

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第4263371号
(P4263371)

(45) 発行日 平成21年5月13日 (2009. 5. 13)

(24) 登録日 平成21年2月20日 (2009. 2. 20)

(51) Int. Cl.

F I

G 0 6 F 17/27 (2006. 01)

G 0 6 F 17/27 E

G 0 6 F 17/30 (2006. 01)

G 0 6 F 17/30 1 7 0 A

G 0 6 F 17/30 2 1 0 A

G 0 6 F 17/30 3 3 0 C

請求項の数 16 (全 14 頁)

(21) 出願番号 特願2000-611158 (P2000-611158)
 (86) (22) 出願日 平成12年4月6日 (2000. 4. 6)
 (65) 公表番号 特表2002-541580 (P2002-541580A)
 (43) 公表日 平成14年12月3日 (2002. 12. 3)
 (86) 国際出願番号 PCT/US2000/009357
 (87) 国際公開番号 W02000/062155
 (87) 国際公開日 平成12年10月19日 (2000. 10. 19)
 審査請求日 平成19年4月6日 (2007. 4. 6)
 審判番号 不服2008-3157 (P2008-3157/J1)
 審判請求日 平成20年2月12日 (2008. 2. 12)
 (31) 優先権主張番号 09/288, 994
 (32) 優先日 平成11年4月9日 (1999. 4. 9)
 (33) 優先権主張国 米国 (US)

早期審理対象出願

(73) 特許権者 501393287
 セミオ コーポレイション
 アメリカ合衆国 カリフォルニア州 94
 402 サン マテオ サウス アムフレ
 ット ブールヴァード 1730 スイ
 ト 101
 (74) 代理人 100059959
 弁理士 中村 稔
 (74) 代理人 100067013
 弁理士 大塚 文昭
 (74) 代理人 100082005
 弁理士 熊倉 禎男
 (74) 代理人 100065189
 弁理士 穴戸 嘉一

最終頁に続く

(54) 【発明の名称】 文書をパーズするシステム及び方法

(57) 【特許請求の範囲】

【請求項 1】

テキストの断片を、文書の特徴付ける 1 つ以上のキーフレーズへとパーズする コン
ピュータシステムであって、

上記システムが、

ブレイクキャラクタ/ワード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、
 ブレイクキャラクタ/ワード間における 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イクキャラクタ/ワードの形式を決定
 する手段と、ブレイクキャラクタ/ワードの決定された形式に基づいて異なるパーズ
 ルールを使用して上記識別されたフレーズをキーフレーズとしてセーブするか又は削除
 する手段とを含み、上記決定されたブレイクキャラクタ/ワードの形式は、ソフト終止ブ
 レイクキャラクタ/ワード、明確なブレイクキャラクタ/ワード及びハード終止ブレイクキ
 ャラクタ/ワードを含む複数の形式の 1 つである、
 ことを特徴とする コンピュータシステム。

【請求項 2】

上記バッファは、更に、上記セーブされた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 コンピュータシステム。

10

20

【請求項 3】

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イクキャラクタ/ワードを識別する手段を含む請求項 2 に記載のコンピュータシステム。

【請求項 4】

上記明確なブレイクキャラクタ/ワードは、句読点から成る請求項 3 に記載のコンピュータシステム。

【請求項 5】

テキストの断片をキーフレーズへとパーズングするためにテキストの断片に適用すべき 1 つ以上のパーズ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 に記載のコンピュータシステム。

10

【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ズングルールと、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ズングルールとを含む請求項 5 に記載のコンピュータシステム。

【請求項 7】

上記フレーズレベルパーズングルールは、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 に記載のコンピュータシステム。

20

【請求項 8】

テキストの断片を、文書の特徴付ける 1 つ以上のキーフレーズへとパーズングするコンピュータにおいて実行されるパーズング方法であって、

上記方法が、

ブレイクキャラクタ/ワード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、ブレイクキャラクタ/ワード間における 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イクキャラクタ/ワードの形式を決定する段階を含み、上記決定されたブレイクキャラクタ/ワードの形式は、ソフト終止ブレイクキャラクタ/ワード、明確なブレイクキャラクタ/ワード及びハード終止ブレイクキャラクタ/ワードを含む複数の形式の 1 つであり、そして

30

上記ブレイクキャラクタ/ワードの決定された形式に基づいて異なるパーズングルールを使用して上記識別されたフレーズをキーフレーズとしてセーブするか又は削除する段階を含む、

ことを特徴とする方法。

【請求項 9】

上記セーブされた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】

40

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イクキャラクタ/ワードを識別する段階を含む請求項 9 に記載の方法。

【請求項 11】

上記明確なブレイクキャラクタ/ワードは、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ズングするためにテキストの断片に適用すべき 1 つ以上のパーズングルールを、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】

50

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ズングルールと、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ズングルールとを含む請求項 1 2 に記載の方法。

【請求項 1 4】

上記フレーズレベルパーズングルールは、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 1 3 に記載の方法。

【請求項 1 5】

テキストの断片を、文書の特徴付ける 1 つ以上のキーフレーズへとパーズングするコンピュータシステムであって、

上記コンピュータシステムが、

ネットワークから受け取ったテキストの断片のワードを記憶するメモリと、アプリケーションを実行するように構成された処理ユニットとを備え、

上記アプリケーションが、

ブレイクキャラクタ/ワード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、ブレイクキャラクタ/ワード間における 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イクキャラクタ/ワードの形式を決定するブレイクキャラクタ/ワード識別器と、上記ブレイクキャラクタ/ワードの決定された形式に基づいて異なるパーズングルールを使用して上記識別されたフレーズをキーフレーズとしてセーブするか又は削除する処理ユニットとを含み、上記決定されたブレイクキャラクタ/ワードの形式は、ソフト終止ブレイクキャラクタ/ワード、ハード終止ブレイクキャラクタ/ワード及び明確なブレイクキャラクタ/ワードを含む複数の形式の 1 つであり、

更に、上記セーブしたキーフレーズを記憶するためのデータベースを備えた、ことを特徴とするコンピュータシステム。

【請求項 1 6】

テキストの断片を、文書の特徴付ける 1 つ以上のキーフレーズへとパーズングするコンピュータにおいて実行されるパーズング方法であって、

上記方法が、

ブレイクキャラクタ/ワード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、ブレイクキャラクタ/ワード間における 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ブレイクキャラクタ/ワード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イクキャラクタ/ワード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イクキャラクタ/ワード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。

【発明の詳細な説明】

【0001】

【技術分野】

本発明は、一般に、文書処理するためのシステム及び方法に係り、より詳細には、文書の文脈を指示する文書内の複数のフレーズを識別するためのシステム及び方法に係る。

【 0 0 0 2 】

【背景技術】

コンピュータデータベースを使用したテキストデータ情報の広範囲な記憶及び検索には種々のファクタが作用している。コンピュータハードドライブのコストの低下に結び付いたハードドライブの記憶容量の急激な増加、及びコンピュータ通信の送信速度の増加がファクタとなっている。更に、コンピュータの処理速度の増加、及びブルテンボードやインターネットのようなコンピュータ通信ネットワークの拡張もファクタとなっている。それ故、人々は、これらデータベースに記憶された大量のテキストデータにアクセスすることができる。しかしながら、大量のテキストデータの記憶及び及びアクセスはテクノロジーによって容易になったが、現在入手できる大量のデータによって新たな問題が生じている。

