

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6687741号
(P6687741)

(45) 発行日 令和2年4月28日 (2020.4.28)

(24) 登録日 令和2年4月6日 (2020.4.6)

(51) Int. Cl.	F I
GO 6 F 16/383 (2019.01)	GO 6 F 16/383
GO 6 F 40/263 (2020.01)	GO 6 F 17/27 6 5 5

請求項の数 7 (全 19 頁)

(21) 出願番号	特願2018-537630 (P2018-537630)	(73) 特許権者	517406065
(86) (22) 出願日	平成29年6月30日 (2017.6.30)		平安科技 (深▲せん▼) 有限公司
(65) 公表番号	特表2019-520616 (P2019-520616A)		PING AN TECHNOLOGY
(43) 公表日	令和1年7月18日 (2019.7.18)		(SHENZHEN) CO., LTD.
(86) 国際出願番号	PCT/CN2017/091360		中国広東省深▲せん▼市福田区福田街道福
(87) 国際公開番号	W02018/201600		安社区益田路5033号平安金融中心23
(87) 国際公開日	平成30年11月8日 (2018.11.8)		楼518000
審査請求日	平成30年9月19日 (2018.9.19)		23F, Ping'an Financi
(31) 優先権主張番号	201710313993.1		al Center, No. 5033 Y
(32) 優先日	平成29年5月5日 (2017.5.5)		itian Road, Fu'an Co
(33) 優先権主張国・地域又は機関	中国 (CN)		munity of Futian S
			treet, Futian Distri
			ct Shenzhen, Guangdo
			ng 518000 China
			最終頁に続く

(54) 【発明の名称】 情報マイニング方法、システム、電子装置及び読み取り可能な記憶媒体

(57) 【特許請求の範囲】

【請求項 1】

情報マイニング方法であって、

予め設定されたデータソースから特定タイプの情報をリアルタイムに、又は定期的に取り得るステップと、

取得した各情報に対して単語分割処理を行い、各情報に対応する各分割単語に品詞タグをつけるステップと、

各情報に対応する各分割単語の順序と品詞に基づき、各情報に対応する各分割単語で所定構造のワードツリーを構築するステップと、

1 件の情報に対応する所定構造のワードツリーが構築された後、当該情報に対応する所定構造のワードツリーに基づき解析することで、当該情報に対応するポイント情報を得るステップと、を含み、

前記所定構造のワードツリーは多階層のノードを含み、第 1 階層のノードは各原情報であり、第 2 階層のノードは分割単語からなる連語であり、第 2 階層のノード以後の各階層のノードは、いずれもその直上階層のノードに対応する直下階層の分割単語又は分割単語からなる連語であり、前記各情報に対応する各分割単語の順序と品詞に基づき、各情報に対応する各分割単語で所定構造のワードツリーを構築する前記ステップは、

各情報に対応する各分割単語から、各所定品詞の目的分割単語を特定するステップ A 1 と、

各情報における各目的分割単語の順序に従って、各第 2 階層のノードに対応する分割単

10

20

語からなる連語を決定するステップA 2と、

分割単語からなる連語に対してさらなる単語分割が不可能な場合、当該分割単語からなる連語を、所属するノードレベルの最終階層のノードとするステップA 3と、

分割単語からなる連語に対してさらなる単語分割が可能な場合、当該分割単語からなる連語における各所定品詞の目的分割単語を特定し、当該分割単語からなる連語に対応する各目的分割単語の順序に従って、当該分割単語からなる連語の直下階層のノードに対応する分割単語、又は分割単語からなる連語を決定するステップA 4と、

各ノードレベルの最終階層のノードに対応する分割単語が決定されるまで、ステップA 3及びステップA 4を繰り返し実行するステップA 5と、を含み、

当該情報に対応する所定構造のワードツリーに基づき解析することで、当該情報に対応するポイント情報を得る前記ステップは、

構築された所定構造のワードツリーに基づき、各所定の第1重要品詞の分割単語と各所定の第2重要品詞の分割単語との間の距離を算出するステップと、

各所定の第1重要品詞の分割単語との距離が最小の所定の第2重要品詞の分割単語をそれぞれ特定し、各所定の第1重要品詞の分割単語と、それとの距離が最小の所定の第2重要品詞の分割単語を当該情報における順序に従い組み合わせて、対応するポイント情報を構成するステップとを含み、

前記ワードツリーに基づく解析は、特定品詞の分割単語を中心的情報として設定し、又は検索履歴記録について統計することで中心的情報に対応する品詞の分割単語を決定し、当該品詞を中心的品詞として設定し、そして当該情報に対応する所定構造のワードツリーから中心的品詞に該当し、且つ所定構造のワードツリーにおけるマスターノードとのノード距離が最小の分割単語を特定して、当該情報に対応するポイント情報とすることを含む、

ことを特徴とする情報マイニング方法。

【請求項2】

取得した各情報に対して単語分割処理を行う前記ステップにおいて、

順方向最大マッチング法に従って、各情報における処理対象文字列を汎用辞書にてマッチングすることで、第1マッチング結果を得て、ただし、前記第1マッチング結果に第1の数の第1連語と第3の数の単一文字が含まれ、

逆方向最大マッチング法に従って、各件情報における処理対象文字列を汎用辞書にてマッチングすることで、第2マッチング結果を得て、ただし、前記第2マッチング結果に第2の数の第2連語と第4の数の単一文字が含まれ、

前記第1の数と前記第2の数が同じで、且つ前記第3の数が前記第4の数以下である場合、前記第1マッチング結果を当該情報の単語分割結果とし、

前記第1の数と前記第2の数が同じで、且つ前記第3の数が前記第4の数より大きい場合、前記第2マッチング結果を当該情報の単語分割結果とし、

前記第1の数と前記第2の数が異なり、且つ前記第1の数が前記第2の数より大きい場合、前記第2マッチング結果を当該情報の単語分割結果とし、

前記第1の数と前記第2の数が異なり、且つ前記第1の数が前記第2の数より小さい場合、前記第1マッチング結果を当該情報の単語分割結果とすることを特徴とする請求項1に記載の情報マイニング方法。

【請求項3】

各情報に対応する各分割単語に品詞タグをつける前記ステップにおいて、

汎用辞書における文字及び単語と品詞とのマッピング関係及び／又は予め設定された文字及び単語と品詞とのそれぞれマッピング関係に基づき、各情報の各分割単語に対応する品詞を決定し、

各情報の各分割単語に対応する品詞を注記することを特徴とする請求項1又は請求項2に記載の情報マイニング方法。

【請求項4】

電子装置であって、

記憶手段と、処理手段と、前記記憶手段に記憶されており、且つ前記処理手段上で動作可能な情報マイニングシステムとを含み、前記情報マイニングシステムが前記処理手段により実行されるとき、

予め設定されたデータソースから特定タイプの情報をリアルタイムに、又は定期的に取得するステップと、

取得した各情報に対して単語分割処理を行い、各情報に対応する各分割単語に品詞タグをつけるステップと、

各情報に対応する各分割単語の順序と品詞に基づき、各情報に対応する各分割単語で所定構造のワードツリーを構築するステップと、

1件の情報に対応する所定構造のワードツリーが構築された後、当該情報に対応する所定構造のワードツリーに基づき解析することで、当該情報に対応するポイント情報を得るステップと、を実現し、

前記所定構造のワードツリーは多階層のノードを含み、第1階層のノードは各原情報であり、第2階層のノードは分割単語からなる連語であり、第2階層のノード以後の各階層のノードは、いずれもその直上階層のノードに対応する直下階層の分割単語又は分割単語からなる連語であり、前記各情報に対応する各分割単語の順序と品詞に基づき、各情報に対応する各分割単語で所定構造のワードツリーを構築する際、

