(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第5559352号
(P5559352)

(45)発行日　平成26年7月23日(2014.7.23)　(24)登録日　平成26年6月13日(2014.6.13)

(51)Int.Cl.　F I
G06F 17/30 (2006.01)　G06F 17/30 220Z
G06F 17/30 210A
G06F 17/30 170A

請求項の数 16　(全 51 頁)

(21)出願番号　特願2012-543372(P2012-543372)
(86)(22)出願日　平成24年5月31日(2012.5.31)
(86)国際出願番号　PCT/JP2012/003586
(87)国際公開番号　WO2013/179346
(87)国際公開日　平成25年12月5日(2013.12.5)
審査請求日　平成24年9月24日(2012.9.24)

(73)特許権者　000003078
株式会社東芝
東京都港区芝浦一丁目１番１号
(73)特許権者　301063496
東芝ソリューション株式会社
神奈川県川崎市幸区堀川町７２番地３４
(74)代理人　100159938
弁理士　砂井　正之
(74)代理人　100149803
弁理士　藤原　康高
(74)代理人　100078019
弁理士　山下　一
(72)発明者　牧野　恭子
東京都港区芝浦一丁目１番１号　東芝ソリューション株式会社内

最終頁に続く

(54)【発明の名称】知見抽出装置、知見更新装置、及びプログラム

(57)【特許請求の範囲】
【請求項１】
電子文書から知見情報を抽出する知見抽出装置であって、
前記電子文書を受信する情報受信部と、
前記知見情報を抽出する対象語と、前記知見情報を抽出する手掛かり語とに基づいて前記電子文書から概念を抽出し、抽出された当該概念と前記対象語を関連付けた知見情報を作成する知見抽出部と、
前記抽出された知見情報を記憶する記憶部と、
前記知見情報が記憶された後、前記電子文書を前記記憶部内の知見情報に基づいて分析する情報分析部と、
前記電子文書に含まれる前記概念もしく前記対象語を検索する電子文書検索部と、
前記検索部による検索の結果、前記電子文書に前記概念が規定数以上記載された場合に前記概念を抽出し、前記電子文書に前記対象語が規定数以上記載された場合に前記対象語に関連付けられた前記知見情報を前記記憶部から抽出する知見登録概念抽出部と、
前記知見登録概念抽出部が抽出した前記対象語もしくは前記概念に関連する数値情報の大きさもしくは変動を評価する評価部と、
前記数値情報の大きさもしくは変動の方向もしくは数値情報の変動の大きさに従って前記概念と前記対象語の関連性の評価数値である確信度を更新する確信度更新部と
を備える知見抽出装置。
【請求項２】

請求項１に記載の知見抽出装置において、
前記確信度が規定値を下回った場合にユーザに判断を求め、ユーザの指示により選択中の前記概念間の関連を削除する知見更新部を備える知見抽出装置。
【請求項３】
前記手掛かり語はあらかじめ強度が設定され
前記知見抽出部は、
前記手掛かり語の強度と前記抽出された概念と前記対象語を関連付けた知見情報関連付けた知見情報を作成する請求項１に記載の知見抽出装置。
【請求項４】
配信されたニュース情報から株取引に関する知見情報を抽出する知見抽出装置であって、
前記ニュース情報を受信する情報受信部と、
前記受信されたニュース情報から経済用語を含むニュース情報を選択する情報選択部と、
前記選択されたニュース情報から前記知見情報を抽出する知見抽出部と、
前記抽出された知見情報を記憶する記憶部と、
前記知見情報が記憶された後、前記情報受信部に受信された全てのニュース情報を前記記憶部内の知見情報に基づいて分析する情報分析部と
を備え、
前記知見抽出部は、
前記選択されたニュース情報が株価変動を評価する株価変動評価語を含むか否かを判定する株価変動評価語判定部と、
前記判定の結果、前記ニュース情報が株価変動評価語を含むとき、当該ニュース情報が銘柄のグループ名を定義するグループ定義語を含むか否かを判定するグループ定義語判定部と、
この判定の結果、前記ニュース情報がグループ定義語を含むとき、当該グループ定義語に基づいて当該ニュース情報からグループ名を抽出するグループ名抽出部と、
前記株価変動評価語判定部による判定の結果、前記ニュース情報が株価変動評価語を含むとき、当該ニュース情報から企業名及び証券コードを含む銘柄を抽出する銘柄抽出部と、
前記株価変動評価語判定部による判定の結果、前記ニュース情報が株価変動評価語を含むとき、当該ニュース情報から株価変動材料を抽出する材料抽出部と、
前記抽出されたグループ名、銘柄及び株価変動材料を組み合わせて前記知見情報を作成する組み合わせ作成部と、
前記作成された知見情報を前記記憶部に書込む書込部と
を備え、
前記知見情報は、前記グループ名、前記銘柄及び前記株価変動材料を含む情報、又は前記グループ名及び前記銘柄からなる第１情報と前記株価変動材料からなる第２情報とを含む情報である知見抽出装置。
【請求項５】
請求項４に記載の知見抽出装置において、
前記知見抽出部は、
前記株価変動材料の表現に対する手掛かり語と、前記株価変動の方向を表す情報と、前記株価変動の強度を表す情報とを関連付けて記憶した材料表現手掛かり辞書記憶部と、
前記抽出された株価変動材料に基づいて、前記材料表現手掛かり辞書を検索し、当該検索された株価変動の方向及び強度を表す各情報を当該株価変動材料に付与する情報付与部と
を更に備える知見抽出装置。
【請求項６】
請求項４に記載の知見抽出装置において、

前記情報受信部は、前記ニュース情報を受信する際に、当該ニュース情報にタイムスタンプを付加し、
前記知見抽出部は、前記選択されたニュース情報から所定の対象を示す表現を含むイベント情報を抽出すると共に、当該イベント情報が抽出されたニュース情報から抽出し、当該抽出したイベント情報及びタイムスタンプを前記知見情報として前記記憶部に書込む処理を実行し、
前記所定の対象を示す表現は、銘柄の企業名を表す企業名表現、又は前記銘柄のグループ名を定義するグループ定義語の直前の名詞を表す表現である知見抽出装置。
【請求項７】
請求項４に記載の知見抽出装置において、
前記知見情報は上位概念の情報と下位概念の情報とによって構成される階層構造であり、
前記記憶部は評価情報と、前記下位概念の情報が前記上位概念の情報に帰属する度合いを示す確信度とを記憶し、
前記評価情報から抽出された、前記上位概念の情報と前記下位概念の情報とに関連する評価情報に基づいて評価値を算出する評価部と、
前記評価値に基づいて、前記確信度を更新する確信度更新部と
を備える知見抽出装置。
【請求項８】
請求項４に記載の知見抽出装置において、
前記記憶部はさらに、評価情報と、前記知見情報に含まれる前記銘柄が前記グループ名に帰属する度合いを示す確信度とを記憶し、
前記ニュース情報に含まれている前記グループ名を確認する確認部と、
前記確認部により、前記ニュース情報に含まれていると確認されたグループ名に基づいて前記知見情報から銘柄を抽出する抽出部と、
前記銘柄に関連する前記評価情報に基づいて評価値を算出する評価部と、
前記評価値に基づいて、前記確信度を更新する確信度更新部と
を備える知見抽出装置。
【請求項９】
請求項７に記載の知見抽出装置において、
前記確信度が規定値を下回った場合に、前記下位概念の情報と前記上位概念の情報とを含む知見情報を前記記憶部から削除する知見更新部を備える知見抽出装置。
【請求項１０】
記憶部を備え、配信された電子文書から知見情報を抽出する知見抽出装置に用いられるプログラムであって、
前記知見抽出装置を、
前記電子文書を受信する情報受信部、
前記知見情報を抽出する対象語と、前記知見情報を抽出する手掛かり語とに基づいて前記電子文書から概念を抽出し、抽出された当該概念を関連付けた知見情報を作成する知見抽出部、
前記抽出された知見情報を記憶する記憶部、
前記知見情報が記憶された後、前記電子文書を前記記憶部内の知見情報に基づいて分析する情報分析部、
前記電子文書に含まれる前記概念もしく前記対象語を検索する電子文書検索部、
前記検索部による検索の結果、前記電子文書に前記概念が規定数以上記載された場合に前記概念を抽出し、前記電子文書に前記対象語が規定数以上記載された場合に前記対象語に関連付けられた前記知見情報を前記記憶部から抽出する知見登録概念抽出部、
前記知見登録概念抽出部が抽出した前記対象語もしくは前記概念に関連する数値情報の大きさもしくは変動を評価する評価部、
前記数値情報の大きさもしくは変動の方向もしくは数値情報の変動の大きさに従って前

記概念と前記対象語の関連性の評価数値である確信度を更新する確信度更新部、
前記確信度が規定値を下回った場合にユーザに判断を求め、ユーザの指示により選択中の前記概念間の関連を削除する知見更新部、
として機能させるためのプログラム。
【請求項１１】
記憶部を備え、配信されたニュース情報から株取引に関する知見情報を抽出する知見抽出装置に用いられるプログラムであって、
前記知見抽出装置を、
前記ニュース情報を受信する情報受信部、
前記受信されたニュース情報から経済用語を含むニュース情報を選択する情報選択部、
前記選択されたニュース情報から前記知見情報を抽出し、当該知見情報を前記記憶部に書込む知見抽出部、
前記知見情報の書込み後、前記情報受信部に受信された全てのニュース情報を前記記憶部内の知見情報に基づいて分析する情報分析部、
として機能させるとともに、
前記知見抽出部を、
前記選択されたニュース情報が株価変動を評価する株価変動評価語を含むか否かを判定する株価変動評価語判定部、
前記判定の結果、前記ニュース情報が株価変動評価語を含むとき、当該ニュース情報が銘柄のグループ名を定義するグループ定義語を含むか否かを判定するグループ定義語判定部、
この判定の結果、前記ニュース情報がグループ定義語を含むとき、当該グループ定義語に基づいて当該ニュース情報からグループ名を抽出するグループ名抽出部、
前記株価変動評価語判定部による判定の結果、前記ニュース情報が株価変動評価語を含むとき、当該ニュース情報から企業名及び証券コードを含む銘柄を抽出する銘柄抽出部、
前記株価変動評価語判定部による判定の結果、前記ニュース情報が株価変動評価語を含むとき、当該ニュース情報から株価変動材料を抽出する材料抽出部と、
前記抽出されたグループ名、銘柄及び株価変動材料を組み合わせて前記知見情報を作成する組み合わせ作成部、
前記作成された知見情報を前記記憶部に書込む書込部と
として機能させ、
前記知見情報は、前記グループ名、前記銘柄及び前記株価変動材料を含む情報、又は前記グループ名及び前記銘柄からなる第１情報と前記株価変動材料からなる第２情報とを含む情報であるプログラム。
【請求項１２】
請求項１１に記載のプログラムにおいて、
前記知見情報は上位概念の情報と下位概念の情報とによって構成される階層構造であり、
前記記憶部は評価情報と、前記下位概念の情報が前記上位概念の情報に帰属する度合いを示す確信度とを記憶し、
前記知見抽出装置を
前記評価情報から抽出された、前記上位概念の情報と前記下位概念の情報とに関連する評価情報に基づいて評価値を算出する評価部、
前記評価値に基づいて、前記確信度を更新する確信度更新部、
として機能させるためのプログラム。
【請求項１３】
上位概念の情報と下位概念の情報とによって構成される階層構造である知見情報と、前記知見情報の評価情報と、前記知見情報に含まれる前記下位概念の情報が前記上位概念の情報に帰属する度合いを示す確信度とを記憶する記憶部と、
前記評価情報から抽出された、前記上位概念の情報と前記下位概念の情報とに関連する

評価情報に基づいて評価値を算出する評価部と、
前記評価値に基づいて、前記確信度を更新する確信度更新部と、
を備える知見更新装置。
【請求項１４】
請求項１３に記載の知見更新装置において、
前記記憶部は前記確信度の更新ルールを記憶し、
前記確信度更新部は、前記評価値と前記更新ルールとに基づいて前記確信度を更新するか否かを判定する知見更新装置。
【請求項１５】
請求項１３に記載の知見更新装置において、
前記確信度に基づいて前記知見情報を更新する知見更新部を備える知見更新装置。
【請求項１６】
上位概念の情報と下位概念の情報とによって構成される階層構造の知見情報と、前記知見情報の評価情報と、前記知見情報に含まれる前記下位概念の情報が前記上位概念の情報に帰属する度合いを示す確信度とを記憶する記憶部を備え、前記知見情報を更新する知見更新装置に用いられるプログラムであって、
前記知見更新装置を、
前記評価情報から抽出された、前記上位概念の情報と前記下位概念の情報とに関連する評価情報に基づいて評価値を算出する評価部、
前記評価値に基づいて、前記確信度を更新する確信度更新部、
として機能させるためのプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明の実施形態は、知見抽出装置、知見更新装置、及びプログラムに関する。
【背景技術】
【０００２】
インターネットなどで提供される大量のデータを有効活用するために、ある概念に関する同義語や、概念間の関連を定義したシソーラスなどを用いて大量のデータからユーザに必要としている情報に関連する情報を選択して、もしくは選択した情報を組み合わせて知見情報を提供する技術がある。
【０００３】
例えば、株取引においては、株変動のきっかけとなるような知見情報を迅速に入手して判断し、取引内容に反映させることが有効である。例えば、通常とは異なる時期に新型のインフルエンザが流行した場合、感染予防に使われるマスクを製造する企業や消毒薬を製造する企業の需要が増加することが容易に予想され、該当企業の株価が上昇する。株取引を行うユーザは、「インフルエンザの通常とは異なる時期の流行」の情報から、需要の増える銘柄を推測し、取引に反映することが有効である。
【０００４】
これを踏まえ、従来、証券情報と合わせてニュースを提供する技術や、ユーザの検索要求に応じて関連ニュースを抽出し、知見情報として提供する技術がある。
【０００５】
上記のような技術として、例えば、株価チャートと関連ニュースを同一画面に表示する技術（第１の技術）や、注目している銘柄の関連銘柄（同業他社、取引先、メインバンク、資本関係あり、など予め設定したもの）のうち、注目している銘柄と同時期に同様の株価変動をした銘柄を関連銘柄として記憶する技術（第２の技術）がある。
【０００６】
また、ユーザの要求をトリガーとして記憶装置に保持されたニュース中の「～関連株」「～特需」などの手がかり語から注目銘柄検索のためのキーワード抽出と銘柄関連付けを行い、ユーザに提示する技術や、一般的なシソーラス構築等の分野において、単語の係り

受けを、複数発生する構文解析結果のうちの出現数に応じた確信度で管理・更新する技術がある（第３の技術）。また、文書群から企業名と共起する関連語を取得してテーブルに保持するとともに、その共起対が株価変動に影響する度合いを予測係数テーブルで保持し、関連語の評価値と、予測係数を、それぞれ管理・更新し、株価予測を行う技術（第４の技術）もある。

【０００７】

しかしながら、株価チャートとニュースを同一画面でユーザに提示する第１の技術では、ニュースに予め銘柄などの知見情報を抽出する対象の語句（以下、知見情報抽出対象語という）が関連付けられていなければならないという問題がある。

【０００８】

また、注目している銘柄と同時期に同様の株価変動をした銘柄を関連銘柄として記憶する第２の技術では、同様の変動をしなかった銘柄が関連付けられないという問題がある。また、注目している銘柄に同業などの明示的な関連のある銘柄を予め設定するため、銘柄をグループ化する範囲が限定されてしまう。

【０００９】

第３の技術は、記憶装置に保持されたニュースからキーワード抽出と銘柄関連付けを実行するが、その活用方法や更新方法が開示されていない。

【００１０】

また、第３の技術は、単語の係り受けの確信度を評価し、以後に入手する新規コーパスで確信度を更新するが、情報源は単独（同種の情報）である。

【００１１】

第４の技術では、企業名の関連語を取得・管理し、その株価変動への影響も管理・更新しているが、株価変動で更新する対象は予測係数のみである。

【先行技術文献】

【特許文献】

【００１２】

【特許文献１】特開２００３－１０８７８５号公報
【特許文献２】特開２００３－１６２６３９号公報
【特許文献３】特開平９－１６０９１５号公報
【特許文献４】特開２０１１－１４１８３３号公報

【発明の概要】

【発明が解決しようとする課題】

【００１３】

本発明が解決しようとする課題は、知見情報抽出対象に関連付けられていない状態で随時配信される電子文書から、当該知見抽出対象語に関する知見情報を抽出する知見抽出装置、知見更新装置、及びプログラムを提供することである。

【課題を解決するための手段】

【００１４】

実施形態の知見抽出装置は、電子文書から知見情報を抽出する知見抽出装置であって、前記電子文書を受信する情報受信部と、前記知見情報を抽出する対象語と、前記知見情報を抽出する手掛かり語とに基づいて前記電子文書から概念を抽出し、抽出された当該概念と前記対象語を関連付けた知見情報を作成する知見抽出部と、前記抽出された知見情報を記憶する記憶部と、前記知見情報が記憶された後、前記電子文書を前記記憶部内の知見情報に基づいて分析する情報分析部と、前記電子文書に含まれる前記概念もしく前記対象語を検索する電子文書検索部と、前記検索部による検索の結果、前記電子文書に前記概念が規定数以上記載された場合に前記概念を抽出し、前記電子文書に前記対象語が規定数以上記載された場合に前記対象語に関連付けられた前記知見情報を前記記憶部から抽出する知見登録概念抽出部と、前記知見登録概念抽出部が抽出した前記対象語もしくは前記概念に関連する数値情報の大きさもしくは変動を評価する評価部と、前記数値情報の大きさもしくは変動の方向もしくは数値情報の変動の大きさに従って前記概念と前記対象語の関連性

の評価数値である確信度を更新する確信度更新部と、前記確信度が規定値を下回った場合にユーザに判断を求め、ユーザの指示により選択中の前記概念間の関連を削除する知見更新部とを備える。

【図面の簡単な説明】

【００１５】

【図１】第１の実施形態における知見抽出装置の構成例を示す図。

【図２】第１の実施形態における第１手掛かり語辞書の例を示す図。

【図３】第１の実施形態における知見抽出対象語辞書の例を示す図。

【図４】第１の実施形態における第２手掛かり語辞書の例を示す図。

【図５】第１の実施形態における電子文書の例を示す図。

【図６】第１の実施形態における電子文書の例を示す図。

【図７】第１の実施形態にしたがうユーザ端末装置の構成例を示す図。

【図８】第１の実施形態における知見抽出部の処理動作を詳細に示す図。

【図９】第１の実施形態における知見抽出部の処理動作の一例を説明するためのフローチャート。

【図１０】第１の実施形態における処理動作の順序の概略を示す図。

【図１１】第１の実施形態における分析対象の電子文書からの第１概念名であるグループ名、銘柄、株取引材料の抽出結果の一例を示す図。

【図１２】第１の実施形態における知見情報の一例を示す図。

【図１３】第１の実施形態における情報分析部の処理動作を詳細に示す図。

【図１４】第１の実施形態における情報分析部の処理動作の一例を説明するためのフローチャート。

【図１５】第１の実施形態における情報分析部の処理動作の一例を説明するためのフローチャート。

【図１６】第２の実施形態における第２手掛かり語辞書の例を示す図。

【図１７】第２の実施形態における電子文書の例を示す図。

【図１８】第２の実施形態における知見情報の例を示す図。

【図１９】第３の実施形態における知見抽出部の処理動作の一例を説明するためのフローチャート。

【図２０】第３の実施形態における知見抽出部の処理結果の一例を示す図。

【図２１】第４の実施形態における知見抽出装置の構成例を示す図。

【図２２】第４の実施形態における知見情報の一例を示す図。

【図２３】第４の実施形態における知見評価更新装置の処理動作の一例を示す図。

【図２４】第４の実施形態における知見評価更新装置の処理動作の一例を説明するためのフローチャート。

【図２５】第４の実施形態における知見評価更新装置の処理動作の一例を説明するためのフローチャート。

【図２６】第４の実施形態における知見評価更新装置の評価情報の一例を示す図。

【図２７】第４の実施形態における知見評価更新装置の確信度更新ルールの一例を示す図。

【図２８】第４の実施形態における知見評価更新装置の確信度更新処理の一例を示す図。

【図２９】第４の実施形態における知見評価更新装置の確信度更新処理の一例を示す図。

【図３０】第４の実施形態における知見評価更新装置の確信度更新の効果の一例を示す図。

【図３１】第５の実施形態における分析対象のニュース情報の例を示す図。

【図３２】第５の実施形態における関連語情報の例を示す図。

【図３３】第５の実施形態における知見抽出装置の構成例を示す図。

【図３４】第６の実施形態における知見抽出装置の構成例を示すブロック図。

【図３５】第６の実施形態における知見抽出装置の第１手掛かり語辞書の例を示す模式図。

【図３６】第６の実施形態における知見抽出装置の知見抽出対象語辞書の例を示す模式図。
【図３７】第６の実施形態における知見抽出装置の第２手掛かり語辞書の例を示す模式図。
【図３８】第６の実施形態における知見抽出装置の電子文書の例を示す模式図。
【図３９】第６の実施形態における知見抽出部の処理動作を詳細に示す模式図。
【図４０】第６の実施形態における情報分析部の処理動作を詳細に示す模式図。
【図４１】第６の実施形態における知見抽出部の処理結果の一例を示す模式図。
【図４２】第６の実施形態における知見評価更新装置の処理結果の一例を示す模式図。
【発明を実施するための形態】
【００１６】
以下、各実施形態の知見抽出装置について図面を用いて説明する。なお、以下の各装置は、装置毎に、ハードウェア構成、又はハードウェア資源とソフトウェアとの組合せ構成のいずれでも実施可能となっている。組合せ構成のソフトウェアとしては、予めネットワーク又は記憶媒体から対応する装置のコンピュータにインストールされ、対応する装置の機能を実現させるためのプログラムが用いられる。
【００１７】
（第１の実施形態）
図１は、第１の実施形態に係る知見抽出装置の構成例を示すブロック図である。本実施形態の知見抽出装置１は、株取引を行うユーザや株取引判断を行うアルゴリズムトレードエンジンなどのシステム（複合イベント処理エンジン、ＣＥＰなどを組み込んだものを含む）への証券市場立会時間中の情報提供に使用可能なものであり、例えばインターネット等で配信されたニュース情報などの電子文書から株取引に関する知見情報を抽出する。抽出結果の、知見情報は情報提供などに使用可能となっている。
【００１８】
具体的には、この知見抽出装置１は、ニュース情報のうちの経済ニュースを受信した際に、例えば、キーワード「インフルエンザ」に関連する銘柄グルーピングの知識や、「インフルエンザ」関連銘柄で株価変動の材料となる「集団感染」「鳥インフルエンザ」の知識を獲得し、これらの知識を知見情報として保持するものである。また、知見抽出装置１は、経済ニュースを含むインターネット配信ニュース、企業情報を受信した際は、保持した知見情報を参照し、ニュース情報に「インフルエンザ」などのグループ、該当グループに所属する銘柄、該当グループの株価変動材料の有無の情報を付与し、アルゴリズムトレードエンジンなどのシステムもしくは株取引を行うユーザのユーザ端末装置に提供可能となっている。すなわち、知見抽出装置１は、銘柄などの知見抽出対象語のグルーピングをするための情報である概念（第１概念）や、第１概念に関連づける情報（第２概念）を含む知見情報を抽出する。
【００１９】
例えば、株取引に関する知見情報を抽出する場合、株の銘柄が所属する業界（例：金融、金属、広告など）が第１概念となるような手掛かり語（第１手掛かり語）を、知見抽出装置１に記憶させる。第１手掛かり語については後述する。第２概念は、第１概念と異なる概念であり、後述する第２手掛かり語に基づいて抽出される。
【００２０】
また、知見抽出装置１は、知見抽出システム１と呼んでもよい。
【００２１】
具体的には知見抽出装置１は、装置全体の制御を司る制御部１０、記憶部１１、情報受信部１２、情報選択部１３、知見抽出部１４、情報分析部１５および入出力インタフェース１６を備え、各部１１～１６がバス１７を介して接続される。
【００２２】
記憶部１１は、例えばハードディスクドライブや不揮発性メモリ装置などのハードウェアで構成される。記憶部１１は制御部１０による実行対象のプログラムを記憶するのに加

