

【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5年12月8日(2023.12.8)

【国際公開番号】WO2022/239174

【出願番号】特願2023-520672(P2023-520672)

【国際特許分類】

G 0 6 F 1 6 / 3 1 (2 0 1 9 . 0 1)

【 F I 】

G 0 6 F 1 6 / 3 1

10

【手続補正書】

【提出日】令和5年9月13日(2023.9.13)

【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】

20

複数の集合に含まれる各集合の個々の要素に対して複数のハッシュ関数を適用して得られる複数のハッシュ値を求める際に、前記複数のハッシュ関数のうちの所定のハッシュ関数によって得られるハッシュ値が一致し、かつ、要素自体が一致する複数の要素に関しては、前記所定のハッシュ関数以外の各ハッシュ関数の計算の重複を排除し、前記各集合の個々の要素に対して前記複数のハッシュ値を求めるハッシュ値計算手段と、

前記複数の集合に含まれるそれぞれの集合に関して、個々のハッシュ関数毎に、ハッシュ値の最小値である最小ハッシュ値を特定する最小ハッシュ値特定手段と、

前記複数の集合から得られる1組以上の集合のペアに関して、ペアをなす2つの集合の類似度を、個々のハッシュ関数に対応する最小ハッシュ値に基づいて導出する類似度導出手段とを備える

30

ことを特徴とする類似度導出システム。

【請求項2】

前記ハッシュ値計算手段は、

前記各集合の個々の要素に対して前記所定のハッシュ関数を適用することによって第1のハッシュ値を計算する第1のハッシュ値計算手段と、

前記各集合の個々の要素のうち、前記第1のハッシュ値が一致し、かつ、要素自体が一致する複数の要素からは1つの要素だけを取り出すとともに、前記複数の要素に該当しない各要素を取り出すことによって、重複なく全ての種類の要素を含む1つの集合である全体集合を生成する全体集合生成手段と、

前記複数の集合に含まれるどの集合のどの要素が前記全体集合のどの要素に該当するかを示すインデックス情報を生成するインデックス情報生成手段と、

40

前記全体集合に属する各要素に対して、前記複数のハッシュ関数のうちの前記所定のハッシュ関数以外の各ハッシュ関数を適用することによって、前記各ハッシュ関数に対応するハッシュ値を計算する第2のハッシュ値計算手段と、

前記インデックス情報に基づいて、前記各集合の個々の要素の前記第1のハッシュ値と、当該第1のハッシュ値に対応する、前記第2のハッシュ値計算手段によって計算されたハッシュ値とを組み合わせることによって、前記各集合の個々の要素に対して、前記複数のハッシュ関数に対応する複数のハッシュ値を決定するハッシュ値決定手段とを含む

請求項1に記載の類似度導出システム。

【請求項3】

50

前記ハッシュ値計算手段は、
 前記複数の集合から、順次、1つの集合を選択する集合選択手段と、
 選択された集合から、順次、1つの要素を選択する要素選択手段と、
 選択された要素に前記所定のハッシュ関数を適用することによって前記選択された要素の第1のハッシュ値を計算する第1のハッシュ値計算手段と、
 前記選択された要素と、第1のハッシュ値が一致し、かつ、要素自体が一致する要素である一致要素が既に選択されているか否かを判定する判定手段と、
 前記一致要素が既に選択されている場合に、前記選択された要素の前記所定のハッシュ関数以外の各ハッシュ関数に対応するハッシュ値を、前記一致要素の前記各ハッシュ関数に対応するハッシュ値と同一であると定め、
 前記一致要素が選択されていない場合に、前記選択された要素の前記所定のハッシュ関数以外の各ハッシュ関数に対応するハッシュ値を計算する第2のハッシュ値計算手段とを含む

