【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18年8月10日(2006.8.10)

【公表番号】特表2002-517039(P2002-517039A)
【公表日】平成14年6月11日(2002.6.11)
【出願番号】特願2000-551333(P2000-551333)
【国際特許分類】
Ｇ０６Ｆ　17/27　　(2006.01)
Ｇ０６Ｆ　17/21　　(2006.01)
【ＦＩ】
Ｇ０６Ｆ　17/27　　Ｅ
Ｇ０６Ｆ　17/21　　５９２Ｊ

【手続補正書】
【提出日】平成18年5月29日(2006.5.29)
【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】　コンピュータシステムにおいてテキストセンテンス中に現れる個々の単語を識別するための方法であって、該方法は、
複数の単語の各々について、
隣接する文字の機能として、前記単語が自然言語テキスト中に現れるか否かの確率の指示を格納するステップと、
前記センテンス中に現れる連続した複数の文字グループの各々について、
重複する可能性のある単語を判定するステップと、
前記格納された指示と隣接する文字とに基づいて確率を確認するステップと、
可能性のある単語であると判定された前記文字グループを確率の指示とともにパーサに提出するステップと
を含むことを特徴とする単語識別方法。
【請求項２】　請求項１に記載の単語識別方法において、確率の指示を有する複数の単語の各々について、前記データ構造は、関連する文字のリストをさらに含むことを特徴とする単語識別方法。
【請求項３】　請求項１に記載の単語識別方法において、前記単語の前にリスト中の文字の１つがある場合、確率の指示は低であることを特徴とする単語識別方法。
【請求項４】　請求項１に記載の単語識別方法において、前記単語の後にリスト中の文字の１つが続く場合、確率の指示は低であることを特徴とする単語識別方法。
【請求項５】　請求項１に記載の単語識別方法において、前記単語がリスト中の文字の１つで始まる場合、確率の指示は０であることを特徴とする単語識別方法。
【請求項６】　請求項１に記載の単語識別方法において、前記単語の後にリスト中の文字の１つが続く場合、確率の指示は０であることを特徴とする単語識別方法。
【請求項７】　請求項１に記載の単語識別方法において、前記自然言語は、中国語であることを特徴とする単語識別方法。
【請求項８】　コンピュータシステムに対し、テキストセンテンス中に現れる個々の単語を識別するための命令を格納するコンピュータ可読媒体において、前記命令は、
複数の単語の各々について、
隣接する文字の機能として、前記単語が自然言語テキスト中に現れるか否かの確率の

指示を格納するステップと、
前記センテンス中に現れる連続した複数の文字グループの各々について、
重複する可能性のある単語を判定するステップと、
前記格納された指示と隣接する文字とに基づいて確率を確認するステップと、
可能性のある単語であると判定された前記文字グループを、確率の指示とともにパーサに提出するステップと
を含むことを特徴とするコンピュータ可読媒体。
【請求項９】 請求項８に記載のコンピュータ可読媒体において、確率の指示を有する複数の単語の各々について、前記データ構造は、関連する文字のリストをさらに含むことを特徴とするコンピュータ可読媒体。
【請求項１０】 請求項８に記載のコンピュータ可読媒体において、前記単語の前にリスト中の文字の１つがある場合、確率の指示は低であることを特徴とするコンピュータ可読媒体。
【請求項１１】 請求項８に記載のコンピュータ可読媒体において、前記単語の後にリスト中の文字の１つが続く場合、確率の指示は低であることを特徴とするコンピュータ可読媒体。
【請求項１２】 請求項８に記載のコンピュータ可読媒体において、前記単語の前にリスト中の文字の１つがある場合、確率の指示は０であることを特徴とするコンピュータ可読媒体。
【請求項１３】 請求項８に記載のコンピュータ可読媒体において、前記単語の後にリスト中の文字の１つが続く場合、確率の指示は０であることを特徴とするコンピュータ可読媒体。
【請求項１４】 請求項８に記載のコンピュータ可読媒体において、前記自然言語は、中国語であることを特徴とするコンピュータ可読媒体。
【請求項１５】 自然言語テキスト中に現れる個々の単語を識別するのに使用する単語分割データ構造を含むコンピュータメモリにおいて、前記データ構造は、
複数の単語の各々について、
隣接する文字の機能として、前記単語が自然言語テキスト中に現れるか否かの確率の指示を含むことを特徴とするコンピュータメモリ。
【請求項１６】 請求項１５に記載のコンピュータメモリにおいて、確率の指示を有する複数の単語の各々について、前記データ構造は、関連する文字のリストをさらに含むことを特徴とするコンピュータメモリ。
【請求項１７】 請求項１５に記載のコンピュータメモリにおいて、前記単語の前にリスト中の文字の１つがある場合、確率の指示は低であることを特徴とするコンピュータメモリ。
【請求項１８】 請求項１５に記載のコンピュータメモリにおいて、前記単語の後にリスト中の文字の１つが続く場合、確率の指示は低であることを特徴とするコンピュータメモリ。
【請求項１９】 請求項１５に記載のコンピュータメモリにおいて、前記単語の前にリスト中の文字の１つがある場合、確率の指示は０であることを特徴とするコンピュータメモリ。
【請求項２０】 請求項１５に記載のコンピュータメモリにおいて、前記単語の後にリスト中の文字の１つが続く場合、確率の指示は０であることを特徴とするコンピュータメモリ。
【請求項２１】 自然言語テキスト中に現れる個々の単語を識別するのに使用する単語分割データ構造を含むコンピュータメモリにおいて、前記データ構造は、
複数の文字の各々について、
前記文字で始まる単語の第２の位置に現れる文字の識別と、
前記文字を含む単語について、
前記単語の長さおよび前記文字に占める前記単語内での前記文字の位置の識別と、

複数の単語の各々について、
前記単語を含む文字のシーケンスが、一連のより短い単語をも含むことがあるか否かの指示と、
隣接する文字の機能として、前記単語が自然言語テキスト中に現れるか否可の確率の指示と
を含むことを特徴とするコンピュータメモリ。
【請求項２２】 請求項２１に記載のコンピュータメモリにおいて、確率の指示を有する複数の単語の各々について、前記データ構造は、関連する文字のリストをさらに含むことを特徴とする記載のコンピュータメモリ。
【請求項２３】 コンピュータシステムにおいてテキストセンテンス中に現れる個々の単語を識別するための方法であって、該方法は、
前記センテンス中に現れる連続した複数の文字グループの各々について、
重複する可能性のある単語を判定するステップと、
隣接する文字の機能として、前記単語が自然言語テキスト中に現れるか否かの重み付けされた指示をアクセスすることによって、可能性のある単語の各々の対する値を確認するステップと、
可能性のある単語であると判定された前記文字グループを確率の指示とともにパーサに提出するステップと
を含むことを特徴とする方法。