え、知見抽出部１４で抽出した知見情報、受信し分析を行った電子文書などを保持する。また、記憶部１１は、後述する第１手掛かり語辞書１１１、知見抽出対象語辞書１１２、第２手掛かり語辞書１１３及び処理途中のデータ等も保持する。

【００２３】

第１手掛かり語辞書１１１は、図２に一例を示すように、第１手掛かり語を識別するコードと、知見情報抽出対象をグルーピングする第１概念の手掛かりとなる第１手掛かり語とが関連付けられて記述されている。すなわち、本実施形態の電子文書はあらかじめ記憶された第１手掛かり語に基づいて第１概念名（グループ名）をつけられる。本実施形態の第１手掛かり語辞書はグループ定義語辞書１１１といえる。

【００２４】

知見抽出対象語辞書１１２は、図３に一例を示すように、本実施形態において電子文書から知見情報を抽出する対象語（知見抽出対象語）である銘柄を識別する証券コードと、銘柄の企業名と、銘柄の業種の大分類を示す東証業種１と、東証業種１内の小分類を示す東証業種２と、銘柄の企業名を表す企業名表現とが関連付けられて記述されている。企業名表現としては、企業名の正称、略称、通称、俗称、愛称などの様々な名称が使用可能となっている。なお、企業名表現は、名称に限らず、各企業の証券コードを用いてもよい。すなわち、本実施形態では知見抽出対象語辞書を銘柄辞書１１２といえる。

【００２５】

第２手掛かり語辞書１１３は、図４に一例を示すように、第２手掛かり語（材料表現手掛かり語）を識別するコードと、後述する知見抽出判定語の表現に対する手掛かり語とが関連付けられて記述されている。すなわち、本実施形態では第２手掛かり語辞書１１３を材料表現手掛かり語辞書１１３といえる。

【００２６】

情報受信部１２は、インターネットで配信されるニュース情報を受信する。図５及び図６はインターネットで配信されるニュース情報の例である。ニュース情報の受信は、配信業者と契約を結び配信を受ける形態でも、一般に検索ロボットもしくはクローラと呼ばれるインターネット検索プログラムにより新規配信されたニュース情報を発見し入手する形態でもよい。本実施形態では、インターネットで配信されるニュース情報の構成要素のうち、図５及び図６に示す「タイムスタンプ」「ニュースソース」「ジャンル」「ニュース見出し」「ニュース本文」の５項目と、情報受信部１２が付与する「ニュースＩＤ」を利用する。「タイムスタンプ」は、配信されたニュース情報に付与された配信日時情報である。「ニュースソース」は、その情報を提供した事業者の名称である。「ジャンル」は、ニュース情報に付与された「経済」「社会」「スポーツ」などのジャンル（分野）を示すタグ情報である。「ジャンル」は、ニュース受信時に付与されていなくてもよい。「ニュース見出し」は配信されたニュース情報の見出し部分の日本語列、「ニュース本文」は配信されたニュース情報の見出しを除く本文部分の日本語列である。

【００２７】

情報選択部１３は、情報受信部１２が受信したニュース情報から、ここでは経済ニュース（経済用語を含むニュース情報）を選択する。経済ニュースであることの判断は、例えば、受信したニュース情報に「ジャンル」が付与されている場合は、「ジャンル」が「経済」「市況」などの経済とその類義語のうちのいずれかの用語を含む表現であることにより判断する。受信したニュース情報に「ジャンル」が付与されていない場合は、例えば、ニュース見出しもしくはニュース本文に「東証」「大証」「日経平均」「株価」「株式市場」「高値」「終値」「出来高」などの経済用語、特に株取引に関する用語のいずれかを含むことにより判断する。すなわち、情報選択部１３は、知見抽出対象語に関連の高い電子文書を選択する。

【００２８】

知見抽出部１４は、情報入力部１３によって選択されたニュース情報から知見情報を抽出するものであり、例えば、後述する知見抽出判定語チェック部１４１、第１手掛かり語チェック部１４２、第１概念抽出部１４３、知見抽出対象語抽出部１４４、第２概念抽出

部１４５、組み合わせ作成部１４６及び第１概念チェック部１４７により、記憶部１１内の各辞書１１１～１１３を用い、知見情報抽出対象の経済ニュースから、知見抽出対象である銘柄をグループ分けする第１概念名、グループに所属する銘柄、株価変動の要因となる材料表現を抽出し、銘柄グルーピングや材料（イベント）の知見情報として記憶部１１に書込む。ここで、知見情報としては、例えば、第１概念名、銘柄及び株価変動材料を含む情報、又は第１概念名及び銘柄からなる第１情報と株価変動材料からなる第２情報とを含む情報が使用可能となっているが、これらの組み合わせには限定されない。なお、知見抽出部１４は情報選択部１３によって選択された情報ではなく記憶部１１に記憶された電子文書から知見情報を抽出することも可能である。

【００２９】

なお、知見抽出部１４は、これら各部１４１～１４７及び各辞書１１１～１１３を用いる方法に限らず、例えば、構文解析、意味解析、文脈解析を用いて銘柄をグループ分けする方法により、知見情報を抽出してもよい。後者の方法では、例えば「インフルエンザの流行を受けて、ＸＸ社がランクイン」のように、図２に示したグループ定義語を含まないニュース情報からも知見情報を抽出可能となっている。

【００３０】

情報分析部１５は、知見情報が記憶部１１に記憶された後、情報受信部１２に（新規に）受信された全てのニュース情報を記憶部１１内の知見情報に基づいて分析するものであり、例えば、後述する知見抽出対象語・第１概念・第２概念抽出部１５１、電子文書割付部１５２及び知見抽出対象別判定部１５３を備えている。また、情報分析部１５は、新規に受信したニュース情報にグルーピングや材料（イベント）の知見情報に該当する情報が存在する場合は、当該ニュース情報に当該知見情報を付与してもよい。情報分析部１５はさらに、新規に受信したニュース情報を含めて一定時間に処理したニュース情報をグループや銘柄にわけ、それぞれのグループや銘柄で統計処理を行い、統計情報を付与してもよい。

【００３１】

一連の処理終了後、情報分析部１５の出力が、ニュース情報の分析結果として入出力インタフェース１６によりユーザ端末装置に提示される。

【００３２】

入出力インタフェース１６は図示しない外部記憶装置とケーブルを介して接続可能であり、この外部記憶装置との間で記憶部１１に記憶される知見情報や分析結果保存データベース（図示せず）にデータを入出力する。

【００３３】

図７は、以上のような知見抽出装置１に接続されるユーザ端末装置の構成例を示すブロック図である。ユーザ端末装置２は、知見抽出装置１を利用するユーザが取り扱う端末装置である。ユーザ端末装置２は、装置全体の処理を司る端末制御部２１、端末記憶部２２、表示部２３、入力部２４、通信インタフェース２５を備え、それぞれがバス２６を介して相互に接続される。

【００３４】

端末記憶部２２は、例えばハードディスクドライブや不揮発性メモリ装置などのハードウェアで構成された記憶装置である。端末記憶部２２は、制御用のプログラムを記憶する。また、端末記憶部２２は、端末制御部２１による各種処理のワークメモリとしても機能する。

【００３５】

表示部２３は、例えば液晶ディスプレイであり、情報分析部１５の出力をユーザに向けて出力する。例えば、新規に受信したニュース情報の配信時刻（タイムスタンプ）、ニュースソース、ニュース情報で話題とされた企業名・第１概念名、ニュース情報に含まれる株取引材料（自己株取得、経営悪化、インフルエンザの集団感染、など）を表形式でユーザに提示する。

【００３６】

入力部２４は例えばキーボードやマウスであり、知見抽出、情報分析およびシステム設定にかかる操作を受け付ける。

【００３７】

通信インタフェース２５は、ケーブルを介して知見抽出装置１と接続され、ユーザ端末装置２においてユーザが指定した条件や知見抽出装置１の処理結果などのやり取りを行う。また、通信インタフェース２５は外部記憶装置とケーブルを介して接続可能であり、この外部記憶装置との間で記憶部１１に記憶する銘柄グルーピングや材料（イベント）の知見情報、分析結果を入出力することもできる。

【００３８】

次に、以上のように構成された知見抽出装置１の動作について説明する。図８は、知見抽出部１４の処理動作の詳細を示す模式図、図９は知見抽出部の処理動作の一例を説明するためのフローチャートである。

【００３９】

なお、以下の説明では、知見抽出部１４の処理を、ニュース情報を受信する都度としているが、この処理は夜間などの情報分析処理を行わない時間帯に、受信済みのニュース情報に対してまとめて実施してもよい。

【００４０】

情報受信部１２が、ニュース配信サイトからの配信を受けるか、ニュース配信サイトにアクセスすることにより、新規のニュース情報を受信する。例えば、図５及び図６のうち、図５に示すニュースＩＤ“１”に該当するニュース情報を受信する。

【００４１】

情報選択部１３は、情報受信部１２が受信した電子文書であるニュース情報から、知見情報を抽出する電子文書として経済ニュースを選択する（ステップＳ１）。経済ニュースであることの判断は、例えば、受信したニュース情報に「ジャンル」が付与されている場合は、「ジャンル」が「経済」「市況」などの経済とその類義語のうちのいずれかを含む表現であることにより判断する。

【００４２】

受信したニュース情報に「ジャンル」が付与されていない場合は、例えば、ニュース見出しもしくはニュース本文に「東証」「大証」「日経平均」「株価」「株式市場」「高値」「終値」「出来高」などの経済用語、特に株取引に関する用語のいずれかを含むことにより判断する。選択されたニュース情報は、情報選択部１３から知見抽出部１４に送出される。

【００４３】

知見抽出部１４においては、知見抽出判定語チェック部１４１が、情報選択部１３に選択されたニュース情報が知見情報を抽出する可能性のある電子文書かどうかを判定する知見抽出判定語を含むか否かを判定する（ステップＳ２）。

【００４４】

本実施形態では知見抽出判定語は株価変動を評価する株価変動評価語であるとする。すなわち、本実施形態では知見抽出判定語チェック部１４１を株価変動評価語チェック部１４１といえる。

【００４５】

例えば、知見抽出判定語チェック部１４１は、受信したニュース情報のニュース見出しもしくはニュース本文に、「高値更新」「強い動き」「反発」「急落」「続伸」「ランクイン」などの株価変動評価語が含まれるか否かを判定する。

【００４６】

ニュース情報に株価変動評価語が含まれる場合、知見抽出部１４は、当該ニュース情報が知見情報を含むニュース情報であると判定して、ニュース情報を第１手掛かり語チェック部１４２に送信する。

【００４７】

なお、「株価変動評価語」は、株価の変動を評価する内容であれば、「株価変動表現語

」、「株価変動語」又は「株価変動定義語」といった他の用語に言い換えてもよい。

【００４８】

第１手掛かり語チェック部１４２は、ステップＳ２の判定の結果、ニュース情報が株価変動評価語を含むとき、当該ニュース情報が銘柄の第１概念名を定義する第１手掛かり語（グループ定義語）を含むか否かを判定する（ステップＳ３）。すなわち、本実施形態の第１手掛かり語チェック部１４２はグループ定義語チェック部１４２といえる。

【００４９】

例えば、第１手掛かり語チェック部１４２は、第１手掛かり語辞書１１１を参照し、受信したニュース情報のニュース見出しもしくはニュース本文にグループ定義語の少なくとも１つが含まれるか否かを判定する。例えば図５に示すニュースＩＤ“１”のニュースは、グループ定義語「関連銘柄」を含む。

【００５０】

受信したニュース情報にグループ定義語が含まれない場合は第１概念抽出部１４３によるステップＳ４の処理を省略する。

【００５１】

受信したニュース情報にグループ定義語が含まれる場合は、第１手掛かり語チェック部１４２は、ニュース情報を第１概念抽出部１４３に送信する。

【００５２】

第１概念抽出部１４３は、ステップＳ３の判定の結果、ニュース情報がグループ定義語を含むとき、当該グループ定義語に基づいて当該ニュース情報から第１概念名を抽出する（ステップＳ４）。

【００５３】

例えば、第１概念抽出部１４３は、グループ定義語を含むニュース情報を受信すると、グループ定義語直前の名詞を第１概念名として抽出する。受信したニュース情報が例えば図５に示すニュースＩＤ“１”の場合は、グループ定義語「関連銘柄」の直前の名詞「インフルエンザ」を第１概念名として抽出する。

【００５４】

第１概念名は名詞としたが、「インフルエンザ」のような単語ではなくて、「新型インフルエンザ」「季節性インフルエンザ」のような複合名詞も抽出対象としてよい。また、グループ定義語を含むが直前が名詞でない場合は、第１概念名を含まないと判断する。

【００５５】

知見抽出対象語抽出部１４４は、知見抽出判定語チェック部１４１によるステップＳ２の判定の結果、ニュース情報が株価変動評価語を含むとき、当該ニュース情報から企業名及び証券コードを含む銘柄を抽出する。

【００５６】

具体的には、知見抽出対象語抽出部１４４は、例えば、ステップＳ３又はＳ４の後、第１概念抽出部１４３の処理を省略して送信されたニュース情報に対して、知見抽出対象語辞書１１２を参照し、ニュース見出しもしくはニュース本文に知見抽出対象名表現が含まれるか否かを判定し（ステップＳ５）、知見抽出対象名表現が含まれる場合に当該知見抽出対象名表現に対応する知見抽出対象語である銘柄を抽出する（ステップＳ６）。すなわち、本実施形態の知見抽出対象語抽出部１４４は銘柄抽出部１４４といえる。

【００５７】

ニュース見出しまたはニュース本文に、知見抽出対象語辞書１１２の「知見抽出対象名表現」列の表現がある場合は、その行の「証券コード」「企業名」に記された知見抽出対象語である銘柄が含まれると判断する。

【００５８】

知見抽出対象語辞書１１２は、１つの証券コードに対して複数行の定義を含んでよい。また、１件のニュース情報から複数の知見抽出対象名表現を抽出してよい。

【００５９】

例えば図５のニュースＩＤ“１”のニュース情報の場合、知見抽出対象名表現「ＲＲボ

ウ」「ｙｙボウＨＤ」「ホヘト薬品」が含まれるため、銘柄「証券コード　１０３１　企業名　ＲＲボウ（株）」「証券コード　１０３３　企業名　ｙｙボウホールディングス（株）」「証券コード　１０４１　企業名　ホヘト薬品（株）」が抽出される。

【００６０】

知見抽出対象語辞書１１２では知見抽出対象名表現を企業名称もしくは企業名の略称としているが、各企業の証券コードも知見抽出対象名表現の一例としてもよい。本実施形態の知見抽出対象語辞書１１２における知見抽出対象名表現は企業名表現といえる。

【００６１】

なお、知見抽出対象語抽出部１４４は、図１０に示すように、ステップＳ２の結果、株価変動評価語を含む場合に銘柄を抽出すればよいので、ステップＳ３，Ｓ４の後に限らず、ステップＳ２の後からステップＳ９の前の間の任意の時点で銘柄を抽出可能である。

【００６２】

第２概念抽出部１４５は、知見抽出判定語チェック部１４１による判定の結果、ニュース情報が知見抽出判定語である株価変動評価語を含むとき、当該ニュース情報から第２概念として株価変動材料を抽出する。具体的には、第２概念抽出部１４５は、株価変動評価語を含むニュース情報のニュース見出しもしくはニュース本文に、第２手掛かり語辞書１１３内の第２手掛かり語の１つ以上が含まれるか否かを判定する（ステップＳ７）。

【００６３】

ステップＳ７およびステップＳ８の処理は、図９のフローチャートではステップＳ６の後としているが、ステップＳ３の処理の前、もしくはステップＳ３からステップＳ６の処理と並列で実施してもよい。

【００６４】

第２概念抽出部１４５は、受信したニュース情報のニュース見出しもしくはニュース本文に、第２手掛かり語辞書１１３内の第２手掛かり語が含まれる場合、その第２手掛かり語と係り受けしている名詞句を第２概念名として抽出する（ステップＳ８）。本実施形態では第２概念名を株価変動材料表現といえる。すなわち、本実施形態では第２概念抽出部１４５は材料抽出部１４５といえる。

【００６５】

例えば、図５に示すニュースＩＤ“１”のニュース本文では、第２文「○県が×市内の学生寮において、新型インフルエンザの集団感染が発生したと発表したことが手掛かり材料となっているようだ。」に第２手掛かり語「手掛かり材料」が含まれる。

【００６６】

第２手掛かり語「手掛かり材料」と係り受けしている名詞句は「新型インフルエンザの集団感染が発生したと発表したこと」である。この名詞句「新型インフルエンザの集団感染が発生したと発表したこと」が株価変動材料（イベント）表現として抽出される。

【００６７】

株価変動材料（イベント）表現は、抽出した名詞句を分析し、「新型インフルエンザの集団感染が発生した」もしくは「新型インフルエンザの集団感染」を抽出対象としてもよい。

【００６８】

また、ステップＳ６、ステップＳ７で抽出する第１概念名「インフルエンザ」を含む「新型インフルエンザ」とその付属語を消去し「集団感染」を抽出対象としてもよい。

【００６９】

さらに、該当のニュース本文が「○県が×市内の学生寮において、新型インフルエンザの集団感染が発生したと発表した。これが手掛かり材料となっているようだ。」のように２つの文にわかれている場合、手掛かり語「手掛かり材料」と係り受けしている代名詞「これ」を抽出した後で、文脈解析により「これ」を前文の「○県が×市内の学生寮において、新型インフルエンザの集団感染が発生したと発表した」に置き換え、置き換えた後の「○県が×市内の学生寮において、新型インフルエンザの集団感染が発生したと発表した」もしくはその分析結果から「新型インフルエンザの集団感染が発生した」、「新型イン

フルエンザの集団感染」、もしくは「集団感染」を抽出対象としてもよい。

【００７０】

ステップＳ３からステップＳ８の処理の終了後、組み合わせ作成部１４６は、処理中のニュース情報から抽出された第１概念名、知見抽出対象語である銘柄及び株価変動材料（第２概念名もしくはイベント）を組み合わせて知見情報を作成し、当該知見情報を記憶部に書き込む（ステップＳ９）。

【００７１】

ニュース情報から抽出された第１概念名が１つである場合は、その第１概念名に、抽出された全ての銘柄（知見抽出対象語）、及び抽出された全ての株価変動材料（第２概念名）を割り付ける。すなわち、１つの第１概念名に対して、知見抽出対象語である銘柄と第２概念である株価変動材料はそれぞれ１つまたは複数が割り付けられることを許容する。

【００７２】

分析対象のニュース情報が図６のニュースＩＤ“３”の場合は、図１１に示すようにニュース本文から複数の６件の第１概念名、８件の銘柄、３件の株価変動材料が抽出される。

【００７３】

このように抽出された第１概念名が複数である場合、文脈解析などの高度な自然言語処理を用いて、第１概念名、第１概念名に係り受けしている銘柄、第１概念名もしくはその第１概念名に係り受けしている銘柄に係り受けしている株価変動材料を組み合せてもよい。

【００７４】

また、簡易的には、ニュース本文を文に区切り、文ごとに第１概念名、銘柄、株価変動材料を組み合せるとしてもよい。さらに、ニュース本文を１文ずつ取り出し、第１概念名と銘柄が揃った時にそこまでに抽出された第１概念名、銘柄、株価変動材料を組み合せて出力し、次の文から新たに第１概念名、銘柄、株価変動材料を組み合せるとしてもよい。

知見情報は第１概念名に対して銘柄と株価変動材料が割り付けられるとしているが、銘柄もしくは株価変動材料がない組み合せが作成されてもよい。

【００７５】

このような処理により、本実施形態では、第１概念名、銘柄（企業名表現）、株価変動材料のセットであるされた知見情報として以下の６セットが抽出される。

【００７６】

（１）第１概念名：石油　銘柄：○×△石〈１００１〉、石油ＶＶ〈１００２〉、
　　材料：１０年４～１２月期で連結経常益２０００億円強が報じられた。

【００７７】

（２）第１概念名：商社　銘柄：ＨＨ商〈１０８０〉。

【００７８】

（３）第１概念名：非鉄金属　銘柄：ＳＳ鉱〈１０５０〉。

【００７９】

（４）第１概念名：繊維　銘柄：センイＪＪ〈１０３０〉。

【００８０】

（５）第１概念名：銀行　銘柄：ＨＨ銀〈１０８２〉。

【００８１】

（６）第１概念名：ハイテク　銘柄：ハイテクＢＢ〈１０６０〉、ＱＱメモリ〈１０６３〉、
　　材料：Ｄ証（Ｄ証券）が投資判断を引き上げ。

【００８２】

組み合せは、第１概念名、銘柄、株価変動材料をセットにするとしたが、組み合わせは第１概念名と銘柄のみとし、株価変動材料は単独で知見情報として保持するとしてもよい。または、株価変動材料は第１概念名ではなく銘柄とセットにして知見情報として保持するとしてもよい。

