

【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6年4月1日(2024.4.1)

【国際公開番号】WO2021/173777
 【公表番号】特表2023-516139(P2023-516139A)
 【公表日】令和5年4月18日(2023.4.18)
 【年通号数】公開公報(特許)2023-072
 【出願番号】特願2022-550704(P2022-550704)
 【国際特許分類】
 G 0 6 F 1 6 / 2 1 5 (2 0 1 9 . 0 1)
 【 F I 】
 G 0 6 F 1 6 / 2 1 5

10

【手続補正書】
 【提出日】令和6年3月19日(2024.3.19)
 【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

データ記録のセット内のデータ記録のフィールド内の値に対するデータ品質ルールを、前記フィールドに関連付けられたラベルに基づいて決定するための方法であって、前記ラベルは、前記フィールドの前記値の特性を示し、前記方法は、データ処理システムによって実装され、

ラベルをデータ記録内の1つ以上のフィールドのセットに関連付けるラベルインデックスを取り出すことであって、前記ラベルは、前記1つ以上のフィールドのセットのそれぞれのフィールドで予想される情報のタイプを識別する、ことと、

30

前記ラベルによって示される前記情報のタイプを、前記ラベルに関連付けられた前記1つ以上のフィールドの値に対する要件を表す属性値のセットに関連付けるデータ辞書にアクセスすることであって、前記要件は、前記1つ以上のフィールドの前記値の論理的又は構文的特性を含む、ことと、

特定のデータ記録のフィールドについて、
前記ラベルインデックスにアクセスすることによって、前記特定のデータ記録の前記フィールドに関連付けられた特定のラベルを識別することと、

前記データ辞書から、前記特定のラベルの属性値を取り出すことであって、前記属性値は、前記フィールドに対する特定の要件を指定する、ことと、

データ品質ルールを生成することであって、前記データ品質ルールは、実行されると、
前記フィールドの値が前記属性値によって表される前記特定の要件を満たすかどうかを検証し、

40

前記特定の要件が満たされているかどうかを示す出力データを生成するように構成されている、ことと、を含む、方法。

【請求項2】

前記データ品質ルールは、
前記要件からの前記フィールド内の前記値の許容偏差、
前記フィールド内の1つ以上の許容値、及び

前記フィールド内の1つ以上の禁止値のうちの1つ以上を示す、請求項1に記載の方法。

【請求項3】

50

前記フィールド内の前記1つ以上の許容値又は前記1つ以上の禁止値は、前記フィールドのフィールド名に関連付けられている、請求項2に記載の方法。

【請求項4】

前記フィールドは第1のフィールドであり、前記フィールド内の前記1つ以上の許容値又は前記1つ以上の禁止値は、前記特定のデータ記録内の第2のフィールド内の値に基づいて決定され、前記第2のフィールドは、前記特定のデータ記録の前記第1のフィールドに結び付けられている、請求項2に記載の方法。

【請求項5】

前記1つ以上の許容値又は前記1つ以上の禁止値は、前記第2のフィールド内の前記値と前記第1のフィールド内の前記値との組み合わせに基づく、請求項4に記載の方法。

10

【請求項6】

前記1つ以上の許容値は、数値関数を満たす値に対応し、前記1つ以上の禁止値は、前記数値関数を満たさない値に対応する、請求項2に記載の方法。

【請求項7】

前記フィールドは第1のフィールドであり、前記データ記録は第2のフィールドを含み、前記データ品質ルールは第1のデータ品質ルールであり、前記方法は、

前記特定のラベルの前記属性値に基づいて、前記第1のフィールドと前記第2のフィールドとの間に関係が存在すると決定することと、

前記第2のフィールドと前記第1のフィールドとの間の前記関係に基づいて、前記第2のフィールドに対する第2のデータ品質ルールを生成することと、を含む、請求項1に記載の方法。

20

【請求項8】

前記関係は、前記第1のフィールドの値に対する前記第2のフィールドの値の依存性、又は前記第2のフィールドの前記値に対する前記第1のフィールドの前記値の依存性を示す、請求項7に記載の方法。

【請求項9】

前記関係は、前記第1のフィールドの値と前記第2のフィールドの値との間の相関関係を示す、請求項7に記載の方法。

【請求項10】

前記第1のフィールド及び前記第2のフィールドのそれぞれの値の前記関係を検証する検証データを取得することを更に含む、請求項7に記載の方法。

30

【請求項11】

前記関係は、前記第1のフィールド及び前記第2のフィールドの値の閾値数について検証される、請求項7に記載の方法。

【請求項12】

前記第2のデータ品質ルールは、前記第1のフィールドの値に基づいて、前記第2のフィールドの値に制約を強制するように構成されている、請求項7に記載の方法。

【請求項13】

前記第1のフィールドと前記第2のフィールドとの間に前記関係が存在すると決定することは、前記第2のフィールドの値が前記第1のフィールドの値によって参照されるキー値を含むと決定することを含み、前記データ品質ルールは、前記第2のフィールドのそれぞれの値が有効なキー値であることを要求するように構成されている、請求項7に記載の方法。