10

【 0 0 0 3 】

特に、大量のデータを有するコンピュータデータベースのテキストデータにアクセスしようとする人は、余計な情報を検索することなく希望の情報を迅速且つ効率的に検索するためにデータ分析システムを必要とする。更に、システムのユーザは、文書全体を実際に見ることなく文書を理解できるように、各大きな文書を、その文書の特徴付ける複数のフレーズ（１つ以上のワード）に凝縮するための効率的なシステムを必要とする。各文書を複数のキーフレーズに凝縮するためのシステムは、パーズングシステム又はパーザとして知られている。

20

【 0 0 0 4 】

１つの典型的なパーザでは、パーザが、文書内でしばしば繰り返されるフレーズを識別するよう試み、そして文書の特徴付けるフレーズをキーフレーズとして識別する。このようなシステムに伴う問題は、文書における各フレーズの繰り返しをカウントしなければならないので、非常に低速なことである。又、このシステムは、多量のメモリも必要とする。パーズされるべきデータの量が増加するにつれて、このパーザの低い速度が受け入れられないものとなる。別の典型的なパーザは、キーフレーズを識別するために３段階のプロセスを実行する。第１に、文書の各ワードに、そのワードのスピーチの部分（即ち、名詞、形容詞、副詞、動詞等）に基づいてタグが指定され、そしてスピーチのある部分、例えば、冠詞や形容詞は、その文書の特徴付けるフレーズのリストから除去される。次いで、ワードの１つ以上のシーケンス（テンプレート）を使用して、文書に何の理解も付加しないフレーズを識別しそして除去することができる。最終的に、スピーチの適当な部分であってテンプレートの１つに入らないフレーズは、文書の特徴付けるキーフレーズとして受け入れられる。しかし、この従来型のパーザも低速であり、パーズされるべきデータの量が増加するにつれて受け入れられない。

30

【 0 0 0 5 】

これら従来の全てのパーザシステムにおいて、パーザは、文書内の特定ワードの特徴（繰り返しの頻度又はスピーチの部分）に基づいて、文書を小さな断片に分断するよう試みる。問題は、言語が一般に容易に分類されるものではないことであり、それ故、従来のパーザは、文書を正確にパーズせず、又は文書をパーズするのに長い時間を必要とする。更に、従来のパーザシステムは、どれも、文書からキーフレーズをパーズングする方法として言語の複雑な特徴を利用しようとするために、非常に低速である。従来のパーザに伴うこれらの問題は、パーズされねばならない文書の数が増加するにつれて、よりシビアになる。今日、パーズされねばならない文書の数、とりわけ、インターネットやワールドワイドウェブのために甚だしい割合で増加し続けている。それ故、これらの従来のパーザは、受け入れられない。従って、従来のパーズングシステムに伴う上記問題及び制約を解消するパーズングシステム及び方法を提供することが要望され、本発明は、この目的に向けられる。

40

【 0 0 0 6 】

【発明の開示】

本発明によれば、センテンス又はパラグラフ内のブレイクキャラクタ又はブレイクワー

50

ド（以下「ブレイクキャラクタ／ワード」と称する。）を用いて文書を複数のキーフレーズにパーズするパーザシステム及び方法が提供される。本発明によるパーザシステムは、非常に高速であり、速度に対して精度を著しく犠牲にしない。文書内のブレイクキャラクタ／ワードは、句読点、ある終止ワード、及び動詞及び冠詞のようなある形式のワードを含む。パーザシステムは、ブレイクキャラクタ／ワードを受け取る前に１つ以上のワードを受け取るバッファを備えている。このバッファがブレイクキャラクタ／ワードを受け取ると、パーザは、ブレイクキャラクタ／ワードの前のフレーズがセーブされるかどうかをブレイクキャラクタ／ワードの形式に基づいて決定する。特に、ブレイクキャラクタ／ワードが句読点である場合には、パーザは、ブレイクキャラクタ／ワードの前の１つ以上のワードをキーフレーズとして保持する。ブレイクキャラクタ／ワードが別の形式のキャラクタである場合には、ブレイクキャラクタ／ワードの前のフレーズがセーブされてもされなくてもよい。フレーズの使命が決定されると、バッファはフラッシュされ、１つ以上のワードの次のシーケンスがバッファに読み込まれ、それもパーズされる。このようにして、文書内の複数のフレーズが、文書のセンテンス及びパラグラフ内のブレイクキャラクタ／ワードに基づいて文書から迅速に抽出される。

【０００７】

従って、本発明によれば、テキストの断片を、文書を特徴付ける１つ以上のフレーズにパーズするシステムが提供される。このシステムは、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と、このバッファに含まれたフレーズを識別するためのパーザとを備え、上記フレーズは、ブレイクキャラクタ／ワード間における２つ以上のワードのシーケンスである。パーザは、更に、識別されたフレーズに続くブレイクキャラクタ／ワードの形式を決定する手段と、ブレイクキャラクタ／ワードの決定された形式に基づいてバッファからキーフレーズをセーブする手段とを含む。キーフレーズは、データベースに記憶される。

【０００８】

【発明を実施するための最良の形態】

本発明は、英語の文書をパーズするシステムに特に適用でき、本発明は、これについて説明する。しかしながら、本発明のシステム及び方法は、他の言語や、テキストデータの種々の異なる断片にも利用できることが明らかである。本発明をよく理解するために、テキスト処理システムについて以下に説明する。

図１は、テキスト処理システム１０のブロック図である。このテキスト処理システム１０は、パーザシステム１２と、クラスタライザー１４と、マップジェネレータ１６と、データベース（ＤＢ）１８とを備えている。このテキスト処理システムは、物語、新聞記事又は文書のようなテキストの１つ以上の断片を受け取り、そして文書内のキーフレーズ間の関係をグラフィ的に示すマップを発生する。テキストの各断片は、パーザシステム１２によって受け取られ、該パーザシステムは、到来するテキストの各断片を処理し、そしてテキストの各断片に対しその断片を特徴付ける１つ以上のキーフレーズを発生する。これらキーフレーズは、データベース１８に記憶される。パーザシステムに関する詳細は、図２ないし５を参照して説明する。テキストの各断片からキーフレーズが抽出されると、クラスタライザー１４は、フレーズ間の関係に基づいてキーフレーズの１つ以上のクラスタを発生する。発生されたクラスタは、データベース１８にも記憶される。マップジェネレータ１６は、発生されたクラスタをデータベース内のテキストの断片に対して使用して、データベースのテキストの種々の断片内のキーフレーズの互いに関係を示すグラフィ的マップを発生し、システムのユーザがテキスト断片のキーフレーズを見ることによりデータベースを通して容易にサーチできるようにする。クラスタライザー及びマップジェネレータの詳細は、本発明の譲受人が所有する参考としてここに取り上げる米国特許出願第０８／８０１，９７０号に開示されている。テキスト処理システムは、クライアントコンピュータがインターネットのような公衆コンピュータネットワークを経てサーバーにアクセスするようなクライアント／サーバー型コンピュータシステムを含む種々の形態で実施することができる。パーザ、クラスタライザー及びマップジェネレータは、テキスト処理システム