各情報に対応する各分割単語から、各所定品詞の目的分割単語を特定し、各情報における各目的分割単語の順序に従って、各第2階層のノードに対応する分割単語からなる連語を決定し、分割単語からなる連語に対してさらなる単語分割が不可能な場合、当該分割単語からなる連語を、所属するノードレベルの最終階層のノードに決定し、分割単語からなる連語に対してさらなる単語分割が可能な場合、当該分割単語からなる連語における各所定品詞の目的分割単語を特定し、当該分割単語からなる連語に対応する各目的分割単語の順序に従って、当該分割単語からなる連語の直下階層のノードに対応する分割単語、又は分割単語からなる連語を決定し、各ノードレベルの最終階層のノードに対応する分割単語が決定されるまで継続し、

当該情報に対応する所定構造のワードツリーに基づき解析することで、当該情報に対応するポイント情報を得る際、

構築された所定構造のワードツリーに基づき、各所定の第1重要品詞の分割単語と各所定の第2重要品詞の分割単語との間の距離を算出し、

各所定の第1重要品詞の分割単語との距離が最小の所定の第2重要品詞の分割単語をそれぞれ特定し、各所定の第1重要品詞の分割単語と、それとの距離が最小の所定の第2重要品詞の分割単語を当該情報における順序に従い組み合わせて、対応するポイント情報を構成し、

前記ワードツリーに基づく解析は、特定品詞の分割単語を中心的情報として設定し、又は検索履歴記録について統計することで中心的情報に対応する品詞の分割単語を決定し、当該品詞を中心的品詞として設定し、そして当該情報に対応する所定構造のワードツリーから中心的品詞に該当し、且つ所定構造のワードツリーにおけるマスターノードとのノード距離が最小の分割単語を特定して、当該情報に対応するポイント情報とすることを含む、

ことを特徴とする電子装置。

【請求項5】

取得した各情報に対する前記単語分割処理において、

順方向最大マッチング法に従って、各情報における処理対象文字列を汎用辞書にてマッチングすることで、第1マッチング結果を得て、ただし、前記第1マッチング結果に第1の数の第1連語と第3の数の単一文字が含まれ、

逆方向最大マッチング法に従って、各件情報における処理対象文字列を汎用辞書にてマッチングすることで、第2マッチング結果を得て、ただし、前記第2マッチング結果に第2の数の第2連語と第4の数の単一文字が含まれ、

前記第1の数と前記第2の数が同じで、且つ前記第3の数が前記第4の数以下である場

10

20

30

40

50

合、前記第 1 マッチング結果を当該情報の単語分割結果とし、

前記第 1 の数と前記第 2 の数が同じで、且つ前記第 3 の数が前記第 4 の数より大きい場合、前記第 2 マッチング結果を当該情報の単語分割結果とし、

前記第 1 の数と前記第 2 の数が異なり、且つ前記第 1 の数が前記第 2 の数より大きい場合、前記第 2 マッチング結果を当該情報の単語分割結果とし、

前記第 1 の数と前記第 2 の数が異なり、且つ前記第 1 の数が前記第 2 の数より小さい場合、前記第 1 マッチング結果を当該情報の単語分割結果とすることを特徴とする請求項 4 に記載の電子装置。

【請求項 6】

各情報に対応する各分割単語に品詞タグをつける際、

汎用辞書における文字及び単語と品詞とのマッピング関係及び／又は予め設定された文字及び単語と品詞とのそれぞれマッピング関係に基づき、各情報の各分割単語に対応する品詞を決定し、各情報の各分割単語に対応する品詞を注記することを特徴とする請求項 4 又は請求項 5 に記載の電子装置。

【請求項 7】

コンピュータ読み取り可能な記憶媒体であって、

当該記憶媒体に少なくとも 1 つのコンピュータ読み取り可能なコマンドが記憶されており、前記少なくとも 1 つのコンピュータ読み取り可能なコマンドが処理手段により実行されることで、

予め設定されたデータソースから特定タイプの情報をリアルタイムに、又は定期的に取得するステップと、

取得した各情報に対して単語分割処理を行い、各情報に対応する各分割単語に品詞タグをつけるステップと、

各情報に対応する各分割単語の順序と品詞に基づき、各情報に対応する各分割単語で所定構造のワードツリーを構築するステップと、

1 件の情報に対応する所定構造のワードツリーが構築された後、当該情報に対応する所定構造のワードツリーに基づき解析することで、当該情報に対応するポイント情報を得るステップと、を実現し、

前記所定構造のワードツリーは多階層のノードを含み、第 1 階層のノードは各原情報であり、第 2 階層のノードは分割単語からなる連語であり、第 2 階層のノード以後の各階層のノードは、いずれもその直上階層のノードに対応する直下階層の分割単語又は分割単語からなる連語であり、前記各情報に対応する各分割単語の順序と品詞に基づき、各情報に対応する各分割単語で所定構造のワードツリーを構築する前記ステップは、

各情報に対応する各分割単語から、各所定品詞の目的分割単語を特定するステップ A 1 と、

各情報における各目的分割単語の順序に従って、各第 2 階層のノードに対応する分割単語からなる連語を決定するステップ A 2 と、

分割単語からなる連語に対してさらなる単語分割が不可能な場合、当該分割単語からなる連語を、所属するノードレベルの最終階層のノードとするステップ A 3 と、

分割単語からなる連語に対してさらなる単語分割が可能な場合、当該分割単語からなる連語における各所定品詞の目的分割単語を特定し、当該分割単語からなる連語に対応する各目的分割単語の順序に従って、当該分割単語からなる連語の直下階層のノードに対応する分割単語、又は分割単語からなる連語を決定するステップ A 4 と、

各ノードレベルの最終階層のノードに対応する分割単語が決定されるまで、ステップ A 3 及びステップ A 4 を繰り返し実行するステップ A 5 と、を含み、

当該情報に対応する所定構造のワードツリーに基づき解析することで、当該情報に対応するポイント情報を得る前記ステップは、

構築された所定構造のワードツリーに基づき、各所定の第 1 重要品詞の分割単語と各所定の第 2 重要品詞の分割単語との間の距離を算出するステップと、

各所定の第 1 重要品詞の分割単語との距離が最小の所定の第 2 重要品詞の分割単語をそ

10

20

30

40

50

れぞれ特定し、各所定の第1重要品詞の分割単語と、それとの距離が最小の所定の第2重要品詞の分割単語を当該情報における順序に従い組み合わせて、対応するポイント情報を構成するステップとを含む、

前記ワードツリーに基づく解析は、特定品詞の分割単語を中心的情報として設定し、又は検索履歴記録について統計することで中心的情報に対応する品詞の分割単語を決定し、当該品詞を中心的品詞として設定し、そして当該情報に対応する所定構造のワードツリーから中心的品詞に該当し、且つ所定構造のワードツリーにおけるマスターノードとのノード距離が最小の分割単語を特定して、当該情報に対応するポイント情報とすることを含む、

コンピュータ読み取り可能な記憶媒体。

10

【発明の詳細な説明】

【技術分野】

【0001】

「関連出願の相互参照」

本願は、パリ条約に基づき、2017年5月5日に出願された中国特許出願第201710313931号の「情報マイニング方法、電子装置及び読み取り可能な記憶媒体」と題する中国特許出願の優先権を主張し、当該出願の全体が参照により本願に援用される。

【0002】

本発明は、コンピュータ技術分野に関し、特に情報マイニング方法、システム、電子装置及び読み取り可能な記憶媒体に関する。

20

【背景技術】

【0003】

現在、情報マイニング及びプッシュ型情報配信分野では、予め設定されたデータソース（例えば、ポータルサイト）から特定タイプの情報（例えば、ニュースタイトル）についてリアルタイムに、又は定期的に分析し選別することで、目的情報を特定するという方法が一般的に用いられている。現行の分析・選別方法は、情報のタイプタグを認識する分類器を予めトレーニングし、トレーニングにより得られる分類器を用いて特定タイプの情報のタイプタグを認識することで、所定のタイプタグに該当する目的情報を認識する。このような分析・選別方法は、所定のタイプタグに該当する目的情報を認識する程度にとどまり、目的情報が指し示すポイント情報について高度なマイニングを行うことができず、目的情報のマイニング及びプッシュ配信における正確性を確保できず、誤りが生じやすい。

