

【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19年7月12日(2007.7.12)

【公開番号】特開2005-352888(P2005-352888A)

【公開日】平成17年12月22日(2005.12.22)

【年通号数】公開・登録公報2005-050

【出願番号】特願2004-174516(P2004-174516)

【国際特許分類】

**G 06 F 17/30 (2006.01)**

**G 06 F 17/21 (2006.01)**

【F I】

G 06 F 17/30 3 2 0 D

G 06 F 17/30 1 7 0 J

G 06 F 17/30 3 5 0 C

G 06 F 17/21 5 5 0 K

【手続補正書】

【提出日】平成19年5月28日(2007.5.28)

【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】

検索語として与えられた用語の表記揺れを抽出するシステムにおいて、

テキスト文書から用語の集合を収集する用語収集部と、

前記用語収集部によって収集された用語の集合の中から前記検索語に類似した用語群を検索する類似用語検索部と、

前記類似用語検索部によって検索された用語群の中から前記検索語の表記揺れを抽出する表記揺れ検索部とを備え、

前記類似用語検索部は、1文字ずつずらした隣接する所定長の部分文字列の共有度を基準にして、比較する2つの用語の類似度を判定し、

前記表記揺れ検索部は、前記検索語との編集距離の総コストが与えられた閾値より小さい用語を前記検索語の表記揺れとして抽出することを特徴とする表記揺れ抽出システム。

【請求項2】

請求項1記載の表記揺れ抽出システムにおいて、前記表記揺れ検索部は、2つの用語の編集距離を、文字の置換、挿入、削除に対して割り当てられたコストを用いて計算することを特徴とする表記揺れ抽出システム。

【請求項3】

請求項1記載の表記揺れ抽出システムにおいて、前記類似用語検索部は、前記検索語との文字列数の相違が許容値内にある用語を対象として前記検索語に類似した用語群を検索することを特徴とする表記揺れ抽出システム。

【請求項4】

請求項1記載の表記揺れ抽出システムにおいて、前記検索語の文字列を1文字ずつずらした部分文字列の索引を作成する索引作成部を有することを特徴とする表記揺れ抽出システム。

【請求項5】

コンピュータを用いて検索語として入力された用語の表記揺れを抽出する方法において

、コンピュータが、  
指定されたテキスト文書から用語の集合を収集する工程と、  
前記収集された用語の集合の中から、1文字ずつずらした隣接する所定長の部分文字列  
の共有度を基準にして、比較する2つの用語の類似度を判定し、前記検索語に類似した用  
語群を検索する類似用語検索工程と、  
検索された前記用語群の中から、前記検索語との編集距離の総コストが与えられた閾値  
より小さい用語を前記検索語の表記揺れとして抽出し、前記検索語の表記揺れを抽出する  
表記揺れ検索工程と、  
を実行する表記揺れ抽出方法。