【００８３】
組み合わせ作成部１４６の処理後、図１２に一例を示す如き、抽出された知見情報は、記憶部１１に書き込まれる（ステップＳ９）。なお、図１２に示すように、本実施形態における第１概念の例として「インフルエンザ」、「石油」、「商社」、「非鉄金属」、「繊維」、「金属」、「ハイテク」などがある。
【００８４】
ステップＳ２で株価変動評価語がないと判定された経済ニュースは、第１概念チェック部１４７により、ニュース見出しもしくはニュース本文に、記憶部１１に書き込まれた知見情報の第１概念名（既存第１概念名）を含むか否かを判定される（ステップＳ１１）。本実施形態では第１概念名はグループ名であるため、第１概念チェック部１４７はグループ名チェック部といえる。
【００８５】
既存第１概念名を含む場合は、知見抽出対象語抽出部１４４が知見抽出対象語辞書１１２を用いて、ニュース見出しとニュース本文に知見抽出対象名表現が含まれるか否かを判定する（ステップＳ１２）。
【００８６】
知見抽出対象名表現が含まれる場合、例えば「参入」「新事業」などの表現がある場合は、組み合せ作成部１４６が抽出された知見抽出対象名表現に対応する銘柄を既存グループに追加する処理のための組み合わせを作成する。
【００８７】
「撤退」などの表現がある場合は、組み合わせ作成部１４６が、既存第１概念に対して抽出された知見抽出対象名表現に対応する銘柄を取り除く処理のための組み合わせを作成する（ステップＳ１３）。
【００８８】
ステップＳ１３で既存第１概念と銘柄の追加もしくは削除処理の組み合わせが作成された場合は、その処理を記憶部１１に対して実行する（ステップＳ９）。
【００８９】
次に、情報分析部１５の動作を図１３、図１４及び図１５によって説明する。
【００９０】
始めに、情報受信部１２は、ニュース配信サイトからの配信を受けるか、ニュース配信サイトにアクセスすることにより、新規のニュース情報を受信する。例えば、図５のニュースＩＤ２に該当するニュース情報を受信する（ステップＳ２１）。受信したニュース情報は知見抽出対象語・第１概念・第２概念抽出部１５１に送信される。
【００９１】
知見抽出対象語・第１概念・第２概念抽出部１５１は、まず、記憶部１１に記憶された知見抽出対象語辞書１１２を参照し、受信したニュース情報のニュース見出しもしくはニュース本文に知見抽出対象名表現があるか否かを判定する（ステップＳ２２）。
【００９２】
知見抽出対象名表現がある場合、知見抽出対象語・第１概念・第２概念抽出部１５１は、抽出された知見抽出対象名表現に対応する企業名・証券コードを抽出する。また、知見抽出対象語・第１概念・第２概念抽出部１５１は、図１２に示す知見情報を参照し、抽出した企業名・証券コードが割り付けられた第１概念名を全て抽出する（ステップＳ２３）。知見抽出対象語・第１概念・第２概念抽出部１５１及び電子文書割付部（ニュース割付部）１５２は、ニュース見出しとニュース本文から抽出された企業名ごとに、その企業が属する各グループに対して、ステップＳ２４・ステップＳ２５の処理を行う。
【００９３】
始めに、知見抽出対象語・第１概念・第２概念抽出部１５１は、図１２に示す知見情報を参照し、その時点で処理中の企業名・その企業が属する第１概念名に割り付けられた第２概念（株価変動材料）を抽出し、ニュース見出し・ニュース本文中に該当の株価変動材料（表現）があるか否かを判定する（ステップＳ２４）。

【００９４】
該当の株価変動材料（表現）がある場合は、処理中の受信ニュース情報にそのグループに関する情報があると判断し、電子文書割付部１５２が、処理中のニュース情報を、図１２に示す知見情報でその第１概念名に対応付けられた企業に割り付ける（ステップＳ２５）。
【００９５】
知見抽出対象語・第１概念・第２概念抽出部１５１は、ステップＳ２２からステップＳ２５の処理の後、もしくは、前、または並列で、受信したニュース情報のニュース見出しもしくはニュース本文に、図１２に示す知見情報内の第１概念名が含まれるか否かを判定する（ステップＳ２６）。
【００９６】
例えば、図６に示すニュースＩＤ“２”のニュース情報では、本文に第１概念名「インフルエンザ」が含まれる。
【００９７】
第１概念名が含まれる場合、知見抽出対象語・第１概念・第２概念抽出部１５１は、当該第１概念名を本文又は知見情報から抽出して電子文書割付部１５２に送出する。
【００９８】
第１概念名が抽出された場合、電子文書割付部１５２は、図１２に示す知見情報でその第１概念名に対応付けられた知見抽出対象語である銘柄を取り出し、処理中の電子文書を該当知見抽出対象語全てに割り付ける（ステップＳ２７）。
【００９９】
ステップＳ２７までの処理でニュース情報を割り付けられた知見抽出対象語ごとに、知見抽出対象別判定部１５３は、以下に記載するステップＳ２８からステップＳ３４の処理を行う。
【０１００】
ステップＳ２８では、ステップＳ２４の第２概念、すなわち株価変動材料（表現）抽出結果を参照し、現在処理中の知見抽出対象語である銘柄が属する第１概念に特有の第２概念が抽出されているか否かを判定する。
【０１０１】
処理中の知見抽出対象語において、当該知見抽出対象語の属する第１概念に関連づいて第２概念が抽出されている場合、当該知見抽出対象語に関する情報が変動する可能性があると判断し、当該知見抽出対象語と第２概念名のセットをアラームに追加する（ステップＳ２９）。
【０１０２】
具体的に記載すると、処理中の銘柄において、当該銘柄の所属グループに対応付けられた株価変動材料が抽出されている場合は、その銘柄の株価が変動する可能性があると判断し、その銘柄と株価変動材料のセットを株価変動材料アラームに追加する。
【０１０３】
記憶部１１に保持された知見情報が「第１概念名・知見抽出対象語」と第２概念名に分けられている場合は、ステップＳ２８は、現在処理中の知見抽出対象語が属するか否かに関わらず記憶部１１に保持された第２概念全てに関して抽出されているか否かを判定する。
【０１０４】
この場合、ステップＳ２９では、知見抽出対象語である銘柄と、ステップＳ２８で抽出された第２概念である株価変動材料のセットを第２概念アラームに追加する。
【０１０５】
知見抽出対象語である銘柄にニュース情報が割り付けられた後で、処理中の銘柄に割り付けられ、記憶部１１に記憶された過去一定期間、例えば３日間のニュース見出し・ニュース本文を参照して急上昇ワードがあるかを判定し（ステップＳ３０）、該当銘柄のニュース情報に急上昇ワードが存在する場合は、銘柄と急上昇ワードのセットを急上昇アラー

ムに追加する（ステップＳ３１）。
【０１０６】
急上昇ワードは、近年、検索サイトなどで公開されているもので、出現単語の新しさ、急上昇を判定する技術が公知となっている。
【０１０７】
さらに、同様に処理中の銘柄に割り付けられ、記憶部１１に記憶された過去一定期間、例えば１時間のニュース情報のニュース見出し・ニュース本文を参照して処理中の受信ニュース情報が過去一定期間のニュース情報と類似・一致しているかを判定し（ステップＳ３２）、類似ニュース情報もしくは一致するニュース情報がある場合は、そのニュース情報対の類似度を求め、多重配信アラームを追加する（ステップＳ３３）。
【０１０８】
以上の処理により出力された第２概念アラーム、急上昇アラーム、多重配信アラームは、情報分析部１５から出力され、アルゴリズムトレードエンジンなどのシステムもしくは株取引を行うユーザ端末装置２に提供される（ステップＳ３４）。
【０１０９】
提供先がアルゴリズムトレードエンジンなどのシステムである場合、アルゴリズムトレードエンジンなどのシステムは、知見抽出装置１から提供されたニュース情報とニュース情報に付与された情報の他に、例えば該当銘柄の現在の株価、注文状況などを入力情報として、現時点で取るべき取引戦略を、株取引を行うユーザ端末装置２を介してユーザに提示する。
【０１１０】
上述したように本実施形態によれば、受信されたニュース情報から経済用語を含むニュース情報を選択し、当該選択されたニュース情報から知見情報を抽出し、当該抽出された知見情報を記憶する構成により、銘柄に関連付けられていない状態で随時配信されるニュース情報から銘柄グルーピングや株価変動材料の知見情報を随時抽出することにより、知見情報を最新の状態にすることができる。
【０１１１】
また、株価変動評価語を含むニュース情報が銘柄の第１概念名を定義するグループ定義語を含むとき、当該グループ定義語に基づいて当該ニュース情報から第１概念名を抽出し、株価変動評価語を含むニュース情報から企業名及び証券コードを含む銘柄を抽出し、株価変動評価語を含むニュース情報から株価変動材料を抽出し、当該抽出された第１概念名、銘柄及び株価変動材料を組み合わせて知見情報を作成し、当該作成された知見情報を記憶部１１に書込む構成により、証券市場で使われる固定の業種ではないグルーピングの知見情報を活用できるので、株価変動予測を有効に支援できる。
【０１１２】
さらに、知見情報が記憶された後、新規に受信された全てのニュース情報を記憶部１１内の知見情報に基づいて分析する構成により、分析結果を、即時にアルゴリズムトレードエンジンなどのシステムもしくは株取引を行うユーザに提供することで、直近の株取引を有効に支援することができる。
【０１１３】
また、提供先がアルゴリズムトレードエンジンなどのシステムである場合、現在の株価や注文状況などの数値情報のみを参照する場合と比較して、豊富な情報をもとにした判定を行うことが可能となる。
【０１１４】
また、本実施形態及び以下の各実施形態は、対象をインターネット配信ニュースに限定しない。例えば、インターネットで配信される他の情報（ブログ、ミニブログ、マイクロブログ、企業発表情報）や、インターネット外に存在する電子文書を対象とすることも可能である。さらに、対象をニュース情報やブログなどの複数種類の情報の組み合わせとすることも可能である。
【０１１５】

なお、ステップＳ３２の類似ニュース情報もしくは一致するニュース情報の判定には、例えば、この出願の出願時に未公開である先願（特願２０１０－２４７５１８）の明細書に記載した処理［１］～［３］を用いてもよい。以下の処理［１］～［３］において「タイトル」の語は「ニュース本文」と読み替えてもよい。また、「ＩＤ付ニュース情報」の語は「ニュース情報」と読み替えてもよい。また、処理［１］～［３］の主体は、前述したステップＳ３２に合わせて「知見抽出対象別判定部１５３」に書き換えている。

【０１１６】

［１］記憶部１１は、単語解析辞書及び同一性判定基準などを予め記憶する。

【０１１７】

同一性判定基準は、送信されるニュース情報における任意の２件のニュース情報が同一ニュースであるか否かを判定する基準を示しており、当該２件のニュース情報に含まれるニュースソース名が互いに一致することと、当該２件のニュース情報に含まれる配信日時の差分を示す配信時間差が基準値（最大の配信時間差）よりも小さいことと、当該２件のニュース情報における２件のタイトルの形態素解析結果から算出される類似度が規定値より高いことと、当該類似度は前記数値情報の有効数字の桁を四捨五入により合わせた後に算出されることとを含んでいる。

【０１１８】

なお、類似度は、例えば、２件のタイトルの形態素解析結果に含まれる自立語の集合全体における当該自立語及び数値情報の個数に対し、当該２件のタイトルの形態素解析結果の両方に含まれる自立語及び数値情報の個数が占める割合である。また、類似度（０以上１以下）の規定値は、０．９程度の高い値が好ましい。また、「２件のタイトルの形態素解析結果から算出される類似度が規定値より高いこと」に代えて、「２件のタイトルの形態素解析結果から抽出される自立語が互いに一致すること」としてもよい。また、同一性判定基準は、同一性判定ルールと読み替えてもよい。

【０１１９】

［２］知見抽出対象別判定部１５３は、記憶部１１内の単語解析辞書を用い、ＩＤ付ニュース情報の単語解析処理を行なう。単語解析処理は例えば一例として、形態素解析技術（公知の技術）を用いる。換言すると、知見抽出対象別判定部１５３は、記憶部１１内のＩＤ付ニュース情報に含まれるタイトルを形態素解析し、得られた形態素解析結果を当該ＩＤ付ニュース情報のニュースＩＤ及びタイトルに付加し、得られた解析結果情報を記憶部１１に書き込む形態素解析機能をもっている。

【０１２０】

なお、以下では形態素解析技術を一例として説明するが、知見抽出対象別判定部１５３での処理は、単語解析辞書を用いない、Ｎグラムなどの形態素解析ではない解析処理を用いて、単語に分解してもよい。つまり、本実施形態は、形態素解析に限定しない手法によりニュース情報を単語に分割し、単語の比較で類似度を判断するものである。

【０１２１】

但し、例えば文字を単位としたＮグラムの場合、品詞付けや自立語であるか否かの判定はできない。従って、形態素解析の他の手段で単語解析処理を行なうときは、「自立語」ではなく「単語」が処理対象となり得る。

【０１２２】

［３］知見抽出対象別判定部１５３は、記憶部１１内の同一性判定基準を満たすか否かに基づいて、記憶部１１内の最新のＩＤ付ニュース情報と過去に書き込んだＩＤ付ニュース情報とが同一ニュースであるか否かを判定する。

【０１２３】

例えば、知見抽出対象別判定部１５３は、最新のＩＤ付ニュース情報におけるタイトルの形態素解析結果から抽出された自立語及び数値情報と、過去に書き込まれたＩＤ付ニュース情報におけるタイトルの形態素解析結果から抽出された自立語及び数値情報とが一致する割合を示す類似度を算出する。但し、類似度を算出する前に、数値情報の有効数字の桁を四捨五入により合わせておく。この類似度が規定値（例、０．９）より高い場合、同

一性判定部１５は、ＩＤ付ニュース情報内のニュースソース名が一致し、さらに、各ニュースの配信時刻の差が基準値（例、５分）以内ならば、同一ニュースである旨を判定する。なお、類似度が高い旨の確認、ニュースソース名の一致確認、配信時刻の差の確認は、任意の順序で実行可能である。また、類似度が高い旨の確認に代えて、自立語が完全一致する旨を確認してもよい。

【０１２４】

続いて、同一性判定の処理［３］について、ニュースＩＤ“３８”のニュース情報と同一性判定対象のニュースＩＤ“３”のニュース情報とを例に挙げて述べる。

【０１２５】

ニュースＩＤ“３８”を含む最新のＩＤ付ニュース情報に対する形態素解析結果からは、数値情報「１５．８％減」１種、自立語「米＜名詞－固有名詞－国＞」「商品販売＜名詞－一般＞」「Ａ社＜名詞－固有名詞－組織＞」「１月＜名詞－副詞可能＞」「リコール＜名詞－サ変接続＞」「問題＜名詞－ナイ形容詞語幹＞」「響く＜動詞－自立＞」７種が抽出される。

【０１２６】

同一性判定対象のニュースＩＤ“３”を含む過去のＩＤ付ニュース情報に対する形態素解析結果からは、数値情報「１５％減＜数値情報＞」１種、自立語「Ａ社＜名詞－固有名詞－組織＞」「１月＜名詞－副詞可能＞」「米＜名詞－固有名詞－国＞」「商品販売＜名詞－一般＞」４種が抽出される。

【０１２７】

ここで、ニュースＩＤ“３８”とニュースＩＤ“３”に対応する数値情報はそれぞれ「１５．８％減」と「１５％減」であり、値が一致しない。

【０１２８】

自立語については、２件の形態素解析結果から抽出される自立語が全部で７種あるのに対して、２件の形態素解析結果の両方から抽出される自立語が４種である。

【０１２９】

この場合、数値情報１種と自立語全７種の合計８種の情報のうち一致するものが４種であることに基づき、一致度が５０％、類似度が０．５と算出される。

【０１３０】

類似度の算出は、さらに「数値情報が含まれ、一致しない場合は類似度０とする」又は「数値情報の一致度と、自立語の一致度の平均を類似度とする」などとして実行してもよい。

【０１３１】

なお、知見抽出対象別判定部１５３による自立語比較は、形態素解析結果をそのまま比較したが、これに限らず、形態素解析結果における動詞・形容詞・形容動詞を原形に変換して比較する処理や、否定の助動詞が続いていた場合には原形に戻す際に否定形の終止形とする処理、などの処理を加えるように変形してもよい。

【０１３２】

（第２の実施形態）

次に、第２の実施形態について説明する。なお、第２の実施形態は、株取引に関する知見情報を抽出する第１の実施形態の知見抽出装置１に関する変形例である。

【０１３３】

株価変動の方向は、新たに得られた情報がポジティブかネガティブであるかとは必ずしも一致しない。株取引を行うユーザたちが事前に予測していた内容と比較して改善方向か改悪方向かが株価変動の方向を決める傾向がある。

【０１３４】

したがって、第１の実施形態では、情報分析部１５の処理において株価変動材料の有無とその表現のみをＣＥＰ（Ｃｏｍｐｌｅｘ　Ｅｖｅｎｔ　Ｐｒｏｃｅｓｓｉｎｇ：複合イベント処理）もしくはユーザへの提供内容とした。しかし、一部の株価変動材料は、株価変動に一定の方向性を与える傾向がある。

【０１３５】
そこで、第２の実施形態では、第２手掛かり語のうち、方向性が明らかであるものには、知見抽出部１４内の材料抽出部１４５が、方向（ポジティブもしくはネガティブ）と、その強度を付与する形態としている。この強度は、例えば整数－５～５の範囲で予め設定される。
【０１３６】
これに伴い、第２概念抽出部１４５は、図４に示した第２手掛かり語辞書１１３に代えて、図１６に示す第２手掛かり語辞書１１３ａを用いる形態としている。
【０１３７】
具体的には、第２概念抽出部１４５は、知見抽出判定語チェック部１４１による判定の結果、電子文書であるニュース情報が株価変動評価語を含むとき、当該ニュース情報から株価変動材料を抽出する前述した機能に加え、当該抽出された株価変動材料に基づいて、記憶部１１内の第２手掛かり語辞書１１３ａを検索し、当該検索された株価変動の方向及び強度を表す各情報を当該株価変動材料に付与する機能を更に備えている。
【０１３８】
なお、第２手掛かり語辞書１１３ａは、手掛かり語を識別するコードと、株価変動材料の表現に対する手掛かり語と、株価変動の方向を表す情報と、株価変動の強度を表す情報とが関連付けられて記述されている。
【０１３９】
次に、以上のように構成された知見抽出装置１の動作を説明する。なお、第１の実施形態の知見抽出装置１と同様の構成・動作については説明を省略する
ステップＳ１～Ｓ６までの動作は、前述同様に実行される。
【０１４０】
第２概念抽出部（以下、材料抽出部と記載する）１４５は、知見抽出判定語チェック部（以下、株価変動評価語チェック部と記載する）１４１による判定の結果、ニュース情報が株価変動評価語を含むとき、当該ニュース情報から株価変動材料を抽出する。具体的には、第２概念抽出部（以下、材料抽出部と記載する）１４５は、株価変動評価語を含むニュース情報のニュース見出しもしくはニュース本文に、第２手掛かり語辞書（以下、材料表現手掛かり語辞書と記載する）１１３ａ内の手掛かり語の１つ以上が含まれるか否かを判定する（ステップＳ７）。
【０１４１】
ステップＳ７およびステップＳ８の処理は、図９のフローチャートではステップＳ６の後としているが、ステップＳ３の処理の前、もしくはステップＳ３からステップＳ６の処理と並列で実施してもよい。
【０１４２】
材料抽出部１４５は、受信したニュース情報のニュース見出しもしくはニュース本文に、材料表現手掛かり語辞書１１３ａ内の手掛かり語が含まれる場合、その手掛かり語に係り受けしている名詞句を株価変動材料として抽出する（ステップＳ８）。
【０１４３】
例えば図１７に示すニュースＩＤ“４”からの知見抽出処理において、材料表現手掛かり語辞書１１３ａに登録された手掛かり語「嫌気」を抽出し、手掛かり語「嫌気」に係り受けしている名詞句「業績下方修正」からなる株価変動材料に、手掛かり語「嫌気」に対応づけられた方向及び強度を付与し、株価変動材料「材料：業績下方修正　方向：ネガティブ　強度：２」が得られる。
【０１４４】
組み合わせ作成部１４６は、前述同様に、知見情報を作成し、当該知見情報を記憶部１１に書き込む（ステップＳ９）。図１８は当該知見情報の一例である。ここで、知見情報は、方向及び強度が付与された株価変動材料「材料：業績下方修正　方向：ネガティブ　強度：２」を含んでいる。知見情報は、第１の実施形態と同様に、組み合せを、第１概念名、銘柄、株価変動材料をセットにしているが、組み合わせは第１概念名と銘柄のみとし

、株価変動材料は単独で知見情報として保持するとしてもよい。または、株価変動材料は第１概念名ではなく銘柄とセットにして知見情報として保持するとしてもよい。

【０１４５】

その後、図１７に示すニュースＩＤ“５”を受信した際は、情報分析部１５は、ニュース本文から株価変動材料「業績下方修正」を抽出した際に、「方向：ネガティブ　強度：２」の情報を、銘柄「１１５２：　ＡＡ製鉄（株）」材料「業績下方修正」に加えてアルゴリズムトレードエンジンなどのシステムもしくは株取引を行うユーザ端末装置２に提供する。

【０１４６】

上述したように本実施形態によれば、株取引に関する知見情報を抽出する知見抽出装置１によって抽出された株価変動材料に基づいて、第２手掛かり語辞書を検索し、当該検索された株価変動の方向及び強度を表す各情報を当該株価変動材料に付与する構成により、第１の実施形態の効果に加え、株価変動材料に方向と強度を加えた情報を提供することで、アルゴリズムトレードエンジンなどのシステムもしくは株取引を行うユーザの速やかな判断をさらに強く支援することができる。

【０１４７】

（第３の実施形態）

次に、第３の実施形態について説明する。

【０１４８】

第３の実施形態は、株取引に関する知見情を抽出する第１の実施形態の知見抽出装置１の変形例であり、知見抽出部１４が抽出する知見情報をある対象に関する材料（イベント）とそのタイムスタンプとしている。

【０１４９】

これに伴い、情報受信部１２は、ニュース情報を受信する前述した機能に加え、ニュース情報を受信する際に、当該ニュース情報にタイムスタンプを付加する機能を更に備えている。

【０１５０】

知見抽出部１４は、情報選択部１３により選択されたニュース情報から知見情報を抽出する前述した機能において、当該選択されたニュース情報から所定の対象を示す表現を含むイベント情報を抽出すると共に、当該イベント情報が抽出されたニュース情報から抽出し、当該抽出したイベント情報及びタイムスタンプを知見情報として記憶部１１に書込む機能をもっている。

【０１５１】

ここで、所定の対象を示す表現は、例えば、銘柄の企業名を表す企業名表現、又は銘柄の第１概念名を定義するグループ定義語の直前の名詞を表す表現である。なお、「銘柄の企業名を表す企業名表現」は、知見抽出対象語辞書（以下、銘柄辞書と記載する）１１２に記述されている。「銘柄の第１概念名を定義するグループ定義語」は、第１手掛かり語辞書(以下、グループ定義後辞書と記載する)１１１に記述されている。