10の中央処理ユニット(図示せず)により実行されるソフトウェアアプリケーションでよい。本発明によるパーザシステム12を以下に詳細に述べる。

【0009】

図2は、本発明によるパーズングシステム12のブロック図である。パーズングシステム12は、バッファ20と、パーザ22と、ルールデータベース(ルールDB)24とを備えている。バッファは、文書である到来するテキストの断片の1つ以上のワードを記憶し、これらワードは、ルールDB24に含まれたルールを使用してパーザ22により分析される。パーザシステム12の出力は、パーズされる文書を特徴付ける1つ以上のフレーズである(各フレーズは、1つ以上のワードを含む)。特に、パーザは、本発明により文書内のブレイクキャラクタ/ワードに基づいて文書内のフレーズを分離する。より詳細には、ブレイクキャラクタ/ワードが識別されるまで1つ以上のワードが文書からバッファに読み込まれる。従って、パーザシステム12は、ブレイクキャラクタ/ワード間にあるフレーズを識別する。次いで、ブレイクキャラクタ/ワードの形式に基づいて、フレーズがキーフレーズとしてセーブされるか又は削除される。パーザシステム12は、例えば、サーバーコンピュータのマイクロプロセッサ(図示せず)によって実行されるソフトウェアの1つ以上の断片として実施されてもよく、サーバーコンピュータは、インターネット、ローカルエリアネットワーク又はワイドエリアネットワークのようなコンピュータネットワークを経て複数のクライアントコンピュータによりアクセスすることができる。パーザ22は、ブレイクキャラクタ/ワードを使用してテキストの断片からキーフレーズを迅速に抽出するのが効果的である。本発明によるブレイクキャラクタ/ワードについて以下

【0010】

ブレイクキャラクタ/ワードは、句読点のような明確なブレイクキャラクタ/ワードと、数字(number)と、数字(number)を含むワードと、終止キャラクタ/ワードとを含む。上記終止キャラクタ/ワードは、更に、ソフト終止ブレイクキャラクタ/ワード又はハード終止ブレイクキャラクタ/ワードとして分類される。これらの異なるブレイクキャラクタ/ワードを各々以下に説明する。明確なブレイクキャラクタ/ワードは、ピリオド、カンマ、セミコロン、コロン、感嘆符、右又は左のかっこ、左又は右の中かっこ、左又は右の大かっこ、リターンキャラクタ又はラインフィードキャラクタのような種々の句読記号を含む。終止キャラクタは、作成されたリストでもよいし、又はスラッシュ(/)及びアンパーサンド記号(&)を含んでもよい。数文字(digit)、文字、外来文字、並びに、アポストロフィー、ダッシュ及び他の終止キャラクタのようなブレイクキャラクタ/ワードによって分離子が定義されてもよい。テキストの断片における種々のワードは、冠詞、接続詞、ハード及びソフトの終止ブレイクキャラクタ/ワード、語学上の指標、構文的分類、例えば、名詞、動詞、不規則動詞、形容詞及び副詞として分類されてもよい。

【0011】

ワードの始めにあるアポストロフィー又はダッシュは、ブレイクキャラクタ/ワードとして処理され(以下参照)、ワードの終りにあるアポストロフィー又はダッシュも、ブレイクキャラクタ/ワードとして処理され、そしてワードの中間にアポストロフィー又はダッシュをもつワードは、バッファ内のフレーズに追加される。全てのハード又はソフトの終止キャラクタ/ワード及び明確なブレイクキャラクタ/ワードは、以下に述べるように処理される。ワードレベルのパーズングにおいては、ワードの第1キャラクタが大文字かどうかテストすることにより適切な名詞が保持される。更に、大文字しかもたないワード及び数字のワードは、全て、バッファに保持される。任意であるが、数字のストリングは、終止キャラクタとして分類及び処理されてもよい。強制的なワードレベルパーズングルールは、次の通りである。フレーズの始めに現れる接続詞ワードは削除されるが、「The」が後に続く接続詞ワードは、バッファに保持される。ハード終止ブレイクキャラクタ/ワードの場合には、そのハード終止ブレイクキャラクタ/ワードに接続された最後のフレーズが削除され、そして残りのバッファが処理される。ソフト終止ブレイクキャラクタ/ワードは、ブレイクキャラクタ/ワードとして処理される。繰り返されるキャラクタ/ワ

ードは、ソフト又はハードの終止キャラクタ／ワードとして処理される。

【 0 0 1 2 】

パーズングに対して望ましくないワードを更に除去するために、ある任意のフレーズレベルのパーズングルールを使用することができる。特に、6個のワードといった所定の長さより長いフレーズが削除されてもよいし、全て大文字のワードをもつフレーズが削除されてもよいし、そして全て数字のワードをもつフレーズが削除されてもよい。上記の全パーズングルールは、図2に示すパーズングルールデータベース24に記憶される。ここで、図3を参照して、パーザシステム12を詳細に説明する。

図3は、文書をパーズングするための本発明の方法40を示すフローチャートである。この方法は、ステップ42において、文書の第1ワードが文書データベース又はサーバーのメモリからバッファにロードされたときに開始される。次いで、パーザは、ステップ44において、ワードがブレイクキャラクタ／ワードであるかどうか決定する。又、パーザは、パーズングプロセスのこの段階において、あるキャラクタ又はワードを削除することもできる。ワードがブレイクキャラクタ／ワードでない場合には、ステップ42へ戻り、文書の次のワードがバッファに読み込まれる。ワードをバッファに読み込むこのプロセスは、ブレイクキャラクタ／ワードに遭遇するまで繰り返され、従って、バッファは、ワードのシーケンスの前にブレイクキャラクタ／ワードを有し且つワードのシーケンスの後にブレイクキャラクタ／ワードを有するワードのシーケンス（フレーズ）を含む。このようにして、文書は、ブレイクキャラクタ／ワードで互いに分離されたフレーズへとパーズされる。

