

【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23年11月4日(2011.11.4)

【公表番号】特表2011-511366(P2011-511366A)

【公表日】平成23年4月7日(2011.4.7)

【年通号数】公開・登録公報2011-014

【出願番号】特願2010-545034(P2010-545034)

【国際特許分類】

G 06 F 17/30 (2006.01)

【F I】

G 06 F 17/30 210 A

G 06 F 17/30 170 A

【手続補正書】

【提出日】平成23年8月26日(2011.8.26)

【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】

複数のデータを処理して、前記データ形式は不明である前記複数のデータとともに含まれている語の特定および検索をするための方法であって、該方法は、

前記識別ステップが、

語を識別するために前記データを検索前に処理するステップ、および

所定の方法で前記語を保存するステップを含む前記データ内の語を識別するステップ、ならびに、

前記検索ステップが、

一致結果を識別するために少なくとも1つの検索語に応答する前記語を検索するステップ、および

前記一致結果のファイルへの保存および前記一致結果の表示の少なくとも1つを行うことにより前記一致結果を処理するステップを含む方法。

【請求項2】

前記識別するステップはさらに前記データの少なくとも一部分について自然な構成の言語の判別を含む請求項1の方法。_____

【請求項3】

前記識別するステップはさらに、

前記データの文字符号化を判別するステップ、および

前記データ内の語群の場所を判別するステップ、

のうちの少なくとも1つを含む請求項1の方法。

【請求項4】

前記識別するステップはさらに、

データファイルタイプを判別するために前記データの少なくとも一部を調べ、および

前記データが、前記データファイルタイプに応答する特別な取り扱いが必要か否かを判別するステップ

を含む請求項1の方法。

【請求項5】

前記識別するステップはさらに前記データについての第1の組の処理パラメータを判別

するステップを含む請求項 1 の方法。

【請求項 6】

前記第1の組の処理パラメータは満足できる結果を生成するかどうかを判別し、前記第1の組の処理パラメータは満足できる結果を生成しない場合、満足できる結果を生成する第2の組の処理パラメータは存在するかどうかを判別するステップ、をさらに含む請求項5の方法。

【請求項 7】

第2の組の処理パラメータが存在するか否かの前記判別が、テキストのセクションおよび前記テキストの言語を探すための前記データの調査、および前記データのすべてまたは一部が圧縮データまたは画像データであるか否かを判断するための前記データのエントロピーの調査、のうち少なくとも1つの実行を含む請求項 6 の方法。

【請求項 8】

前記識別するステップはさらに、識別された語と語参照とを関連付けるステップを含む請求項 1 の方法。

【請求項 9】

線形記憶法および索引付き記憶法のうち少なくとも1つの方法による前記語を保存するステップを含む請求項 1 の方法。

【請求項 10】

前記語が線形記憶法を使用して保存された場合、前記検索は線形検索法を使用して実施され、

前記語が索引付き記憶法を使用して保存された場合、前記検索は索引付き検索法を使用して実施される前記検索に以下の通り前記保存に応答する前記語の検索を含む請求項1の方法。

【請求項 11】

データ形式が不明である複数のデータに含まれている語を識別するための方法であって、

前記データの少なくとも一部分の自然な構成の言語を判別し、前記データを含む語を識別するために、前記自然な言語に応答する前記データを検索前に処理し、および線形記憶法および索引付き記憶法のうち少なくとも1つの方法を使用し前記語を保存するステップを含む方法。

【請求項 12】

前記データの文字符号化の判別、および

前記データ内の語群の場所の判別

のうち少なくとも1つをさらに含む請求項 11 の方法。

【請求項 13】

データファイルタイプを判別するために前記データの少なくとも一部を調べ、および前記データは、前記データファイルタイプに応答する特別な取り扱いが必要か否かを判別するステップ

