

【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21年9月24日(2009.9.24)

【公表番号】特表2009-503739(P2009-503739A)

【公表日】平成21年1月29日(2009.1.29)

【年通号数】公開・登録公報2009-004

【出願番号】特願2008-525156(P2008-525156)

【国際特許分類】

G 0 6 F 17/27 (2006.01)

G 0 6 F 17/30 (2006.01)

G 0 6 F 17/21 (2006.01)

【F I】

G 0 6 F 17/27 Z

G 0 6 F 17/30 2 2 0 Z

G 0 6 F 17/30 1 7 0 A

G 0 6 F 17/21 5 5 0 A

【手続補正書】

【提出日】平成21年8月3日(2009.8.3)

【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】

複数のドキュメントの中で定義を識別する方法であって、  
 複数のテキスト単位を入力テキストとして受け取るステップと、  
 前記複数のテキスト単位のうち、どのテキスト単位がキューフレーズを含むか識別する  
 ステップと、

前記複数のテキスト単位のうち、キューフレーズを含まないと識別された全てのテキスト  
 単位を、定義を含む可能性のある検討対象から除外するステップと、

キューフレーズを含むと識別されたテキスト単位それぞれについて、キューフレーズを  
 含むテキスト単位が定義を含むかどうかを判断するために、前記キューフレーズの周辺で  
 局所的な解析を実行するステップであって、前記キューフレーズの周辺で局所的な解析を  
 実行するステップは、前記キューフレーズの第 1 の所定のウィンドウ内で前記テキスト単  
 位の左フレーズを検索することと、前記キューフレーズの第 2 の所定のウィンドウ内で前  
 記テキスト単位の右フレーズを検索することを含む、該局所的な解析を実行するステッ  
 プと、

スコアリングルールのセットに基づいて、正および負のスコアリングを前記左フレーズ  
 および前記右フレーズに適用して、フレーズスコアリングを得るステップと、

前記左フレーズのスコアリングと前記右フレーズのスコアリングを結合して、前記テキ  
 スト単位の全体のスコアを得るステップと、

前記キューフレーズを含む前記テキスト単位が定義を含んでいるかどうかを、前記テキ  
 スト単位の前記全体のスコアに応じて判断するステップと

を含むことを特徴とする方法。

【請求項 2】

前記左フレーズおよび前記右フレーズを検索することは、構文上のフレーズを検索する  
 ことをさらに含むことを特徴とする請求項 1 に記載の方法。

**【請求項 3】**

前記キューフレーズの第 1 の所定のウィンドウ内で前記テキスト単位の左フレーズを検索すること、および前記キューフレーズの第 2 の所定のウィンドウ内で前記テキスト単位の右フレーズを検索することは、

前記キューフレーズの前記第 1 の所定のウィンドウの前、および該ウィンドウ内にある前記テキスト単位の中のフレーズを検索すること

を含むことを特徴とする請求項 2 に記載の方法。

**【請求項 4】**

前記左フレーズおよび前記右フレーズが、前記キューフレーズの前記第 1 の所定のウィンドウおよび前記第 2 の所定のウィンドウ内の前記テキスト単位の中で識別されたかどうかを判断するステップと、

前記左フレーズおよび前記右フレーズが、前記キューフレーズの前記第 1 の所定のウィンドウおよび前記第 2 の所定のウィンドウ内の前記テキスト単位の中で識別されなかった場合は、前記キューフレーズを含む前記テキスト単位は定義を含まないと判断するステップと

をさらに含むことを特徴とする請求項 1 に記載の方法。

**【請求項 5】**

前記左フレーズおよび前記右フレーズが、前記キューフレーズの前記第 1 の所定のウィンドウおよび前記第 2 の所定のウィンドウ内の前記テキスト単位の中で識別されたと判定される場合は、

定義を含まない前記テキスト単位を示す手法で除外ルールのセットのいずれかが前記左フレーズまたは前記右フレーズに適用されるかどうかを判断するステップと

定義を含まない前記テキスト単位を示す手法で前記除外ルールのセットのいずれかが前記左フレーズまたは前記右フレーズに適用されると判断された場合、前記キューフレーズを含む前記テキスト単位は、定義を含まないと判断するステップと