【 0 0 1 3 】

ブレイクキャラクタ／ワードに遭遇した場合には、パーザは、ステップ46において、そのブレイクキャラクタ／ワードが明確なブレイクキャラクタ／ワードであるかどうか決定し、そして明確なブレイクキャラクタ／ワードが存在する場合には、ステップ48においてそのブレイクキャラクタ／ワードを削除し、バッファに含まれたフレーズを抽出する。バッファから抽出されたフレーズは、将来の使用のためにデータベースに記憶される。次いで、ステップ50において、バッファがフラッシュされてバッファからワードが除去され、そしてバッファは、ステップ42及び44において、別のブレイクキャラクタ／ワードが識別されるまで新たなワードをそこにロードし始める。ステップ46に戻ると、ブレイクキャラクタ／ワードが明確なブレイクキャラクタ／ワードでない場合に、パーザは、ステップ52において、ブレイクキャラクタ／ワードがソフト終止ブレイクキャラクタ／ワードであるかどうか決定する。ブレイクキャラクタ／ワードがソフト終止ブレイクキャラクタ／ワードである場合には、ステップ54において、ソフト終止ブレイクキャラクタ／ワードが削除されそしてバッファのフレーズがデータベースに記憶され、ステップ50においてバッファがフラッシュされ、そしてバッファには文書から新たなワードが補給される。ブレイクキャラクタ／ワードがソフト終止ブレイクキャラクタ／ワードでない（即ちブレイクキャラクタ／ワードがハード終止ブレイクキャラクタ／ワードである）場合には、そのハード終止ブレイクキャラクタ／ワード及びバッファのフレーズがステップ56において削除され、バッファはステップ50においてフラッシュされ、そしてステップ42及び44において文書から新たなワードが補給される。このように、文書からのフレーズは、本発明により、ブレイクキャラクタ／ワード及びブレイクキャラクタ／ワードの形式を使用して抽出され、フレーズが互いに分離されると共に、どのフレーズをデータベースにセーブすべきか決定される。本発明によるパーザは、従来のシステムのように文書の各ワードの特徴を分析してキーフレーズを識別するように試みるものではなく、従来のパーザより非常に迅速に且つ従来のパーザと同程度の精度で文書からフレーズを抽出するものである。次に、図4及び図5A - 5Lを参照して本発明によるパーザの動作例を説明する。

【 0 0 1 4 】

図4は、本発明によるパーズングシステムによりパーズされる文書60の一例であり、一方、図5A - 5Lは、図4に示した文書60のパーズング中におけるバッファの動作を

示す。この例では、文書は、短い電子ニュースストーリーであるが、パーザは、他のテキスト断片からフレーズを抽出することもできる。実際に、本発明によるパーザは、1 Mバイトデータ/秒までの速度で種々の形式の文書からフレーズを抽出することができる。図示された特定のストーリーは、NECにより開発された新規な「蛇状」ロボットについて述べている。図5A-Lは、上記ストーリーに対する本発明によるバッファの動作をテーブル68において示す。より詳細には、このテーブルの第1の欄70は、バッファに読み込まれた現在ワードを含み、第2の欄72は、本発明のパーザによるワードの形式の決定を含み、第3の欄74は、特定の時間におけるバッファの内容を含み、第4の欄76は、ワードインデックス（即ち文書から抽出されるフレーズ）を含み、そして第5の欄78は、パーズングプロセスに関するコメントを含む。

10

【0015】

図5Aに示すように、バッファに読み込まれた第1ワードは、ストーリーの始めにある一連のアスタリスクであり、これは、パーザによりブレイクキャラクタ/ワード（句読点）と分類され、バッファから削除される。次のワードは、「Computer」であり、これは、ブレイクキャラクタ/ワードではないのでバッファに入力され、そして次のワード「Select」も、ブレイクキャラクタ/ワードではないのでバッファに入力される。従って、バッファは、区分80に示すように、フレーズ「Computer Select」を含む。文書における次のワードは、カンマであり、パーザによりブレイクキャラクタ/ワードとして分類される。ブレイクキャラクタ/ワードは、句読点（明確な中断）であるから、ワードインデックス欄76に示すように、バッファのワードがデータベースにセーブされ、そしてバッファはフラッシュされる。ここで、新たなワードがバッファに読み込まれてパーズされる。バッファに読み込まれた次のワードは、「October」であり、これは、日付に関連したものであるからハード終止ブレイクキャラクタ/ワードであり、削除される。バッファにより受け取られる次のワードは「1995」であり、これは数字であるからブレイクキャラクタ/ワードであり、これも削除される。バッファにより受け取られる次のワードは「COPYRIGHT」であり、これは終止キャラクタ/ワードとして識別され、削除される。次のワードは「Newsbytes」であり、これは、ブレイクキャラクタ/ワードではないので、バッファに記憶される。次のワード「Inc.」も、バッファに記憶される。次のワードは、ブレイクキャラクタ/ワードのピリオドであり、従って、バッファの内容「Newsbytes Inc.」が、ワードインデックス欄に示すようにデータベースにセーブされ、ブレイクキャラクタ/ワードが削除されそしてバッファがフラッシュされる。

20

30

【0016】

バッファによって受け取られる次の2つのワードは、「1995」と、一連のアスタリスクであり、これらは両方ともブレイクキャラクタ/ワードであり削除される。バッファにより受け取られる次の2つのワードは、「Newsbytes」及び「Newsbytes」であり、これらは両方ともバッファに記憶される。次に受け取られるワードは、ハード終止ブレイクキャラクタ/ワードの「August」であり、従って、バッファの内容及びハード終止ブレイクキャラクタ/ワードは削除される。バッファにより受け取られる次の3つのワードは、全てブレイクキャラクタ/ワード（即ち数字又は句読点）であり、削除される。次のワードは、区画82に数字を含むワードであり、これは、バッファに記憶されるが、次のキャラクタがブレイクキャラクタ/ワードであるときには削除される。というのは、バッファは単一のワードしか含まないからである。図5B-5Lから明らかなように、パーズングプロセスは、文書全体にわたって続けられ、ワードインデックス欄76に示すように、キーフレーズのリストが文書から抽出されて、データベースにセーブされる。

40

【0017】

要約すれば、文書又はテキストの断片を特徴付けるフレーズが、本発明により文書から迅速に抽出される。本発明は、文書又はテキストの断片におけるブレイクキャラクタ/ワードを使用して、フレーズを互いに分離し、そして文書に対するキーフレーズを抽出する。上記の例では、抽出されたフレーズ、例えば、「Newsbytes Inc.」、「snake-like robot」、「NEC Corporation」、「robotically controlled electronic snake」、「disast

50

er relief work」及び「world's first active universal joint」は、これらキーフレーズのみを見る人が、文書全体を見なくても文書の内容を理解できるようにする。本発明によるパージングシステムは、従来の他のパージングシステムよりも非常に迅速にキーフレーズの抽出を実行し、これは、パージングに用いられるテキストデータ及び文書の全量が、例えばインターネットユーザの急増により指数関数的な割合で増加したときに、重要となる。

以上、本発明の特定の実施形態を参照して詳細に説明したが、当業者であれば、本発明の精神及び範囲から逸脱せずに種々の変更がなされ得ることが明らかであろう。従って、本発明の範囲は、特許請求の範囲のみによって限定されるものとする。

