

【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 24 年 5 月 10 日 (2012.5.10)

【公表番号】特表 2011-516989 (P2011-516989A)

【公表日】平成 23 年 5 月 26 日 (2011.5.26)

【年通号数】公開・登録公報 2011-021

【出願番号】特願 2011-504031 (P2011-504031)

【国際特許分類】

G 0 6 F 17/30 (2006.01)

G 0 6 F 13/00 (2006.01)

【 F I 】

G 0 6 F 17/30 3 5 0 C

G 0 6 F 17/30 3 8 0 E

G 0 6 F 13/00 5 4 0 E

【手続補正書】

【提出日】平成 24 年 3 月 12 日 (2012.3.12)

【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】

コンピュータにより実装される関連度システムであって、

1 つまたは複数のプロセッサと、

前記 1 つまたは複数のプロセッサに結合されたメモリであって、複数の命令を記憶し、当該複数の命令は前記 1 つまたは複数のプロセッサにより実行されたときに、当該 1 つまたは複数のプロセッサに、

クエリ文字列に基づいた検索結果として受信した文書から文書情報を抽出する処理であって、前記文書情報は、複合語を有するユニフォームリソースロケータ (URL) 含む、抽出する処理と、

前記複合語を複数の別個の用語に分割する処理と、

当該複数の別個の用語の少なくとも 1 つを用語の辞書中で見つける処理と、

抽出された前記文書情報に基づきターゲットデータの文字列を生成する処理であって、当該ターゲットデータ文字列は前記辞書中で見つけられた、前記複数の別個の用語の 1 つを含む、生成する処理と、

前記ターゲットデータ文字列と前記クエリ文字列の間の編集距離を計算する処理であって、順位付けの結果の一部として文書の関連度を決定する際に、前記編集距離が使用される、計算する処理と

を実行させることを特徴とするシステム。

【請求項 2】

前記文書情報は、タイトル情報、URL 情報、クリック情報、またはアンカーテキストのうちの少なくとも 1 つを含むことを特徴とする請求項 1 に記載のシステム。

【請求項 3】

URL に対する前記編集距離を計算するために、インデックス時間において前記文書情報の複合語が分割されることを特徴とする請求項 1 に記載のシステム。

【請求項 4】

前記複数の命令は前記 1 つまたは複数のプロセッサに、さらに、インデックス時間にお

いて前記文書情報のアンカーテキストをフィルタリングしてアンカーテキストの上位順位付けセットを計算する処理を実行させることを特徴とする請求項 1 に記載のシステム。

【請求項 5】

前記文書情報は、タイトル文字、アンカー文字、クリック文字、または URL 文字のうちの少なくとも 1 つを含み、前記文書情報と、BM25F 関数、クリック距離、ファイルタイプ、言語および URL 深さの生の入力特徴とに基づいて前記文書の関連度を計算するニューラルネットワークをさらに有することを特徴とする請求項 1 に記載のシステム。

【請求項 6】

前記編集距離は、前記データ文字列と前記クエリ文字列の間の近接性を高めるために、用語の挿入および削除に基づいて計算されることを特徴とする請求項 1 に記載のシステム。

【請求項 7】

前記編集距離は、前記データ文字列と前記クエリ文字列の間の近接性を高めるために、用語の挿入および削除に関連付けられているコストに基づいて計算されることを特徴とする請求項 1 に記載のシステム。

【請求項 8】

文書の関連度を決定するコンピュータ実施の方法であって、  
検索プロセスの一部としてクエリ文字列をプロセッサにより受信するステップと、  
ユニフォームリソースロケータ (URL) を、前記検索プロセス中に返された文書から前記プロセッサにより抽出するステップと、  
前記 URL の複合語を別個の複数の用語に分割することにより前記 URL からターゲットデータ文字列を前記プロセッサにより生成するステップと、  
前記ターゲットデータ文字列と前記クエリ文字列の間の編集距離を前記プロセッサにより計算するステップと、  
前記編集距離に基づいて関連度スコアを前記プロセッサにより計算するステップと  
を備えることを特徴とする方法。