をさらに含むことを特徴とする請求項 4 に記載の方法。

**【請求項 6】**

定義を含まない前記テキスト単位を示す手法で前記除外ルールのセットのいずれもが、で、前記左フレーズまたは前記右フレーズに適用されないと判断された場合、

スコアリングルールのセットに基づいて、正および負のスコアリングを前記左フレーズおよび前記右フレーズに適用し、前記左フレーズおよび前記右フレーズについてフレーズのスコアリングを得るステップ

をさらに含むことを特徴とする請求項 5 に記載の方法。

**【請求項 7】**

前記複数のテキスト単位を入力テキストとして受け取るステップは、複数のセンテンスを前記入力テキストとして受け取ることをさらに含むことを特徴とする請求項 1 に記載の方法。

**【請求項 8】**

前記テキスト単位の中の語に、定義のコンテキストの中で最も一般的な品詞に基づいた品詞 ( P O S ) タグでタグ付けするステップをさらに含むことを特徴とする請求項 1 に記載の方法。

**【請求項 9】**

前記テキスト単位の中の語に、定義のコンテキストの中で最も一般的な品詞に基づいた P O S タグでタグ付けするステップは、

前記複数のテキスト単位のうち、どのテキスト単位がキューフレーズを含むかを識別するのに使用されたパスと同じ処理パスにおいて、前記テキスト単位の中の語に P O S タグでタグ付けすること

をさらに含むことを特徴とする請求項 8 に記載の方法。

**【請求項 10】**

複数のドキュメントの中で定義を識別するための定義抽出システムであって、

複数のテキスト単位を入力として受け取り、前記複数のテキスト単位のうち、どのテキスト単位がキューフレーズを含むか識別するように構成された字句スキャナコンポーネントであって、前記複数のテキスト単位のうちのキューフレーズを含まない全てのテキスト単位を、検討対象から除外するように構成された字句スキャナコンポーネントと、

前記字句スキャナコンポーネントによってキューフレーズを含むと判断されたテキスト単位に対して、前記テキスト単位が定義を包含するかどうかを判断するために、前記キューフレーズの周辺で局所的な解析を実行するように構成された計算コンポーネントと、

スコアリングルールのセットに基づいて、前記キューフレーズの所定のウィンドウ内で前記テキスト単位の左フレーズおよび右フレーズに、正および負のスコアリングを適用して、フレーズスコアリングを得るように構成されたスコアリングコンポーネントであって、前記テキスト単位の全体のスコアを得るために前記左フレーズと前記右フレーズの前記スコアリングを結合し、前記テキスト単位の全体のスコアに応じて、前記キューフレーズを含む前記テキスト単位が、定義を含むかどうか判断するスコアリングコンポーネントとを備えることを特徴とする定義抽出システム。

【請求項 1 1】

前記字句スキャナコンポーネントは、各テキスト単位の語に、定義のコンテキストの中でそれらの最も一般的な品詞（POS）でタグ付けするようにさらに構成されたことを特徴とする請求項 1 0 に記載の定義抽出システム。

【請求項 1 2】

前記字句スキャナコンポーネントは、前記複数のテキスト単位のうち、どのテキスト単位がキューフレーズを含むか識別するのに使用されたパスと同じ処理パスにおいて、各テキスト単位の前記語に、それぞれの最も一般的なPOSでタグ付けするようにさらに構成されたことを特徴とする請求項 1 1 に記載の定義抽出システム。

【請求項 1 3】

キューフレーズを含むと判断されたテキスト単位の前記キューフレーズの前記所定のウィンドウ内で左フレーズおよび右フレーズを識別するように構成されたフレーズ識別コンポーネントをさらに備え、

前記計算コンポーネントは、前記識別された左および右フレーズを使用して前記局所的な解析を実行するようにさらに構成されたことを特徴とする請求項 1 2 に記載の定義抽出システム。