【図面の簡単な説明】

10

【図 1】 テキスト処理システムのブロック図である。

【図 2】 本発明によるパージングシステムのブロック図である。

【図 3】 文書をパージングするための本発明の方法を示すフローチャートである。

【図 4】 本発明のパージングシステムによりパーズされるべき文書の一例を示す図である。

【図 5 A】 図 4 に示す文書に対する本発明のパージングバッファの動作を説明する図である。

【図 5 B】 図 4 に示す文書に対する本発明のパージングバッファの動作を説明する図である。

【図 5 C】 図 4 に示す文書に対する本発明のパージングバッファの動作を説明する図である。

20

【図 5 D】 図 4 に示す文書に対する本発明のパージングバッファの動作を説明する図である。

【図 5 E】 図 4 に示す文書に対する本発明のパージングバッファの動作を説明する図である。

【図 5 F】 図 4 に示す文書に対する本発明のパージングバッファの動作を説明する図である。

【図 5 G】 図 4 に示す文書に対する本発明のパージングバッファの動作を説明する図である。

【図 5 H】 図 4 に示す文書に対する本発明のパージングバッファの動作を説明する図である。

30

【図 5 I】 図 4 に示す文書に対する本発明のパージングバッファの動作を説明する図である。

【図 5 J】 図 4 に示す文書に対する本発明のパージングバッファの動作を説明する図である。

【図 5 K】 図 4 に示す文書に対する本発明のパージングバッファの動作を説明する図である。

【図 5 L】 図 4 に示す文書に対する本発明のパージングバッファの動作を説明する図である。

【図 1】

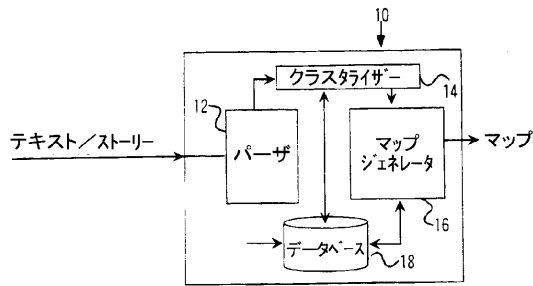


FIGURE 1

【図 2】

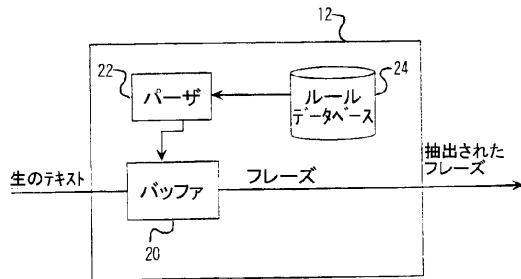


FIGURE 2

【図 4】

COMPUTER SELECT, OCTOBER 1995
COPYRIGHT NEWSBYTES INC. 1995

NEWSBYTES
NEWSBYTES AUGUST 31, 1995 PNEW08310001

NEC DEVELOPS "SNAKE-LIKE" ROBOT.

AUTHOR
MARK, JEREMY

FULL TEXT

TOKYO, JAPAN, 1995 AUG 31 (NB) - NEC CORPORATION HAS DEVELOPED A ROBOTICALLY CONTROLLED ELECTRIC SNAKE THAT OFFERS FAR MORE MOVEMENT THAN PREVIOUSLY DESIGNED ROBOTS. THE COMPANY SAYS THAT THE MUCH GREATER FREEDOM OF MOVEMENT "MAKES IT PERFECT" FOR EVERYTHING FROM INDUSTRIAL TO DISASTER RELIEF WORK. THE SECRET OF THE NEW DEVICE LIES IN A REVOLUTIONARY NEW TYPE OF UNIVERSAL JOINT. THE SECRET OF THE NEW DEVICE LIES IN A REVOLUTIONARY NEW TYPE OF UNIVERSAL JOINT. PREVIOUSLY JOINTS IN THE BODY OF ROBOT HAVE BEEN RESTRICTED TO MOVEMENT IN JUST ONE PLANE, EITHER LEFT AND RIGHT OR UP AND DOWN. BUT THE TOKYO-BASED COMPANY SAYS IT HAS SUCCEEDED IN DEVELOPING THE WORLD'S FIRST ACTIVE UNIVERSAL JOINT. CONTROLLED BY TWO MOTORS, THE JOINT ALLOWS FULL FREEDOM OF MOVEMENT IN ALL PLANES AT EACH OF THE SIX JOINTS ALONG THE ROBOT'S BODY, ALLOWING THE UNIT TO CRAWL INTO PLACES PREVIOUSLY INACCESSIBLE.

AT THE HEART OF THE ROBOT IS A COMPUTER PROCESSING UNIT THAT RECEIVES SIGNALS FROM THE OPERATOR'S HANDSET AND CONTROLS MOVEMENT. THE CONTROLLER CAN INSTRUCT THE COMPUTER TO CONTROL ALL THE JOINTS IN HARMONY OR SPECIFY INDIVIDUAL CONTROL OF EACH JOINT IF NECESSARY.

A VIDEO CAMERA AT THE HEAD OF THE ROBOT SENDS SIGNALS BACK TO THE OPERATOR WHO CAN USE THEM TO STEER THE UNIT AND ALSO TO EXAMINE PLACES INACCESSIBLE TO HUMANS.

THE ENTIRE DEVICE IS 1.4 METERS LONG AND MEASURES 42 MILLIMETERS IN DIAMETER. IT WEIGHS 4.6 KILOGRAMS.

THE AS-YET UNNAMED DEVICE IS NOT YET COMMERCIALY AVAILABLE. NEC'S MARK PEARCE TOLD NEWSBYTES. "IT WILL BE A COUPLE OF YEARS BEFORE EVERYTHING IS SORTED OUT AND IT'S READY TO BE SOLD. WE HAVE TO INCREASE THE SPEED AMONGST OTHER THINGS," HE SAID.

NEC SAYS TYPICAL APPLICATION FOR SUCH A ROBOT COULD BE INVESTIGATION OF COMPLEX PIPEWORK OR AS AN AID TO SEARCH TEAMS IN DISASTER HIT AREAS WHERE THE DEVICE COULD CRAWL THROUGH THE RUBBLE OF COLLAPSED BUILDINGS.

(MARTYN WILLIAMS/19950830/PRESS CONTACT: MARK PEARCE, NEC CORPORATION, TEL +81-3-3457-7249. INTERNET E-MAIL MAKU10-22150@ALADDIN.NEC.CO.JP/NEC950831/PHOTO)