30

【発明の概要】

【発明が解決しようとする課題】

【0004】

本発明は、ポイント情報を効率的にマイニングするために、情報マイニング方法、システム、電子装置及び読み取り可能な記憶媒体を提供することを目的とする。

【課題を解決するための手段】

【0005】

上記の目的を達成すべく、本発明の第1態様は、情報マイニング方法を提供し、前記方法は、

40

予め設定されたデータソースから特定タイプの情報をリアルタイムに、又は定期的に取得するステップと、

取得した各情報に対して単語分割処理を行い、各情報に対応する各分割単語に品詞タグをつけるステップと、

各情報に対応する各分割単語の順序と品詞に基づき、各情報に対応する各分割単語で所定構造のワードツリーを構築するステップと、

1件の情報に対応する所定構造のワードツリーが構築された後、当該情報に対応する所定構造のワードツリーに基づき解析することで、当該情報に対応するポイント情報を得るステップと、を含む。

50

【0006】

本発明の第2の態様は、情報マイニングシステムを提供し、前記情報マイニングシステムは、

予め設定されたデータソースから特定タイプの情報をリアルタイムに、又は定期的に取得する取得モジュールと、

取得した各情報に対して単語分割処理を行い、各情報に対応する各分割単語に品詞タグをつける単語分割モジュールと、

各情報に対応する各分割単語の順序と品詞に基づき、各情報に対応する各分割単語で所定構造のワードツリーを構築する構築モジュールと、

1件の情報に対応する所定構造のワードツリーが構築された後、当該情報に対応する所定構造のワードツリーに基づき解析することで、当該情報に対応するポイント情報を得る解析モジュールと、を含む。

10

【0007】

本発明の第3の態様は、電子装置を提供し、前記電子装置は、記憶手段と、処理手段と、前記記憶手段に記憶されており、且つ前記処理手段上で動作可能な情報マイニングシステムとを含み、前記情報マイニングシステムが前記処理手段により実行されるとき、

予め設定されたデータソースから特定タイプの情報をリアルタイムに、又は定期的に取得するステップと、

取得した各情報に対して単語分割処理を行い、各情報に対応する各分割単語に品詞タグをつけるステップと、

20

各情報に対応する各分割単語の順序と品詞に基づき、各情報に対応する各分割単語で所定構造のワードツリーを構築するステップと、

1件の情報に対応する所定構造のワードツリーが構築された後、当該情報に対応する所定構造のワードツリーに基づき解析することで、当該情報に対応するポイント情報を得るステップと、を実現する。

【0008】

本発明の第4の態様は、コンピュータ読み取り可能な記憶媒体を提供し、当該コンピュータ読み取り可能な記憶媒体に少なくとも1つのコンピュータ読み取り可能なコマンドが記憶されており、前記少なくとも1つのコンピュータ読み取り可能なコマンドが処理手段により実行されることで、

30

予め設定されたデータソースから特定タイプの情報をリアルタイムに、又は定期的に取得する動作と、

取得した各情報に対して単語分割処理を行い、各情報に対応する各分割単語に品詞タグをつける動作と、

各情報に対応する各分割単語の順序と品詞に基づき、各情報に対応する各分割単語で所定構造のワードツリーを構築する動作と、

1件の情報に対応する所定構造のワードツリーが構築された後、当該情報に対応する所定構造のワードツリーに基づき解析することで、当該情報に対応するポイント情報を得る動作と、を実現する。

【発明の効果】

40

【0009】

本発明による情報マイニング方法、システム、電子装置及び読み取り可能な記憶媒体は、データソースから取得した特定タイプの情報に対して単語分割を行い各分割単語に対して品詞注記を行い、各分割単語の順序と品詞に基づき所定構造のワードツリーを構築し、構築された所定構造のワードツリーに基づき解析することで、当該情報に対応するポイント情報を得るものである。本発明は、取得された情報に対して単語分割を行い、各分割単語の品詞に基づき所定構造のワードツリーを構築し、所定構造のワードツリーでマイニングして情報における各分割単語間の関係性を決定することでポイント情報を取得できるため、情報に対する高度なマイニングを実現し、情報からポイント情報を正確に得ることができる。

50

【図面の簡単な説明】

【0010】

【図1】本発明の情報マイニング方法の好ましい実施例を実現する環境の概略図である。

【図2】本発明の情報マイニングシステム10の一実施例における機能モジュールの概略図である。

【図3】本発明の情報マイニング方法の一実施例における所定構造のワードツリーの概略図である。

【図4】本発明の情報マイニング方法の一実施例のフローチャートである。

【発明を実施するための形態】

【0011】

10

本発明の解決しようとする技術的課題、技術的解決手段及び有益な効果をより明瞭にするために、以下に実施例及び図面を踏まえて、本発明の目的の達成、機能上の特徴及び利点について説明する。ただし、本明細書に記載の実施例は、本発明を説明するためのものに過ぎず、本発明を限定するためのものではない。

【0012】

図1は、本発明の情報マイニング方法の好ましい実施例を実施する環境の概略図である。図1によると、前記環境は、電子装置1と、端末機器2とを含む。電子装置1は、インターネットや近距離通信技術などの適切な技術により、端末機器2とデータ交換を行うことができる。

【0013】

20

端末機器2は、キーボード、マウス、リモートコントローラ、タッチパッド、又は音声制御機器などの手段によりユーザとマンマシンインタラクションを行うあらゆる電子製品、例えば、PC、タブレットPC、スマートフォン、パーソナルデジタルアシスタント(Personal Digital Assistant、PDA)、ゲーム機、インターネットプロトコル放送テレビ(Internet Protocol Television、IPTV)、ウェアラブル端末を含むが、これらに限定されない。

【0014】

電子装置1は、予め設定された又は記憶されているコマンドに従って、数値計算及び／又は情報処理を自動的に行う機器である。電子装置1には、コンピュータ、単一のネットワークサーバ、複数のネットワークサーバからなるサーバ群、もしくはクラウドコンピューティングに基づく大量のホストやネットワークサーバからなるクラウド型機器が挙げられ、クラウドコンピューティングとは、分散コンピューティングの1種であり、疎結合されている複数のコンピュータクラスターからなる大型仮想コンピュータを指すものである。

30

【0015】

本実施例では、電子装置1が、システムバスを介して相互に通信可能に接続される記憶手段11と、処理手段12と、ネットワークインタフェース13とを含むが、これらに限定されない。図1は要素11と、要素12と、要素13とを備える電子装置1を示しているが、必ずしも示された全ての要素で実施するとは限らず、代替案として、より多い又はより少ない数の要素で実施してもよい。

40

【0016】

このうち、記憶手段11はメモリ及び少なくとも1つのタイプの読み取り可能な記憶媒体を含む。メモリは、電子装置1の動作時のキャッシュとして機能し、読み取り可能な記憶媒体は、例えば、フラッシュメモリ、ハードディスク、マルチメディアカード、カード型記憶手段などの不揮発性記憶媒体である。いくつかの実施例において、読み取り可能な記憶媒体は、電子装置1の内部記憶手段、例えば、電子装置1のハードディスクであり、別の実施例において、当該不揮発性記憶媒体は、電子装置1の外部記憶手段、例えば、電子装置1に搭載される外付けハードディスク、スマートメディアカード(Smart(登録商標)Media Card、SMC)、セキュアデジタル(Secure Digital、SD)カード、フラッシュカード(Flash Card)などである。本実

50

施例では、記憶手段 11 の読み込み可能な記憶媒体が一般的に、電子装置 1 にインストールされるオペレーティングシステム及び各種アプリケーションソフトウェア、例えば、本発明の一実施例における情報マイニングシステム 10 のプログラムコードなどを記憶するために用いられる。記憶手段 11 は、すでに出力された又は出力する各種データを一時的に記憶するためにも用いられる。