【０１５２】

次に、以上のように構成された知見抽出装置１の動作を説明する。図１９は、知見抽出部１４の処理動作の一例を示すフローチャートである。

【０１５３】

情報受信部１２がニュース情報を受信すると（ステップＳ４１）、知見抽出部１４はグループ定義語辞書１１１及び銘柄辞書１１２などを参照し、ユーザによって選択されている対象（企業が関連づけられた第１概念）を示す表現が含まれるか否かを判定する（ステップＳ４２）。

【０１５４】

ユーザによって選択されている対象を示す表現が含まれる場合、知見抽出部１４は、その対象を示す表現を含む名詞句を、第２概念（イベント）として抽出する（ステップＳ４３）。

【０１５５】
知見抽出部１４は、抽出した第２概念を、対象名、受信情報のタイムスタンプと合わせて記憶部１１に追加的に書込む（ステップＳ４４）。
【０１５６】
以下、情報分析部１５は、前述同様に、ステップＳ２１～Ｓ３４の動作を実行する。
【０１５７】
図２０は、知見情報の一例を示す図である。図５に示すニュースＩＤ“１”，“２”から第１概念名「インフルエンザ」に関して処理を行った例である。
【０１５８】
抽出する知識（第２概念）はユーザによって選択された対象表現を含む名詞句としたが、この名詞句中から複合名詞を選択したり、固有名詞（地名など）を選択したり、急上昇ワードを選択して第２概念とし、第２手掛かり語を「○県」「集団感染」「△市」やその組み合わせとしてもよい。
【０１５９】
第１概念「インフルエンザ」で抽出した第２概念を、図１２に示した如き、第１の実施形態で取得した知見情報にしたがって企業に展開し、図２０に示す知見情報で「対象」を「インフルエンザ」から「１０３３：ｙｙボウホールディングス（株）」に置き換えた情報を合わせて知見情報として書き込んでもよい。
【０１６０】
さらに、図２０に示すように、類似度の高いニュース情報の配信数を「類似情報数」として合わせて知見情報としてもよい。
【０１６１】
上述したように本実施形態によれば、ニュース情報を受信する際に、ニュース情報にタイムスタンプを付加し、選択されたニュース情報から所定の対象を示す表現を含むイベント情報を抽出すると共に、当該抽出したイベント情報及びタイムスタンプを知見情報として記憶部１１に書込む構成により、第１の実施形態の効果に加え、ある対象（第１概念）に関する第２概念をタイムスタンプ情報のある時系列で抽出し保持することで、同様の事象が発生した際に、近い将来の予想に有効な知見情報を保持しユーザ端末装置２に提示することができる。
【０１６２】
なお、第３の実施形態は、受信する情報を電子メールとし、対象をプロジェクトとすることで、企業活動のうちのプロジェクトに関するリスクを示す知見情報を抽出することも可能である。さらに、受信する情報を情報機器のシステムログとし、対象をコンピュータネットワークシステムとすることで、コンピュータネットワークシステムの障害検知に関する知見情報の抽出をすることも可能である。
【０１６３】
（第４の実施形態）
次に、第４の実施形態について説明する。
【０１６４】
図２１を参照して、第４の実施形態の知見抽出装置１００の構成について説明する。なお、第１の実施形態乃至第３の実施形態に記載の構成については省略する。
【０１６５】
図２１に示すように、本実施形態の知見抽出装置１００は記憶部１１の評価情報１１４と確信度更新ルール１１５、及び知見評価更新装置１８を備える。評価情報１１４と確信度更新ルール１１５、及び確信度については後述する。
【０１６６】
知見評価更新装置１８は、上位概念チェック部１８１と、下位概念抽出部１８１と、評価部１８３と、確信度更新部１８４と、知見更新部１８５を備える。
【０１６７】
上位概念チェック部１８１は、知見情報に含まれるグループ名毎に、あらかじめ設定し

た期間に記憶部１１に新規登録されたニュース情報中で、当該グループ名が記載されたニュースの有無と配信数をチェックする。すなわち、本実施形態の知見情報は階層構造の情報であり、グループ名を上位概念とし、銘柄を下位概念とする。

【０１６８】

下位概念抽出部１８２は、上位概念チェック部１８２によってチェックされたグループ名の下位概念である、知見情報に含まれる銘柄を全て抽出する。また、下位概念抽出部１８２は、上位概念チェック部１８１がチェックしたニュース情報から当該銘柄が記載されたニュースの有無と配信数をチェックする。

【０１６９】

評価部１８３は、記憶部１１に記憶された評価情報１１４に基づいて、下位概念抽出部１８２が抽出した銘柄毎に評価値を算出する。評価情報１１４は、例えば株価情報、ニュース、株の出来高、会社の業績や従業員数などである。

【０１７０】

確信度更新部１８４は、評価部１８３が算出した評価値と、確信度更新ルール１５とに基づいて、後述する確信度を更新する。

【０１７１】

知見更新部１８５は、確信度更新部１８４による更新結果に基づいて、記憶部１１に記憶された知見情報を更新する。

【０１７２】

ここで、株価変動が、各銘柄が関連するニュースの発生の影響を受けることは前述のとおりであるが、影響の大きさは銘柄や、グループ名（キーワード）により異なる。したがって、グループ名と銘柄の組み合わせに加えて、その銘柄が該当グループ名のニュース発生時に受ける影響の強さ、すなわち、その銘柄が該当グループに所属する確信度の情報を合わせて保持することが有効である。

【０１７３】

したがって、第４の実施形態では、図１２に示した知見情報に代えて、図２２に示すように、確信度と確信度更新履歴とを含む確信度情報を加えた知見情報を保持する形態としている。すなわち、本発明の確信度とは、下位概念が上位概念に帰属する（関連する）度合いを示す数値であり、背景技術に記載した第３の技術における確信度とは異なる。知見情報における確信度が高いほど、下位概念が上位概念に帰属する度合いが高い、すなわち関連性が強い情報であると判定できる。

【０１７４】

なお、図２２では図１２の株価変動材料の情報を省略しているが、株価変動材料を加えた構成としてもよい。または、グループ名と株価変動材料の組合せを図２２の知見情報とは別テーブルで管理する構成としてもよい。さらに、株価変動材料に加えて、第２の実施形態および図１８に示すように、方向、強度を合わせて管理してもよい。また、確信度情報に含まれる情報は確信度のみでもよい。

【０１７５】

図２２の知見情報では、第１の実施形態、第２の実施形態、第３の実施形態で用いた図１２の知見情報と同様にグループ名、銘柄の２種の情報を保持する。さらに、その銘柄が該当グループに所属する確信度の情報を加えた３種の情報を必須とする。これらに加えて、グループ名を細分化する下位グループ名、該当企業から発表された、該当事業（グループ）に参入予定の時期である事業参入時期、該当企業から発表された、該当事業から撤退予定の時期である事業撤退時期、過去一定回数までの確信度更新履歴、そのグループ名と銘柄の関係を新規抽出した際のグループ定義語を記録するグループ定義語、などの情報を合わせて保持してもよい。本実施形態では、確信度更新履歴を２つ前の履歴まで保持する。

【０１７６】

次に、図２２の知見情報の確信度を保持し更新する知見評価更新装置１８を備えた知見抽出装置１００の動作を説明する。図２３は、知見評価更新装置１８の処理動作（以下、

知見評価更新処理という）の詳細を示す模式図、図２４及び図２５は知見評価更新処理の一例を説明するためのフローチャートである。

【０１７７】

以下、本実施形態の知見抽出装置１００において、図２２の４行目の知見情報に関する処理を行う場合について説明する。図２２の４行目の知見情報は、「第１概念名：インフルエンザ、下位グループ名：インフルエンザ薬、銘柄：ｆｆ医薬＜１０４２＞、確信度：１、事業参入時期：２００７年１０月１９日、確信度更新履歴１（前回の確信度更新履歴）：２００８年9月１０日　＋１、グループ定義後：関連」、である。

【０１７８】

第１概念名（以下、グループ名と記載する）、銘柄（企業名表現）、株価変動材料のセットの抽出は、第１の実施例と同様に実行される。グループ名と銘柄の組み合わせが新規に抽出された際は、確信度の初期値として、例えば１．０を付与する。また、本実施形態の確信度の最小値は０とし、後述する確信度更新処理により、確信度が０を下回る結果となった場合は０に修正する。また、本実施形態の確信度の最大値は５．０とする。確信度更新処理により、確信度が５．０を超える結果となった場合は、例外処理により、５．０を大きく超えることがないように調整してもよい。例外処理は、例えば、前日の確信度が５．０より大、かつ、今回の更新でも確信度が増える状況の場合は、確信度を０．０５のみ増やすこととし、前日の確信度が４．９５未満、かつ今回の更新で確信度が５．０を超える場合に、更新後の確信度を５．０とする、という処理を行う。

【０１７９】

なお、以下の説明では、知見評価更新処理を、証券市場終了後の夕方もしくは夜間に１回実行するとしているが、この処理は、１日１回ではなく、証券市場の前場と後場の終了時にそれぞれ実施してもよい。

【０１８０】

知見評価更新処理が起動されると、上位概念チェック部１８１は図２２の知見情報からグループ名を１つ選択する（ステップＳ５１）。ここでは、グループ名「インフルエンザ」が選択される。続いて、上位概念チェック部１８１は、その日に記憶部１１に新規登録されたニュース中で、ステップＳ５１で選択されたグループ名が記載されたニュースの有無を確認する（ステップＳ５２）。

【０１８１】

なお、本実施形態では、ニュース件数の集計は、例えば営業日の１５：００などの所定の時刻ごとに行う。休日に配信されたニュースは、翌営業日のニュースに加算してもよい。また、ステップＳ５２においては、グループ名と下位グループ名の両方が記載されたニュースの有無を確認してもよい。

【０１８２】

選択中のグループ名が記載されたニュースがない場合（ステップＳ５２がＮｏ）、ステップＳ５３～ステップＳ６３の処理を省略し、ステップＳ６４に進む。ステップＳ６４については後述する。

【０１８３】

選択中のグループ名が記載されたニュースがある場合（ステップＳ５２がＹｅｓ）、下位概念抽出部１８２は、図２２の知見情報のうち、グループ名が選択中のものと同一である知見情報に含まれる銘柄を抽出する（ステップＳ５３）。評価部１８３は、下位概念抽出部１８２によって選択された銘柄から１つを選ぶ（ステップＳ５４）。ここでは銘柄「ｆｆ医薬＜１０４２＞」が選択される。

【０１８４】

続いて、評価部１８３は、下位概念抽出部１８２によって選択された銘柄と、その銘柄に対応するグループ名と同時に含み、かつ事業撤退に関連するあらかじめ定めた語句（「撤退」「売却」など）がないニュースがあるか否かを判定する（ステップＳ５５）。

【０１８５】

下位概念抽出部１８２によって選択された銘柄と、その銘柄に対応するグループ名と同

時に含み、かつ事業撤退に関連する語句がないニュースがない場合（ステップＳ５５はＮｏ）、ステップＳ５６、Ｓ５７の処理を省略し、ステップＳ５８に進む。

【０１８６】

下位概念抽出部１８２によって選択された銘柄と、その銘柄に対応するグループ名と同時に含み、かつ事業撤退に関連する語句がないニュースがある場合（ステップＳ５５はＹｅｓ）、評価部１８３は、当該グループ名を含むニュース件数が規定数以上かどうかを確認する（ステップＳ５６）。本実施形態ではステップＳ５６での規定数を３とする。

【０１８７】

当該グループ名を含むニュース件数が規定数以上の場合（ステップＳ５６がＹｅｓ）、評価部１８３は、評価値として、当該銘柄の「出来高変化率」を算出し、算出した出来高変化率が後述する確信度の更新条件に合致するか否かを確認する（ステップＳ５７）。

【０１８８】

この場合、本実施形態の評価情報１１４には株の銘柄に関する日ごとの出来高が含まれる。銘柄がｆｆ医薬＜１０４２＞の出来高の一例を図２６に示す。

【０１８９】

ステップＳ５７で評価部１５８は、この評価情報１１４を用いて「出来高変化率」を以下の式で算出する。

【０１９０】

（１）　出来高変化率＝（当日の出来高）／（直近５営業日の出来高の平均）

上記の式（１）では、直近５営業日の平均を用いているが、５営業日に限らないことは言うまでもない。

【０１９１】

なお、ステップＳ５７における更新条件は例えば「あらかじめ設定した閾値以上であれば更新する」である。また、更新条件は、当該知見抽出装置１００に含まれる記憶部１１に含まれる情報に基づいて動的に変化させてもよい。更新条件の動的な変化とは、例えば、評価情報１１４としてニュース情報を用いる場合、ニュース情報に含まれる当該銘柄の関連ニュースの件数が所定の数よりも多い場合に、更新条件に含まれる閾値を所定の単位で上昇させるようなことが考えられる。

【０１９２】

本実施形態では、更新条件は「出来高変化率＞１．１もしくは出来高変化率＜１．０に当てはまるか否か」とする。

【０１９３】

ステップＳ５５～ステップＳ５７の判定は、確信度を更新するかを判定するための処理の一例であり、判定をさらに詳細に分けることや、判定基準として別の指標を用いることも考えられる。

【０１９４】

当該グループ名を含むニュース件数が規定数未満の場合（ステップＳ５６がＮｏ）、もしくは評価部１８３が算出した評価値が、確信度の更新条件に合致しない場合（ステップＳ５７がＮｏ）、ステップＳ５８～ステップＳ６２の処理を省略し、ステップＳ６３の処理に進む。

【０１９５】

評価部１８３が算出した評価値が、確信度の更新条件に合致する場合（ステップＳ５７がＹｅｓ）、確信度更新部１８４が当該銘柄と対応付けて記憶部１１に記憶された確信度を更新する（ステップＳ５８）。なお、確信度更新部１８４による確信度の更新は、記憶部１１に確信度更新ルール１１５をあらかじめ設定しておき、このルールに従って更新する。

【０１９６】

本実施形態における確信度更新ルール１１５の一例を図２７に示す。図２７に示す確信度更新ルール１１５は、記憶部１１に格納されている。

【０１９７】

図２７の確信度更新ルール１１５は、「ステップＳ５５の判定がＹｅｓ、かつ出来高変化率≦１．１の場合、確信度に０．０５を加える」，「ステップＳ５５の判定がＹｅｓ、かつ出来高変化率＞１．１の場合、確信度に（出来高変化率－１．０）×０．５を確信度に加える」，「ステップＳ５６がＹｅｓ、かつ出来高変化率＞１．１の場合、（ニュース変化率－出来高変化率）×０．５を確信度から減らす」，「ステップＳ５６がＹｅｓ、かつ出来高変化率＜１．０の場合、（ニュース件数変化率－出来高変化率）×０．５を確信度に加える」である。

【０１９８】

ニュース件数変化率は、以下の式に基づいて算出する。

【０１９９】

（２）　ニュース件数変化率＝（当日の該当グループ名を含むニュース件数に当日の確信度を乗じたもの）／（直近５営業日の該当グループ名を含むニュース件数に該当日の確信度を乗じたものの平均）

また、本実施形態におけるインフルエンザが含まれるニュースの件数を日ごとに示したグラフを図２８に示す。

【０２００】

本実施形態では、ステップＳ５５のＹｅｓからステップＳ５８に進んだ場合と、ステップＳ５７のＹｅｓからステップＳ５８に進んだ場合で更新ルールが異なる。また、例えば「出来高変化率＜１．０の場合（１－出来高変化率）×０．５を前営業日の確信度から減らす」というような更新ルールでもよい。なお、確信度の前日値は例えば図２２の知見情報に含まれる確信度更新履歴から算出する。または、確信度自体を履歴として保持してもよい。

【０２０１】

図２９に、本実施形態の確信度に関するグラフの一例を示す。

【０２０２】

続いて、知見更新部１８５が、図２２の知見情報を更新する。具体的には、知見更新部１８５は、更新後の確信度が、当該銘柄を含む知見情報を記憶部１１から削除する条件（以下、削除条件という）を満たすか否かを判定する（ステップＳ５９）。本実施形態では、例えば、確信度が０．３以下になった場合は、知見更新部１８５は、削除条件を満たすと判定する。なお、この判定は、最新の確信度の数値で判定するのではなく、更新履歴で例えば減少が５回連続した場合に削除条件を満たすとするとしてもよい。

【０２０３】

更新後の確信度が０．３より大きい場合（ステップＳ５９がＮｏ）、すなわち、グループからの削除条件を満たさない場合、知見更新部１８５はステップＳ６１～ステップＳ６２の処理省略し、ステップＳ６３に進む。

【０２０４】

更新後の確信度が０．３以下である場合（ステップＳ５９がＹｅｓ）、すなわち、グループからの削除条件を満たす場合、知見更新部１８５は、ユーザにアラームを提示する（ステップＳ６０）。このアラームとは、ユーザに選択中の銘柄を選択中のグループから削除するか否かを確認するためのものであり、例えば入出力インタフェース１６に表示される。

【０２０５】

ユーザが、表示されたアラームを参照し、当該知見情報を削除するか否かを、入出力インタフェースを用いて入力すると、知見更新部１８５は、入力された情報が削除を指示する情報であるかを判定する（ステップＳ６１）。

【０２０６】

削除が指示されない場合（ステップＳ６１がＮｏ）、知見更新部１８５はステップＳ６２の処理を省略し、ステップＳ６３の処理に進む。

【０２０７】

削除が指示された場合（ステップＳ６１がＹｅｓ）、知見更新部１８５は、選択された

銘柄をグループから削除する（ステップＳ６２）。すわなち、選択中のグループと選択中の銘柄の組み合わせに該当する知見情報を記憶部から削除する。なお、銘柄のグループからの削除は、この一連の処理のみで行うのではなく、図２２の知見情報に記載された事業撤退時期の情報に従い、撤退時期に到達した銘柄を削除する処理を別途行ってもよい。

【０２０８】

続いて、知見更新部１８５は、選択中のグループに未処理の銘柄があるか否かを判定する（ステップＳ６３）。未処理の銘柄がある場合（ステップＳ６３がＹｅｓ）、ステップＳ５４に戻って未処理の銘柄のうちの１つについて同様の処理を行う。

【０２０９】

選択中のグループで未処理の銘柄がなくなった場合（ステップＳ６３がＮｏ）もしくは、ステップＳ５２がＮｏの場合、知見更新部１８５は、知見情報に保持されたグループで未処理のものがあるか否かを判定する（ステップＳ６４）。

【０２１０】

未処理のグループがある場合（ステップＳ６４がＹｅｓ）、ステップＳ５１に戻って未処理のグループのうちの１つについて同様の処理を行う。未処理のグループがない場合８ステップＳ６４がＮｏ）、処理を終了する。

【０２１１】

上述したように本実施形態によれば、知見を抽出した情報源以外の情報（株価変動や出来高など）に基づいて、知見の有効性を評価し更新する構成により、第１の実施形態の効果に加え、知見の確かさを複数の情報源に基づき高く維持できる。

【０２１２】

なお、第４の実施形態において、図２４、図２５の処理の判定基準は、前記に限定するものではない。例えば、ステップＳ５２の選択中のグループ名を含むニュースがあるか否かの判定は、選択中のグループ名を含むニュース数が前日以上であるか否かの判定に置き換え、もしくは選択中のグループ名が急上昇ワードに該当するか否かの判定に置き換えてもよい。

【０２１３】

また、ステップＳ５７における選択中の銘柄の出来高変化率が更新条件を満たすか否かの判定は、選択中の銘柄の株価変動や売買代金、日経平均などの株式市場全体の傾向を示す指標と比較した株価変動傾向による判定に置き換えてもよい。

【０２１４】

例えば、ステップＳ５７において株価変動を用いて判定する場合、評価部１８３は、評価値としてその銘柄の直前の営業日の株価変動、例えばさらにその前の営業日からの株価変動の割合を求める。評価部１８３はこのとき、例えば株価変動が５％以上の上昇ならば選択中の銘柄が選択中のグループに帰属することは適切であると判定する。また、評価部１８３は、株価変動が５％未満の下落もしくは５％未満の上昇の場合は、銘柄のグループ帰属の適切判断を保留する。また、評価部１８３は、株価変動が５％以上の下落の場合は、選択中の銘柄が選択中のグループに帰属することは不適切であると判定する。この株価変動の判定は、第２の実施形態および図１８に示したように、ニュースに記載された材料表現とその方向も加味し、上昇（ポジティブ）と下落（ネガティブ）のいずれがグループに帰属することの判断に適切であるかを決定した上で行ってもよい。

【０２１５】

また、各銘柄に関する処理は、所属するグループごとに実施する形態としたが、ある銘柄が同日に複数のグループで処理対象となる場合は、例外的な処理を実施してもよい。例外的な処理とは、例えば、［４］もしくは［５］の処理である。

【０２１６】

［４］同日に複数のグループで処理対象となる銘柄は、確信度更新を実施しない。

【０２１７】

［５］同日に複数のグループで処理対象となる銘柄は、所属するグループごとに、他の所属銘柄の変動を参照する。他の銘柄と同じ方向・同程度の数値変動をした場合は、その

グループでの確信度更新処理を実施する。

【０２１８】

また、本実施形態では、確信度による評価更新の対象をグループ名と銘柄の関係としたが、確信度は、図２２の知見情報に保持されたグループ定義語の評価に用いてもよい。すなわち、知見情報に保持されたグループ定義語と対応する確信度の情報をグループ定義語ごとに整理し、確信度の平均や最大値が規定値未満となったグループ定義語を有効ではないと判定して図２のグループ定義語辞書（第１手掛かり語辞書）から削除してもよい。

【０２１９】

また、確信度の情報は、知見情報もしくはグループ定義語の評価更新に用いるものではなく、一般ニュースに含まれるグループ名を数える際に重み付けに用いてもよい。すなわち、例えばグループ名「インフルエンザ」を含むニュースが５０件ある場合、確信度が１．０の銘柄には５０件、確信度が０．８の銘柄には４０件のニュース配信があったものとして件数を割り付けるようにしてもよい。

【０２２０】

図３０は、本実施例の効果の一例を示す図である。図３０には、ある東証一部上場銘柄について、その銘柄名を含むニュース件数（銘柄名のみ）、銘柄名もしくはその銘柄が所属するグループ名（図３０の例では約２００のグループがあるとする）の少なくとも一つを含むニュース件数（銘柄名ＯＲグループ名（確信度なし））、グループ名を含むニュース件数をそのグループ名の確信度で補正した件数と銘柄名を含むニュース件数の合計（銘柄名ＯＲグループ名（確信度補正なし））について、ニュース件数と該当銘柄の出来高の相関係数を求めたものが示されている。

