【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4年1月25日(2022.1.25)

【国際公開番号】WO2021/038887
【出願番号】特願2021-541969(P2021-541969)
【国際特許分類】
G06F 16/383(2019.01)
G06F 16/93(2019.01)
G06F 16/31(2019.01)
G06F 16/90(2019.01)
【FI】
G06F 16/383
G06F 16/93
G06F 16/31
G06F 16/90 100

【手続補正書】
【提出日】令和3年10月22日(2021.10.22)
【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】
検索対象文書に含まれる単語の集合と、単語の語意の近さを示す語間情報とに基づき、前記単語の集合に含まれる単語それぞれに対して、所定の単語を基準として当該単語に対して語意の近い順に近い値を割り当てるハッシュ関数を生成し、
生成した前記ハッシュ関数に基づいて複数の前記検索対象文書それぞれの要約情報を算出し、
生成した前記ハッシュ関数に基づいて入力文書の要約情報を算出し、
算出した前記検索対象文書の要約情報と、前記入力文書の要約情報との間の比較に基づいて、複数の前記検索対象文書の中から前記入力文書に類似する文書を検索する、
処理をコンピュータが実行することを特徴とする類似文書検索方法。
【請求項２】
前記算出する処理のそれぞれは、生成した前記ハッシュ関数において前記検索対象文書または前記入力文書に含まれる単語の集合の単語それぞれに割り当てられた値の中で最小の値をハッシュ値として算出する、
ことを特徴とする請求項１に記載の類似文書検索方法。
【請求項３】
前記生成する処理は、前記所定の単語を選び直して前記ハッシュ関数を生成する処理を繰り返すことで前記ハッシュ関数を複数生成し、
前記算出する処理のそれぞれは、生成した複数の前記ハッシュ関数によって前記検索対象文書または前記入力文書に含まれる単語の集合から算出される複数のハッシュ値を含むベクトルを要約情報として算出する、
ことを特徴とする請求項２に記載の類似文書検索方法。
【請求項４】
算出した複数の前記検索対象文書それぞれの要約情報と、前記検索対象文書とを対応付けた索引情報を生成する処理をさらにコンピュータが実行し、

前記検索する処理は、前記入力文書の要約情報と、前記索引情報において前記検索対象文書と対応付けられた要約情報との間の比較を行う、
ことを特徴とする請求項１に記載の類似文書検索方法。
【請求項５】
検索対象文書に含まれる単語の集合と、単語の語意の近さを示す語間情報とに基づき、前記単語の集合に含まれる単語それぞれに対して、所定の単語を基準として当該単語に対して語意の近い順に近い値を割り当てるハッシュ関数を生成し、
生成した前記ハッシュ関数に基づいて複数の前記検索対象文書それぞれの要約情報を算出し、
生成した前記ハッシュ関数に基づいて入力文書の要約情報を算出し、
算出した前記検索対象文書の要約情報と、前記入力文書の要約情報との間の比較に基づいて、複数の前記検索対象文書の中から前記入力文書に類似する文書を検索する、
処理をコンピュータに実行させることを特徴とする類似文書検索プログラム。
【請求項６】
検索対象文書に含まれる単語の集合と、単語の語意の近さを示す語間情報とに基づき、前記単語の集合に含まれる単語それぞれに対して、所定の単語を基準として当該単語に対して語意の近い順に近い値を割り当てるハッシュ関数を生成するハッシュ関数生成部と、
生成した前記ハッシュ関数に基づいて複数の前記検索対象文書それぞれの要約情報を算出する第１の算出部と、
生成した前記ハッシュ関数に基づいて入力文書の要約情報を算出する第２の算出部と、
算出した前記検索対象文書の要約情報と、前記入力文書の要約情報との間の比較に基づいて、複数の前記検索対象文書の中から前記入力文書に類似する文書を検索する検索部と、
を有することを特徴とする類似文書検索装置。
【請求項７】
検索対象文書に含まれる単語の集合と、単語の語意の近さを示す語間情報とに基づき、前記単語の集合に含まれる単語それぞれに対して、所定の単語を基準として当該単語に対して語意の近い順に近い値を割り当てるハッシュ関数を生成し、
生成した前記ハッシュ関数に基づいて複数の前記検索対象文書それぞれの要約情報を算出し、
算出した複数の前記検索対象文書それぞれの要約情報を探索する索引情報を生成する、
処理をコンピュータが実行することを特徴とする索引情報作成方法。
【請求項８】
検索対象文書に含まれる単語の集合と、単語の語意の近さを示す語間情報とに基づき、前記単語の集合に含まれる単語それぞれに対して、所定の単語を基準として当該単語に対して語意の近い順に近い値を割り当てるハッシュ関数を生成し、
生成した前記ハッシュ関数に基づいて複数の前記検索対象文書それぞれの要約情報を算出し、
算出した複数の前記検索対象文書それぞれの要約情報を探索する索引情報を生成する、
処理をコンピュータに実行させることを特徴とする索引情報作成プログラム。
【請求項９】
検索対象文書に含まれる単語の集合と、単語の語意の近さを示す語間情報とに基づき、前記単語の集合に含まれる単語それぞれに対して、所定の単語を基準として当該単語に対して語意の近い順に近い値を割り当てるハッシュ関数を生成するハッシュ関数生成部と、
生成した前記ハッシュ関数に基づいて複数の前記検索対象文書それぞれの要約情報を算出する算出部と、
算出した複数の前記検索対象文書それぞれの要約情報を探索する索引情報を生成する索引情報生成部と、
を有することを特徴とする索引情報作成装置。