【0017】

いくつかの実施例では、処理手段 12 が、マイクロプロセッサ、マイクロコントローラ、データプロセッサなどを 1 つ又は複数を含んでもよい。処理手段 12 は一般的に、電子装置 1 の動作、例えば、端末機器 2 とのデータ交換又は通信に関連する制御や処理などの実行を制御するために用いられる。本実施例では、処理手段 12 が、記憶手段 11 に記憶

10

【0018】

ネットワークインタフェース 13 は、無線ネットワークインタフェース、又は有線ネットワークインタフェースを含んでもよく、ネットワークインタフェース 13 は一般的に、電子装置 1 とその他の電子機器の間に通信接続を確立するために用いられる。本実施例では、ネットワークインタフェース 13 は主に、電子装置 1 と 1 つ又は複数の端末機器 2 を接続させることで、電子装置 1 と 1 つ又は複数の端末機器 2 の間にデータを伝送するチャンネルと通信接続を確立するために用いられる。

【0019】

20

情報マイニングシステム 10 は、記憶手段 11 に記憶されている少なくとも 1 つのコンピュータ読み取り可能なコマンドを含み、当該少なくとも 1 つのコンピュータ読み取り可能なコマンドが処理手段 12 により実行されることで、本発明の各実施例の方法を実現する。また、後述するように、当該少なくとも 1 つのコンピュータ読み取り可能なコマンドは、その各部分で実現する機能の違いにより、異なる論理モジュールに分かれてもよい。

【0020】

一実施例として、情報マイニングシステム 10 が処理手段 12 により実行されるとき、次の動作を実現する。まず、端末機器 2 における予め設定されたデータソースから特定タイプの情報をリアルタイムに、又は定期的に取得する。そして取得した各情報に対して単語分割処理を行い、各情報に対応する各分割単語に品詞タグをつける。各情報に対応する各分割単語の順序と品詞に基づき、各情報に対応する各分割単語で所定構造のワードツリーを構築する。1 件の情報に対応する所定構造のワードツリーが構築された後、当該情報に対応する所定構造のワードツリーに基づき解析することで、当該情報に対応するポイント情報を得て、そして端末機器 2 に送信してユーザに表示する。

30

【0021】

一実施例として、情報マイニングシステム 10 は、記憶手段 11 に記憶されており、記憶手段 11 に記憶されている少なくとも 1 つのコンピュータ読み取り可能なコマンドを含み、当該少なくとも 1 つのコンピュータ読み取り可能なコマンドが処理手段 12 により実行されることで、本発明の各実施例の方法を実現する。また、後述するように、当該少なくとも 1 つのコンピュータ読み取り可能なコマンドは、その各部分で実現する機能の違いにより、異なる論理モジュールに分かれてもよい。

40

【0022】

図 2 は、本発明の情報マイニングシステム 10 の好ましい実施例における機能モジュールの概略図である。本実施例では、情報マイニングシステム 10 を 1 つ又は複数のモジュールに分け、前記 1 つ又は複数のモジュールを記憶手段 11 に記憶し、1 つ又は複数の処理手段（本実施例では処理手段 12）により実行することで本発明を実現することができる。例えば、図 2 では、情報マイニングシステム 10 を、取得モジュール 01 と、単語分割モジュール 02 と、構築モジュール 03 と、解析モジュール 04 とに分ける。上記の各モジュールは、一連のコンピュータプログラムブロックを含み、これらのコンピュータプログラムブロックが処理手段 12 により実行されることで、本発明の各実施例における対

50

応する機能を実現する。以下に取得モジュール01、単語分割モジュール02、構築モジュール03、解析モジュール04の機能について詳しく説明する。

【0023】

取得モジュール01は、予め設定されたデータソースから特定タイプの情報をリアルタイムに、又は定期的に取得する。例えば、クローラなどのツールを用いて、予め設定されたデータソース（例えば、ポータルサイト、掲示板など）から特定タイプの情報（例えば、ニュースタイトル、インデックス情報、概要情報など）をリアルタイムに、又は定期的に取得する。

【0024】

単語分割モジュール02は、取得した各情報に対して単語分割処理を行い、各情報に対応する各分割単語に品詞タグをつける。データソースから特定タイプの各情報を取得した後、取得した各情報に対して単語分割処理を行うことができる。各情報に対して単語分割処理を行うとき、例えば、文字列マッチングによる単語分割方法を用いることができ、例えば、順方向最大マッチング法により、1件の情報における文字列に対して左から右へ単語分割を行い、すなわち左から右へ単語分割対象情報における連続文字を辞書にてマッチングし、マッチングに成功したら1つの単語を切り分け、もしくは、逆方向最大マッチング法により、1件の情報における文字列に対して右から左へ単語分割を行い、すなわち単語分割対象情報の末尾から走査し、右から左へ単語分割対象情報における連続文字を辞書とマッチングし、マッチングに成功したら1つの単語を切り分け、もしくは、最短経路による単語分割方法により、1件の情報における文字列から切り分けられる単語数が最少となることを前提に単語分割を行い、もしくは、双方向最大マッチング法により、順方向と逆方向から同時にマッチングして単語分割を行う。又は、意味による単語分割方法を用いて、各情報に対して単語分割処理を行うこともでき、意味による単語分割法は、機器による音声判断に基づく単語分割方法であり、構文情報と意味情報を利用して多義の状況処理して単語分割を行うものである。この他に、統計的手法による単語分割方法を用いて単語分割処理を行うこともでき、すなわち当該ユーザの検索履歴記録、又は一般ユーザの検索履歴記録における連語に対して統計し、出現する頻度が高い2つの連続文字を特定できれば、この2つの連続文字を連語として単語分割を行うことができる。取得した各情報に対する単語分割処理が完了すると、各情報に対応する各分割単語（連語及び単一文字を含む）に品詞タグをつけることができる。例えば、前記品詞は、実詞（名詞、動詞、形容詞、数量詞、代詞など）及び虚詞（副詞、介詞（前置詞）、接続詞、助詞、感嘆詞、擬音語など）を含む。

【0025】

構築モジュール03は、各情報に対応する各分割単語の順序と品詞に基づき、各情報に対応する各分割単語で所定構造のワードツリーを構築する。

【0026】

解析モジュール04は、1件の情報に対応する所定構造のワードツリーが構築された後、当該情報に対応する所定構造のワードツリーに基づき解析することで、当該情報に対応するポイント情報を得る。

【0027】

各情報に対応する各分割単語に対して品詞注記を行った後、各情報における各分割単語の順序と各単語分割に注記された品詞に基づき、各情報に対応する各分割単語で所定構造のワードツリーを構築する。例えば、所定構造のワードツリーにおける品詞によって対応するノード階層を設定し、1件の情報における各分割単語を異なるノードとすることで、当該情報に対応する所定構造のワードツリーを構築する。又は、異なる品詞の分割単語を組み合わせて連語を構成して各分割単語と共に異なるノード階層を形成することで、当該情報に対応する所定構造のワードツリーを構築する。1件の情報に対応する所定構造のワードツリーが構築された後、当該情報に対応する所定構造のワードツリーに基づき解析することで、当該情報に対応するポイント情報を得る。例えば、特定品詞の分割単語を中心的情報として設定し、又は検索履歴記録について統計することで中心的情報に対応する品

10

20

30

40

50

詞の分割単語を決定し、当該品詞を中心的品詞として設定し、そして当該情報に対応する所定構造のワードツリーから中心的品詞に該当し、且つ所定構造のワードツリーにおけるマスターノードとのノード距離が最小の分割単語を特定して、当該情報に対応するポイント情報とする。又は、複数の中心的品詞を設定し、当該情報に対応する所定構造のワードツリーから中心的品詞に該当する複数の分割単語を特定し、且つ所定構造のワードツリーにおける中心的品詞に該当する複数の分割単語間のノード距離が最小の分割単語の組み合わせを特定して、当該分割単語の組み合わせに対応する情報を当該情報のポイント情報とする。