【０２２１】

相関係数とは、２つの確率変数の間の相関、すなわち類似度の度合いを示す統計学的指標であり、－１から＋１の間の実数値をとる。１に近いときは２つの確率変数には正の相関があるといい、－１に近ければ負の相関があるという。０に近いときはもとの確率変数の相関は弱い。相関係数と相関関係は、例えば相関係数　が「０．０～±０．２」であれば、相関関係は「ほとんど相関がない」、相関係数が「±０．２～±０．４」であれば、相関関係は「やや相関がある」、相関係数が「±０．４～±０．７」であれば、相関関係は「相関がある」、相関係数が「±０．７～±０．９」であれば、相関関係は「強い相関がある」、相関係数が「±０．９～±１．０」であれば、相関関係は「きわめて強い相関がある」、とみなせる。

【０２２２】

図３０に示した、本実施形態の知見評価更新処理を適用した「銘柄名ＯＲグループ名（確信度補正あり）」の相関係数は、銘柄名のみ、および銘柄名ＯＲグループ名（確信度なし）の相関係数よりも高い値になっており、より強い相関がある知見情報を得ることができていると言える。

【０２２３】

なお、本実施形態の知見評価更新装置１８では、株取引に関する知見情報以外の知見情報を評価・更新することも可能であり、例えば商品に関する知見情報（以下、商品知見情報という）の知見評価更新処理が可能である。

【０２２４】

この場合、商品知見情報は、企業の告知、ブログ、マイクロブログ、などのインターネットで配信される情報から抽出される。また、このとき、商品知見情報に含まれる下位概念の情報を商品名とし、上位概念の情報をグループ名とし、さらに商品知見情報の評価値を商品の売り上げ情報の変化率とする。この売り上げ情報の変化率により商品知見情報を評価・更新することで、商品のグループ名と売り上げとの相関がより強い知見情報を抽出することも可能である。

【０２２５】

（第５の実施形態）

次に、第５の実施形態について説明する。

【０２２６】
第１の実施形態乃至第４の実施形態では、ニュース情報にグループの情報があるか否かを判定する際に、グループ名のみを抽出すべき表現としていた。しかし、一般ニュースでは、経済ニュースで用いられるグループ名ではない関連語が記載される場合がある。例えば、図２９に示すような経済ニュースからは、グループ名「防衛（関連）」、銘柄「ａａ重工〈７１９１〉」、「ｂｂ重工業〈７１９２〉」、「ｃｃ工業〈７１９３〉」の組み合わせが抽出される。なお、社名の横に示されている各数字は、架空の証券コードである。
【０２２７】
上記の場合、グループ名は「防衛」であるが、一般ニュースで報じられる内容は、図２９の各経済ニュースの前半部分の「政府が自衛隊機の民間転用を進める方針を固めた」や「朝鮮半島情勢緊迫化」であり、グループ名「防衛」は記載されない可能性が高い。
【０２２８】
そこで、第５の実施形態では、図１２の知見情報に加えて、図３２に示す関連語情報１１６を用いる。図３２に示すように、本実施形態の関連語情報１１６は、知見Ｎｏ．とグループ名（第１概念名）と関連語とを対応付けたテーブルとして記憶部１１に保持される。なお、関連語はあらかじめユーザが登録してもよいし、ニュース情報にグループ名と同時にでてくる単語を関連語として抽出して登録してもよい。
【０２２９】
ここで、図３３を参照して、第５の実施形態の知見抽出装置１０１の構成について説明する。第１の実施形態乃至第４の実施形態と同じ図番が付与されている各構成は、第１の実施形態乃至第４の実施形態と同じ動作を行なうものとする。
【０２３０】
図３３に示すように、第５の実施形態の知見抽出装置１０１は、第４の実施形態の知見抽出装置１００に加えて、記憶部１１に関連語情報１１６を保持する。なお、本実施形態の知見抽出装置１０１は、第１乃至第３の実施形態の知見抽出装置１に関連語情報１１６を加えた構成としてもよい。
【０２３１】
本実施形態の知見抽出装置１０１の処理の一例について、図１４、図２４及び図２５を参照して説明する。
【０２３２】
本実施形態の知見抽出装置１０１は、図１４もしくは図２４及び図２５に示す処理において、ニュース情報に含まれるグループ名の有無を判定する際に（図１４のステップＳ２４、もしくは図２４のステップＳ５２）、図３２に示す関連語情報１１６に含まれる関連語をグループ名と同等の表現としてチェック対象に加える。すなわち本実施形態の知見抽出装置１０１は、ステップＳ５１で選択されたグループ名と同一のグループ名が関連語情報１１６に保持されているか検索し、保持されている場合、このグループ名もしくはこのグループ名に対応する関連語がニュース情報に含まれるか否かを判定する。
【０２３３】
このとき、１つのニュースにグループ名と関連語が記載されている場合は、グループ名が記載されたニュース１件相当と数える。関連語は、第１の実施形態で説明した株価変動材料から、他のグループには登録されていない表現を選んでもよい。また、グループ名を含むニュース群と他のニュース群で出現頻度に有意な差のある名詞句を抽出してもよい。さらに、図１２の知見情報と図２２の知見情報と図３２の関連語情報１１６は、グループ名で結合できる範囲で、別の分割形態で保持してもよい。
【０２３４】
本実施形態によれば、一般ニュースでは記載されることが少ないグループ名に加えて関連語を抽出対象とすることで、一般ニュースの銘柄への影響をさらに適切に評価することができる。
【０２３５】
（第６の実施形態）

次に、第６の実施形態における知見抽出装置について説明する。第６の実施形態における知見抽出装置は、第３の実施形態と同様に、電子文書を受信する際に、当該電子文書にタイムスタンプを付加する機能を更に備え、知見抽出部１４が抽出する知見情報をある対象に関する材料（イベント）とそのタイムスタンプとする。

【０２３６】

図３４を参照して、第６の実施形態の知見抽出装置１０２の構成について説明する。なお、上述の実施形態に記載の構成については省略する。

【０２３７】

図３４に示すように、本実施形態の知見抽出装置１０２は知見評価更新装置１９を備える。

【０２３８】

知見評価更新装置１９は、図２１に示した知見抽出装置１００の知見評価更新装置１８における、上位概念チェック部１８１に変えて電子文書検索部１８６を備え、下位概念抽出部１８２に変えて知見登録概念抽出部１８７を備える。

【０２３９】

電子文書検索部１８６は、知見情報に含まれる第１概念名に、あらかじめ設定した期間に記憶部１１に新規登録された電子文書中で、当該第１概念名が記載された電子文書の有無と配信数をチェックする。

【０２４０】

知見登録概念抽出部１８７は、電子文書検索部１８６によってチェックされた第１概念名と関連付けられた、知見情報に含まれる知見抽出対象語を全て抽出する。また、知見登録概念抽出部１８７は、電子文書検索部１８６がチェックした電子文書から当該知見抽出対象語が記載された電子文書の有無と配信数をチェックする。

【０２４１】

第６の実施形態においては、電子文書をインターネットで配信されるマイクロブログ（ミニブログ、つぶやきブログとも称し、Ｔｗｉｔｔｅｒ（登録商標）が代表例）の記事とし、第１概念をテレビ番組、知見抽出対象語をテレビ番組もしくはテレビ番組の登場人物もしくはテレビ番組の中のコーナーもしくはテレビ番組中で取り上げられる商品などの番組中に現れるもの、第２概念をテレビ番組もしくは知見抽出対象語に対する評価表現とする。すなわち、本実施形態では図２１に示した第４の実施形態における知見抽出装置１００のようにニュース情報から株取引に関する知見情報を抽出するのではなく、電子文書であるマイクロブログの記事のうち、テレビ番組に言及する内容のものを受信した際に、例えば、ドラマＡに関連する知見抽出対象語（登場人物、取り上げられるもの、ストーリー展開、スタッフなど）をグルーピングする知識や、ドラマＡもしくはその知見抽出対象語に関してよく使われる評価表現「ワロタ」「泣けた」「ぱねぇ」の知識を獲得し、これらの知識を知見情報として抽出する知見抽出装置１０２について説明する。

【０２４２】

なお、「ワロタ」とは「笑った」、もしくは「笑える」という意味で使用される語句であり、「ぱねぇ」とは、「半端なことではない」の意味で使用される語句である。また、知見抽出装置１０２は、マイクロブログを含むインターネット配信の電子文書、例えばニュース、一般ブログ、テレビ局発表情報を受信した際は、保持した知見情報を参照し、電子文書に「ドラマＡ」などの番組名、該当番組に所属する知見抽出対象語、該当番組もしくは知見抽出対象語の評価表現の有無の情報を付与し、番組制作者などのユーザの端末装置に提供可能となっている。

【０２４３】

すなわち、本実施形態の知見抽出装置１０２は、第４の実施形態に示した知見抽出装置１０１の変形例である。

【０２４４】

図３５に、本実施形態における第１手掛かり語辞書１１１の一例を示す。図３５に示すように、第１手掛かり語を識別するコードと、第１概念名、本実施形態においては番組名

を定義する第１手掛かり語とが、関連付けられて記述されている。

【０２４５】

第１手掛かり語辞書１１１は、本実施形態においては、図３５に一例を示すように、第１手掛かり語を識別するコードと、第１概念名、本実施形態においては番組名を定義する第１手掛かり語とが、関連付けられて記述されている。具体的には、番組名などの記事内容を定義する第１手掛かり語は、記事内容の先頭文字の「＃」であるとする。

【０２４６】

本実施形態における知見抽出対象語辞書１１２は、図３６に一例を示すように、知見抽出対象語、本実施形態においてはテレビ番組中に現れるものについて、知見抽出対象語を識別するコードと、知見抽出対象語が関連付けられた番組名と、知見抽出対象語のマイクロブログ記事中に現れる別表記である知見抽出対象語表現とが関連付けられて記述されている。知見抽出対象語表現としては、知見抽出対象語名の正称、略称、通称、俗称、愛称などの様々な名称が使用可能となっている。知見抽出対象語辞書１１２は、まず、テレビ局の公開情報に基づいて構築してもよい。その際、他の番組にも登場する出演者に関しては番組名を特定せず、番組に固有の役名などについては番組名を記載するとしてもよい。さらに、知見抽出対象語辞書は、知見情報を抽出する際に、第１概念が記載されたマイクロブログ記事に現れる別の名詞もしくは複合名詞を知見抽出対象語候補として記録し、ユーザの確認後に知見抽出対象語として追加登録するものとしてもよい。

【０２４７】

第２手掛かり語辞書１１３ａは、図３７に一例を示すように、第２手掛かり語を識別するコードと、第２概念、すなわち本実施形態においては知見抽出対象語表現に対す第２手掛かり語とが関連付けられて記述されている。本実施形態においては、第２手掛かり語は感情表現である顔文字およびそれに類する表現とする。第２手掛かり語辞書１１３ａは、第２の実施形態と同様に、感情の方向と、その感情の強度が付与された形態とする。なお、強度は、感情が強いほど大きな値で示される。

【０２４８】

図３８に、本実施形態で情報受信部１２が受信する電子文書であるマイクロブログの記事の一例を示す。図３８に示すように、本実施形態では、インターネットで配信されるマイクロブログ記事の構成要素のうち、図３８に示す「タイムスタンプ」「ユーザ」「引用情報」「本文」の４項目と、情報受信部１２が付与する「記事ＩＤ」を利用する。

【０２４９】

「タイムスタンプ」は、配信されたマイクロブログ記事に付与された発信日時情報である。「ユーザ」は、その情報を発信したユーザの識別名称である。「引用情報」は、他のユーザが発信したマイクロブログ記事をそのまま引用し再発信する際などに付与される、元の記事の発信ユーザの識別名称である。「記事本文」は配信されたマイクロブログ記事の本文の文字列である。

【０２５０】

情報選択部１３は、情報受信部１２が受信したマイクロブログの記事から、知見抽出対象とする電子文書、すなわち本実施形態ではテレビ番組に関するマイクロブログ記事を選択する。テレビ番組に関するマイクロブログであることの判断は、例えば、マイクロブログの記事内容の先頭に第１手掛かり語「＃」に続いてテレビ番組名もしくはテレビ局名が記載されていることで判断する。第１手掛かり語がないマイクロブログ記事の場合は、本文に、テレビ番組名、テレビ局名、もしくは図３６の知見抽出対象語辞書１１２に記載された知見抽出対象語のいずれかが含まれることにより判断する。

【０２５１】

次に、以上のように構成された知見抽出装置１０２の動作について説明する。図３９は知見抽出装置１０２における知見抽出部１４の処理動作の詳細を示す模式図である。知見抽出部の処理動作は、第１の実施形態と同様に、図９のフローチャートに従う。

【０２５２】

なお、以下の説明では、知見抽出部１４の処理を、電子文書を受信する都度としている

が、この処理は夜間などの情報分析処理を行わない時間帯に、受信済みのマイクロブログの記事に対してまとめて実施してもよい。

【０２５３】

情報受信部１２が、マイクロブログの掲載されたサイトからの配信を受けるか、マイクロブログの掲載されたサイトにアクセスすることにより、新規のマイクロブログを受信する。例えば、図３８に示す記事ＩＤ“３”に該当するマイクロブログを受信する。

【０２５４】

情報選択部１３は、情報受信部１２が受信したマイクロブログの記事から、テレビ番組に関するマイクロブログ記事を選択する（ステップＳ１）。テレビ番組に関するマイクロブログであることの判断は、例えば、記事内容の先頭文字「＃」に続いてにテレビ番組名もしくはテレビ局名が記載されていることで判断する。記事内容の先頭文字「＃」がないマイクロブログ記事の場合は、本文に、テレビ番組名、テレビ局名、もしくは図３６の知見抽出対象語辞書に記載された知見抽出対象語のいずれかが含まれることにより判断する。選択されたマイクロブログ記事は、情報選択部１３から知見抽出部１４に送出される。

【０２５５】

知見抽出部１４においては、知見抽出判定語チェック部１４１が、情報選択部１３に選択されたマイクロブログ記事が知見抽出判定語を含むか否かを判定する（ステップＳ２）。本実施形態の場合は、ステップＳ１の処理とステップＳ２の処理の判断基準を同一とし、いずれかの処理ステップのみ実施するものとしてもよい。

【０２５６】

知見抽出判定語が含まれる場合は、知見情報が含まれるマイクロブログ記事として、マイクロブログ記事を第１手掛かり語チェック部１４２に送信する。

【０２５７】

第１手掛かり語チェック部１４２は、ステップＳ２の判定の結果、マイクロブログ記事が知見抽出判定語を含むとき、当該マイクロブログ記事が第１概念名、すなわち、本実施形態においては番組名を定義する第１手掛かり語を含むか否かを判定する（ステップＳ３）。

【０２５８】

例えば、第１手掛かり語チェック部１４２は、第１手掛かり語辞書１１１を参照し、受信したマイクロブログ記事の本文に第１手掛かり語の少なくとも一つが含まれるか否かを判定する。例えば図３８に示す記事ＩＤ“３”のニュースは、第１手掛かり語「＃」を含む。

【０２５９】

受信したマイクロブログ記事に第１手掛かり語が含まれない場合は第１概念抽出部１４３によるステップＳ４の処理を省略する。

【０２６０】

受信したマイクロブログ記事に第１手掛かり語が含まれる場合は、第１手掛かり語チェック部１４２は、マイクロブログ記事を第１概念抽出部１４３に送信する。

【０２６１】

第１概念抽出部１４３は、ステップＳ３の判定の結果、マイクロブログ記事が第１手掛かり語を含むとき、当該第１手掛かり語に基づいて当該マイクロブログ記事から番組名を抽出する（ステップＳ４）。

【０２６２】

例えば、第１概念抽出部１４３は、第１手掛かり語を含むマイクロブログの記事を受信すると、第１手掛かり語直後の名詞もしくは複合名詞を番組名として抽出する。受信したマイクロブログの記事が例えば図３８に示すニュースＩＤ“３”の場合は、第１手掛かり語「＃」の直後の複合名詞「ドラマＡ」を番組名として抽出する。

【０２６３】

本実施形態の場合は、例えば、番組名として手掛かり語直後の名詞もしくは複合名詞を抽出した後で、既存の番組名の一覧（図示せず）と比較し、抽出した名詞もしくは複合名

詞が番組名の正式名称もしくは略称である場合のみ番組名として採用し、それ以外の場合は、番組名を含まないと判断してもよい。

【０２６４】

一方、知見抽出対象語抽出部１４４は、知見抽出判定語チェック部１４１による判定の結果、マイクロブログの記事が知見抽出判定語を含むとき、当該マイクロブログ記事から知見抽出対象語名を抽出する。

【０２６５】

具体的には、知見抽出対象語抽出部１４４は、例えば、ステップＳ３又はＳ４の後、第１概念抽出部１４３の処理を省略して送信されたマイクロブログ記事に対して、知見抽出対象語辞書１１２を参照し、マイクロブログ記事本文に知見抽出対象語表現が含まれるか否かを判定し（ステップＳ５）、知見抽出対象語表現が含まれる場合に当該知見抽出対象語表現に対応する知見抽出対象語を抽出する（ステップＳ６）。

【０２６６】

マイクロブログ記事本文に、知見抽出対象語辞書１１２の「知見抽出対象語表現」列の表現がある場合は、その行の「知見抽出対象語名」に記された知見抽出対象語が含まれると判断する。

【０２６７】

知見抽出対象語辞書１１２は、一つの知見抽出対象語に対して複数行の定義もしくは知見抽出対象語表現列に複数の表現を含んでよい。また、１件のマイクロブログ記事から複数の知見抽出対象語表現を抽出してよい。

【０２６８】

例えば図３８の記事ＩＤ“３”のマイクロブログ記事の場合、知見抽出対象語表現「ｍさん」が含まれるため、知見抽出対象語「コード　Ｂ００１　知見抽出対象語名　役名Ｍ」が抽出される。

【０２６９】

第２概念抽出部１４５は、知見抽出判定語チェック部１４１による判定の結果、マイクロブログ記事が知見抽出判定語を含むとき、当該マイクロブログ記事から第２概念、すなわち、本実施形態においては評価表現を抽出する。具体的には、第２概念抽出部１４５は、知見抽出判定語を含むマイクロブログ記事本文に、第２手掛かり語辞書１１３ａ内の手掛かり語の一つ以上が含まれるか否かを判定する（ステップＳ７）。

【０２７０】

ステップＳ７およびステップＳ８の処理は、図９のフローチャートではステップＳ６の後としているが、第１の実施形態と同様に、ステップＳ３の処理の前、もしくはステップＳ３からステップＳ６の処理と並列で実施してもよい。

【０２７１】

第２概念抽出部１４５は、受信したマイクロブログ記事の見出しもしくは本文に、第２手掛かり語辞書１１３ａ内の第２手掛かり語が含まれる場合、その手掛かり語と共起している語句を評価表現として抽出する（ステップＳ８）。

【０２７２】

例えば、図３８に示す記事ＩＤ“３”の本文では、「ｍさん万能すぎワロタｗｗ＃ドラマＡ」に第２手掛かり語「ｗｗ」が含まれる。

【０２７３】

第２手掛かり語「ｗｗ」と共起している語句、例えば直前に置かれた語句は「ワロタ」（一般辞書にはない未知語）である。この未知語「ワロタ」が評価表現として抽出される。

【０２７４】

評価表現は、処理中の１件の記事で手掛かり語直前の語句とせずに、同じ第２手掛かり語を含む多数のマイクロブログ記事から全ての単語を抽出し、該当の第２手掛かり語を含まない記事の集合と比較して統計的に共起する確率が高い語句を評価表現とするとしてもよい。例えば図３８の記事ＩＤ７と記事ＩＤ８は、異なる第２手掛かり語「ｗｗ」および

「（；＿；）の直前の語句が「ヤバい」で一致し、どちらの第２手掛かり語と共起する確率も同じなので、評価表現としない。

【０２７５】

抽出された評価表現「ワロタ」には、第２手掛かり語辞書１１３ａに登録された第２手掛かり語「ｗｗ」に対応づけられた方向及び強度を付与し、評価表現「評価表現：ワロタ　方向：笑い　強度：１」が得られる。

【０２７６】

ステップＳ３からステップＳ８の処理の終了後、組み合わせ作成部１４６は、処理中のマイクロブログ記事から抽出された番組名、知見抽出対象語及び評価表現を組み合わせて知見情報を作成し、当該知見情報を記憶部に書き込む（ステップＳ９）。ここで、知見情報は、方向及び強度が付与された評価表現「評価表現：ワロタ　方向：笑い　強度：１」を含んでいる。評価表現は、第１の実施形態と同様に、方向や強度が付与されない形式を選択してもよい。

【０２７７】

マイクロブログ記事から抽出された番組名が１つである場合は、その番組名に、抽出された全ての知見抽出対象語、抽出された全ての評価表現を割り付ける。すなわち、１つの番組名に対して、知見抽出対象語と評価表現はそれぞれ１つまたは複数が割り付けられることを許容する。抽出された番組名が複数である場合は、第１の実施形態と同様に、文脈解析や簡易的な判別処理により、番組名、番組名に係り受けしている知見抽出対象語、番組名もしくはその番組名に係り受けしている知見抽出対象語である銘柄に係り受けしている評価表現を判別して組み合せてもよい。さらに、知見情報は番組名に対して知見抽出対象語と評価表現が割り付けられるとしているが、知見抽出対象語もしくは評価表現がない組み合せが作成されてもよい。

【０２７８】

このような処理により、番組名、知見抽出対象語名、評価表現のセットとして以下の知見が抽出され、記憶部１１に書き込まれる。

【０２７９】

番組名　：　ドラマＡ
番組に属する知見抽出対象語　：　役名Ｍ
評価表現　：　ワロタ　　方向　：　笑い　　強度　：　１

組み合せは、番組名、知見抽出対象語、評価表現をセットにするとしたが、第１の実施形態と同様に、組み合わせは番組名と知見抽出対象語のみとし、評価表現は単独で知見情報として保持するとしてもよい。または、評価表現は番組名ではなく知見抽出対象語とセットにして知見情報として保持するとしてもよい。

【０２８０】

ステップＳ２で知見抽出判定語がないと判定されたマイクロブログ本文は、第１概念チェック部１４７により、マイクロブログ記事本文に、記憶部１１に書き込まれた知見情報の番組名を含むか否かを判定される（ステップＳ１１）。

【０２８１】

番組名を含む場合は、知見抽出対象語抽出部１４４が知見抽出対象語辞書１１２を用いて、マイクロブログ本文に知見抽出対象語表現が含まれるか否かを判定する（ステップＳ１２）。

【０２８２】

知見抽出対象語表現が含まれる場合、例えば「採用」「登場」などの表現がある場合は、組み合せ作成部１４６が抽出された知見抽出対象語表現に対応する知見抽出対象語を番組に追加する処理のための組み合せを作成する。

