11、文ID=1で特定される文、および、文書ID=12、文ID=1で特定される文）との間の含意度が、第1の実施形態に記載された方法を用いて計算され、計算された含意度が事前に記憶されているものとする。図11（b）に示す例では、含意度はそれぞれ1.0と0.67である。ここで、基準値が0.5と定められている場合、いずれの含意度も基準値以上であるため、いずれの判定対象文も仮説を含意すると判定されることになる。判定された結果は、図11（c）に例示するように、共に仮説を含意する判定対象文として含意判定結果記憶部25に記憶される。

[0106] ステップS 2の後、仮説含意文追加手段13は、含意判定手段11の判定結果をもとに、仮説を含意する判定対象文を仮説含意文として追加し、その内容を新事実性判定手段10に通知する（図9におけるステップS 4）。こ

のとき、仮説含意文追加手段 13 は、仮説記憶部 21 と仮説含意文書記憶部 22 に、判定対象文の内容を記憶させてもよい。

[0107] 例えば、含意判定手段 11 が、仮説と判定対象文との間の含意度が 0.7 以上の判定対象文を仮説含意文と判定した場合、図 11 (b) より、仮説含意文追加手段 13 は、含意度が 1.0 である文書 ID=11、文 ID=1 で特定される判定対象文を仮説含意文書記憶部 22 に記憶させる。また、仮説含意文追加手段 13 は、仮説記憶部 21 にも、仮説 ID=2 で特定される仮説を含意する文書 ID=11、文 ID=1 の内容を記憶させる。この結果、仮説含意文書記憶部 22 の内容は、図 10 に例示する初期状態から、図 12 に例示する状態に変更される。

[0108] ステップ S4 において、仮説含意文が追加された場合（図 9 のステップ S5 における YES）、仮説が新事実か否かを判定するステップ S1 以降の処理を繰り返す。一方、仮説含意文が追加されていない場合（ステップ S5 における NO）、処理を終了する。

[0109] 図 12 に例示する変更された仮説記憶部 21 と、仮説含意文書記憶部 22 とを用いて、再度、ステップ S1 及びステップ S2 を実行した結果を図 13 に示す。一回目との違いは、仮説 ID=2 で特定される仮説が、文書 ID=11、文 ID=1 において新事実と判定されている点である。これは、文書 ID=11、文 ID=1 で特定される仮説含意文が、第 1 の実施形態で示す手がかり (1) の条件に一致するためである。

[0110] 以上のように、本実施形態によれば、第 1 の実施形態における効果に加え、含意文検索性能をさらに向上できる。その理由は、含意判定手段 11 が仮説を含意すると判定した判定対象文を仮説含意文追加手段 13 が仮説含意文として新事実性判定手段 10 に通知するためである。すなわち、含意判定手段 11 が与えられた仮説を判定対象文が含意すると判定した場合、仮説含意文追加手段 13 がその判定対象文を仮説含意文として追加するためである。

[0111] 例えば、初期状態で、図 10 に例示する内容が記憶装置 2 に記憶されているとする。一回目の処理では、仮説 ID=2 で特定される仮説を、文書 ID

= 1 2、文 I D = 1 の判定対象文が含意すると判定される。しかし、これは本来誤りである。このような誤りが発生した場合であっても、一回目の処理の結果、文書 I D = 1 1、文 I D = 1 で特定される判定対象文が仮説含意文になる。そのため、二回目の処理を実行することで、仮説 I D = 2 で特定される仮説を、文書 I D = 1 2、文 I D = 1 の判定対象文が含意しないと修正できる。