【0028】

本実施例では、データソースから取得した特定タイプの情報に対して単語分割を行い各単語分割に対して品詞注記を行い、各分割単語の順序と品詞に基づき所定構造のワードツリーを構築し、構築された所定構造のワードツリーに基づき解析することで、当該情報に対応するポイント情報を得る。取得した情報に対して単語分割を行い、さらに各分割単語の品詞に基づき所定構造のワードツリーを構築し、所定構造のワードツリーでマイニングして情報における各分割単語間の関係性を決定することでポイント情報を取得できるため、情報に対する高度なマイニングを実現し、情報からポイント情報を正確に得ることができる。

【0029】

さらに、別の実施例として、当該情報に対応する所定構造のワードツリーに基づき解析して、当該情報に対応するポイント情報を得た後、解析モジュール04において、さらに、予めトレーニングされた分類器を用いて、当該情報のポイント情報に対応するタイプタグを認識し、認識されたタイプタグが所定のタイプタグに該当する場合、当該情報の全てのコンテンツ及び／又は当該情報の全てのコンテンツのリンクアドレスを所定の端末にプッシュ配信する。例えば、ユーザがスポーツ情報に興味を持っている場合、予めタイプタグを「スポーツ」に設定し、データソースから取得した情報を解析してポイント情報を得た後、当該情報のポイント情報に対応するタイプタグを認識し、認識されたタイプタグが「スポーツ」タグであれば、当該情報をユーザが興味を持っている情報として判断し、当該情報の全てのコンテンツ及び／又は当該情報の全てのコンテンツのリンクアドレスを所定の端末（例えばユーザの携帯電話、タブレットPCなど）にプッシュ配信することで、目的情報に対する効率的なマイニング及び正確なプッシュ配信を実現する。

【0030】

さらに、別の実施例として、単語分割モジュール02は、さらに、

順方向最大マッチング法に従って、各情報における処理対象文字列を汎用辞書にてマッチングすることで、第1マッチング結果を得て、

逆方向最大マッチング法に従って、各件情報における処理対象文字列を汎用辞書にてマッチングすることで、第2マッチング結果を得る。ただし、前記第1マッチング結果に第1の数の第1連語が含まれ、前記第2マッチング結果に第2の数の第2連語が含まれ、前記第1マッチング結果に第3の数の単一文字が含まれ、前記第2マッチング結果に第4の数の単一文字が含まれる。

前記第1の数と前記第2の数が同じで、且つ前記第3の数が前記第4の数以下である場合、前記第1マッチング結果（連語及び単一文字を含む）を出力し、前記第1の数と前記第2の数が同じで、且つ前記第3の数が前記第4の数より大きい場合、前記第2マッチング結果（連語及び単一文字を含む）を出力し、前記第1の数と前記第2の数が異なり、且つ前記第1の数が前記第2の数より大きい場合、前記第2マッチング結果（連語及び単一文字を含む）を出力し、前記第1の数と前記第2の数が異なり、且つ前記第1の数が前記第2の数より小さい場合、前記第1マッチング結果（連語及び単一文字を含む）を出力する。

【0031】

本実施例では、双方向マッチング法により、取得した各情報に対して単語分割処理を行い、順方向と逆方向から同時に単語分割を行ってマッチングすることにより、各情報の処

10

20

30

40

50

理対象文字列における組み合わせコンテンツの関連性を分析し、一般的に、連語の方はポイント情報を表す確率がより高く、すなわち連語の方はポイント情報を表現するのにより適切である。よって、順方向と逆方向から同時に単語分割を行ってマッチングすることで、単一文字の数がより少なく、連語の数がより多い分割単語マッチング結果を得て情報の単語分割結果とすることにより、単語分割及び情報マイニングにおける正確性を高めることができる。

【0032】

さらに、別の実施例として、単語分割モジュール02は、さらに、汎用辞書における文字及び単語と品詞とのマッピング関係（例えば、汎用辞書においては、運動場に対応する品詞が名詞である）及び／又は予め設定された文字及び単語と品詞とのマッピング関係（例えば、予め設定された文字及び単語と品詞とのマッピング関係においては、運動場に対応する品詞が常用名詞である）に基づき、各情報の各分割単語に対応する品詞を決定し、各情報の各分割単語に対応する品詞を注記する。ただし、予め設定された文字及び単語と品詞とのマッピング関係に基づく品詞注記の優先度は、汎用辞書における文字及び単語と品詞とのマッピング関係よりも高い。例えば、汎用辞書によると、運動場に対応する品詞が名詞であり、予め設定された文字及び単語と品詞とのマッピング関係によると、運動場に対応する品詞が常用名詞である場合、注記するとき、先に予め設定された文字及び単語と品詞とのマッピング関係に基づき注記し、すなわち運動場の品詞に常用名詞を注記する。

【0033】

さらに、別の実施例として、前記所定構造のワードツリーは、複数階層のノードを含み、第1階層のノードは原情報であり、第2階層のノードは分割単語からなる連語であり、第2階層のノード以後の各階層のノードは、いずれもその直上階層のノードに対応する直下階層の分割単語又は分割単語からなる連語である。

構築モジュール03において、さらに、各情報に対応する各分割単語から、各所定品詞の目的分割単語を特定し、各情報における各目的分割単語の順序に従って、各第2階層のノードに対応する分割単語からなる連語を決定する。分割単語からなる連語に対してさらなる単語分割が不可能な場合、当該分割単語からなる連語を、所属するノードレベルの最終階層のノードとし、分割単語からなる連語に対してさらなる単語分割が可能な場合、当該分割単語からなる連語における各所定品詞の目的分割単語を特定し、当該分割単語からなる連語に対応する各目的分割単語の順序に従って、各ノードレベルの最終階層のノードに対応する分割単語が決定されるまで、当該分割単語からなる連語の直下階層のノードに対応する分割単語、又は分割単語からなる連語を決定する。

【0034】

解析モジュール04において、さらに、構築された所定構造のワードツリーに基づき、各所定の第1重要品詞の分割単語と各所定の第2重要品詞の分割単語との間の距離を算出し、各所定の第1重要品詞の分割単語との距離が最小の所定の第2重要品詞の分割単語をそれぞれ特定し、各所定の第1重要品詞の分割単語と、それとの距離が最小の所定の第2重要品詞の分割単語を当該情報における順序に従い組み合わせて、対応するポイント情報を構成する。

【0035】

一つの具体的な実施形態として、図3に示すように、当該情報は「我去操場＊（あしへんに易）足球了（私は運動場へサッカーをしに行く）」である場合、対応する単語分割結果は「我 去 操場 ＊（あしへんに易）足球」であり、品詞注記結果は「我（私）／代詞 去（行く）／動詞 操場（運動場）／常用名詞 ＊（あしへんに易）足球（サッカーをする）／常用名詞」である。図3は当該情報「我 去 操場 ＊（あしへんに易）足球」で構築された所定構造のワードツリーを示している。当該所定構造のワードツリーは複数階層のノードを含み、第1階層のノードは原情報であり、第2階層のノードは分割単語からなる連語（例えば、名詞連語、動詞連語及び句点などの区切り記号）である。本実施例では、各情報に対応する各分割単語から、各所定品詞（例えば、名詞、動詞など）の目

