

【公報種別】特許法第17条の2の規定による補正の掲載

【部門区分】第6部門第3区分

【発行日】平成26年10月16日(2014.10.16)

【公開番号】特開2014-135045(P2014-135045A)

【公開日】平成26年7月24日(2014.7.24)

【年通号数】公開・登録公報2014-039

【出願番号】特願2013-247330(P2013-247330)

【国際特許分類】

G 06 F 17/30 (2006.01)

【F I】

G 06 F 17/30 2 1 0 D

G 06 F 17/30 3 5 0 C

G 06 F 17/30 4 1 9 A

【手続補正書】

【提出日】平成26年9月2日(2014.9.2)

【手続補正1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】

未分類の入力文書と複数の分野に分類された少なくとも1以上の文書との類似度、及び前記分野分類された文書が属する分野の数を用いて、前記未分類の入力文書及び前記分野分類された文書間の総類似度を求める第一の分野判定手段と、

前記類似度を用いることで、前記分野分類された文書が属するそれぞれの分野における、前記未分類の入力文書及び前記分野分類された文書間の総類似度を求める第二の分野判定手段と、

前記第一の分野判定手段によって求まる文書間の総類似度に対する第二の分野判定手段によって求まる分野の文書間の総類似度から、前記入力文書を分野に分類するためのスコアを求める第三の分野判定手段と、

を備えたことを特徴とする文書分類装置。

【請求項2】

前記分野は、階層的に分類された分野であり、

前記第一の分野判定手段は、前記類似度、及び前記階層的に分類された分野の数を用いて、前記未分類の入力文書及び前記分野分類された文書間の総類似度を求め、

前記第二の分野判定手段は、前記類似度を用いることで、前記分野分類された文書が属する階層的に分類された分野のうち、各階層のそれぞれの分野における、前記未分類の入力文書及び前記分野分類された文書間の総類似度を求ることを特徴とする請求項1に記載の文書分類装置。

【請求項3】

前記未分類の入力文書は、複数の入力文書であり、

前記入力文書のそれぞれに対して、それぞれの分野における前記スコアを用いて、前記入力文書間の類似度を求める第四の分野判定手段を備えたことを特徴とする請求項1または2に記載の情報処理装置。

【請求項4】

前記第三の分野判定手段は、前記未分類の入力文書における単語の頻出度を求めてスコアとし、

前記第四の分野判定手段は、前記入力文書のそれぞれに対して、それぞれの分野における前記スコア及び前記単語の頻出度から求まるスコアを用いて、前記入力文書間の類似度を求ることを特徴とする請求項3に記載の情報処理装置。

【請求項5】

前記第四の分野判定手段は、前記入力文書のそれぞれに対して、それぞれの分野における前記スコアを用いて求まる前記入力文書間の類似度と、前記単語の頻出度から求まる前記分野に応じたスコアを用いた前記入力文書間の類似度と、から前記入力文書間の類似度を求ることを特徴とする請求項3または4に記載の情報処理装置。

【請求項6】

前記第四の分野判定手段によって求まる入力文書間の類似度を用いた階層的クラスタリングによって、前記入力文書間の関係を、階層を有するクラスタとして出力する出力手段を備えたことを特徴とする請求項4または5に記載の情報処理装置。

【請求項7】

前記第二の分野判定手段は、前記階層の深さに応じた各階層のそれぞれの分野における文書間の総類似度を求ることを特徴とする請求項2乃至6の何れか1項に記載の情報処理装置。

【請求項8】

文書分類装置における未分類の文書を分野分類する文書分類方法であって、

前記文書分類装置の第一の分野判定手段は、未分類の入力文書と複数の分野に分類された少なくとも1以上の文書との類似度、及び前記分野分類された文書が属する分野の数を用いて、前記未分類の入力文書及び前記分野分類された文書間の総類似度を求める第一の分野判定ステップ、

前記文書分類装置の第二の分野判定手段は、前記類似度を用いることで、前記分野分類された文書が属するそれぞれの分野における、前記未分類の入力文書及び前記分野分類された文書間の総類似度を求める第二の分野判定ステップ、