RECORD #
17 337 353

FIGURE 4

【図 3】

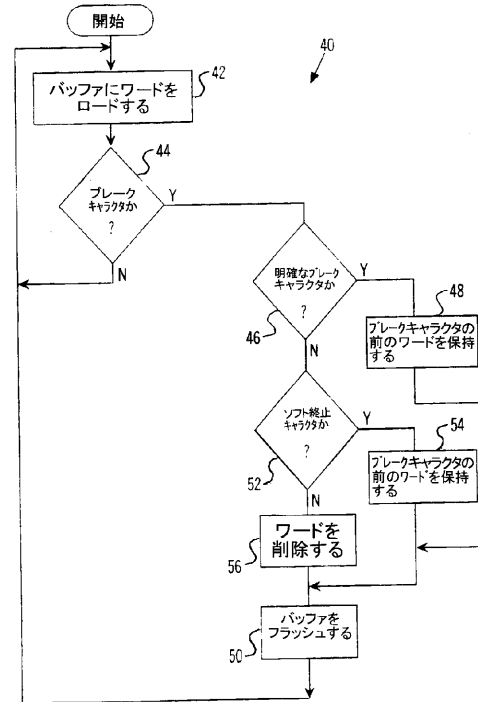


FIGURE 3

【図 5 A】

読み取られたワード	ワードの形式	バッファ	ワードインデックス	コメント
.....	BREAK			
COMPUTER	COMPUTER	COMPUTER		
SELECT	SELECT	COMPUTER		
.....	BREAK	FLUSH	COMPUTER	
OCTOBER	STOP		SELECT	
1995	BREAK			
COPYRIGHT	STOP			
NEWSBYTES	NEWSBYTES	NEWSBYTES		
INC.	INC.	NEWSBYTES		
.....	BREAK	FLUSH	NEWSBYTES INC.	
1995	BREAK			
.....	BREAK			
NEWSBYTES	NEWSBYTES	NEWSBYTES		
NEWSBYTES	NEWSBYTES	NEWSBYTES		
NEWSBYTES	NEWSBYTES	NEWSBYTES		
AUGUST	STOP	FLUSH	SELECT	
31	BREAK			
.....	BREAK			
1995	BREAK			
(PNEW08310001)	(PNEW08310001)	(PNEW08310001)		WORDS CONTAINING NUMBERS
.....	BREAK	FLUSH		
NEC	NEC	NEC		
DEVELOPS	ACTIVE VERB	FLUSH		
SNAKE LIKE	SNAKE-LIKE	SNAKE-LIKE		
ROBOT	ROBOT	SNAKE-LIKE		
.....	BREAK	FLUSH	SNAKE-LIKE	
AUTHOR	AUTHOR	AUTHOR		
OR	BREAK	FLUSH		
MARK	MARK	FLUSH		
.....	BREAK	FLUSH		
JEREMY	JEREMY	JEREMY		
OR	BREAK	FLUSH		
FULL	FULL	FULL		
TEXT	TEXT	FULL TEXT		
OR	BREAK	FLUSH	FULL TEXT	
TOKYO	TOKYO	TOKYO		

FIGURE 5A

【図 5 B】

読み取られたワード	ワードの形式	バツファ	ワードインデックス	コメント
	BREAK	FLUSH		
JAPAN	JAPAN	JAPAN		
	BREAK	FLUSH		
1995	BREAK			
AUG	STOP			
31	BREAK			
K	BREAK			
ING	STOP			
7	BREAK			
NEC	NEC	NEC		
CORPORATION	CORPORATION	NEC CORPORATION		
HAS	STOP	FLUSH	NEC CORPORATION	
DEVELOPED	DEVELOPED	DEVELOPED		
A	STOP	FLUSH		
ROBOTICALLY	ROBOTICALLY	ROBOTICALLY		
CONTROLLED	CONTROLLED	ROBOTICALLY CONTROLLED		
ELECTRONIC	ELECTRONIC	ROBOTICALLY CONTROLLED ELECTRONIC		
SNAKE	SNAKE	ROBOTICALLY CONTROLLED ELECTRONIC SNAKE		
THAT	STOP	FLUSH		
OFFERS	ACTIVE VERB			
FAR	STOP			
MORE	STOP			
MOVEMENT	MOVEMENT	MOVEMENT		
THAT	STOP	FLUSH		
PREVIOUSLY	PREVIOUSLY	PREVIOUSLY		
DESIGNED	DESIGNED	DESIGNED		
ROBOTS	ROBOTS	PREVIOUSLY DESIGNED ROBOTS		
	BREAK	FLUSH	PREVIOUSLY DESIGNED ROBOTS	
THE	STOP			
COMPANY	COMPANY	COMPANY		
SAYS	STOP	FLUSH		
THAT	STOP			
THE	STOP			

FIGURE 5B

【図 5 C】

読み取られたワード	ワードの形式	バツファ	ワードインデックス	コメント
MUCH	STOP			
GREATER	STOP			
FREEDOM	FREEDOM	FREEDOM		
OF	CONNECTOR	FREEDOM OF		
MOVEMENT	MOVEMENT	FREEDOM OF MOVEMENT		
MAKES	STOP	FLUSH	FREEDOM OF MOVEMENT	
IT	STOP			
PERFECT	PERFECT	PERFECT		
FOR	STOP	FLUSH		
EVERYTHING	STOP			
FROM	STOP			
INDUSTRIAL	INDUSTRIAL	INDUSTRIAL		
TO	STOP	FLUSH		
DISASTER	DISASTER	DISASTER		
RELIEF	RELIEF	DISASTER RELIEF		
WORK	WORK	DISASTER RELIEF WORK		
	BREAK	FLUSH	DISASTER RELIEF WORK	
THE	STOP			
SECRET	SECRET	SECRET		
OF	CONNECTOR	SECRET OF		
THE	THE	SECRET OF THE		
NEW	NEW	SECRET OF THE NEW		
DEVICE	DEVICE	SECRET OF THE NEW DEVICE		
LIES	ACTIVE VERB	FLUSH	SECRET OF THE NEW DEVICE	
IN	STOP			
A	STOP			
REVOLUTIONARY	REVOLUTIONARY	REVOLUTIONARY		
NEW	NEW	REVOLUTIONARY NEW		
TYPE	TYPE	REVOLUTIONARY NEW TYPE		
OF	CONNECTOR	REVOLUTIONARY NEW TYPE OF		
UNIVERSAL	UNIVERSAL	REVOLUTIONARY NEW TYPE OF UNIVERSAL		
JOINT	JOINT	REVOLUTIONARY NEW TYPE OF UNIVERSAL JOINT		

FIGURE 5C

【図 5 D】

読み取られたワード	ワードの形式	バツファ	ワードインデックス	コメント
	BREAK	FLUSH	REVOLUTIONARY NEW TYPE OF UNIVERSAL JOINT	
PREVIOUSLY	PREVIOUSLY	PREVIOUSLY		
JOINTS	ACTIVE VERB	FLUSH		
IN	STOP			
THE	STOP			
BODY	BODY	BODY		
OF	CONNECTOR	OF		
ROBOT	ROBOT	BODY OF ROBOT		
HAVE	STOP	FLUSH	BODY OF ROBOT	
BEEN	STOP			
RESTRICTED	RESTRICTED	RESTRICTED		
TO	STOP	FLUSH		
MOVEMENT	MOVEMENT	MOVEMENT		
IN	STOP	FLUSH		
JUST	STOP			
ONE	STOP			
PLANE	PLANE	PLANE		
	BREAK	FLUSH		
EITHER	STOP			
LEFT	LEFT	LEFT		
AND	CONNECTOR	LEFT AND		
RIGHT	RIGHT	LEFT AND RIGHT		
OR	STOP	FLUSH	LEFT AND RIGHT	
UP	STOP			
AND	CONNECTOR			
DOWN	STOP			
BUT	BREAK			
THE	STOP			
TOKYO-BASED	TOKYO-BASED	TOKYO-BASED		
COMPANY	COMPANY	FLUSH COMPANY		
SAYS	STOP	FLUSH		
IT	STOP			
HAS	STOP			
SUCCEEDED	SUCCEEDED	SUCCEEDED		
IN	STOP	FLUSH		
DEVELOPING	DEVELOPING	DEVELOPING		
THE	STOP	FLUSH		
WORLD'S	WORLD'S	WORLD'S		
FIRST	FIRST	WORLD'S FIRST		