10

20

30

40

50

的分割単語を特定し、当該情報における各目的分割単語の順序に従って、各第2階層のノードに対応する分割単語からなる連語を決定する。第2階層のノード以後の各階層のノードは、いずれもその直上階層のノードに対応する直下階層の分割単語又は分割単語からなる連語であり、第3階層ノードは第2階層のノードの分割単語又は分割単語からなる連語である。図3に示すように、当該情報に対する品詞注記結果は「我／代詞 去／動詞 操場／常用名詞 *（あしへんに易）足球／常用名詞」であり、当該情報における各分割単語の順序に従って、左から右へ第2階層のノードを決定し、予め第2階層のノードを、名詞連語、動詞連語などを含む分割単語からなる連語に設定し、当該情報においては、左から右へ、「我」が代詞で名詞連語に該当するため、「我」を第2階層のノードに決定し、「我」に続く「去」、「操場」、「*（あしへんに易）足球」は、動詞連語「去操場*（あしへんに易）足球」を構成し、よって、「去操場*（あしへんに易）足球」を第2階層のノードに決定する。これにより、当該情報の所定構造のワードツリーにおける第2階層のノードは「我」、「去操場*（あしへんに易）足球」を含む。また、第2階層のノード「我」はさらなる単語分割が不可能であるため、当該分割単語からなる連語を所属するノードレベルの最終階層のノードに決定する。第2階層のノードである動詞連語「去操場*（あしへんに易）足球」は、さらなる単語分割が可能であるため、当該第2階層のノードである「去操場*（あしへんに易）足球」の分割単語又は分割単語からなる連語を第3階層ノードとすることができ、当該第3階層のノードは、動詞「去」及び名詞連語「操場*（あしへんに易）足球」を含み、名詞連語「操場*（あしへんに易）足球」をさらに第4階層のノード「操場」、「*（あしへんに易）足球」に分割することができる。分割単語からなる連語に対してさらなる分割単語が可能な場合、当該分割単語からなる連語における各所定品詞（例えば、名詞、動詞など）の目的分割単語を特定し、当該分割単語からなる連語に対応する各目的分割単語の順序に従って、当該分割単語からなる連語の直下階層のノードに対応する分割単語又は分割単語からなる連語を決定し、分割単語からなる連語に対してさらなる単語分割が不可能な場合、当該分割単語からなる連語を、所属するノードレベルの最終階層のノードとする。

【0036】

構築された所定構造のワードツリーに基づき、各第1重要品詞（例えば、動詞）の分割単語と各第2重要品詞（例えば、名詞）の分割単語との距離を算出し、各第1重要品詞の分割単語と各第2重要品詞の分割単語の間におけるノード数を前記距離とする。ただし、第1重要品詞、第2重要品詞は、必要に応じて設定し、又は当該ユーザの検索履歴記録における中心的情報に一般的に対応する品詞に基づき設定する。各第1重要品詞の分割単語との距離が最小の第2重要品詞の分割単語をそれぞれ特定し、各第1重要品詞の分割単語と、それとの距離が最小の第2重要品詞の分割単語を当該情報における順序に従い組み合わせ、対応するポイント情報を構成する。例えば、図3に示すように、「去操場（運動場へ行く）」、「去*（あしへんに易）足球（サッカーをしに行く）」を、マイニングにより得られた当該情報「我去操場*（あしへんに易）足球」に対応するポイント情報とする。

【0037】

本発明は、さらに情報マイニング方法を提供する。

【0038】

図4は、本発明の情報マイニング方法の一実施例のフローチャートである。

【0039】

一実施例として、当該情報マイニング方法はステップS10と、ステップS20と、ステップS30と、ステップS40とを含む。

【0040】

ステップS10において、予め設定されたデータソースから特定タイプの情報をリアルタイムに、又は定期的に取得する。例えば、クローラなどのツールを用いて、予め設定されたデータソース（例えば、ポータルサイト、掲示板など）から特定タイプの情報（例えば、ニュースタイトル、インデックス情報、概要情報など）をリアルタイムに、又は定期

的に取得する。

【0041】

ステップS20において、取得した各情報に対して単語分割処理を行い、各情報に対応する各分割単語に品詞タグをつける。

【0042】

データソースから特定タイプの各情報を取得した後、取得した各情報に対して単語分割処理を行うことができる。各情報に対して単語分割処理を行うとき、例えば、文字列マッチングによる単語分割方法を用いることができ、例えば、順方向最大マッチング法により、1件の情報における文字列に対して左から右へ単語分割を行い、すなわち左から右へ単語分割対象情報における連続文字を辞書にてマッチングし、マッチングに成功したら1つの単語を切り分け、もしくは、逆方向最大マッチング法により、1件の情報における文字列に対して右から左へ単語分割を行い、すなわち単語分割対象情報の末尾から走査し、右から左へ単語分割対象情報における連続文字を辞書とマッチングし、マッチングに成功したら1つの単語を切り分け、もしくは、最短経路による単語分割方法により、1件の情報における文字列から切り分けられる単語数が最少となることを前提に単語分割を行い、もしくは、双方向最大マッチング法により、順方向と逆方向から同時にマッチングして単語分割を行う。又は、意味による単語分割方法を用いて、各情報に対して単語分割処理を行うこともでき、意味による単語分割法は、機器による音声判断に基づく単語分割方法であり、構文情報と意味情報を利用して多義の状況进行处理して単語分割を行うものである。この他に、統計的手法による単語分割方法を用いて単語分割処理を行うこともでき、すなわち当該ユーザの検索履歴記録、又は一般ユーザの検索履歴記録における連語に対して統計し、出現する頻度が高い2つの連続文字を特定できれば、この2つの連続文字を連語として単語分割を行う。

【0043】

取得した各情報に対する単語分割処理が完了すると、各情報に対応する各分割単語（連語及び単一文字を含む）に品詞タグをつけることができる。例えば、前記品詞は、実詞（名詞、動詞、形容詞、数量詞、代詞など）及び虚詞（副詞、介詞（前置詞）、接続詞、助詞、感嘆詞、擬音語など）を含む。

【0044】

ステップS30において、各情報に対応する各分割単語の順序と品詞に基づき、各情報に対応する各分割単語で所定構造のワードツリーを構築する。

【0045】

ステップS40において、1件の情報に対応する所定構造のワードツリーが構築された後、当該情報に対応する所定構造のワードツリーに基づき解析することで、当該情報に対応するポイント情報を得る。

【0046】

各情報に対応する各分割単語に対して品詞注記を行った後、各情報における各分割単語の順序と各単語分割に注記された品詞に基づき、各情報に対応する各分割単語で所定構造のワードツリーを構築する。例えば、所定構造のワードツリーにおける品詞によって対応するノード階層を設定し、1件の情報における各分割単語を異なるノードとすることで、当該情報に対応する所定構造のワードツリーを構築する。又は、異なる品詞の分割単語を組み合わせて連語を構成して各分割単語と共に異なるノード階層を形成することで、当該情報に対応する所定構造のワードツリーを構築する。1件の情報に対応する所定構造のワードツリーが構築された後、当該情報に対応する所定構造のワードツリーに基づき解析することで、当該情報に対応するポイント情報を得る。例えば、特定品詞の分割単語を中心的情報として設定し、又は検索履歴記録について統計することで中心的情報に対応する品詞の分割単語を決定し、当該品詞を中心的品詞として設定し、そして当該情報に対応する所定構造のワードツリーから中心的品詞に該当し、且つ所定構造のワードツリーにおけるマスターノードとのノード距離が最小の分割単語を特定して、当該情報に対応するポイント情報とする。又は、複数の中心的品詞を設定し、当該情報に対応する所定構造のワード

ツリーから中心的品詞に該当する複数の分割単語を特定し、且つ所定構造のワードツリーにおける中心的品詞に該当する複数の分割単語間のノード距離が最小の分割単語の組み合わせを特定して、当該分割単語の組み合わせに対応する情報を当該情報のポイント情報とする。

【0047】

本実施例では、データソースから取得した特定タイプの情報に対して単語分割を行い各単語分割に対して品詞注記を行い、各分割単語の順序と品詞に基づき所定構造のワードツリーを構築し、構築された所定構造のワードツリーに基づき解析することで、当該情報に対応するポイント情報を得る。取得した情報に対して単語分割を行い、さらに各分割単語の品詞に基づき所定構造のワードツリーを構築し、所定構造のワードツリーでマイニングして情報における各分割単語間の関係性を決定することでポイント情報を取得できるため、情報に対する高度なマイニングを実現し、情報からポイント情報を正確に得ることができる。

