

【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】令和 1 年 5 月 30 日 (2019.5.30)

【公開番号】特開 2018-77611 (P2018-77611A)

【公開日】平成 30 年 5 月 17 日 (2018.5.17)

【年通号数】公開・登録公報 2018-018

【出願番号】特願 2016-217884 (P2016-217884)

【国際特許分類】

G 0 6 F 16/00 (2019.01)

G 0 6 F 16/30 (2019.01)

【F I】

G 0 6 F 17/30 4 1 4 B

G 0 6 F 17/30 1 7 0 A

【手続補正書】

【提出日】平成 31 年 4 月 17 日 (2019.4.17)

【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】

複数の文書データを記憶する第 1 文書記憶部であって、前記複数の文書データのそれぞれは、オリジナル文書に記載された互いに異なる複数の単語が、互いに異なる固定長のコードに変換されたものである第 1 文書記憶部と、

複数の単語を指定する検索要求であって、かつ、それら複数の単語が存在すべき範囲を単語数で指定する検索要求を受け付ける受付部と、

検索要求で指定された複数の単語のそれぞれに対応する固定長のコードを取得する取得部と、

検索要求で指定された単語数基準の範囲と、固定のコード長とに応じて、文字数基準の範囲を導出する導出部と、

前記取得部により取得された複数の単語のコードと、前記導出部により導出された文字数基準の範囲とを条件とする近傍検索を実行し、前記第 1 文書記憶部に記憶された複数の文書データの中から前記条件を満たす文書データを抽出する検索部と、

を備え、

前記固定長のコードの先頭には、先頭以外では使用されない特殊値が設定されることを特徴とする文書検索装置。

【請求項 2】

複数の文書データを記憶する第 1 文書記憶部であって、前記複数の文書データのそれぞれは、オリジナル文書に記載された互いに異なる複数の単語が、互いに異なる固定長のコードに変換されたものである第 1 文書記憶部と、

複数の単語を指定する検索要求であって、かつ、それら複数の単語が存在すべき範囲を単語数で指定する検索要求を受け付ける受付部と、

検索要求で指定された複数の単語のそれぞれに対応する固定長のコードを取得する取得部と、

検索要求で指定された単語数基準の範囲と、固定のコード長とに応じて、文字数基準の範囲を導出する導出部と、

前記取得部により取得された複数の単語のコードと、前記導出部により導出された文字

数基準の範囲とを条件とする近傍検索を実行し、前記第 1 文書記憶部に記憶された複数の文書データの中から前記条件を満たす文書データを抽出する検索部と、

を備え、

前記固定長のコードの先頭には、先頭以外では使用されない複数の特殊値のうちいずれかが設定されることを特徴とする文書検索装置。

【請求項 3】

複数の文書データを記憶する第 1 文書記憶部であって、前記複数の文書データのそれぞれは、オリジナル文書に記載された互いに異なる複数の単語が、互いに異なる固定長のコードに変換されたものである第 1 文書記憶部と、

前記第 1 文書記憶部に記憶された複数の文書データに対応する変換前の文書データである複数のオリジナル文書データを記憶する第 2 文書記憶部と、

複数の単語を指定する検索要求であって、かつ、それら複数の単語が存在すべき範囲を単語数で指定する検索要求を受け付ける受付部と、

検索要求で指定された複数の単語のそれぞれに対応する固定長のコードを取得する取得部と、

検索要求で指定された単語数基準の範囲と、固定のコード長とに応じて、文字数基準の範囲を導出する導出部と、

前記取得部により取得された複数の単語のコードと、前記導出部により導出された文字数基準の範囲とを条件とする近傍検索を実行し、前記第 1 文書記憶部に記憶された複数の文書データの中から前記条件を満たす文書データを抽出し、前記第 1 文書記憶部から抽出した文書データに対応するオリジナル文書データを前記第 2 文書記憶部から抽出する検索部と、

を備え、

前記第 1 文書記憶部に記憶される文書データは、オリジナル文書データに記載された単語に予め対応付けられたコードが記録されたものであり、オリジナル文書データに記載された或る単語に対応するコードが未定義の場合、その単語に関するコードは前記文書データに記録されず、

検索要求で指定された複数の単語のうち一部の単語に対応するコードが未定義の場合、前記検索部は、対応するコードが存在する単語のコードに基づく近傍検索を実行し、当該近傍検索で抽出した 1 つ以上のオリジナル文書データの中から、前記一部の単語を含むオリジナル文書データを抽出することを特徴とする文書検索装置。