FIGURE 5D

【図 5 E】

読み取られたワード	ワードの形式	バツファ	ワードインデックス	コメント
ACTIVE	ACTIVE	WORLD'S FIRST ACTIVE		
UNIVERSAL	UNIVERSAL	WORLD'S FIRST ACTIVE UNIVERSAL		
JOINT	JOINT	WORLD'S FIRST ACTIVE UNIVERSAL JOINT		
	BREAK	FLUSH	WORLD'S FIRST ACTIVE UNIVERSAL JOINT	
CONTROLLED	CONTROLLED	CONTROLLED		
BY	STOP	FLUSH		
TWO	STOP			
MOTORS	MOTORS	MOTORS		
	BREAK	FLUSH		
THE	STOP			
JOINT	JOINT	JOINT		
ALLOWS	ACTIVE VERB	FLUSH		
FULL	FULL	FULL		
FREEDOM	FREEDOM	FULL FREEDOM		
OF	CONNECTOR	FULL FREEDOM OF		
MOVEMENT	MOVEMENT	FULL FREEDOM OF MOVEMENT		
IN	STOP	FLUSH	FULL FREEDOM OF MOVEMENT	
ALL	STOP			
PLANES	ACTIVE VERB			
AT	STOP			
EACH	STOP			
OF	CONNECTOR			
THE	STOP			
SIX	ACTIVE VERB			
JOINTS	STOP			
ALONG	STOP			
THE	STOP			
ROBOT'S	ROBOT'S	ROBOT'S		
BODY	BODY	ROBOT'S BODY		
	BREAK	FLUSH	ROBOT'S BODY	
ALLOWING	ALLOWING	ALLOWING		
THE	STOP	FLUSH		
UNIT	UNIT	UNIT		
TO	STOP	FLUSH		
CRAWL	CRAWL	CRAWL		

FIGURE 5E

【図 5 F】

読み取られたワード	ワードの形式	バッファ	ワードインデックス	コメント
INTO	STOP	FLUSH		
PLACES	ACTIVE VERB			
PREVIOUSLY	PREVIOUSLY			
INACCESSIBLE	INACCESSIBLE	PREVIOUSLY		
		INACCESSIBLE		
	IBREAK	FLUSH	PREVIOUSLY	
			INACCESSIBLE	
AT	STOP			
THE	STOP			
HEART	HEART	HEART		
OF	CONNECTOR	HEART OF		
THE	THE	HEART OF THE		
ROBOT	ROBOT	HEART OF THE		
		ROBOT		
IS	STOP	FLUSH	HEART OF THE	
			ROBOT	
A	STOP			
COMPUTER	COMPUTER	COMPUTER		
PROCESSING	PROCESSING	COMPUTER		
		PROCESSING		
UNIT	UNIT	COMPUTER		
		PROCESSING UNIT		
THAT	STOP	FLUSH	COMPUTER	
			PROCESSING	
			UNIT	
RECEIVES	ACTIVE VERB			
SIGNALS	SIGNALS	SIGNALS		
FROM	STOP	FLUSH		
THE	STOP			
OPERATOR'S	OPERATOR'S	OPERATOR'S		
HANDSET	HANDSET	OPERATOR'S		
		HANDSET		
AND	CONNECTOR	OPERATOR'S		
		HANDSET AND		
CONTROLS	ACTIVE VERB	FLUSH	OPERATOR'S	
			HANDSET	
MOVEMENT	MOVEMENT	MOVEMENT		
BREAK	BREAK	FLUSH		
THE	STOP			
CONTROLLER	CONTROLLER	CONTROLLER		
CAN	STOP	FLUSH		
INSTRUCT	INSTRUCT	INSTRUCT		
THE	STOP	FLUSH		
COMPUTER	COMPUTER	COMPUTER		
TO	STOP	FLUSH		
CONTROL	CONTROL	CONTROL		

FIGURE 5F

【図 5 G】

読み取られたワード	ワードの形式	バッファ	ワードインデックス	コメント
ALL	STOP	FLUSH		
THE	STOP			
JOINTS	ACTIVE VERB			
IN	STOP			
HARMONY	HARMONY	HARMONY		
OR	STOP	FLUSH		
SPECIFY	SPECIFY	SPECIFY		
INDIVIDUAL	INDIVIDUAL	INDIVIDUAL		
CONTROL	CONTROL	INDIVIDUAL		
		CONTROL		
OF	CONNECTOR	INDIVIDUAL		
		CONTROL OF		
EACH	STOP	FLUSH	INDIVIDUAL	
			CONTROL	
JOINT	JOINT	JOINT		
IF	STOP	FLUSH		
NECESSARY	NECESSARY	NECESSARY		
		BREAK		
A	STOP			
VIDEO	VIDEO	VIDEO		
CAMERA	CAMERA	VIDEO CAMERA		
AT	STOP	FLUSH	VIDEO CAMERA	
THE	STOP			
HEAD	HEAD	HEAD		
OF	CONNECTOR	HEAD OF		
THE	THE	HEAD OF THE		
ROBOT	ROBOT	HEAD OF THE		
		ROBOT		
SENDS	ACTIVE VERB	FLUSH	HEAD OF THE	
			ROBOT	
SIGNALS	SIGNALS	SIGNALS		
BACK	STOP	FLUSH		
TO	STOP			
THE	STOP			
OPERATOR	OPERATOR	OPERATOR		
WHO	STOP	FLUSH		
CAN	STOP			
USE	STOP			
THEM	STOP			
TO	STOP			
STEER	STEER	STEER		
THE	STOP	FLUSH		
UNIT	UNIT	UNIT		
AND	CONNECTOR	UNIT AND		
ALSO	STOP	FLUSH		
TO	STOP	CONTROL		

FIGURE 5G

【図 5 H】

読み取られたワード	ワードの形式	バッファ	ワードインデックス	コメント
EXAMINE	EXAMINE	EXAMINE		
PLACES	ACTIVE VERB	FLUSH		
INACCESSIBLE	INACCESSIBLE	INACCESSIBLE		
TO	STOP	FLUSH		
HUMANS	HUMANS	HUMANS		
	BREAK	FLUSH		
THE	STOP			
ENTIRE	ENTIRE	ENTIRE		
DEVICE	DEVICE	ENTIRE DEVICE		
IS	STOP	FLUSH	ENTIRE DEVICE	
	BREAK			
	BREAK			
	STOP			
METERS	ACTIVE VERB			
LONG	LONG	LONG		
AND	CONNECTOR	LONG AND		
MEASURES	ACTIVE VERB	FLUSH		
22	BREAK			
MILLIMETERS	MILLIMETERS	MILLIMETERS		
IN	STOP	FLUSH		
DIAMETER	DIAMETER	DIAMETER		
	BREAK	FLUSH		
IT	STOP			
WEIGHS	WEIGHS	WEIGHS		
A	BREAK	FLUSH		
	BREAK			
	BREAK			
KILOGRAMS	KILOGRAMS	KILOGRAMS		
	BREAK	FLUSH		
THE	STOP			
AS-YET	AS-YET	AS-YET		
UNNAMED	UNNAMED	AS-YET UNNAMED		
DEVICE	DEVICE	AS-YET UNNAMED		
		DEVICE		
IS	STOP	FLUSH	AS-YET	
			UNNAMED	
DEVICE			DEVICE	
NOT	STOP			
YET	STOP			
COMMERCIALLY	COMMERCIALLY	COMMERCIALLY		
AVAILABLE	AVAILABLE	COMMERCIALLY		
		AVAILABLE		
	BREAK	FLUSH	COMMERCIALLY	
			AVAILABLE	