前記文書分類装置の第三の分野判定手段は、前記第一の分野判定手段によって求まる文書間の総類似度に対する第二の分野判定手段によって求まる分野の文書間の総類似度から、前記入力文書を分野に分類するためのスコアを求める第三の分野判定ステップ、
を含むことを特徴とする文書分類方法。

【請求項9】

文書分類装置で読み取実行可能なプログラムであって、

前記文書分類装置を、

未分類の入力文書と複数の分野に分類された少なくとも1以上の文書との類似度、及び前記分野分類された文書が属する分野の数を用いて、前記未分類の入力文書及び前記分野分類された文書間の総類似度を求める第一の分野判定手段と、

前記類似度を用いることで、前記分野分類された文書が属するそれぞれの分野における、前記未分類の入力文書及び前記分野分類された文書間の総類似度を求める第二の分野判定手段と、

前記第一の分野判定手段によって求まる文書間の総類似度に対する第二の分野判定手段によって求まる分野の文書間の総類似度から、前記入力文書を分野に分類するためのスコアを求める第三の分野判定手段と、

して機能させることを特徴とするプログラム。

【手続補正2】

【補正対象書類名】明細書

【補正対象項目名】0011

【補正方法】変更

【補正の内容】

【0011】

上記課題を解決するための第1の発明は、未分類の入力文書と複数の分野に分類された少なくとも1以上の文書との類似度、及び前記分野分類された文書が属する分野の数を用

いて、前記未分類の入力文書及び前記分野分類された文書間の総類似度を求める第一の分野判定手段と、前記類似度を用いることで、前記分野分類された文書が属するそれぞれの分野における、前記未分類の入力文書及び前記分野分類された文書間の総類似度を求める第二の分野判定手段と、前記第一の分野判定手段によって求まる文書間の総類似度に対する第二の分野判定手段によって求まる分野の文書間の総類似度から、前記入力文書を分野に分類するためのスコアを求める第三の分野判定手段と、を備えたことを特徴とする。

【手続補正3】

【補正対象書類名】明細書

【補正対象項目名】0012

【補正方法】変更

【補正の内容】

【0012】

上記課題を解決するための第2の発明は、文書分類装置における未分類の文書を分野分類する文書分類方法であって、前記文書分類装置の第一の分野判定手段は、未分類の入力文書と複数の分野に分類された少なくとも1以上の文書との類似度、及び前記分野分類された文書が属する分野の数を用いて、前記未分類の入力文書及び前記分野分類された文書間の総類似度を求める第一の分野判定ステップ、前記文書分類装置の第二の分野判定手段は、前記類似度を用いることで、前記分野分類された文書が属するそれぞれの分野における、前記未分類の入力文書及び前記分野分類された文書間の総類似度を求める第二の分野判定ステップ、前記文書分類装置の第三の分野判定手段は、前記第一の分野判定手段によって求まる文書間の総類似度に対する第二の分野判定手段によって求まる分野の文書間の総類似度から、前記入力文書を分野に分類するためのスコアを求める第三の分野判定ステップ、を含むことを特徴とする。

【手続補正4】

【補正対象書類名】明細書

【補正対象項目名】0013

【補正方法】変更

【補正の内容】

【0013】

上記課題を解決するための第3の発明は、文書分類装置で読み取実行可能なプログラムであって、前記文書分類装置を、未分類の入力文書と複数の分野に分類された少なくとも1以上の文書との類似度、及び前記分野分類された文書が属する分野の数を用いて、前記未分類の入力文書及び前記分野分類された文書間の総類似度を求める第一の分野判定手段と、前記類似度を用いることで、前記分野分類された文書が属するそれぞれの分野における、前記未分類の入力文書及び前記分野分類された文書間の総類似度を求める第二の分野判定手段と、前記第一の分野判定手段によって求まる文書間の総類似度に対する第二の分野判定手段によって求まる分野の文書間の総類似度から、前記入力文書を分野に分類するためのスコアを求める第三の分野判定手段と、して機能させることを特徴とする。