【請求項 4】

複数の文書データを記憶する第 1 文書記憶部であって、前記複数の文書データのそれぞれは、オリジナル文書に記載された互いに異なる複数の単語が、互いに異なる固定長のコードに変換されたものである第 1 文書記憶部と、

複数の単語を指定する検索要求であって、かつ、それら複数の単語が存在すべき範囲を単語数で指定する検索要求を受け付ける受付部と、

検索要求で指定された複数の単語のそれぞれに対応する固定長のコードを取得する取得部と、

検索要求で指定された単語数基準の範囲と、固定のコード長とに応じて、文字数基準の範囲を導出する導出部と、

前記取得部により取得された複数の単語のコードと、前記導出部により導出された文字数基準の範囲とを条件とする近傍検索を実行し、前記第 1 文書記憶部に記憶された複数の文書データの中から前記条件を満たす文書データを抽出する検索部と、

を備え、

前記第 1 文書記憶部に記憶される文書データは、オリジナル文書データに記載された単語に予め対応付けられたコードが記録されたものであり、オリジナル文書データに記載された或る単語に対応するコードが未定義の場合、その単語は複数の単文字単語に分解され、複数の単文字単語に予め対応付けられたコードが前記文書データに記録され、

検索要求で指定された複数の単語のうち一部の単語に対応するコードが未定義の場合、

前記検索部は、対応するコードが存在する単語のコードと、前記一部の単語を分解した複数の単文字単語のコードとに基づく近傍検索を実行することを特徴とする文書検索装置。

【請求項 5】

前記導出部は、検索要求で指定された単語数基準の範囲を、前記一部の単語を分解した単文字単語の個数に応じて拡大し、拡大した単語数基準の範囲に応じて文字数基準の範囲を導出することを特徴とする請求項 4 に記載の文書検索装置。

【請求項 6】

前記検索部は、前記近傍検索により抽出した 1 つ以上の文書データを母集団とした新たな近傍検索であって、前記複数の単文字単語のコードが隣接することを条件とする新たな近傍検索をさらに実行することを特徴とする請求項 4 または 5 に記載の文書検索装置。

【請求項 7】

前記第 1 文書記憶部に記憶される文書データは、オリジナル文書データに記載された互いに関連性を有する複数の単語を、共通のコードへ変換したものである請求項 1 から 6 のいずれかに記載の文書検索装置。

【請求項 8】

複数の文書データを記憶する文書記憶部であって、前記複数の文書データのそれぞれは、オリジナル文書に記載された互いに異なる複数の単語が、互いに異なる固定長のコードに変換されたものである文書記憶部にアクセス可能なコンピュータが、

複数の単語を指定する検索要求であって、かつ、それら複数の単語が存在すべき範囲を単語数で指定する検索要求を受け付けるステップと、

検索要求で指定された複数の単語のそれぞれに対応する固定長のコードを取得するステップと、

検索要求で指定された単語数基準の範囲と、固定のコード長とに応じて、文字数基準の範囲を導出するステップと、

前記取得するステップで取得された複数の単語のコードと、前記導出するステップで導出された文字数基準の範囲とを条件とする近傍検索を実行し、前記文書記憶部に記憶された複数の文書データの中から前記条件を満たす文書データを抽出するステップと、

を実行し、

前記固定長のコードの先頭には、先頭以外では使用されない特殊値が設定されることを特徴とする文書検索方法。

【請求項 9】

複数の文書データを記憶する文書記憶部であって、前記複数の文書データのそれぞれは、オリジナル文書に記載された互いに異なる複数の単語が、互いに異なる固定長のコードに変換されたものである文書記憶部にアクセス可能なコンピュータに、

複数の単語を指定する検索要求であって、かつ、それら複数の単語が存在すべき範囲を単語数で指定する検索要求を受け付ける機能と、

検索要求で指定された複数の単語のそれぞれに対応する固定長のコードを取得する機能と、

検索要求で指定された単語数基準の範囲と、固定のコード長とに応じて、文字数基準の範囲を導出する機能と、

前記取得する機能により取得された複数の単語のコードと、前記導出する機能により導出された文字数基準の範囲とを条件とする近傍検索を実行し、前記文書記憶部に記憶された複数の文書データの中から前記条件を満たす文書データを抽出する機能と、

を実現させ、

前記固定長のコードの先頭には、先頭以外では使用されない特殊値が設定されることを特徴とするコンピュータプログラム。