FIGURE 5H

【図 5 I】

読み取られたワード	ワードの形式	バッファ	ワードインデックス	コメント
NEC'S	NEC'S	NEC'S		
MARK	MARK	NEC'S MARK		
PEARCE	PEARCE	NEC'S MARK		
		PEARCE		
TOLD	ACTIVE VERB (PAST TENSE)	FLUSH TOLD		
NEWSBYTES	NEWSBYTES	FLUSH		
	BREAK	FLUSH		
IT	STOP			
WILL	STOP			
BE	STOP			
A	STOP			
COUPLE	COUPLE	COUPLE		
OF	CONNECTOR	COUPLE OF		
YEARS	YEARS	COUPLE OF YEARS		
BEFORE	STOP	FLUSH	COUPLE OF	
			YEARS	
EVERYTHING	STOP			
IS	STOP			
SORTED	SORTED	SORTED		
OUT	STOP	FLUSH		
AND	CONNECTOR			
IT'S	STOP			
READY	READY	READY		
TO	STOP	FLUSH		
BE	STOP			
SOLD	SOLD	SOLD		
	BREAK	FLUSH		
WE	STOP			
HAVE	STOP			
TO	STOP			
INCREASE	INCREASE	INCREASE		
THE	STOP	FLUSH		
SPEED	SPEED	SPEED		
AMONGST	STOP	FLUSH		
OTHER	STOP			
THINGS	THINGS	THINGS		
	BREAK	FLUSH		
HE	STOP			
SAID	STOP			
	BREAK			
NEC	NEC	NEC		
SAYS	STOP	FLUSH		
TYPICAL	TYPICAL	TYPICAL		

FIGURE 5I

【図 5 J】

読み取られたワード	ワードの形式	バッファ	ワードインデックス	コメント
APPLICATIONS	APPLICATIONS	TYPICAL APPLICATIONS	TYPICAL APPLICATIONS	
FOR	STOP	FLUSH		
SUCH	STOP			
A	STOP			
ROBOT	ROBOT	ROBOT		
COULD	STOP	FLUSH		
BE	STOP			
INVESTIGATION	INVESTIGATION	INVESTIGATION		
OF	CONNECTOR	INVESTIGATION OF		
COMPLEX	COMPLEX	INVESTIGATION OF COMPLEX		
PIPEWORK	PIPEWORK	INVESTIGATION OF COMPLEX PIPEWORK		
OR	STOP	FLUSH	INVESTIGATION OF COMPLEX PIPEWORK	
AS	STOP			
AN	STOP			
AID	AID	AID		
TO	STOP	FLUSH		
SEARCH	SEARCH	SEARCH		
TEAMS	ACTIVE VERB	FLUSH		
IN	STOP			
DISASTER	DISASTER	DISASTER		
HIT	HIT	DISASTER HIT		
AREAS	AREAS	DISASTER HIT AREAS		
WHERE	STOP	FLUSH	DISASTER HIT AREAS	
THE	STOP			
DEVICE	DEVICE	DEVICE		
COULD	STOP	FLUSH		
CRAWL	CRAWL	CRAWL		
THROUGH	STOP	FLUSH		
THE	STOP			
RUBBLE	RUBBLE	RUBBLE		
OF	CONNECTOR	RUBBLE OF		
COLLAPSED	COLLAPSED	RUBBLE OF COLLAPSED		

FIGURE 5J

【図 5 K】

読み取られたワード	ワードの形式	バッファ	ワードインデックス	コメント
BUILDINGS	BUILDINGS	RUBBLE OF COLLAPSED BUILDINGS		
	BREAK	FLUSH	RUBBLE OF COLLAPSED BUILDINGS	
	BREAK			
MARTYN WILLIAMS/19950830/PRESS	MARTYN WILLIAMS/19950830/PRESS	MARTYN WILLIAMS/19950830/PRESS	MARTYN WILLIAMS/19950830/PRESS	WORDS CONTAINING NUMBERS
CONTACT	CONTACT	FLUSH CONTACT		
	BREAK	FLUSH		
MARK	MARK	MARK		
PEARCE	PEARCE	MARK PEARCE		
	BREAK	FLUSH	MARK PEARCE	
NEC CORPORATION	NEC CORPORATION	NEC CORPORATION		
	BREAK	FLUSH	NEC CORPORATION	
TEL +81-3-3798-6511	STOP			
	BREAK			
	BREAK			
FAX +81-3-3457-7249	STOP			
	BREAK			
INTERNET	INTERNET	INTERNET		
E MAIL	STOP	FLUSH		
MAKU 10 22150@ALADDIN	MAKU 10 22150@ALADDIN	MAKU 10 22150@ALADDIN		WORDS CONTAINING NUMBERS WORDS CONTAINING INTERNET CODES
	BREAK	FLUSH		
NEC	NEC	NEC		
	BREAK	FLUSH		
CO	CO	CO		
	BREAK	FLUSH		
JP/NEC950831/P HOTO	JP/NEC950831/P HOTO	JP/NEC950831/PHOTO		WORDS CONTAINING NUMBERS
	BREAK	FLUSH		
	BREAK			

FIGURE 5K

【図 5 L】

読み取られたワード	ワードの形式	バッファ	ワードインデックス	コメント
RECORD	RECORD	RECORD		
#	BREAK	FLUSH		
17	BREAK			
337	BREAK			
353	BREAK			

FIGURE 5L

フロントページの続き

(74)代理人 100084009

弁理士 小川 信夫

(74)代理人 100086771

弁理士 西島 孝喜

(74)代理人 100084663

弁理士 箱田 篤

(72)発明者 ヴォーゲル クロード

アメリカ合衆国 カリフォルニア州 94404 フォスター シティー バウンティー ドライ
ヴ #203 733

合議体

審判長 田口 英雄

審判官 菅原 浩二

審判官 飯田 清司

(56)参考文献 特開平8-305730(JP,A)

内山 恵三, 中村 正規、重要キーワード抽出方式とその活用方法、情報処理学会研究報告
91-DBS-84、日本、1991.07.18発行、第91巻、第65号、151~161
頁

(58)調査した分野(Int.Cl., DB名)

G06F17/27

G06F17/30