をさらに含む請求項 11 の方法。

【請求項 14】

前記語と語参照とを関連付けるステップをさらに含む請求項 11 の方法。

【請求項 15】

データ形式が不明である複数のデータに含まれている識別された語を検索するための方法であって、

少なくとも1つの検索語を受信するステップ、

一致結果を識別するために少なくとも1つの検索語に応答する前記語の検索であって、前記検索が、前記語の完全一致検索またはファジー検索を平行して実行するために構成された複数の検索エンジンによって実行される検索ステップ、および、

前記一致結果のファイルへの保存および前記一致結果の表示の少なくとも1つを行うことにより前記一致結果を処理するステップを含む方法。

【請求項16】

前記語が線形記憶法を使用して保存された場合、前記検索は線形検索法を使用して実施され、

前記語が索引付き記憶法を使用して保存された場合、前記検索は索引付き検索法を使用して実施される前記検索ステップが前記の識別された語の保存方法に応答する前記識別された語を検索するステップを含む請求項15の方法。

【請求項17】

データファイル内に含まれている複数のデータの検索および索引付けをする方法を実施するためのシステムであって、該システムは、

データファイルを受信する手段、

データファイルを保存する手段、および

データ形式は不明である複数のデータとともに含まれている語を識別および検索するために複数のデータを処理する方法を実行する手段であって、該方法は、

前記識別ステップが、

語を識別するために前記データを検索前に処理するステップ、および

所定の方法で前記語を保存するステップを含む前記データ内の語を識別するステップを含み、ならびに、

前記検索ステップが、

一致結果を識別するために少なくとも1つの検索語に応答する前記語を検索するステップ、および

前記一致結果のファイルへの保存および前記一致結果の表示の少なくとも1つを行うことにより前記一致結果を処理するステップを含む方法

を実施する手段を含むシステム。

【請求項18】

データ形式は不明である複数のデータとともに含まれている語を識別および検索するために、前記複数のデータを処理する方法を実施するためのコンピュータ実行可能な命令を持つ、コンピュータ可読記憶媒体であって、該方法は、

前記識別ステップが、

語を識別するために前記データを検索前に処理するステップ、および

所定の方法で前記語を保存するステップ

を含む前記データ内の語を識別するステップ、ならびに、

前記検索ステップが、

一致結果を識別するために少なくとも1つの検索語に応答する前記語を検索するステップ、および

前記一致結果のファイルへの保存および前記一致結果の表示の少なくとも1つを行うことにより前記一致結果を処理するステップを含む

方法を実施するためのコンピュータ実行可能な命令を持つ、コンピュータ可読記憶媒体。

【請求項19】

データ形式は不明である複数のデータとともに含まれている語を識別および検索するためのシステムであって、

入力装置、

メモリ装置、

索引処理装置、

前記入力装置および前記メモリ装置との信号通信における処理装置、および前記メモリ装置に結合された索引プロセッサを含み、

前記処理装置は、

データを受信し、

前記データをプロセッサへ分配し、

前記プロセッサを使用する前記データのコンテンツ内の語を識別し、
前記語の場所の記録することにより語参照を生成し、
前記語参照についてのハッシュ値を計算し、
前記ハッシュ値を使用し構造的な方法で前記語参照を保存し、
前記参照および前記ハッシュ値を前記メモリ内にある少なくとも1つのテーブルへ転送し
、
一致結果を識別するために少なくとも1つの検索語に応答する前記語を検索し、
および、
前記一致結果のファイルへの保存および前記一致結果の表示の少なくとも1つを行うこと
により前記一致結果を処理するように構成された処理装置
を含むシステム。

【請求項 20】

データを読み取るよう構成されたデータリーダー、
前記データリーダーに結合され、前記データのコンテンツを判別するよう構成されたデータプロセッサ、
前記データプロセッサに結合され、かつ前記データの索引付けをするよう構成された索引プロセッサであって、前記索引プロセッサは、前記データ内の語を検出し、また前記語からハッシュ値を生成するよう構成されている検索／検出プロセッサを含み、および
語参照が前記語に応答して生成され、かつ前記語参照および前記ハッシュ値をテーブルに転送するように構成されたメモリに保存される前記索引プロセッサに結合されたメモリを含む、検索および索引付けシステム。