【請求項 9】

前記編集距離を計算するステップの一部として用語挿入を前記プロセッサにより使用するステップと、前記ターゲットデータ文字列を生成するために前記クエリ文字列に用語を挿入する挿入コストを前記プロセッサにより計算するステップであって、前記コストは重み付けパラメータとして表される、計算するステップをさらに備えることを特徴とする請求項 8 に記載の方法。

【請求項 10】

前記編集距離を計算するステップの一部として用語削除を前記プロセッサにより使用するステップと、前記データ文字列を生成するために前記クエリ文字列の用語を削除する削除コストを前記プロセッサにより計算するステップであって、前記コストは重み付けパラメータとして表される、計算するステップをさらに備えることを特徴とする請求項 8 に記載の方法。

【請求項 11】

前記編集距離を計算するステップの一部として位置コストを前記プロセッサにより計算するステップであって、前記位置コストは前記ターゲットデータ文字列内の 1 つまたは複数の用語位置の用語挿入および用語削除に関連付けられている、ステップをさらに備えることを特徴とする請求項 8 に記載の方法。

【請求項 12】

前記編集距離を計算する全体コストを計算するため、前記ターゲットデータ文字列の文字と前記クエリ文字列の文字の間で照合プロセスを前記プロセッサにより実行するステップをさらに備えることを特徴とする請求項 8 に記載の方法。

【請求項 13】

インデックス時間において前記データ文字列の URL の複合語を前記プロセッサにより

分割することを特徴とする請求項 8 に記載の方法。

【請求項 14】

前記文書内の出現度数に基づいてアンカーテキストの上位順位付けセットを見出すために前記データ文字列のアンカーテキストを前記プロセッサによりフィルタリングするステップをさらに備えることを特徴とする請求項 8 に記載の方法。

【請求項 15】

前記セット内のアンカーテキストの編集距離を前記プロセッサにより計算するステップをさらに備えることを特徴とする請求項 14 に記載の方法。

【請求項 16】

前記編集距離を計算するステップから導かれたスコアを、変換関数の適用後に 2 層ニューラルネットワークに前記プロセッサにより入力するステップであって、タイトル情報、アンカー情報、クリック情報、または URL 情報のうちの少なくとも 1 つおよびその他の生の入力特徴に関連付けられている前記編集距離の計算に基づいて前記スコアが生成される、入力するステップをさらに備えることを特徴とする請求項 8 に記載の方法。

【請求項 17】

文書の関連度を計算するコンピュータ実施の方法であって、

クエリ文字列を検索プロセスの一部として前記プロセッサにより処理して文書の結果セットを返すステップと、

前記結果セットの文書から抽出された文書情報に基づいてターゲットデータ文字列を生成するステップであって、前記文書情報は複合語を有する URL 情報を含み、前記複合語を複数の別個の用語に分割しおよび用語の辞書中の前記複数の別個の用語の少なくとも 1 つを見つけるステップを有する、生成するステップと、

前記ターゲットデータ文字列と前記クエリ文字列の間の編集距離を、用語挿入、用語削除および用語位置に基づいて、前記プロセッサにより、計算するステップと、

前記編集距離に基づいて関連度スコアを前記プロセッサにより、計算するステップであって、前記関連度スコアは前記結果セットの前記文書を順位付けするために使用される、計算するステップと

を備えることを特徴とする方法。

【請求項 18】

用語挿入、用語削除、用語位置のそれぞれに関連付けられているコストを前記プロセッサにより、計算するステップと、前記コストを前記関連度スコアの計算に前記プロセッサにより、組み込むステップとをさらに備えることを特徴とする請求項 17 に記載の方法。

【請求項 19】

インデックス時間において前記 URL の複合語を前記プロセッサにより、分割するステップと、インデックス時間において前記アンカーテキスト情報を前記プロセッサにより、フィルタリングして前記文書内のアンカーテキストの出現度数に基づいて前記アンカーテキストの上位順位付けセットを前記プロセッサにより、見出すステップとをさらに備えることを特徴とする請求項 17 に記載の方法。

【請求項 20】

ソース URL 文字列内の出現の順序でクエリ用語の文字列を構築するために前記クエリ文字列の用語の出現を前記プロセッサにより、読み取るステップと、前記用語の間のスペースを単語マークで前記プロセッサにより、充填するステップとをさらに備えることを特徴とする請求項 17 に記載の方法。