【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12月4日(2008.12.4)

【公開番号】特開2008-251003(P2008-251003A)
【公開日】平成20年10月16日(2008.10.16)
【年通号数】公開・登録公報2008-041
【出願番号】特願2008-63725(P2008-63725)
【国際特許分類】
Ｇ０６Ｆ　17/30　(2006.01)
Ｇ０６Ｆ　17/27　(2006.01)
Ｇ０６Ｆ　17/21　(2006.01)
【ＦＩ】
Ｇ０６Ｆ　17/30　２２０Ａ
Ｇ０６Ｆ　17/30　１７０Ａ
Ｇ０６Ｆ　17/27　Ｚ
Ｇ０６Ｆ　17/21　５５０Ａ

【手続補正書】
【提出日】平成20年6月16日(2008.6.16)
【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】
テキストの１つ以上の断片を処理するテキスト処理システムにおいて、
メモリ及び処理ユニットを有するコンピュータを備え、
上記テキストの各断片を上記コンピュータのメモリから受け取って処理し、そしてテキストの各断片についてその断片を特徴付ける１つ以上のキーフレーズを抽出するパーザシステムを備え、
上記抽出したキーフレーズを記憶するデータベースを備え、
上記パーザシステムは、上記コンピュータの処理ユニットによって実行され、
上記パーザシステムは：
テキストの断片の１つ以上のワードを記憶するバッファを有し；
パージングルールを記憶するルールデータベースを有し；
上記パージングルールに基づいて各テキストの断片を分析し、１つ以上のキーフレーズ出力するバーザを有し、各キーフレーズは１つ以上のワードを含み；
上記パーザは、文書内のフレーズを、その文書内にあるブレークキャラクタに基づいて分離するように構成され、上記ブレークキャラクタは、明確な中断、数字、数字を含むワード、ソフト終止ワード及びハード終止ワードを含み；そして
上記パーザは、上記コンピュータの処理ユニットによって実行されて、
文書の１つのワードを上記コンピュータのメモリからバッファへロードする段階と、
上記ワードがブレークキャラクタかどうかを判定する段階と、
上記ワードがブレークキャラクタでない場合に、上記ロードする段階に戻って文書の次のワードをバッファへロードする段階と、
上記ワードがブレークキャラクタでない場合に、前後にブレークキャラクタを有するワードのシーケンスから成るフレーズを上記バッファに記憶する段階と、

上記ブレークキャラクタが明確なブレークキャラクタ、ソフト終止ワード又はハード終止ワードかどうかを判定する段階と、
上記ブレークキャラクタが明確なブレークキャラクタ又はソフト終止ワードである場合に、上記明確なブレークキャラクタ又はソフト終止ワードを削除し、バッファに記憶されたフレーズを抽出し、該抽出されたフレーズをキーフレーズとしてデータベースに記憶する段階と、
上記ブレークキャラクタかハード終止ワードである場合に、上記ハード終止ワードとバッファに記憶されたフレーズを削除する段階と、
上記バッファをフラッシュする段階と、
上記ロードする段階に戻る段階と、
から成る動作を行う；
ことを特徴とするテキスト処理システム。
【請求項２】
文書のテキストをその文書を特徴付ける１つ以上のキーフレーズへパージングするテキストパージングシステムにおいて、
上記テキストの１つ以上のワードを記憶するバッファを有し；
パージングルールを記憶するルールデータベースを有し；
上記パージングルールに基づいて上記テキストを分析して上記テキストの１つ以上のキーフレーズを識別するようにプログラムによってプログラムされたプロセッサを有し；
上記キーフレーズを記憶するフレーズデータベースを有し；
上記プロセッサは：
（１）上記テキストのワードを上記バッファへロードし、
（２）上記バッファにロードされたワードが所定のブレークキャラクタの１つかどうかを判定し、
（３）上記ワードが上記所定のブレークキャラクタの１つであると判定された場合に、上記所定のブレークキャラクタの１つに対応する上記ルールデータベースのパージングルールを参照して、そのブレークキャラクタの後に続くバッファのワードがキーフレーズであるかどうかを判定し、そして
（４）上記ワードがキーフレーズであると判定された場合に、上記キーフレーズを上記フレーズデータベースへセーブする、
ように構成されたことを特徴とするテキストパージングシステム。
【請求項３】
上記プロセッサは、更に、上記キーフレーズが上記フレーズデータベースにセーブされたときに上記バッファをフラッシュするように構成された請求項２に記載のテキストパージングシステム。
【請求項４】
上記所定のブレークキャラクタは、ソフト終止ブレークキャラクタ、明確な終止ブレークキャラクタ及びハード終止ブレークキャラクタから成る請求項２に記載のテキストパージングシステム。
【請求項５】
上記プロセッサが上記ワードをハード終止ブレークキャラクタであると判定した場合に、上記対応するパージングルールによって上記プロセッサが、上記ハード終止ブレークキャラクタの後に続くワードを上記バッファから削除する請求項４に記載のテキストパージングシステム。
【請求項６】
上記明確なブレークキャラクタは、句読点から成る請求項４に記載のテキストパージングシステム。
【請求項７】
上記ルールデータベースは、バッファの各ワードに適用されるワードレベルパージングルールと、バッファの２つ以上のワードに適用されるフレーズレベルパージングルールと