10

【0048】

さらに、別の実施例として、当該情報に対応する所定構造のワードツリーに基づき解析して、当該情報に対応するポイント情報を得た後、さらに、予めトレーニングされた分類器を用いて、当該情報のポイント情報に対応するタイプタグを認識し、認識されたタイプタグが所定のタイプタグに該当する場合、当該情報の全てのコンテンツ及び／又は当該情報の全てのコンテンツのリンクアドレスを所定の端末にプッシュ配信する。例えば、ユーザがスポーツ情報に興味を持っている場合、予めタイプタグを「スポーツ」に設定し、データソースから取得した情報を解析してポイント情報を得た後、当該情報のポイント情報に対応するタイプタグを認識し、認識されたタイプタグが「スポーツ」タグであれば、当該情報をユーザが興味を持っている情報として判断し、当該情報の全てのコンテンツ及び／又は当該情報の全てのコンテンツのリンクアドレスを所定の端末（例えばユーザの携帯電話、タブレットPCなど）にプッシュ配信することで、目的情報に対する効率的なマイニング及び正確なプッシュ配信を実現する。

20

【0049】

さらに、別の実施例として、ステップS20で取得した各情報に対して単語分割処理を行うプロセスにおいて、

順方向最大マッチング法に従って、各情報における処理対象文字列を汎用辞書にてマッチングすることで、第1マッチング結果を得て、

30

逆方向最大マッチング法に従って、各件情報における処理対象文字列を汎用辞書にてマッチングすることで、第2マッチング結果を得る。ただし、前記第1マッチング結果に第1の数の第1連語が含まれ、前記第2マッチング結果に第2の数の第2連語が含まれ、前記第1マッチング結果に第3の数の単一文字が含まれ、前記第2マッチング結果に第4の数の単一文字が含まれる。

前記第1の数と前記第2の数が同じで、且つ前記第3の数が前記第4の数以下である場合、前記第1マッチング結果（連語及び単一文字を含む）を出力し、前記第1の数と前記第2の数が同じで、且つ前記第3の数が前記第4の数より大きい場合、前記第2マッチング結果（連語及び単一文字を含む）を出力し、前記第1の数と前記第2の数が異なり、且つ前記第1の数が前記第2の数より大きい場合、前記第2マッチング結果（連語及び単一文字を含む）を出力し、前記第1の数と前記第2の数が異なり、且つ前記第1の数が前記第2の数より小さい場合、前記第1マッチング結果（連語及び単一文字を含む）を出力する。

40

【0050】

本実施例では、双方向マッチング法により、取得した各情報に対して単語分割処理を行い、順方向と逆方向から同時に単語分割を行ってマッチングすることにより、各情報の処理対象文字列における組み合わせコンテンツの関連性を分析し、一般的に、連語の方はポイント情報を表す確率がより高く、すなわち連語の方はポイント情報を表現するのにより適切である。よって、順方向と逆方向から同時に単語分割を行ってマッチングすることで

50

、単一文字の数がより少なく、連語の数がより多い分割単語マッチング結果を得て情報の単語分割結果とすることにより、単語分割及び情報マイニングにおける正確性を高めることができる。

【0051】

さらに、別の実施例として、ステップS20で各情報に対応する各分割単語に品詞タグをつけるプロセスにおいて、汎用辞書における文字及び単語と品詞とのマッピング関係（例えば、汎用辞書においては、操場（運動場）に対応する品詞が名詞である）及び／又は予め設定された文字及び単語と品詞とのマッピング関係（例えば、予め設定された文字及び単語と品詞とのマッピング関係においては、操場に対応する品詞が常用名詞である）に基づき、各情報の各分割単語に対応する品詞を決定し、各情報の各分割単語に対応する品詞を注記する。ただし、予め設定された文字及び単語と品詞とのマッピング関係に基づく品詞注記の優先度は、汎用辞書における文字及び単語と品詞とのマッピング関係よりも高い。例えば、汎用辞書によると、操場に対応する品詞が名詞であり、予め設定された文字及び単語と品詞とのマッピング関係によると、操場に対応する品詞が常用名詞である場合、注記するとき、先に予め設定された文字及び単語と品詞とのマッピング関係に基づき注記し、すなわち操場の品詞に常用名詞を注記する。

10

【0052】

さらに、別の実施例として、前記所定構造のワードツリーは、複数階層のノードを含み、第1階層のノードは原情報であり、第2階層のノードは分割単語からなる連語であり、第2階層のノード以後の各階層のノードは、いずれもその直上階層のノードに対応する直下階層の分割単語又は分割単語からなる連語である。ステップS30は、各情報に対応する各分割単語から、各所定品詞の目的分割単語を特定するステップA1と、各情報における各目的分割単語の順序に従って、各第2階層のノードに対応する分割単語からなる連語を決定するステップA2（具体的には、次の目的分割単語前の語をその前の目的分割単語の分割単語からなる連語とし、最後の目的分割単語及びその後の語を最後の分割単語からなる連語とする）と、分割単語からなる連語に対してさらなる単語分割が不可能な場合、当該分割単語からなる連語を、所属するノードレベルの最終階層のノードとするステップA3と、分割単語からなる連語に対してさらなる単語分割が可能な場合、当該分割単語からなる連語における各所定品詞の目的分割単語を特定し、当該分割単語からなる連語に対応する各目的分割単語の順序に従って、当該分割単語からなる連語の直下階層のノードに対応する分割単語、又は分割単語からなる連語を決定するステップA4と、各ノードレベルの最終階層のノードに対応する分割単語が決定されるまで、ステップA3及びステップA4を繰り返し実行するステップA5と、を含む。

20

30

【0053】

ステップS40は、構築された所定構造のワードツリーに基づき、各所定の第1重要品詞の分割単語と各所定の第2重要品詞の分割単語との間の距離を算出し、さらに、各所定の第1重要品詞の分割単語との距離が最小の所定の第2重要品詞の分割単語をそれぞれ特定し、各所定の第1重要品詞の分割単語と、それとの距離が最小の所定の第2重要品詞の分割単語を当該情報における順序に従い組み合わせて、対応するポイント情報を構成する。

40

【0054】

一つの具体的な実施形態として、図3（本発明の情報マイニング方法の一実施例における所定構造のワードツリーの概略図）に示すように、当該情報は「我去操場＊（あしへんに易）足球了（私は、運動場にサッカーをしに行った）」である場合、対応する単語分割結果は「我去操場＊（あしへんに易）足球」であり、品詞注記結果は「我／代詞 去／動詞 操場／常用名詞 ＊（あしへんに易）足球／常用名詞」である。図3は当該情報「我去操場＊（あしへんに易）足球」で構築された所定構造のワードツリーを示している。当該所定構造のワードツリーは複数階層のノードを含み、第1階層のノードは原情報であり、第2階層のノードは分割単語からなる連語（例えば、名詞連語、動詞連語及び句点などの区切り記号）である。本実施例では、各情報に対応する各分割単語から、各所定

50

品詞（例えば、名詞、動詞など）の目的分割単語を特定し、当該情報における各目的分割単語の順序に従って、各第2階層のノードに対応する分割単語からなる連語を決定する。第2階層のノード以後の各階層のノードは、いずれもその直上階層のノードに対応する直下階層の分割単語又は分割単語からなる連語であり、第3階層ノードは第2階層のノードの分割単語又は分割単語からなる連語である。図3に示すように、当該情報に対する品詞注記結果は「我／代詞 去／動詞 操場／常用名詞 ＊（あしへんに易）足球／常用名詞」であり、当該情報における各分割単語の順序に従って、左から右へ第2階層のノードを決定し、予め第2階層のノードを、名詞連語、動詞連語などを含む分割単語からなる連語に設定し、当該情報においては、左から右へ、「我」が代詞で名詞連語に該当するため、「我」を第2階層のノードに決定し、「我」に続く「去」、「操場」、「＊（あしへんに易）足球」は、動詞連語「去操場＊（あしへんに易）足球」を構成し、よって、「去操場＊（あしへんに易）足球」を第2階層のノードに決定する。これにより、当該情報の所定構造のワードツリーにおける第2階層のノードは「我」、「去操場＊（あしへんに易）足球」を含む。また、第2階層のノード「我」はさらなる単語分割が不可能であるため、当該分割単語からなる連語を所属するノードレベルの最終階層のノードに決定する。第2階層のノードである動詞連語「去操場＊（あしへんに易）足球」は、さらなる単語分割が可能であるため、当該第2階層のノードである「去操場＊（あしへんに易）足球」の分割単語又は分割単語からなる連語を第3階層ノードとすることができ、当該第3階層のノードは、動詞「去」及び名詞連語「操場＊（あしへんに易）足球」を含み、名詞連語「操場＊（あしへんに易）足球」をさらに第4階層のノード「操場」、「＊（あしへんに易）足球」に分割することができる。分割単語からなる連語に対してさらなる分割単語が可能な場合、当該分割単語からなる連語における各所定品詞（例えば、名詞、動詞など）の目的分割単語を特定し、当該分割単語からなる連語に対応する各目的分割単語の順序に従って、当該分割単語からなる連語の直下階層のノードに対応する分割単語又は分割単語からなる連語を決定し、分割単語からなる連語に対してさらなる単語分割が不可能な場合、当該分割単語からなる連語を、所属するノードレベルの最終階層のノードとする。