[0112] 次に、本発明の最小構成を説明する。図 1 4 は、本発明による含意判定装置の最小構成の例を示すブロック図である。本発明による含意判定装置は、与えられた仮説（例えば、仮説 H）を含意する文である仮説含意文（例えば、仮説含意文 T__H）で、その仮説が初めて明らかにされた事実を示す新事実であるか否かを、その仮説含意文に記載された特定の表現（例えば、手があり（1）～（5）など）をもとに判定する新事実性判定手段 8 1（例えば、新事実性判定手段 1 0）と、与えられた仮説を、その仮説を含むか否かが判定される文である判定対象文（例えば、判定対象文 T）が含意するか否かを判定する含意判定手段 8 2（例えば、含意判定手段 1 1）とを備えている。

[0113] 含意判定手段 8 2 は、与えられた仮説が仮説含意文で新事実であると判定された場合に、判定対象文の発生日時（例えば、図 3（c）における日時）とその仮説を含意する仮説含意文の発生日時（例えば、図 3（b）における日時）とを比較して、与えられた仮説を判定対象文が含意するか否かを判定する。

[0114] そのような構成により、含意判定の性能を向上させることができる。

[0115] 具体的には、含意判定手段 8 2 は、判定対象文の発生日時が与えられた仮説を含意する仮説含意文の発生日時よりも古い場合、その仮説を判定対象文が含意しないと判定してもよい。また、含意判定手段 8 2 は、判定対象文の発生日時が与えられた仮説を含意する仮説含意文の発生日時よりも古い場合、判定対象文がその仮説を含意する度合いを示す含意度にペナルティを付与し、予め定められた基準値とペナルティ付与後の含意度とを比較して、与え

られた仮説を判定対象文が含意するか否かを判定してもよい。

[0116] また、上記の実施形態の一部又は全部は、以下の付記のようにも記載されるが、以下には限られない。

[0117] (付記 1) 与えられた仮説を含意する文である仮説含意文で、当該仮説が初めて明らかにされた事実を示す新事実であるか否かを、当該仮説含意文に記載された特定の表現をもとに判定する新事実性判定手段と、与えられた仮説を、当該仮説を含むか否かが判定される文である判定対象文が含意するか否かを判定する含意判定手段とを備え、前記含意判定手段は、与えられた仮説が仮説含意文で新事実であると判定された場合に、前記判定対象文の発生日時と当該仮説を含意する仮説含意文の発生日時とを比較して、与えられた仮説を前記判定対象文が含意するか否かを判定することを特徴とする含意判定装置。

[0118] (付記 2) 含意判定手段は、判定対象文の発生日時が与えられた仮説を含意する仮説含意文の発生日時よりも古い場合、当該仮説を前記判定対象文が含意しないと判定する付記 1 記載の含意判定装置。

[0119] (付記 3) 含意判定手段は、判定対象文の発生日時が与えられた仮説を含意する仮説含意文の発生日時よりも古い場合、判定対象文が当該仮説を含意する度合いを示す含意度の値を低減させ若しくは予め定められた基準値の値を増加させ又は含意度の値を低減させるとともに前記基準値の値を増加させ、前記基準値と前記含意度とを比較して、与えられた仮説を前記判定対象文が含意するか否かを判定する付記 1 記載の含意判定装置。

[0120] (付記 4) 含意判定手段は、判定対象文の発生日時が与えられた仮説を含意する仮説含意文の発生日時よりも古い場合、含意度を一定の値または一定の割合低減させ、前記含意度が予め定められた基準値未満である場合、与えられた仮説を判定対象文が含意しないと判定する付記 3 記載の含意判定装置。

[0121] (付記 5) 含意判定手段は、判定対象文の発生日時が与えられた仮説を含意する仮説含意文の発生日時よりも古い場合、予め定められた基準値を一定の値または一定の割合増加させ、含意度が前記基準値未満である場合、与えら

れた仮説を判定対象文が含意しないと判定する付記 3 記載の含意判定装置。

[0122] (付記 6) いずれの判定対象文にも含意されない仮説を特定する新規仮説判定手段を備えた付記 1 から付記 5 のうちのいずれか 1 つに記載の含意判定装置。