を含む請求項２に記載のテキストパージングシステム。
【請求項８】
上記フレーズレベルパージングルールは、所定数のワードよりも長いフレーズを削除する１つ以上のルール、大文字のワードのみをもつフレーズを削除するルール及び数字を表すワードのみをもつフレーズを削除するルールから成る請求項７に記載のテキストパージングシステム。
【請求項９】
文書のテキストをその文書を特徴付ける１つ以上のキーフレーズへとパージングするテキストパージングシステムを制御する方法において、
上記システムは：
上記テキストの１つ以上のワードを記憶するバッファと；
パージングルールを記憶するルールデータベースと；
上記パージングルールに基づいて上記テキストを分析して上記テキストの１つ以上のキーフレーズを識別するようにプログラムによってプログラムされたプロセッサと；
上記キーフレーズを記憶するフレーズデータベースと；
を有し、
上記方法は：
上記テキストのワードを上記バッファへロードするように上記プロセッサに命令し；
上記バッファにロードされたワードが所定のブレークキャラクタの１つかどうかを判定するように上記プロセッサに命令し；
上記ワードが上記所定のブレークキャラクタの１つであると判定された場合に、上記所定のブレークキャラクタの１つに対応する上記ルールデータベースのパージングルールを参照して、そのブレークキャラクタの後に続くバッファのワードがキーフレーズであるかどうかを判定するように上記プロセッサに命令し；そして
上記ワードがキーフレーズであると判定された場合に、上記キーフレーズを上記フレーズデータベースへセーブするように上記プロセッサに命令する；
という段階を含む、
ことを特徴とする方法。
【請求項１０】
上記キーフレーズが上記フレーズデータベースにセーブされたときに上記バッファをフラッシュするように上記プロセッサに命令する段階を更に含む請求項９に記載の方法。
【請求項１１】
上記所定のブレークキャラクタは、ソフト終止ブレークキャラクタ、明確な終止ブレークキャラクタ及びハード終止ブレークキャラクタから成る請求項９に記載の方法。
【請求項１２】
上記ワードがハード終止ブレークキャラクタであると判定された場合に、そのハード終止ブレークキャラクタに対応する上記パージングルールデータベースのパージングルールを参照して、そのハード終止ブレークキャラクタの後に続くワードを上記バッファから削除するように上記プロセッサに命令する段階を更に含む請求項１１に記載の方法。
【請求項１３】
上記明確なブレークキャラクタは、句読点から成る請求項１１に記載の方法。
【請求項１４】
上記ルールデータベースは、バッファの各ワードに適用されるワードレベルパージングルールと、バッファの２つ以上のワードに適用されるフレーズレベルパージングルールとを含む請求項９に記載の方法。
【請求項１５】
上記フレーズレベルパージングルールは、所定数のワードよりも長いフレーズを削除する１つ以上のルール、大文字のワードのみをもつフレーズを削除するルール及び数字を表すワードのみをもつフレーズを削除するルールから成る請求項１４に記載の方法。
【請求項１６】

文書のテキストをその文書を特徴付ける１つ以上のキーフレーズへとパージングするテキストパージング方法をコンピュータに実行させるプログラムにおいて、
上記コンピュータは：
上記テキストの１つ以上のワードを記憶するバッファと；
パージングルールを記憶するルールデータベースと；
上記パージングルールに基づいて上記テキストを分析して上記テキストの１つ以上のキーフレーズを識別するようにプログラムによってプログラムされたプロセッサと；
上記キーフレーズを記憶するフレーズデータベースと；
を有し、
上記方法は：
上記テキストのワードを上記バッファへロードするように上記プロセッサに命令し；
上記バッファにロードされたワードが所定のブレークキャラクタの１つかどうかを判定するように上記プロセッサに命令し；
上記ワードが上記所定のブレークキャラクタの１つであると判定された場合に、上記所定のブレークキャラクタの１つに対応する上記ルールデータベースのパージングルールを参照して、そのブレークキャラクタの後に続くバッファのワードがキーフレーズであるかどうかを判定するように上記プロセッサに命令し；そして
上記ワードがキーフレーズであると判定された場合に、上記キーフレーズを上記フレーズデータベースへセーブするように上記プロセッサに命令する；
という段階を含む、
ことを特徴とするプログラム。