

【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20年1月31日(2008.1.31)

【公開番号】特開2002-169834(P2002-169834A)

【公開日】平成14年6月14日(2002.6.14)

【出願番号】特願2000-353475(P2000-353475)

【国際特許分類】

G 06 F 17/30 (2006.01)

【F I】

G 06 F 17/30 350 C

G 06 F 17/30 170 A

G 06 F 17/30 210 A

【手続補正書】

【提出日】平成19年12月7日(2007.12.7)

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】特許請求の範囲

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】 入力文書から重要な文書セグメントを抽出する方法であって、前記入力文書に出現する用語を検出するステップと、前記入力文書を適当な単位の文書セグメントに区分けするステップと、前記文書セグメントに出現する用語の出現頻度に応じた値を成分とする文書セグメントのベクトルを生成するステップと、

前記文書セグメントベクトルのランクがRで示される平方和行列の固有ベクトルおよび固有値を求めるステップと、

前記複数の固有ベクトルから重要度判定に使用する複数Lの固有ベクトルを選択するステップと、

選択されたそれぞれの固有ベクトルへのそれぞれの文書セグメントベクトルの射影値の重み付き自乗和を求めるステップと、を含み、

それぞれの文書セグメントベクトルの射影値の自乗和に基づいて重要度の大きい文書セグメントを選択する、コンピュータによる重要文抽出方法。

【請求項2】 K個の用語が現れる入力文書がN個の文書セグメントに分割され、 d_{ni} がi番目の用語の文書セグメントnにおける出現頻度に応じた値として、n番目の文書セグメントベクトル d_n ($n=1, \dots, N$)を $(d_{n1}, d_{n2}, \dots, d_{nK})$ により表し、 d_n の射影後のベクトルを $z_n = (z_{n1}, z_{n2}, \dots, z_{nL})$ とすると、m番目の固有ベクトルへの射影値は、 $z_{nm} = z_n^T d_m$ (d_m はm次の固有ベクトル、 T は転置を表す)で与えられ、L次の部分空間への射影値の自乗和は、

【数1】

$$\sum_{m=1}^L z_{nm}^2$$

または、

【数2】

$$\sum_{m=1}^L \lambda_m z_{nm}^2 \quad (\lambda_m \text{は } m \text{番目の固有ベクトルの固有値})$$

により与えられる請求項 1 に記載の方法。

【請求項 3】 前記固有ベクトルおよび固有値は、平方和行列

【数 3】

$$A_{ab} = \sum_{n=1}^N d_{na} d_{nb}$$

から求められ、そのランク R に対して前記 L は、文書の中心概念を表す部分空間の次数を示し、

【数 4】

$$\sum_{m=1}^L \lambda_m / \sum_{m=1}^R \lambda_m$$

により与えられる前記入力文書の全体的概念に対する中心概念の割合に基づいて L の値を決める請求項 2 に記載の方法。

【請求項 4】 入力文書から質問文と関連性をもつ文書セグメントを抽出する方法であって、

前記入力文書に出現する用語を検出し、前記入力文書を適当な単位の文書セグメントに区分けし、該文書セグメントに出現する用語の出現頻度に応じた値を成分とする文書セグメントのベクトルを生成するステップと、

前記文書セグメントベクトルの平方和行列の固有ベクトルおよび固有値を求めて部分空間を決定するステップと、

前記質問文に出現する用語を検出し、該用語の出現頻度に応じた値を成分とする質問文のベクトルを生成し、前記部分空間に射影するステップと、

前記文書セグメントのベクトルのそれぞれを前記部分空間に射影し、前記質問文と文書セグメントとの関連度を算出するステップと、

を含む、コンピュータによる文書セグメントの抽出方法。

【請求項 5】 前記文書セグメントのベクトル d_n の前記部分空間への射影ベクトルを z_n 、前記質問文のベクトル q を前記部分空間に射影したベクトルを y とすると、 n 番目の文書セグメントと質問文との関連度 g_n は、 y と z_n の内積 $y^t z_n$ (t は転置) に基づいて求められる請求項 4 に記載の方法。

【請求項 6】 前記質問文のベクトルを q とし、 m 番目の前記固有ベクトルを ϕ_m とし、 L を前記部分空間の次数とするとき、 m 次の固有ベクトルに対する重み s_m を $(\phi_m^t q)^2$ の関数で定義し、文書セグメント n に対する関連度 g_n を

【数 5】

$$g_n = \sum_{m=1}^L s_m (\phi_m^t d_n)^2$$

で求める請求項 4 に記載の方法。

【請求項 7】 与えられた 2 つの入力文書の類似性を判定する文書間類似性判定方法であって、

それぞれの入力文書に出現する用語を検出するステップと、

それぞれの入力文書を適当な単位の文書セグメントに区分けするステップと、

前記文書セグメントに出現する用語の出現頻度に応じた値を成分とする文書セグメントのベクトルを生成するステップと、

それぞれの入力文書について、それぞれの入力文書に含まれる前記文書セグメントのベクトルのすべての組み合わせについて内積の自乗を求めるステップと、を含み、

前記内積の自乗の和に基づいて入力文書間の類似性を判定する、コンピュータによる文書間類似性判定方法。

【請求項 8】 与えられた 2 つの入力文書の類似性を判定する文書間類似性判定方法であって、

一方の入力文書の文書セグメントベクトルの平方和行列の固有ベクトルおよび固有値を算出するステップと、

前記固有ベクトルから固有値の大きいものを基底ベクトルとして選択するステップと、

他方の文書の文書セグメントベクトルと前記基底ベクトルとの内積の自乗を求め、重み付けして加えあわせるステップと、を含み、

加え合わせられた内積の自乗の和から前記2つの入力文書の間の類似性を判定する、コンピュータによる文書間類似性判定方法。

【請求項9】 与えられた2つの入力文書の類似性を判定する文書間類似性判定方法であって、

前記2つの入力文書のそれぞれについて文書セグメントベクトルの平方和行列の固有ベクトルおよび固有値を算出するステップと、

前記2つの入力文書のそれぞれについて前記固有ベクトルから固有値の大きいものを基底ベクトルとして選択するステップと、

選択された基底ベクトルの組み合わせについて内積の自乗を求め、重み付けして加えあわせるステップと、を含み、

加え合わせられた内積の自乗の和から入力文書間の類似性を判定する、コンピュータによる文書間類似性判定方法。