【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6月7日(2007.6.7)

【公表番号】特表2002-541580(P2002-541580A)
【公表日】平成14年12月3日(2002.12.3)
【出願番号】特願2000-611158(P2000-611158)
【国際特許分類】
Ｇ０６Ｆ　17/27　(2006.01)
Ｇ０６Ｆ　17/30　(2006.01)
【ＦＩ】
Ｇ０６Ｆ　17/27　Ｅ
Ｇ０６Ｆ　17/30　１７０Ａ
Ｇ０６Ｆ　17/30　２１０Ａ
Ｇ０６Ｆ　17/30　３３０Ｃ

【手続補正書】
【提出日】平成19年4月6日(2007.4.6)
【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】　テキストの断片を、文書を特徴付ける１つ以上のフレーズへとパージングするシステムにおいて、
ブレークキャラクタが識別されるまでテキストの断片からの１つ以上のワードを読み込むためのバッファを備え、
上記バッファに含まれたフレーズを識別するためのパーザを備え、上記フレーズは、ブレークキャラクタ間における２つ以上のワードのシーケンスであり、
上記パーザは、識別されたフレーズに続くブレークキャラクタの形式を決定する手段と、ブレークキャラクタの決定された形式に基づいて異なるパージングルールを使用して上記識別されたフレーズを処理する手段とを含み、上記ブレークキャラクタの形式は、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のうちの１つから成る、
ことを特徴とするシステム。
【請求項２】　上記バッファは、更に、キーフレーズがデータベースに記憶されるか又はバッファのフレーズが削除されたときにバッファをフラッシュする手段を含む請求項１に記載のシステム。
【請求項３】　上記ブレークキャラクタの形式は、更に、ハード終止ブレークキャラクタから成り、上記決定手段は、フレーズがバッファから削除されてデータベースに記憶されないことを指示するハード終止ブレークキャラクタを識別する手段を含む請求項２に記載のシステム。
【請求項４】　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３に記載のシステム。
【請求項５】　テキストの断片をキーフレーズへとパージングするためにテキストの断片に適用すべき１つ以上のパージングルールを記憶するルールデータベースを更に備えた請求項１に記載のシステム。
【請求項６】　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、テキストの断片に選択的に適用される任意のフレーズレベル

パージングルールとを含む請求項５に記載のシステム。
【請求項７】　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の１つ以上のルールから成る請求項６に記載のシステム。
【請求項８】　テキストの断片を、文書を特徴付ける１つ以上のフレーズへとパージングする方法において、
ブレークキャラクタが識別されるまでテキストの断片からの１つ以上のワードをバッファに読み込む段階を含み、
バッファに含まれたフレーズを識別する段階を含み、このフレーズは、ブレークキャラクタ間における２つ以上のワードのシーケンスであり、
その識別されたフレーズに続くブレークキャラクタの形式を決定する段階を含み、上記ブレークキャラクタの形式は、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のうちの１つから成り、そして
上記ブレークキャラクタの決定された形式に基づいて異なるパージングルールを使用して上記識別されたフレーズを処理する段階を含む、
ことを特徴とする方法。
【請求項９】　キーフレーズがデータベースに記憶されるか又はバッファのフレーズが削除されたときにバッファをフラッシュする段階を更に含む請求項８に記載の方法。
【請求項１０】　上記ブレークキャラクタの形式は、更に、ハード終止ブレークキャラクタから成り、上記決定する段階は、フレーズがバッファから削除されてデータベースに記憶されないことを指示するハード終止ブレークキャラクタを識別する段階を含む請求項９に記載の方法。
【請求項１１】　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１０に記載の方法。
【請求項１２】　テキストの断片をキーフレーズへとパージングするためにテキストの断片に適用すべき１つ以上のパージングルールを、ルールデータベースに記憶する段階を更に含む請求項８に記載の方法。
【請求項１３】　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、テキストの断片に選択的に適用される任意のフレーズレベルパージングルールとを含む請求項１２に記載の方法。
【請求項１４】　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の１つ以上のルールから成る請求項１３に記載の方法。
【請求項１５】　テキストの断片を、文書を特徴付ける１つ以上のフレーズへとパージングするシステムにおいて、
ブレークキャラクタが識別されるまでテキストの断片からの１つ以上のワードを読み込むためのバッファを備え、
上記バッファに含まれたフレーズを識別するためのパーザを備え、上記フレーズは、ブレークキャラクタ間における２つ以上のワードのシーケンスであり、
上記パーザは、識別されたフレーズに続くブレークキャラクタの形式を決定するブレークキャラクタ識別器と、上記ブレークキャラクタの決定された形式に基づいて異なるパージングルールを使用して上記識別されたフレーズを処理する処理ユニットとを含み、上記ブレークキャラクタの形式は、ソフト終止ブレークキャラクタ、ハード終止ブレークキャラクタ及び明確な終止ブレークキャラクタのうちの１つから成り、
更に、上記キーフレーズを記憶するためのデータベースを備えた、
ことを特徴とするシステム。
【請求項１６】　テキストの断片を、文書を特徴付ける１つ以上のフレーズへとパージングする方法において、