【0055】

構築された所定構造のワードツリーに基づき、各第1重要品詞（例えば、動詞）の分割単語と各第2重要品詞（例えば、名詞）の分割単語との距離を算出し、各第1重要品詞の分割単語と各第2重要品詞の分割単語の間におけるノード数を前記距離とする。ただし、第1重要品詞、第2重要品詞は、必要に応じて設定し、又は当該ユーザの検索履歴記録における中心的情報に一般的に対応する品詞に基づき設定する。各第1重要品詞の分割単語との距離が最小の第2重要品詞の分割単語をそれぞれ特定し、各第1重要品詞の分割単語と、それとの距離が最小の第2重要品詞の分割単語を当該情報における順序に従い組み合わせ、対応するポイント情報を構成する。例えば、図3に示すように、「去操場」、「去＊（あしへんに易）足球」を、マイニングにより得られた当該情報「我去操場＊（あしへんに易）足球」に対応するポイント情報とする。

【0056】

本発明は、さらに、コンピュータ読み取り可能な記憶媒体を提供し、当該コンピュータ読み取り可能な記憶媒体に情報マイニングシステムが記憶されており、当該情報マイニングシステムが少なくとも1つの処理手段により実行されることで、前記少なくとも1つの処理手段が、上記実施例の情報マイニングシステム方法のステップを実行し、当該情報マイニングシステム方法のステップS10、ステップS20、ステップS30などの詳細はすでに記載したため、ここでは説明を省略する。

【0057】

本明細書における用語「含む」、又はこれに類する意味の他の用語は、非排他的な包括を意味し、これにより、一連の要素を含むプロセス、方法、物品、もしくは装置は、これらの要素だけでなく、明記していない他の要素も含み、又はこのようなプロセス、方法、物品、もしくは装置に固有の要素をさらに含む。よって、特段の限定がない限り、文言「～を含む」などにより限定される要素は、当該要素を含むプロセス、方法、物品又は装置

に同一の要素がさらに存在する場合を排除しない。

【0058】

当業者は、実施形態に関する上記の記載により、上記の実施例の方法は、ソフトウェアと、必要な汎用ハードウェアプラットフォームとを組み合わせる方式でも、ハードウェアでも実現できるが、多くの場合には、前者がより好ましい実施形態であることを理解できるであろう。これにより、本発明の技術的解決手段の趣旨又は従来技術について改良した部分を、ソフトウェア製品の形で表現でき、当該コンピュータソフトウェア製品は、1つの記憶媒体（例えば、ROM／RAM、磁気ディスク、CDROM）に記憶されており、且つ、端末機器（例えば、携帯電話、コンピュータ、サーバ、空調装置、又はネットワーク機器）により本発明の各実施例に記載の方法を実行するための複数のコマンドを含む。

10

【0059】

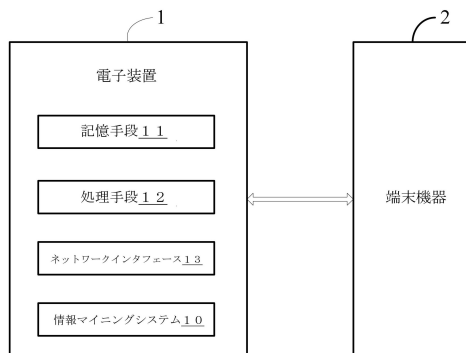
以上、図面を参照しながら本発明の好ましい実施例について説明したが、これをもって本発明の特許請求の範囲を限定するものではなく、本発明の実施例に付与した番号は、記載上の便宜を図るためのものに過ぎず、実施例に優劣をつけるものではない。また、フローチャートでは論理的な順序を示しているが、記載されたステップを異なる順序で実行してもよい。

【0060】

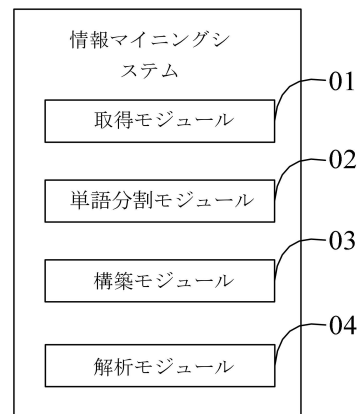
当業者であれば、本発明の範囲及び趣旨から逸脱しない範囲において、種々の変更を加えて本発明を実現することができ、例えば、一つの実施例の特徴を別の実施例に組み込むことで、別の実施例を得ることができる。本発明の趣旨に対して行う変更、均等な置換、又は改良などであれば、いずれも本発明の特許請求の範囲に含まれるものとする。

20

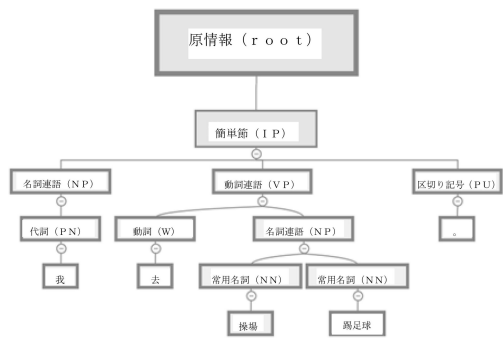
【図1】



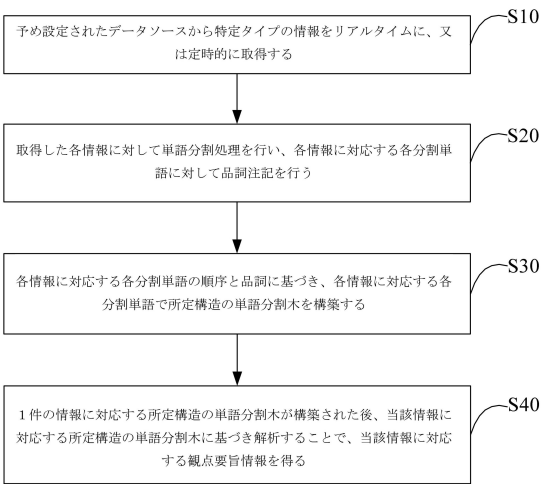
【図2】



【図 3】



【図 4】



フロントページの続き

(74)代理人 110002262

T R Y 国際特許業務法人

(72)発明者 金 戈

中国広東省深▲せん▼市福田区八卦嶺八卦三路平安大厦

(72)発明者 徐 亮

中国広東省深▲せん▼市福田区八卦嶺八卦三路平安大厦

(72)発明者 肖 京

中国広東省深▲せん▼市福田区八卦嶺八卦三路平安大厦

審査官 甲斐 哲雄

(56)参考文献 特開 2 0 0 6 - 0 3 1 1 4 3 (J P , A)

中国特許出願公開第 1 0 2 9 1 5 2 9 9 (C N , A)

特開 2 0 0 4 - 1 1 0 2 0 0 (J P , A)

(58)調査した分野(Int.Cl., D B 名)

G 0 6 F 1 6 / 0 0 - 1 6 / 9 5 8

G 0 6 F 4 0 / 2 6 3