【０２８３】

「降板」などの表現がある場合は、組み合わせ作成部１４６が、番組に対して抽出された知見抽出対象語表現に対応する知見抽出対象語を取り除く処理のための組合せを作成する（ステップＳ１３）。

【０２８４】
ステップＳ１３で番組と知見抽出対象語の追加もしくは削除処理の組み合わせが作成された場合は、その処理を記憶部１１に対して実行する（ステップＳ９）。
【０２８５】
なお、知見抽出対象語の抽出は、知見抽出対象語辞書１１２を用いずに、番組名を含むマイクロブログ記事本文から名詞句を抽出し知見抽出対象語とするとしてもよい。この場合、抽出した名詞句は、新たに知見抽出対象語として知見抽出対象語辞書１１２に追加するものとしてもよい。さらに、ステップＳ１１からＳ１３の処理と判断は、１件のマイクロブログのみで行わずに、例えばある規定の数を超える発信があった場合にのみ記憶部１１に対する実行対象としてもよい。または、ステップＳ１１からＳ１３までの処理は、マイクロブログ本文に対して行うのではなく、テレビ局やニュース配信業者から配信された電子文書のみに対して実行するものとしてもよい。
【０２８６】
次に、情報分析部１５の動作を図３９、図１４及び図１５によって説明する。
【０２８７】
始めに、情報受信部１２は、マイクロブログの掲載されたサイトからの配信を受けるか、マイクロブログの掲載されたサイトにアクセスすることにより、新規のマイクロブログ記事を受信する。例えば、図３８の記事ＩＤ“６”に該当するマイクロブログ記事を受信する（ステップＳ２１）。受信したマイクロブログ記事は知見抽出対象語・第１概念・第２概念抽出部１５１に送信される。
【０２８８】
知見抽出対象語・第１概念・第２概念抽出部１５１は、まず、記憶部１１に記憶された知見抽出対象語辞書１１２を参照し、受信したマイクロブログ記事に知見抽出対象語表現があるか否かを判定する（ステップＳ２２）。
【０２８９】
知見抽出対象語表現がある場合、知見抽出対象語・第１概念・第２概念抽出部１５１は、抽出された知見抽出対象語表現に対応する知見抽出対象語名・コードを抽出する。また、知見抽出対象語・第１概念・第２概念抽出部１５１は、図３６に示す知見抽出対象語辞書１１２を知見情報として参照し、抽出した知見抽出対象語名が割り付けられた第１概念、すなわち本実施形態においては番組名を全て抽出する（ステップＳ２３）。知見抽出対象語・第１概念・第２概念抽出部１５１及び電子文書割付部１５２は、マイクロブログ記事本文から抽出された知見抽出対象語名ごとに、その知見抽出対象語が属する各番組に対して、ステップＳ２４・ステップＳ２５の処理を行う。
【０２９０】
始めに、知見抽出対象語・第１概念・第２概念抽出部１５１は、これに先立つ処理によって抽出された、図４０中に示す知見情報を参照し、その時点で処理中の知見抽出対象語名・その知見抽出対象語が属する番組名に割り付けられた第２概念、すなわち本実施形態においては評価表現を抽出し、マイクロブログ本文中に該当の評価表現があるか否かを判定する（ステップＳ２４）。
【０２９１】
該当の評価表現がある場合は、処理中の受信マイクロブログ記事にその番組に関する情報と評価があると判断し、電子文書割付部１５２が、処理中のマイクロブログ記事を、知見抽出対象語及び番組に割り付ける（ステップＳ２５）。本実施形態においては、第１の実施形態とは異なり、番組名に属する全ての知見抽出対象語への展開はしないとしてもよい。
【０２９２】
知見抽出対象語・第１概念・第２概念抽出部１５１は、ステップＳ２２からステップＳ２５の処理の後、もしくは、前、または並列で、受信したマイクロブログ記事本文に、図１２に示す知見情報内の第１概念名が含まれるか否かを判定する（ステップＳ２６）。
【０２９３】

例えば、図３８に示す記事ＩＤ“６”のマイクロブログ記事では、本文に番組名は含まれていない。この場合は、知見抽出対象語・第１概念・第２概念抽出部１５１は、抽出された知見抽出対象語「役名Ｍ」が図３６の知見抽出対象語辞書１２で関連付けられた番組名「ドラマＡ」を抽出して電子文書割付部１５２に送出する。

【０２９４】

第１概念名が抽出された場合、電子文書割付部１５２は、図１２に示す知見情報でその番組名に対応付けられた知見抽出対象語を取り出し、処理中のマイクロブログ記事を該当銘柄全てに割り付けることもできる（ステップＳ２７）。本実施形態ではこの処理を省略し、マイクロブログ記事は、本文から抽出された知見抽出対象語のみに割り付けるものとする。

【０２９５】

ステップＳ２７までの処理でマイクロブログ記事を割り付けられた知見抽出対象語ごとに、知見抽出対象別判定部１５３は、以下に記載するステップＳ２８からステップＳ３４の処理を行う。

【０２９６】

ステップＳ２８では、ステップＳ２４の第２概念、すなわち本実施形態では評価表現抽出結果を参照し、現在処理中の知見抽出対象語が属する番組の評価表現が抽出されているか否かを判定する。

【０２９７】

処理中の知見抽出対象語で、処理中の所属番組に対応付けられた評価表現が抽出されている場合は、その知見抽出対象語もしくは番組への反応が変動する可能性があると判断し、その番組、知見抽出対象語と、評価表現のセットをアラームに追加する（ステップＳ２９）。

【０２９８】

記憶部１１に保持された知見情報が「番組名・知見抽出対象語」と評価表現に分けられている場合は、ステップＳ２８は、現在処理中の知見抽出対象語や番組が属するか否かに関わらず記憶部１１に保持された評価表現全てに関して抽出されているか否かを判定する。

【０２９９】

この場合、ステップＳ２９では、知見抽出対象語と、ステップＳ２８で抽出された評価表現のセットをアラームに追加する。

【０３００】

知見抽出対象語にマイクロブログ記事が割り付けられた後で、処理中の知見抽出対象語もしくは番組に割り付けられ、記憶部１１に記憶された過去一定期間、例えば３日間のマイクロブログ本文を参照して急上昇ワードがあるかを判定し（ステップＳ３０）、処理中の知見抽出対象語もしくは番組のマイクロブログ記事に急上昇ワードが存在する場合は、番組、知見抽出対象語と急上昇ワードのセットを急上昇アラームに追加する（ステップＳ３１）。

【０３０１】

さらに、同様に処理中の知見抽出対象語もしくは番組に割り付けられ、記憶部１１に記憶された過去一定期間、例えば１時間のマイクロブログ記事本文を参照して処理中の受信マイクロブログ記事が過去一定期間のマイクロブログ記事と類似・一致しているかを判定し（ステップＳ３２）、類似マイクロブログ記事もしくは一致するマイクロブログ記事がある場合は、そのマイクロブログ記事対の類似度を求め、多重配信アラームを追加する（ステップＳ３３）。

【０３０２】

以上の処理により出力されたアラーム、急上昇アラーム、多重配信アラームは、情報分析部１５から出力され、ユーザ端末装置２に提供される（ステップＳ３４）。

【０３０３】

上述したように本実施形態によれば、受信されたマイクロブログ記事からテレビ番組、

知見抽出対象語（番組に取り上げられるものなど）、評価表現の知見を抽出して保持することで、一般の辞書にはない表現も知見として活用できる。

【０３０４】

さらに、知見情報が記憶された後、新規に受信された全てのマイクロブログ記事を記憶部１１内の知見情報に基づいて分析する構成により、分析結果を、即時にユーザに提供することで、視聴者の反応の把握を有効に支援することができる。

【０３０５】

さらに、知見抽出部１４は、情報選択部１３により選択されたマイクロブログ記事から知見情報を抽出する前述した機能において、第３の実施形態と同様に、当該選択されたマイクロブログ記事から所定の対象を示す表現を含むイベント情報を抽出すると共に、当該イベント情報が抽出されたマイクロブログ記事から抽出し、当該抽出したイベント情報及びタイムスタンプを知見情報として記憶部１１に書込む機能をもっている。

【０３０６】

ここで、所定の対象を示す表現は、例えば、知見抽出対象語、すなわち本実施形態においては知見抽出対象語辞書１１２に記述された知見抽出対象語表現、又は第１概念、すなわち本実施形態においては番組名を表す表現である。なお、「知見抽出対象語の番組名を定義する第１手掛かり語」は、第１手掛かり語辞書１１１に記述されている。

【０３０７】

知見抽出部１４の処理は、第３の実施形態と同様に、図１８のフローチャートに従う。

【０３０８】

情報受信部１２がマイクロブログ記事を受信すると（ステップＳ４１）、知見抽出部１４は第１手掛かり語辞書１１１及び知見抽出対象語辞書１１２などを参照し、ユーザによって選択されている対象（知見抽出対象語、番組名など）を示す表現が含まれるか否かを判定する（ステップＳ４２）。

【０３０９】

ユーザによって選択されている対象を示す表現が含まれる場合、知見抽出部１４は、その対象を示す表現を含む名詞句を、材料（イベント）として抽出する（ステップＳ４３）。

【０３１０】

知見抽出部１４は、抽出した材料（イベント）を、対象名、受信情報のタイムスタンプと合わせて記憶部１１に追加的に書込む（ステップＳ４４）。

【０３１１】

以下、情報分析部１５は、前述同様に、ステップＳ２１～Ｓ３４の動作を実行する。

【０３１２】

図４１は、知見情報の一例を示す図である。図３８に示すニュースＩＤ“３”，“６”から知見抽出対象語名「役名Ｍ」に関して処理を行った例である。

【０３１３】

抽出する知識はユーザによって選択された対象表現を含む名詞句としたが、この名詞句中から複合名詞を選択したり、固有名詞を選択したり、急上昇ワードを選択して材料（イベント）としてもよい。

【０３１４】

知見抽出対象語名「役名Ｍ」で抽出した材料（イベント）を、図３６の知見抽出対象語辞書１１２にしたがって番組に割り付け、図４１に示す知見情報で「対象」を「役名Ｍ」から「ドラマＡ」に置き換えた情報を合わせて知見情報として書き込んでもよい。

【０３１５】

さらに、図４１に示すように、類似度の高いマイクロブログ記事の配信数を「類似情報数」として合わせて知見情報としてもよい。

【０３１６】

上述したように本実施形態によれば、マイクロブログ記事を受信する際に、マイクロブログ記事にタイムスタンプを付加し、選択されたマイクロブログ記事から所定の対象を示

す表現を含むイベント情報を抽出すると共に、当該イベント情報が抽出されたマイクロブログ記事から抽出し、当該抽出したイベント情報及びタイムスタンプを知見情報として記憶部１１に書込む構成により、第３の実施形態と同様に、ある対象に関する材料（イベント）をタイムスタンプ情報のある時系列で抽出し保持することで、同様の事象が発生した際に、近い将来の予想に有効な知見情報を保持しユーザ端末装置２に提示することができる。

【０３１７】

続いて、知見評価更新装置１９の処理について説明する。

【０３１８】

番組がコメディタッチである場合は「笑い」の評価は前向きととらえることができるが、番組がシリアスなドラマである場合は、「笑い」は失笑であり、低評価であることが推測される。したがって、番組ごとに、評価表現と、番組の評価の高さ、本実施形態では確信度の情報を合わせて保持することが有効である。

【０３１９】

したがって、本実施形態では、図４２に示すように、確信度と確信度更新履歴とを含む確信度情報を加えた知見情報を保持する形態としている。すなわち、本実施形態の確信度とは、第２概念と第１概念の関連する度合いを示す数値である。

【０３２０】

なお、図４２では第１概念である番組名、第２概念である評価表現、知見抽出対象語の情報を保持しているが、番組名もしくは知見抽出対象語のいずれかの情報は保持しないとしてもよい。

【０３２１】

知見評価更新装置１９の処理は、本願の出願時点で未公開である特願２０１２－１５７００の処理に従って行うことができる。本実施形態の場合、数値情報として、例えば番組の視聴率を用いることができる。知見評価更新処理は、マイクロブログ記事の配信が少ない時間帯に１日１回の頻度で実施しても、分析対象としている番組終了から一定時間を経た後にそれぞれ実施してもよい。

【０３２２】

本実施形態によれば、知見を抽出した情報源以外の情報（視聴率など）に基づいて、第１概念と第２概念の組合せの有効性を評価し更新する構成により、知見の確かさや第１概念もしくは知見抽出対象語への影響度を複数の情報源に基づき高く維持できる。

【０３２３】

確信度の情報は、知見情報もしくは第１手掛かり語もしくは第２手掛かり語の関連性の評価更新に用いるものではなく、マイクロブログ記事をはじめとする電子文書に含まれる番組名もしくは知見抽出対象語を数える際に重み付けに用いてもよい。すなわち、例えば番組名「ドラマＡ」に属する知見抽出対象語「登場人物Ｘ」を含むマイクロブログ記事が５０件ある場合、確信度が１．０の場合は５０件、確信度が０．８の場合は４０件のマイクロブログ記事配信があったものとして件数を割り付けるようにしてもよい。または、確信度の低い評価表現の数が多くなった場合は、視聴者からの評価が下がっていると判断する知見としてもよい。

【０３２４】

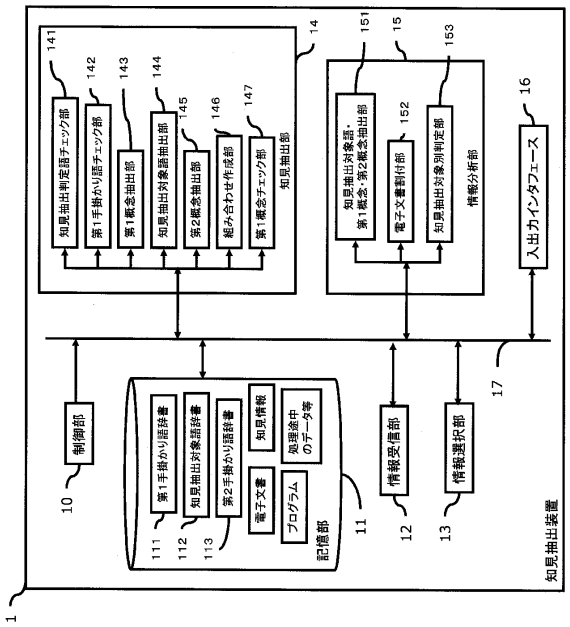
以上説明した少なくとも１つの実施形態の知見抽出装置によれば、知見抽出対象語に関連付けられていない状態で随時配信される電子文書から知見抽出対象語に関する知見情報を随時抽出することが可能である。また、上記実施形態の少なくとも１つの知見抽出装置は、知見情報を最新の状態にすることができる。

【０３２５】

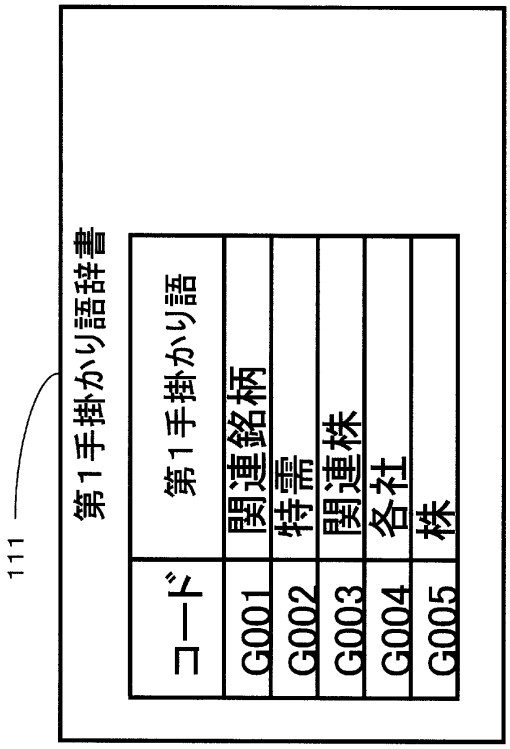
なお、上記の各実施形態に記載した手法は、コンピュータに実行させることのできるプログラムとして、磁気ディスク（フロッピー（登録商標）ディスク、ハードディスクなど）、光ディスク（ＣＤ－ＲＯＭ、ＤＶＤなど）、光磁気ディスク（ＭＯ）、半導体メモリなどの記憶媒体に格納して頒布することもできる。

【０３２６】
また、この記憶媒体としては、プログラムを記憶でき、かつコンピュータが読み取り可能な記憶媒体であれば、その記憶形式は何れの形態であってもよい。
【０３２７】
また、記憶媒体からコンピュータにインストールされたプログラムの指示に基づきコンピュータ上で稼働しているＯＳ（オペレーティングシステム）や、データベース管理ソフト、ネットワークソフト等のＭＷ（ミドルウェア）等が上記実施形態を実現するための各処理の一部を実行してもよい。
【０３２８】
さらに、各実施形態における記憶媒体は、コンピュータと独立した媒体に限らず、ＬＡＮやインターネット等により伝送されたプログラムをダウンロードして記憶または一時記憶した記憶媒体も含まれる。
【０３２９】
また、記憶媒体は１つに限らず、複数の媒体から上記の各実施形態における処理が実行される場合も本発明における記憶媒体に含まれ、媒体構成は何れの構成であってもよい。
【０３３０】
なお、各実施形態におけるコンピュータは、記憶媒体に記憶されたプログラムに基づき、上記の各実施形態における各処理を実行するものであって、パソコン等の１つからなる装置、複数の装置がネットワーク接続されたシステム等の何れの構成であってもよい。
【０３３１】
また、各実施形態におけるコンピュータとは、パソコンに限らず、情報処理機器に含まれる演算処理装置、マイコン等も含み、プログラムによって本発明の機能を実現することが可能な機器、装置を総称している。
【０３３２】
なお、本発明のいくつかの実施形態を説明したが、これらの実施形態は、例として提示したものであり、発明の範囲を限定することは意図していない。これら新規な実施形態は、その他の様々な形態で実施されることが可能であり、発明の要旨を逸脱しない範囲で、種々の省略、置き換え、変更を行うことができる。これら実施形態やその変形は、発明の範囲や要旨に含まれるとともに、特許請求の範囲に記載された発明とその均等の範囲に含まれる。
【符号の説明】
【０３３３】
１，１００，１０１，１０２…知見抽出装置、２…ユーザ端末装置、１０…制御部、１１…記憶部、１２…情報受信部、１３…情報選択部、１４…知見抽出部、１５…情報分析部、１６…入出力インタフェース、１７，２６…バス、２１…端末制御部、２２…端末記憶部、２３…表示部、２４…入力部、２５…通信インタフェース、１１１…第１手掛かり語辞書、１１２…知見抽出対象語辞書、１１３，１１３ａ…第２手掛かり語辞書、１４１…知見抽出判定語チェック部、１４２…第１手掛かり語チェック部、１４３…第１概念抽出部、１４４…知見抽出対象語抽出部、１４５…第２概念抽出部、１４６…組み合わせ作成部、１４７…第１概念チェック部、１５１…知見抽出対象語・第１概念・第２概念抽出部、１５２…電子文書割付部、１５３…知見抽出対象別判定部、１８、１９…知見評価更新装置、１８１…電子文書検索部、１８２…知見登録概念抽出部、１８３…評価部、１８４…確信度更新部、１８５…知見更新部

【図１】



【図２】



111

第１手掛かり語辞書

コード	第１手掛かり語
G001	関連銘柄
G002	特需
G003	関連株
G004	各社
G005	株

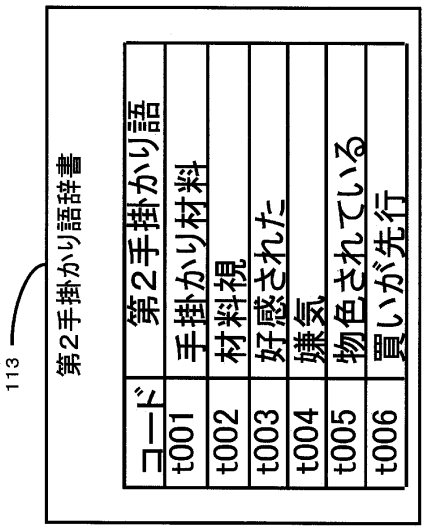
【図３】

112

知見抽出対象語辞書

証券コード	知見抽出対象語（企業名）	東証業種1（大分類）	東証業種2（小分類）	知見抽出対象名表現（企業名表現）
1031	RRボウ	製造業	繊維製品	RRボウ
1035	zzボウ(株)	製造業	繊維製品	zzボウ
1045	イロハニ薬品産業(株)	製造業	医薬品	イロハニ
1045	イロハニ薬品産業(株)	製造業	医薬品	イロハニ薬
1045	イロハニ薬品産業(株)	製造業	医薬品	イロハニ薬品
1045	イロハニ薬品産業(株)	製造業	医薬品	イロハニ薬品産業
1042	ff医薬(株)	製造業	医薬品	ff医薬
1041	ホヘト薬品(株)	製造業	医薬品	ホヘト薬品
1033	yyボウホールディングス(株)	商業	卸売業	yyボHD
1033	yyボウホールディングス(株)	商業	卸売業	yyボウ
1033	yyボウホールディングス(株)	商業	卸売業	yyボウHD
1033	yyボウホールディングス(株)	商業	卸売業	yyボウホールディングス