[0123] (付記 7) 与えられた仮説を判定対象文が含意すると判定された場合、当該判定対象文を仮説含意文として追加する仮説含意文追加手段を備えた付記 1 から付記 6 のうちのいずれか 1 つに記載の含意判定装置。

[0124] (付記 8) 与えられた仮説を含意する文である仮説含意文で、当該仮説が初めて明らかにされた事実を示す新事実であるか否かを、当該仮説含意文に記載された特定の表現をもとに判定し、与えられた仮説が前記仮説含意文で新事実であると判定された場合に、与えられた仮説を含むか否かが判定される文である判定対象文の発生日時と当該仮説を含意する仮説含意文の発生日時とを比較して、与えられた仮説を前記判定対象文が含意するか否かを判定することを特徴とする含意判定方法。

[0125] (付記 9) 判定対象文の発生日時が与えられた仮説を含意する仮説含意文の発生日時よりも古い場合、当該仮説を前記判定対象文が含意しないと判定する付記 8 記載の含意判定方法。

[0126] (付記 10) 判定対象文の発生日時が与えられた仮説を含意する仮説含意文の発生日時よりも古い場合、判定対象文が当該仮説を含意する度合いを示す含意度の値を低減させ若しくは予め定められた基準値の値を増加させ又は含意度の値を低減させるとともに前記基準値の値を増加させ、前記基準値と前記含意度とを比較して、与えられた仮説を前記判定対象文が含意するか否かを判定する付記 8 記載の含意判定方法。

[0127] (付記 11) コンピュータに、与えられた仮説を含意する文である仮説含意文で、当該仮説が初めて明らかにされた事実を示す新事実であるか否かを、当該仮説含意文に記載された特定の表現をもとに判定する新事実性判定処理、および、与えられた仮説を、当該仮説を含むか否かが判定される文である判定対象文が含意するか否かを判定する含意判定処理を実行させ、前記含意

判定処理で、与えられた仮説が仮説含意文で新事実であると判定された場合に、前記判定対象文の発生日時と当該仮説を含意する仮説含意文の発生日時とを比較させ、与えられた仮説を前記判定対象文が含意するか否かを判定させるための含意判定プログラム。

[0128] (付記 1 2) コンピュータに、含意判定処理で、判定対象文の発生日時が与えられた仮説を含意する仮説含意文の発生日時よりも古い場合、当該仮説を前記判定対象文が含意しないと判定させる付記 1 1 記載の含意判定プログラム。

[0129] (付記 1 3) コンピュータに、含意判定処理で、判定対象文の発生日時が与えられた仮説を含意する仮説含意文の発生日時よりも古い場合、判定対象文が当該仮説を含意する度合いを示す含意度の値を低減させ若しくは予め定められた基準値の値を増加させ又は含意度の値を低減させるとともに前記基準値の値を増加させ、前記基準値と前記含意度とを比較させて、与えられた仮説を前記判定対象文が含意するか否かを判定させる付記 1 1 記載の含意判定プログラム。

[0130] 以上、実施形態及び実施例を参照して本願発明を説明したが、本願発明は上記実施形態および実施例に限定されるものではない。本願発明の構成や詳細には、本願発明のScope内で当業者が理解し得る様々な変更をすることができる。

[0131] この出願は、2010年10月21日に提出された日本特許出願2010-236548を基礎とする優先権を主張し、その開示の全てをここに取り込む。

産業上の利用可能性

[0132] 本発明の含意判定装置は、文書検索システムにおける検索結果の絞り込みや、文書要約システムにおける要約の候補文の選択に利用できる。

符号の説明

- [0133] 1 データ処理装置
10 新事実性判定手段

- 1 1 含意判定手段
- 1 2 新規仮説判定手段
- 1 3 仮説含意文追加手段
- 2 記憶装置
- 2 0 判定対象文書記憶部
- 2 1 仮説記憶部
- 2 2 仮説含意文書記憶部
- 2 3 新事実性判定結果記憶部
- 2 4 含意度記憶部
- 2 5 含意判定結果記憶部
- 2 6 新規仮説記憶部

