

【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12月27日(2007.12.27)

【公開番号】特開2007-25834(P2007-25834A)

【公開日】平成19年2月1日(2007.2.1)

【年通号数】公開・登録公報2007-004

【出願番号】特願2005-203799(P2005-203799)

【国際特許分類】

G 06 F 17/30 (2006.01)

G 06 F 17/21 (2006.01)

G 06 Q 50/00 (2006.01)

A 61 B 5/00 (2006.01)

【F I】

G 06 F 17/30 320D

G 06 F 17/21 550J

G 06 F 17/30 170A

G 06 F 17/60 126A

A 61 B 5/00 G

【手続補正書】

【提出日】平成19年11月12日(2007.11.12)

【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】

複数文字列からなるテキストデータを入力する入力手段と、

文字種や単語列のパターンにより定義される文字列分解規則を格納する手段と、

所定の統制用語集を格納する手段と、

前記文字列分解規則に基づき前記統制用語集内の用語をあらかじめ構成素ごとに切断し、該切断された部分文字列が含まれる統制用語を候補用語とする部分文字列辞書を作成する手段と、作成された該部分文字列辞書を格納する手段と、

前記テキストデータ中に出現する前記部分文字列辞書の文字列を全て検索するテキスト検索手段と、

検索された前記部分文字列のうち、共通の前記候補用語を持つ文字列同士を結合し、結合文字列のテキスト上での位置関係と、前記候補用語上での位置関係に基づき完成度を計算する手段と、結合文字列の候補用語と完成度を記録する手段とを備え、

記録された文字列集合の中から、前記完成度の和、及びテキストの被覆率が所定値以上の文字列の集合を選択する手段を備え、選択された前記文字列集合うち、統制用語としての完成度が所定の閾値以上のものを出力手段に出力することを特徴とする表記ゆれ処理システム。

【請求項2】

請求項1に記載の表記ゆれシステムにおいて、

前記部分文字列の完成度は、結合される2つの文字列の間に存在するテキスト文字列に基づき、あらかじめ定義された文字列コスト表のコストおよび文字数に従って完成度を計算することを特徴とする表記ゆれシステム。

【請求項3】

請求項 1 に記載の表記ゆれシステムにおいて、

前記部分文字列辞書は、数量表現、単位表現等の文字列のパターンもしくは直接文字列を登録した重み付け定義に従った完成度を記録しており、結合される文字列の完成度を、各部分文字列の完成度の和により計算することを特徴とする表記ゆれシステム。

【請求項 4】

請求項 1 に記載表記ゆれシステムにおいて、

前記テキスト検索手段は、ハミング距離があらかじめ指定した値以下もしくは、編集距離があらかじめ指定した値以下の近似部分文字列を検索することを特徴とする表記ゆれシステム。