【図４】

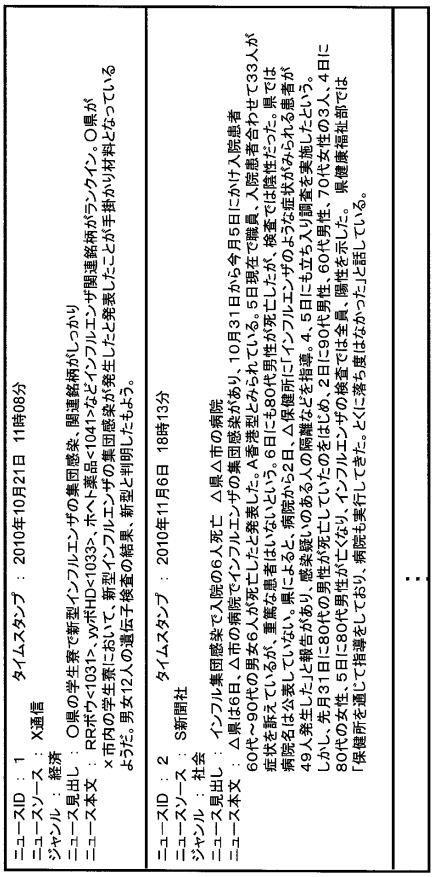


113

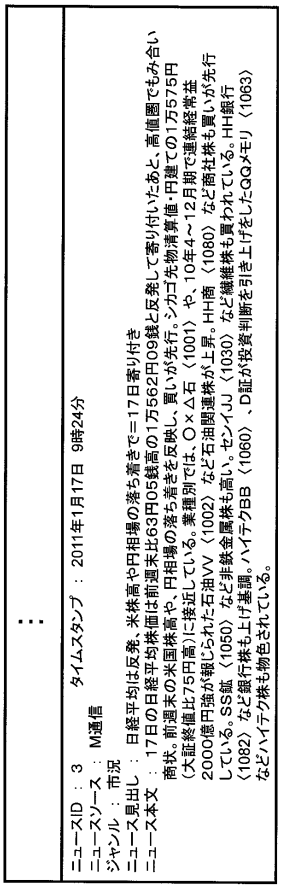
第２手掛かり語辞書

コード	第２手掛かり語
t001	手掛かり材料
t002	材料視
t003	好感された
t004	嫌気
t005	物色されている
t006	買いが先行

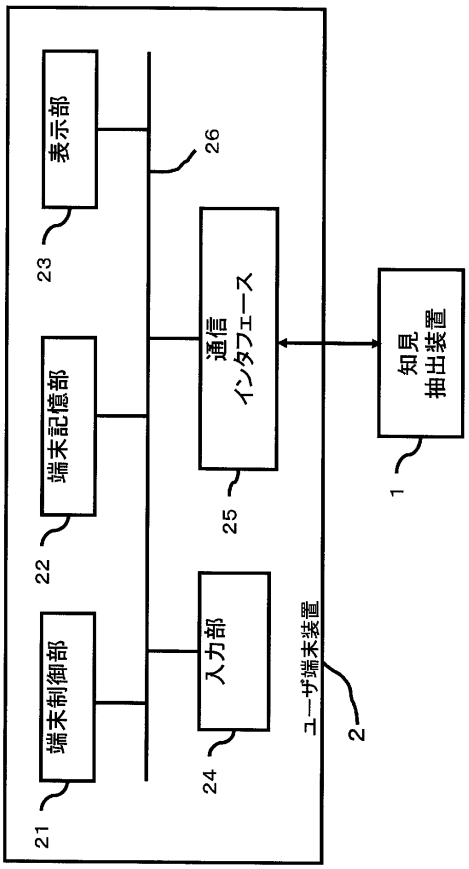
【図５】


ニュースID ： 1　　タイムスタンプ ： 2010年10月21日 11時08分
ニュースソース ： X通信
ジャンル ： 経済
ニュース見出し ： ○県の学生寮で新型インフルエンザの集団感染、関連銘柄がしっかり
ニュース本文 ： RRボウ<1031>、yyボHD<1033>、ホヘト薬品<1041>などインフルエンザ関連銘柄がランクイン。○県が×市内の学生寮において、新型インフルエンザの集団感染が発生したと発表したことが手掛かり材料となっているようだ。男女12人の遺伝子検査の結果、新型と判明したもよう。
ニュースID ： 2　　タイムスタンプ ： 2010年11月6日 18時13分
ニュースソース ： S新聞社
ジャンル ： 社会
ニュース見出し ： インフル集団感染で入院の6人死亡　△県△市の病院
ニュース本文 ： △県は6日、△市の病院でインフルエンザの集団感染があり、10月31日から今月5日にかけ入院患者60代～90代の男女6人が死亡したと発表した。A香港型とみられている。5日現在で職員、入院患者合わせて33人が症状を訴えているが、重篤な患者はいないという。6日にも80代男性が死亡したが、検査では陰性だった。県では病院名は公表していない。県によると、病院から2日、△保健所に「インフルエンザのような症状がみられる患者が49人発生した」と報告があり、感染疑いのある人の隔離などを指導。4、5日にも立ち入り調査を実施したという。
しかし、先月31日に80代の男性が死亡していたのをはじめ、2日に90代男性、60代男性、70代女性の3人、4日に80代の女性、5日に80代男性が亡くなり、インフルエンザの検査では全員、陽性を示した。県健康福祉部では「保健所を通して指導をしており、病院も実行してきた。とくに落ち度はなかった」と話している。
：
：

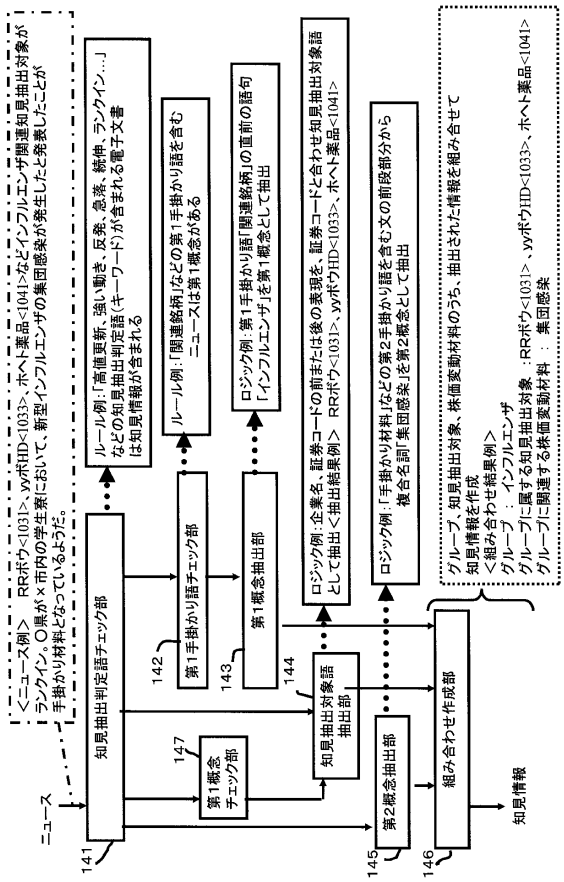
【図６】


：
：
ニュースID ： 3　　タイムスタンプ ： 2011年1月17日 9時24分
ニュースソース ： M通信
ジャンル ： 市況
ニュース見出し ： 日経平均は反発、米株高や円相場の落ち着きで＝17日寄り付き
ニュース本文 ： 17日の日経平均株価は前週末比63円05銭高の1万562円09銭と反発して寄り付いたあと、高値圏でもみ合い商状。前週末の米国株高や、円相場の落ち着きを反映し、買いが先行。シカゴ先物清算値・円建ての1万575円（大証終値比75円高）に接近している。業種別では、○×△石〈1001〉や、10年4～12月期で連結経常益2000億円強が報じられた石油VV〈1002〉など石油関連株が上昇。HH商〈1080〉など商社株も買い が先行している。SS鉱〈1050〉など非鉄金属株も高い。センイJJ〈1030〉など繊維株も買われている。HH銀行〈1082〉など銀行株も上げ基調。ハイテクBB〈1060〉、D証が投資判断を引き上げたQQメモリ〈1063〉などハイテク株も物色されている。

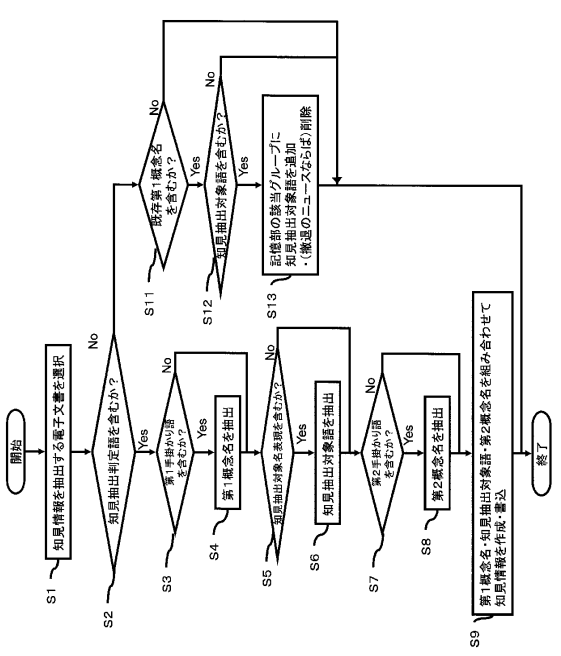
【図７】


端末制御部
21
端末記憶部
22
表示部
23
入力部
24
通信
インタフェース
25
26
ユーザ端末装置
2
知見
抽出装置
1

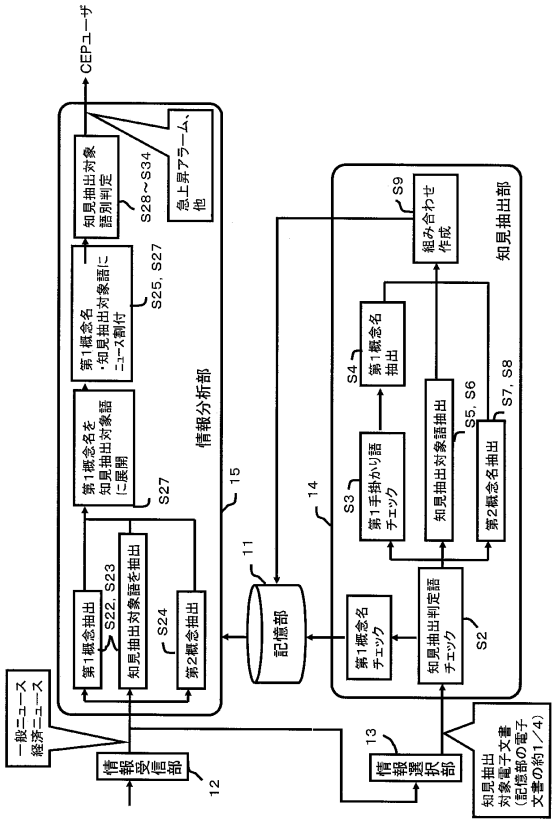
【図８】


ニュース
＜ニュース例＞　RRボウ<1031>、yyボHD<1033>、ホヘト薬品<1041>などインフルエンザ関連知見抽出対象が
ランクイン。○県が×市内の学生寮において、新型インフルエンザの集団感染が発生したと発表したことが
手掛かり材料となっているようだ。
知見抽出判定語チェック部
141
ルール例：「高値更新、強い動き、反発、急落、続伸、ランクイン…」などの知見抽出判定語（キーワード）が含まれる電子文書は知見情報が含まれる
第1手掛かり語チェック部
142
ルール例：「関連銘柄」などの第1手掛かり語を含むニュースは第1概念がある
第1概念抽出部
143
ロジック例：第1手掛かり語「関連銘柄」の直前の語句「インフルエンザ」を第1概念として抽出
144
第1概念
チェック部
147
知見抽出対象語
抽出部
ロジック例：企業名、証券コードの前または後の表現を、証券コードと合わせ知見抽出対象語として抽出＜抽出結果例＞　RRボウ<1031>、yyボウHD<1033>、ホヘト薬品<1041>
第2概念抽出部
145
ロジック例「手掛かり材料」などの第2手掛かり語を含む文の前段部分から複合名詞「集団感染」を第2概念として抽出
組み合わせ作成部
146
グループ、知見抽出対象、株価変動材料のうち、抽出された情報を組み合せて知見情報を作成
＜組み合わせ結果例＞
グループ　：　インフルエンザ
グループに属する知見抽出対象　：RRボウ<1031>、yyボウHD<1033>、ホヘト薬品<1041>
グループに関連する株価変動材料　：　集団感染
知見情報

【図９】



【図１０】



【図１１】

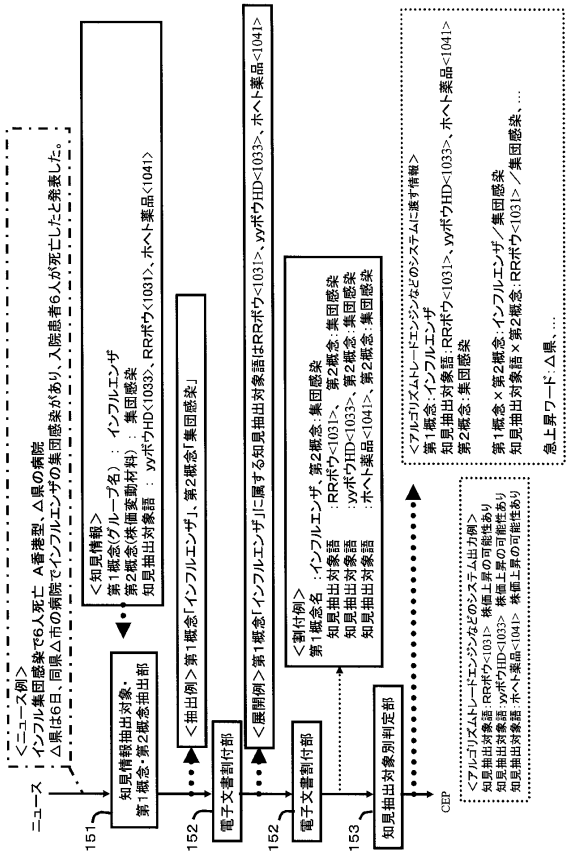
17日の日経平均株価は前週末比63円05銭高の1万562円09銭と反発して寄り付いたあと、高値圏でもみ合い商状。前週末の米国株高や、円相場の落ち着きを*材料1*反映し、買いが先行。シカゴ先物清算値・円建ての1万575円（大証終値比75円高）に接近している。業種別では、○×△石 〈1001〉 や、10年4～12月期で*銘柄1*・*材料2*連結経常益2000億円強が報じられた石油VV 〈1002〉 など石油関連株が上昇。*銘柄2*・*グループ1*HH商 〈1080〉 など商社株も買いが先行している。SS鉱 〈1050〉 など*銘柄3*・*グループ2*・*銘柄4*非鉄金属株も高い。センイJJ 〈1030〉 など繊維株も買われている。HH銀行 〈1082〉*グループ3*・*銘柄5*・*グループ4*・*銘柄6*など銀行株も上げ基調。ハイテクBB 〈1060〉、D証が投資判断を引き上げた*グループ5*・*銘柄7*・*材料3*QQメモリ 〈1063〉 などハイテク株も物色されている。*銘柄8*・*グループ6*

【図１２】

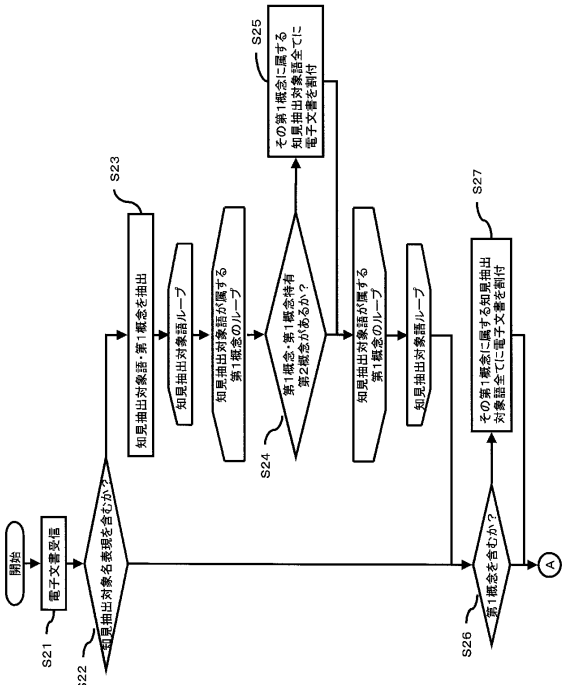
知見情報

知見No.	第1概念名	知見抽出対象語（銘柄）	第2概念（株価変動材料）
1	インフルエンザ	1033:yyボウホールディングス(株); 1031:RRホウ(株); 1041:ホヘト薬品(株)	集団感染
2	石油	1001:○×△石油(株); 1002:石油VVホールディングス(株)	連結経常益2000億円強
3	商社	1080:HH商事(株)	
4	非鉄金属	1050:SS鉱山(株)	
5	繊維	1030:(株)センイJJ	
6	銀行	1082:(株)HH銀行フィナンシャル・グループ	
7	ハイテク	1060:ハイテクニックBB(株); 1063:QQメモリ(株)	投資判断を引き上げ

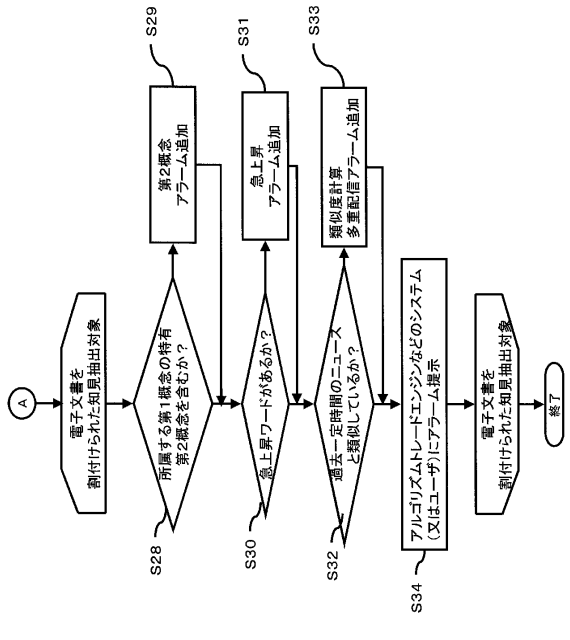
【図１３】



【図１４】



【図１５】



【図１６】

113a

第2手掛かり語辞書

コード	第2手掛かり語	方向	強度
t001	手掛かり材料		
t002	材料視		
t003	好感された	ポジティブ	2
t004	嫌気	ネガティブ	2
t005	物色されている		
t006	買いが先行	ポジティブ	1

【図１７】

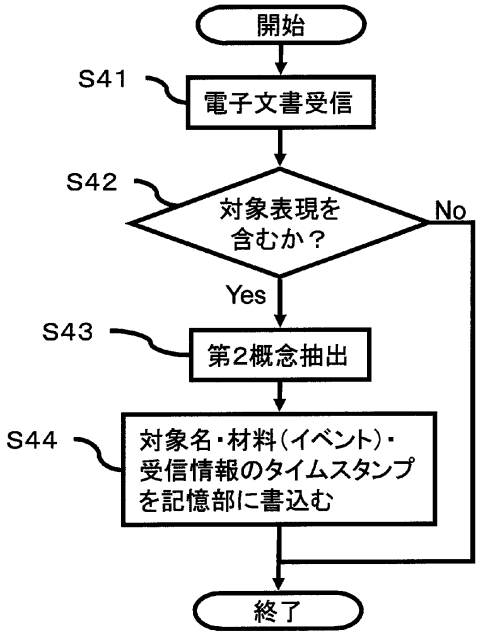
ニュースID：4　　タイムスタンプ：2011年11月17日 9時45分
ニュースソース：X通信
ジャンル：日本株
ニュース見出し：チリヌルヲが大幅安＝業績下方修正を嫌気
ニュース本文：東証マザーズ上場のチリヌルヲ(1123)が前営業日比6400円安の7万3000円と大幅安で推移している。

ニュースID：5　　タイムスタンプ：2011年1月28日 20時04分
ニュースソース：S新聞社
ジャンル：経済総合
ニュース見出し：AA鉱が利益予想を減額、原料価格上昇や価格転嫁の遅れで
ニュース本文：AA製鉄<1152>は28日、2011年3月期の連結経常利益予想を従来予想の2500億円から2200億円に12%下方修正した。原材料価格の上昇に加え、β国の大洪水の影響で追加コスト負担が加わる一方、鋼材価格への転嫁が遅れるため。同社のD山E夫社長は決算会見で、11月頃から上昇し始めた原材料価格は、β国の大洪水の影響で一段高になっており、「1−3月期は10−12月期に比べ相当な減益になる」と語った。同社計画では1−3月期の経常利益は10−12月期に比べ約4割減の317億円となる。主力の製鉄事業の経常利益が同約7割減の139億円に落ち込むことが響く。鋼材価格については「上昇局面には入っているが、当社としては3月までに取れる分がほとんどない」と　述べ、年度内は販売価格の改善が見込めないことも業績下方修正の一因と指摘した。

【図１８】

知見No.	第1概念	知見抽出対象語	株価変動材料	方向	強度
10		1123:チリヌルヲ	業績下方修正	ネガティブ	2

【図１９】

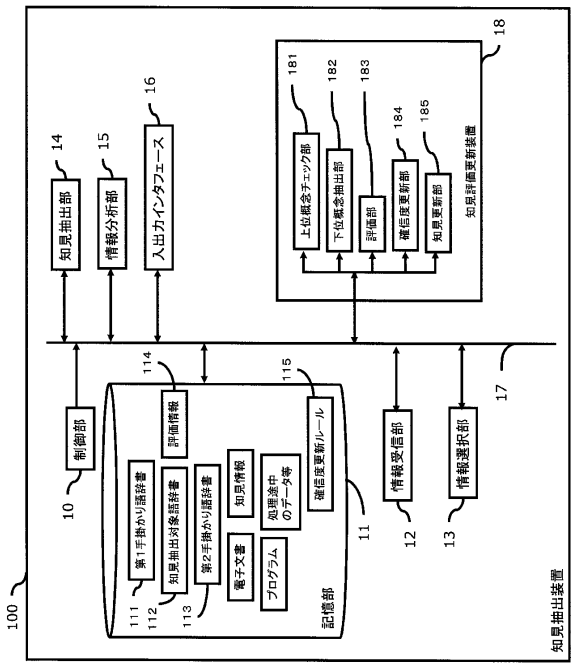


【図２０】

知見情報

対象（第1概念）	第2概念	タイムスタンプ	類似情報数
インフルエンザ	○県の学生寮で新型インフルエンザの集団	2010年10月21日 11時08分	30
インフルエンザ	△市の病院でインフルエンザの集団感染があり、10月31日から今月5日にかけ入院患者	2010年11月6日 18時13分	50