ブレークキャラクタが識別されるまでテキストの断片からの１つ以上のワードをバッファに読み込む段階を含み、
バッファに含まれたフレーズを識別する段階を含み、このフレーズは、ブレークキャラクタ間における２つ以上のワードのシーケンスであり、
バッファにおいてフレーズの後に明確な終止ブレークキャラクタが識別されるのに応答してそのフレーズをキーフレーズとしてセーブする段階を含み、
バッファにおいてフレーズの後にソフト終止ブレークキャラクタが識別されるのに応答してそのフレーズをキーフレーズとしてセーブする段階を含み、
バッファにおいてフレーズの後にハード終止ブレークキャラクタが識別されるのに応答してそのフレーズをバッファから削除する段階を含み、
フレーズが削除又はセーブされるとワードをバッファからフラッシュする段階を含む、
ことを特徴とする方法。

【請求項１７】　テキストの断片を、文書を特徴付ける１つ以上のフレーズへとパージングするシステムにおいて、
ブレークキャラクタが識別されるまでテキストの断片からの１つ以上のワードをバッファに読み込む手段を備え、
バッファに含まれたフレーズを識別する手段を備え、このフレーズは、ブレークキャラクタ間における２つ以上のワードのシーケンスであり、
バッファにおいてフレーズの後に明確な終止ブレークキャラクタ又はソフト終止ブレークキャラクタが識別されるのに応答してそのフレーズをキーフレーズとしてセーブする手段を備え、
バッファにおいてフレーズの後にハード終止ブレークキャラクタが識別されるのに応答してそのフレーズをバッファから削除する手段を備え、
フレーズが削除又はセーブされるとワードをバッファからフラッシュする手段を備えた、
ことを特徴とするシステム。

【請求項１８】　テキストの断片を、文書を特徴付ける１つ以上のフレーズへとパージングするシステムにおいて、
文書に含まれたフレーズを識別するように構成された第１モジュールを備え、このフレーズは、ブレークキャラクタ間における２つ以上のワードのシーケンスであり、
識別されたフレーズに続くブレークキャラクタの形式を決定するように構成された第２モジュールを備え、
上記ブレークキャラクタの形式に基づいて異なる処理段階を使用して上記識別されたフレーズを処理するように構成された第３モジュールを備えた、
ことを特徴とするシステム。

【請求項１９】　上記第３モジュールは、更に、文書においてフレーズの後に明確な終止ブレークキャラクタ又はソフト終止ブレークキャラクタが識別されたときにその識別されたフレーズをセーブする手段と、バッファにおいてフレーズの後にハード終止ブレークキャラクタが識別されるのに応答してそのフレーズをバッファから削除する手段とを含む請求項１８に記載のシステム。

【請求項２０】　キーフレーズがデータベースに記憶されるか又はバッファのフレーズが削除されたときにバッファをフラッシュする手段を更に含む請求項１９に記載のシステム。

【請求項２１】　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２０に記載のシステム。

【請求項２２】　上記テキストの断片をキーフレーズへとパージングするためにテキストの断片に適用すべき１つ以上のパージングルールを記憶するルールデータベースを更に備えた請求項１８に記載のシステム。

【請求項２３】　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、テキストの断片に選択的に適用される任意のフレーズレベ

ルパージングルールとを含む請求項２２に記載のシステム。

【請求項２４】　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の１つ以上のルールから成る請求項２３に記載のシステム。

【請求項２５】　上記異なる処理段階は、フレーズをセーブすること及びフレーズを削除することを含む請求項１８に記載のシステム。