請求の範囲

- [請求項1] 与えられた仮説を含意する文である仮説含意文で、当該仮説が初めて明らかにされた事実を示す新事実であるか否かを、当該仮説含意文に記載された特定の表現をもとに判定する新事実性判定手段と、
- 与えられた仮説を、当該仮説を含むか否かが判定される文である判定対象文が含意するか否かを判定する含意判定手段とを備え、
- 前記含意判定手段は、与えられた仮説が仮説含意文で新事実であると判定された場合に、前記判定対象文の発生日時と当該仮説を含意する仮説含意文の発生日時とを比較して、与えられた仮説を前記判定対象文が含意するか否かを判定することを特徴とする含意判定装置。
- [請求項2] 含意判定手段は、判定対象文の発生日時が与えられた仮説を含意する仮説含意文の発生日時よりも古い場合、当該仮説を前記判定対象文が含意しないと判定する
- 請求項 1 記載の含意判定装置。
- [請求項3] 含意判定手段は、判定対象文の発生日時が与えられた仮説を含意する仮説含意文の発生日時よりも古い場合、判定対象文が当該仮説を含意する度合いを示す含意度の値を低減させ若しくは予め定められた基準値の値を増加させ又は含意度の値を低減させるとともに前記基準値の値を増加させ、前記基準値と前記含意度とを比較して、与えられた仮説を前記判定対象文が含意するか否かを判定する
- 請求項 1 記載の含意判定装置。
- [請求項4] 含意判定手段は、判定対象文の発生日時が与えられた仮説を含意する仮説含意文の発生日時よりも古い場合、含意度を一定の値または一定の割合低減させ、前記含意度が基準値未満である場合、与えられた仮説を判定対象文が含意しないと判定する
- 請求項 3 記載の含意判定装置。
- [請求項5] 含意判定手段は、判定対象文の発生日時が与えられた仮説を含意す

る仮説含意文の発生日時よりも古い場合、予め定められた基準値を一定の値または一定の割合増加させ、含意度が前記基準値未満である場合、与えられた仮説を判定対象文が含意しないと判定する

請求項 3 記載の含意判定装置。

[請求項6] いずれの判定対象文にも含意されない仮説を特定する新規仮説判定手段を備えた

請求項 1 から請求項 5 のうちのいずれか 1 項に記載の含意判定装置

。

[請求項7] 与えられた仮説を判定対象文が含意すると判定された場合、当該判定対象文を仮説含意文として追加する仮説含意文追加手段を備えた

請求項 1 から請求項 6 のうちのいずれか 1 項に記載の含意判定装置

。

[請求項8] 与えられた仮説を含意する文である仮説含意文で、当該仮説が初めて明らかにされた事実を示す新事実であるか否かを、当該仮説含意文に記載された特定の表現をもとに判定し、

与えられた仮説が前記仮説含意文で新事実であると判定された場合に、与えられた仮説を含むか否かが判定される文である判定対象文の発生日時と当該仮説を含意する仮説含意文の発生日時とを比較して、与えられた仮説を前記判定対象文が含意するか否かを判定する

ことを特徴とする含意判定方法。

[請求項9] コンピュータに、

与えられた仮説を含意する文である仮説含意文で、当該仮説が初めて明らかにされた事実を示す新事実であるか否かを、当該仮説含意文に記載された特定の表現をもとに判定する新事実性判定処理、および、

与えられた仮説を、当該仮説を含むか否かが判定される文である判定対象文が含意するか否かを判定する含意判定処理を実行させ、

前記含意判定処理で、与えられた仮説が仮説含意文で新事実である

と判定された場合に、前記判定対象文の発生日時と当該仮説を含意する仮説含意文の発生日時とを比較させ、与えられた仮説を前記判定対象文が含意するか否かを判定させる

ための含意判定プログラム。