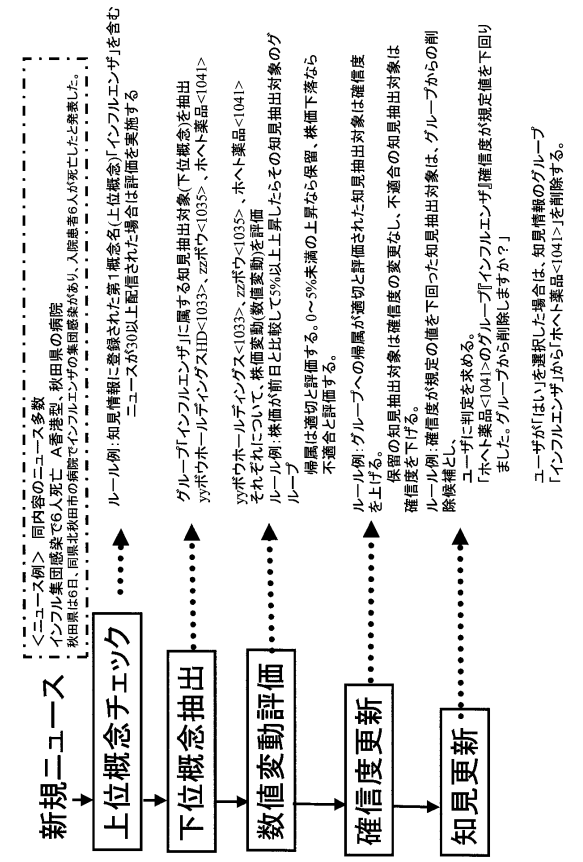
【図２１】



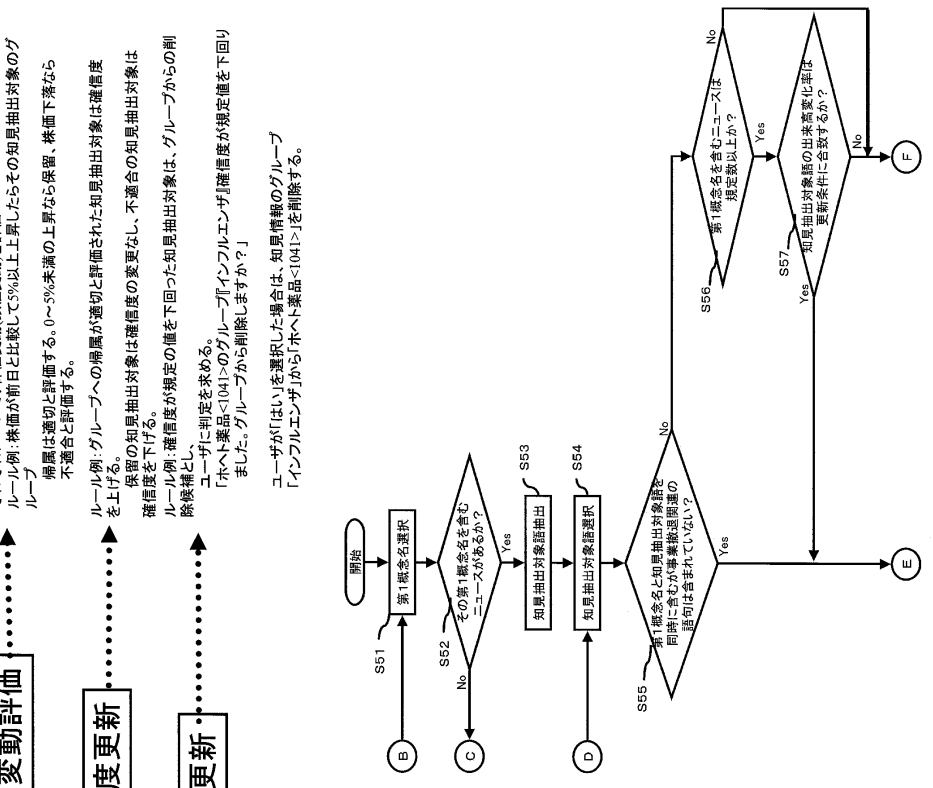
【図２２】

第1概念名	下位グループ名	知見抽出対象(語句)	確信度	事業参入時期	事業撤退時期	確信度更新履歴1	確信度更新履歴2	第1手掛かり語
インフルエンザ	ワクチン	yyポウ<1035>	1.3			2008年1月4日 +0.1	2008年12月7日 +0.1	国産
インフルエンザ	マスク	yyポウ<1035>	1.1			2008年1月4日 +0.1		国産株
インフルエンザ	消毒薬・除菌剤	ホヘト薬品<1041>	0.6		2011年3月31日	2008年1月4日 −0.1		会社
インフルエンザ	インフルエンザ薬	ヒ素薬<1042>	1	2007年10月19日		2008年9月10日 +1		開通
予防接種	消毒薬・除菌剤	ホヘト薬品<1041>	1.2			2008年9月4日 +0.1		経術

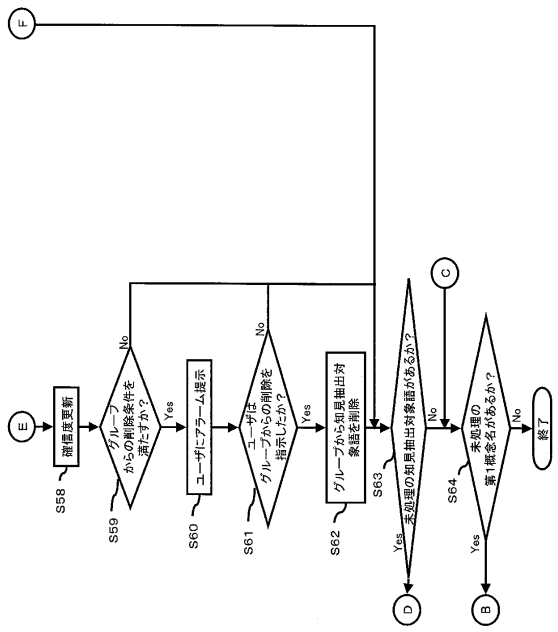
【図２３】



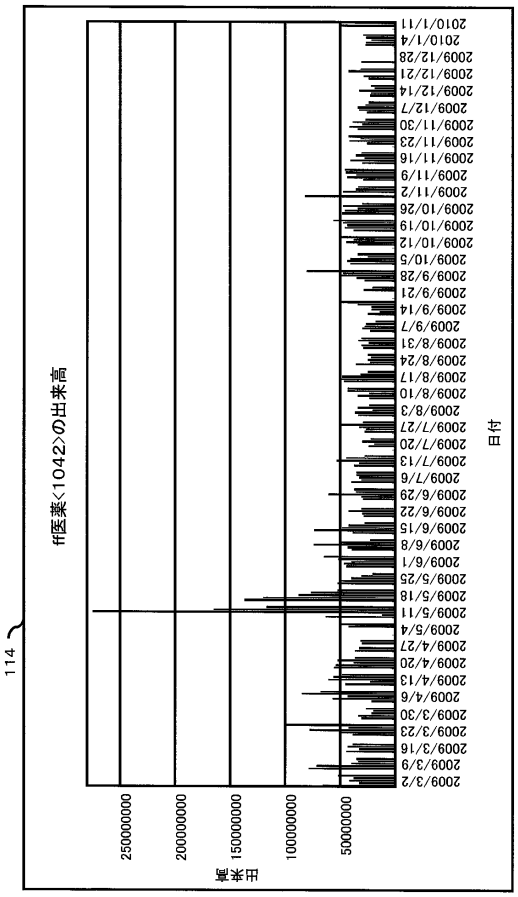
【図２４】



【図２５】



【図２６】

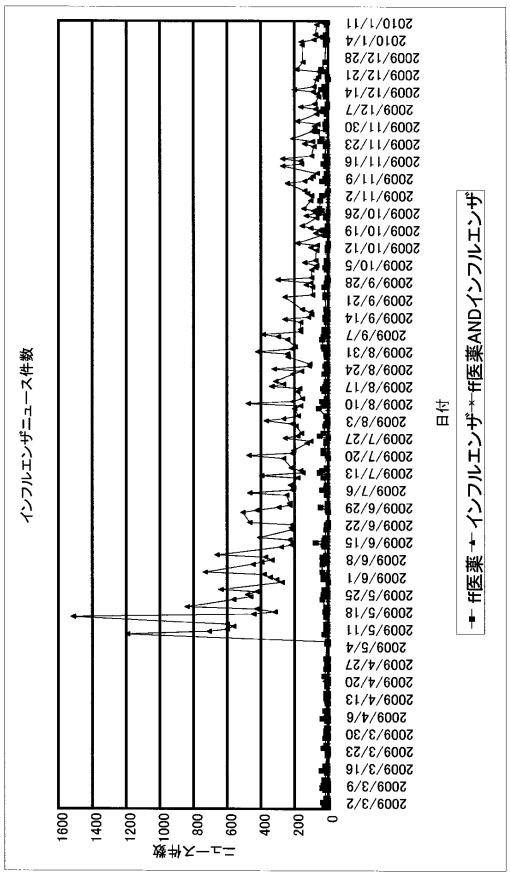


【図２７】

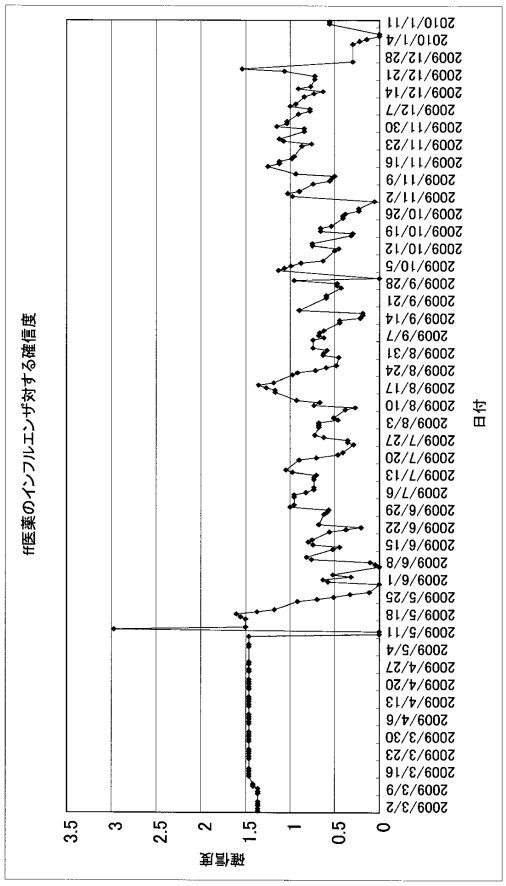
115

1	ステップS55の判定がYes、かつ出来高変化率≦1.1の場合、確信度に0.05を加える
2	ステップS55の判定がYes、かつ出来高変化率＞1.1の場合、 確信度に(出来高変化率－1.0)×0.5を確信度に加える
3	ステップS56の判定がYes、かつ出来高変化率＞1.1の場合、 (ニュース変化率－出来高変化率)×0.5を確信度から減らす
4	ステップS56の判定がYes、かつ出来高変化率＜1.0の場合、 (ニュース変化率－出来高変化率)×0.5を確信度に加える

【図２８】



【図２９】



【図３０】

相関係数	対象とするニュース		
評価期間	銘柄名のみ	銘柄名ORグループ名（確信度なし）	銘柄名ORグループ名（確信度補正あり）
2010/8/31-2011/3/10	0.024	0.011	0.364

【図３１】

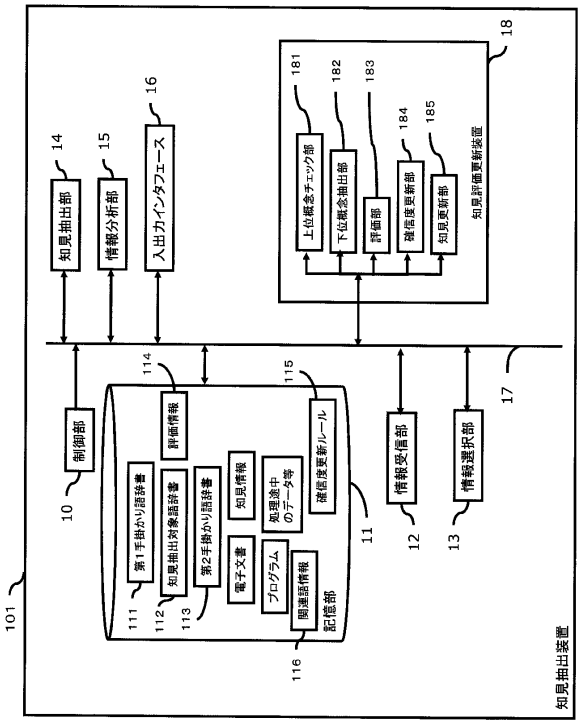
政府が自衛隊機の民間転用を進める方針を固めたと報じられており、 aa重工〈7191〉、bb重工業〈7192〉、cc工業〈7193〉など 防衛関連の一角もしっかり。	朝鮮半島情勢緊迫化で防衛関連株に物色の矛先が向かっている。

【図３２】

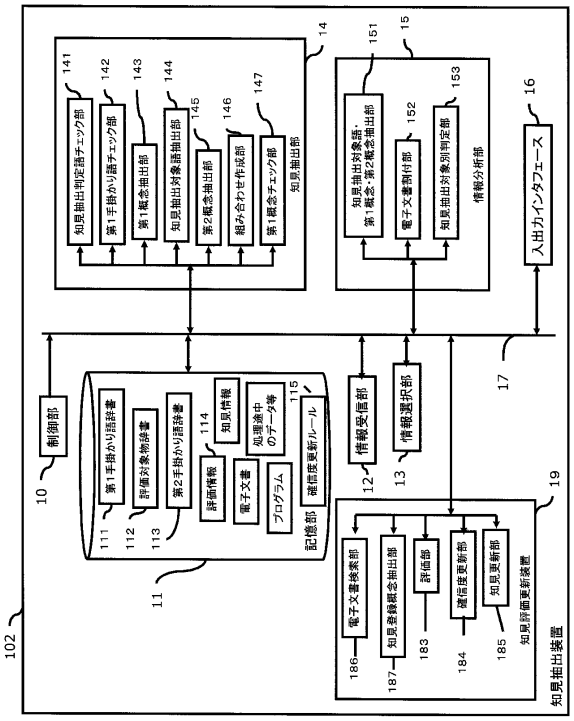
116

知見No.	グループ名	関連語
10	防衛	自衛隊、情勢緊迫化

【図３３】



【図３４】



【図３５】

111

第1手掛かり語辞書

コード	第一手掛かり語
G001	#

【図３６】

112

知見抽出対象語辞書

コード	知見抽出対象語	番組名	知見抽出対象語表現
A001	女優P		女優P、P、p、M
A002	俳優Q		俳優Q、Q、q
A003	役名Mのコート	ドラマA	Mのコート
A004	役名Mのせりふ	ドラマA	Mのせりふ
B001	役名M	ドラマA	mさん
B001	役名M	ドラマA	mさま

【図３７】

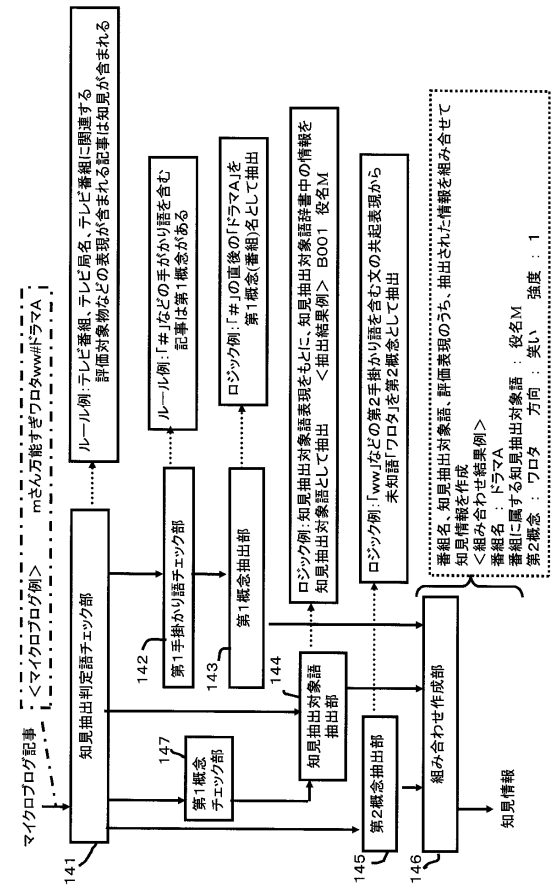
113a 第2手掛かり語辞書

コード	第2手掛かり語	方向	強度
t001	(;_;)	泣き	1
t002	(T_T)	泣き	2
t003	(ToT)	泣き	2
t004	(*_*)	驚き	1
t005	(°□°)	驚き	1
t006	(°口°)	驚き	1
t007	(°Д°)	驚き	1
t008	(^_^;)	焦り	1
t009	(^□^;)	焦り	1
t010	(°Д°;)	焦り	2
t011	(^_^;)	笑い	1
t012	(^∀^)	笑い	2
t013	(^O^)	笑い	2
t014	(^O^)	笑い	2
t015	(^ω^)	笑い	1
t016	(´∀`)	笑い	1
t017	(・∀・)	笑い	1
t018	www	笑い	1
t019	wwww	笑い	2
t020	wwwwww	笑い	3
t021	(ノД`)	怒り	1
t022	(^へ^)	怒り	1

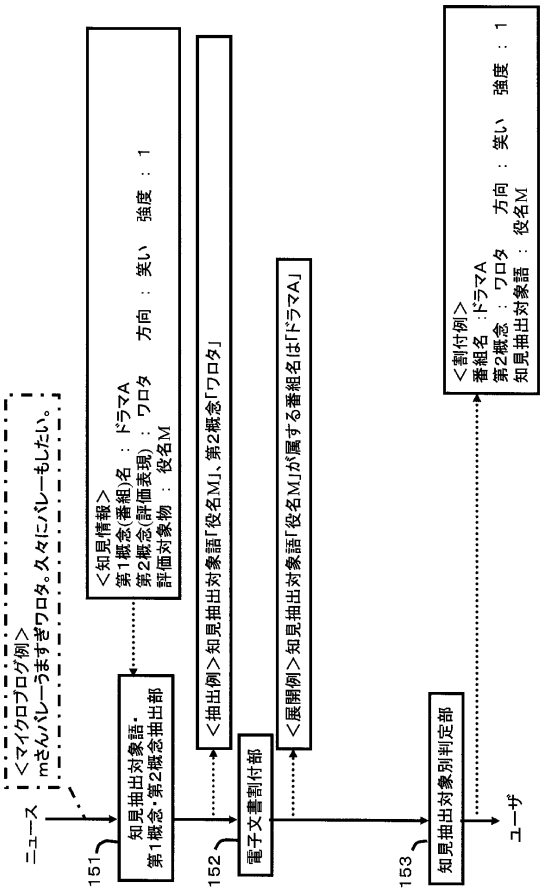
【図３８】

記事ID	タイムスタンプ	ユーザ	引用情報	本文
1	2011/11/28 20:19	aaa		Qの演技シリアスな場面なのに臭すぎてワロタwww#ドラマB
2	2011/11/28 22:11	bbb		mさんスコすぎワロタwww#ドラマA
3	2011/11/28 22:12	ccc		mさん万能すぎワロタww#ドラマA
4	2011/11/28 22:12	ddd	ccc	mさん万能すぎワロタww#ドラマA
5	2011/11/28 22:12	eee		mさん凄すぎワロタwww#ドラマA
6	2011/11/28 22:20	fff		mさんパレーうまずぎワロタ。久々にバレーもしたい。
7	2011/11/28 22:22	ggg		mさんヤバいww#ドラマA
8	2011/11/28 22:34	hhh		mさんヤバい(;;)#ドラマA

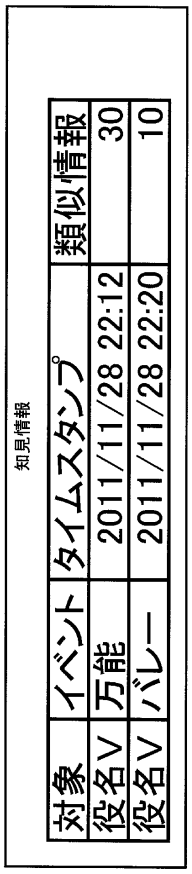
【図３９】



【図４０】



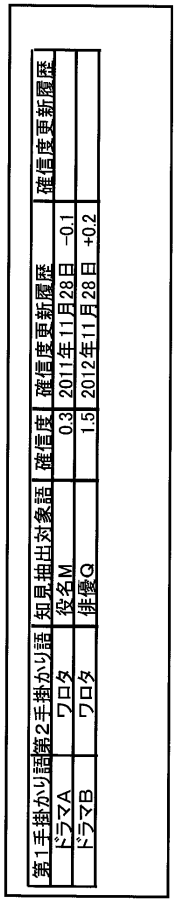
【図４１】



知見情報

対象	イベント	タイムスタンプ	類似情報
役名V	万能	2011/11/28 22:12	30
役名V	バレー	2011/11/28 22:20	10

【図４２】



第1手掛かり語	第2手掛かり語	知見抽出対象語	確信度	確信度更新履歴	確信度更新履歴
ドラマA	ワロタ	役名M	0.3	2011年11月28日	−0.1
ドラマB	ワロタ	俳優Q	1.5	2012年11月28日	+0.2

フロントページの続き

(72)発明者　櫻井　茂明
東京都港区芝浦一丁目１番１号　東芝ソリューション株式会社内
(72)発明者　松本　茂
東京都港区芝浦一丁目１番１号　東芝ソリューション株式会社内
(72)発明者　磯部　庄三
東京都港区芝浦一丁目１番１号　東芝ソリューション株式会社内
(72)発明者　西　一嘉
東京都港区芝浦一丁目１番１号　東芝ソリューション株式会社内
(72)発明者　齋藤　佳美
東京都港区芝浦一丁目１番１号　東芝ソリューション株式会社内
(72)発明者　鈴木　裕之
東京都港区芝浦一丁目１番１号　東芝ソリューション株式会社内
(72)発明者　正岡　良規
東京都港区芝浦一丁目１番１号　東芝ソリューション株式会社内

審査官　鈴木　和樹

(56)参考文献　特開２０１１－１４１８３３（ＪＰ，Ａ）
特開２００２－２５１５９０（ＪＰ，Ａ）
特開２０１２－０９９００１（ＪＰ，Ａ）
酒井浩之、外１名，経済新聞記事内容の個々の企業におけるインパクトの判定，情報処理学会研究報告（２００６－ＦＩ－８４），日本，社団法人情報処理学会，２００６年　9月１３日，第２００６巻，第９４号，ｐ．４３－５０
小原和博、外２名，イベント知識と選択強調学習を用いたニューラル多変量予測，電気学会論文誌Ｃ，日本，社団法人電気学会，１９９６年　6月２０日，第１１６－Ｃ巻，第７号，ｐ．８０２－８０８
西村康成、外１名，時系列データと文章データベースからの知識抽出に関する研究，ＦＩＴ２００５　第４回情報科学技術フォーラム　一般講演論文集　第２分冊　データベース　自然言語　人工，２００５年　8月２２日，ｐ．１１９－１２０
中村健二、外５名，株価情報と新聞記事を用いた企業の評判情報の抽出に関する研究，第７０回（平成２０年）全国大会講演論文集（１）　アーキテクチャ　ソフトウェア科学・工学　データベース，２００８年　3月１３日，ｐ．１－６３７～１－６３８
張へい、外１名，株価データに基づく新聞記事の評価，２００８年度人工知能学会全国大会（第２２回）論文集　［ＣＤ－ＲＯＭ］，２００８年，ｐ．１－３（１Ｅ２－４）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　　１７／３０
Ｇ０６Ｆ　　１９／００