10

請求項1に記載の類似度導出システム。

【請求項4】

前記所定のハッシュ関数は、前記複数のハッシュ関数の中で最も値域が広いハッシュ関数である

請求項1から請求項3のうちのいずれか1項に記載の類似度導出システム。

【請求項5】

前記類似度導出手段は、
 個々のハッシュ関数に対応する最小ハッシュ値同士が一致している数を、ペアをなす2つの集合の類似度として定める

20

請求項1から請求項4のうちのいずれか1項に記載の類似度導出システム。

【請求項6】

複数の集合に含まれる各集合の個々の要素に対して複数のハッシュ関数を適用して得られる複数のハッシュ値を求める際に、前記複数のハッシュ関数のうちの所定のハッシュ関数によって得られるハッシュ値が一致し、かつ、要素自体が一致する複数の要素に関しては、前記所定のハッシュ関数以外の各ハッシュ関数の計算の重複を排除し、前記各集合の個々の要素に対して前記複数のハッシュ値を求めるハッシュ値計算処理と、

前記複数の集合に含まれるそれぞれの集合に関して、個々のハッシュ関数毎に、ハッシュ値の最小値である最小ハッシュ値を特定する最小ハッシュ値特定処理と、

30

前記複数の集合から得られる1組以上の集合のペアに関して、ペアをなす2つの集合の類似度を、個々のハッシュ関数に対応する最小ハッシュ値に基づいて導出する類似度導出処理とを含む

ことを特徴とする類似度導出方法。

【請求項7】

前記ハッシュ値計算処理は、
 前記各集合の個々の要素に対して前記所定のハッシュ関数を適用することによって第1のハッシュ値を計算する第1のハッシュ値計算処理と、

前記各集合の個々の要素のうち、前記第1のハッシュ値が一致し、かつ、要素自体が一致する複数の要素からは1つの要素だけを取り出すとともに、前記複数の要素に該当しない各要素を取り出すことによって、重複なく全ての種類の要素を含む1つの集合である全体集合を生成する全体集合生成処理と、

40

前記複数の集合に含まれるどの集合のどの要素が前記全体集合のどの要素に該当するかを示すインデックス情報を生成するインデックス情報生成処理と、

前記全体集合に属する各要素に対して、前記複数のハッシュ関数のうちの前記所定のハッシュ関数以外の各ハッシュ関数を適用することによって、前記各ハッシュ関数に対応するハッシュ値を計算する第2のハッシュ値計算処理と、

前記インデックス情報に基づいて、前記各集合の個々の要素の前記第1のハッシュ値と、当該第1のハッシュ値に対応する、前記第2のハッシュ値計算処理で計算されたハッシュ

50

ハッシュ値とを組み合わせることによって、前記各集合の個々の要素に対して、前記複数のハッシュ関数に対応する複数のハッシュ値を決定するハッシュ値決定処理とを含む

請求項 6 に記載の類似度導出方法。

【請求項 8】

前記ハッシュ値計算処理は、

前記複数の集合から、順次、1つの集合を選択する集合選択処理と、

選択された集合から、順次、1つの要素を選択する要素選択処理と、

選択された要素に前記所定のハッシュ関数を適用することによって前記選択された要素の第 1 のハッシュ値を計算する第 1 のハッシュ値計算処理と、

前記選択された要素と、第 1 のハッシュ値が一致し、かつ、要素自体が一致する要素である一致要素が既に選択されているか否かを判定する判定処理と、

前記一致要素が既に選択されている場合に、前記選択された要素の前記所定のハッシュ関数以外の各ハッシュ関数に対応するハッシュ値を、前記一致要素の前記各ハッシュ関数に対応するハッシュ値と同一であると定め、

前記一致要素が選択されていない場合に、前記選択された要素の前記所定のハッシュ関数以外の各ハッシュ関数に対応するハッシュ値を計算する第 2 のハッシュ値計算処理とを含む

請求項 6 に記載の類似度導出方法。

【請求項 9】

コンピュータに、

複数の集合に含まれる各集合の個々の要素に対して複数のハッシュ関数を適用して得られる複数のハッシュ値を求める際に、前記複数のハッシュ関数のうちの所定のハッシュ関数によって得られるハッシュ値が一致し、かつ、要素自体が一致する複数の要素に関しては、前記所定のハッシュ関数以外の各ハッシュ関数の計算の重複を排除し、前記各集合の個々の要素に対して前記複数のハッシュ値を求めるハッシュ値計算処理、

前記複数の集合に含まれるそれぞれの集合に関して、個々のハッシュ関数毎に、ハッシュ値の最小値である最小ハッシュ値を特定する最小ハッシュ値特定処理、および、

前記複数の集合から得られる 1 組以上の集合のペアに関して、ペアをなす 2 つの集合の類似度を、個々のハッシュ関数に対応する最小ハッシュ値に基づいて導出する類似度導出処理

を実行させるための類似度導出プログラム。

【請求項 10】

前記コンピュータに、

前記ハッシュ値計算処理で、

前記各集合の個々の要素に対して前記所定のハッシュ関数を適用することによって第 1 のハッシュ値を計算する第 1 のハッシュ値計算処理、

前記各集合の個々の要素のうち、前記第 1 のハッシュ値が一致し、かつ、要素自体が一致する複数の要素からは 1 つの要素だけを取り出すとともに、前記複数の要素に該当しない各要素を取り出すことによって、重複なく全ての種類の要素を含む 1 つの集合である全体集合を生成する全体集合生成処理、

前記複数の集合に含まれるどの集合のどの要素が前記全体集合のどの要素に該当するかを示すインデックス情報を生成するインデックス情報生成処理、

前記全体集合に属する各要素に対して、前記複数のハッシュ関数のうちの前記所定のハッシュ関数以外の各ハッシュ関数を適用することによって、前記各ハッシュ関数に対応するハッシュ値を計算する第 2 のハッシュ値計算処理、および、

前記インデックス情報に基づいて、前記各集合の個々の要素の前記第 1 のハッシュ値と、当該第 1 のハッシュ値に対応する、前記第 2 のハッシュ値計算処理で計算されたハッシュ値とを組み合わせることによって、前記各集合の個々の要素に対して、前記複数のハッシュ関数に対応する複数のハッシュ値を決定するハッシュ値決定処理を実行させる

請求項 9 に記載の類似度導出プログラム。

10

20

30

40

50

【手続補正 2】

【補正対象書類名】明細書

【補正対象項目名】0001

【補正方法】変更

【補正の内容】

【0001】

本発明は、集合同士の類似度を導出する類似度導出システム、類似度導出方法、および、類似度導出プログラムに関する。

【手続補正 3】

【補正対象書類名】明細書

【補正対象項目名】0008

【補正方法】変更

【補正の内容】

【0008】

そこで、本発明は、集合の類似度を導出する際に、ハッシュ値の計算量を低減させることができる類似度導出システム、類似度導出方法、および、類似度導出プログラムを提供することを目的とする。

【手続補正 4】

【補正対象書類名】明細書

【補正対象項目名】0011

【補正方法】変更

【補正の内容】

【0011】

本発明による類似度導出プログラムは、コンピュータに、複数の集合に含まれる各集合の個々の要素に対して複数のハッシュ関数を適用して得られる複数のハッシュ値を求める際に、その複数のハッシュ関数のうちの所定のハッシュ関数によって得られるハッシュ値が一致し、かつ、要素自体が一致する複数の要素に関しては、所定のハッシュ関数以外の各ハッシュ関数の計算の重複を排除し、各集合の個々の要素に対して複数のハッシュ値を求めるハッシュ値計算処理、複数の集合に含まれるそれぞれの集合に関して、個々のハッシュ関数毎に、ハッシュ値の最小値である最小ハッシュ値を特定する最小ハッシュ値特定処理、および、複数の集合から得られる1組以上の集合のペアに関して、ペアをなす2つの集合の類似度を、個々のハッシュ関数に対応する最小ハッシュ値に基づいて導出する類似度導出処理を実行させる。

10

20

30

40

50