40

【請求項14】

前記第1のフィールド及び前記第2のフィールドはそれぞれ数値を含み、前記第1のフィールドと前記第2のフィールドとの間に前記関係が存在すると決定することは、前記第1のフィールドの値を前記第2のフィールドの値に結び付ける数値関数を決定することを含む、請求項7に記載の方法。

【請求項15】

前記第1のフィールドと前記第2のフィールドとの間に前記関係が存在すると決定する

50

ことは、機械学習プロセスによって構成された少なくとも1つの分類子を使用することを
含む、請求項7に記載の方法。

【請求項16】

前記特定のラベルに関連付けられた前記属性値が、前記フィールドが前記特定のデータ
記録の一次キー値を含むことを示すと決定することと、

前記一次キー値が前記フィールド内でそれぞれ固有であることを要求するようにデータ
品質ルールを構成することと、を含む、請求項1に記載の方法。

【請求項17】

前記特定のラベルの前記属性値は、前記フィールド内の前記値の平均、前記値の最大長
、前記値の最小長、前記値のデータタイプ、及び前記値のフォーマットのうちの少なくと
も1つを表す、請求項1に記載の方法。

10

【請求項18】

前記特定のデータ記録は第1のデータ記録であり、前記方法は、前記第1のデータ記録
とは異なる第2のデータ記録内の前記特定のラベルに関連付けられた別のフィールドに前
記データ品質ルールを適用することを含む、請求項1に記載の方法。

【請求項19】

前記方法は、前記データ品質ルール及び前記特定のラベルに関連付けるデータを記憶す
ることを含む、請求項1に記載の方法。

【請求項20】

前記データ品質ルールの承認を要求するためのデータを生成することと、
前記データ品質ルールの承認を示す承認データを取得することに対応して、前記データ
品質ルールを承認することと、を更に含む、請求項1に記載の方法。

20

【請求項21】

前記データ品質ルールを生成することは、
前記フィールド内の値の履歴傾向を決定することと、
前記フィールド内の前記値の前記履歴傾向に基づいて、要件を生成することと、を含む
、請求項1に記載の方法。

【請求項22】

前記フィールドに対する前記データ品質ルールを生成することは、
前記フィールド内の値の履歴平均を識別することと、
前記フィールド内の前記値の前記履歴平均に基づいて、要件を生成することと、を含む
、請求項1に記載の方法。

30

【請求項23】

1つ以上のデータセット内の値に対するデータ品質ルールを決定するための方法であっ
て、前記方法は、データ処理システムによって実装され、

前記データ処理システムによって、1つ以上のデータセットのデータ記録に含まれるフ
ィールドに記憶された値をプロファイルすることと、

前記データ処理システムによって、1つ以上の分類子を前記プロファイルされた値に適
用することと、

前記1つ以上の分類子を適用することに基づいて、前記フィールドの前記値の論理的又
は構文的特性を示す1つ以上の属性を識別することであって、前記1つ以上の属性のそれ
ぞれは、前記1つ以上の分類子のそれぞれの出力に基づく、対応の信頼レベルを有する、
ことと、

40

前記信頼レベルが閾値レベルを満たす、前記識別された属性のうちの1つ以上を前記フ
ィールドに関連付けることと、

前記フィールドに関連付けられた前記1つ以上の属性に基づいて、前記フィールドに含
まれる値に対する1つ以上の制約を決定することと、

前記決定された1つ以上の制約に基づいて、前記データセットの前記フィールドに対す
るデータ品質ルールを決定することであって、前記データ品質ルールは、前記フィール
ドの値と前記1つ以上の制約との間の関係を示す、ことと、を含む、方法。

50

【請求項 24】

前記データ品質ルールは、
前記1つ以上の制約からの前記フィールド内の前記値の許容偏差、
前記フィールド内の1つ以上の許容値、及び
前記フィールド内の1つ以上の禁止値のうちの1つ以上を示す、請求項23に記載の方法。

【請求項 25】

前記フィールドは第1のフィールドであり、前記データ記録は第2のフィールドを含み、
前記データ品質ルールは第1のデータ品質ルールであり、前記方法は、
前記フィールドに関連付けられた前記1つ以上の属性に基づいて、前記第1のフィールドと前記第2のフィールドとの間に関係が存在すると決定することと、
前記第2のフィールドと前記第1のフィールドとの間の前記関係に基づいて、前記第2のフィールドに対する第2のデータ品質ルールを生成することと、を含む、請求項23に記載の方法。

10

【請求項 26】

前記フィールドに関連付けられた前記1つ以上の属性が、前記フィールドが特定のデータ記録の一次キー値を含むことを示すと決定することと、
前記一次キー値が前記フィールド内でそれぞれ固有であることを要求するようにデータ品質ルールを構成することと、を含む、請求項23に記載の方法。

【請求項 27】

前記フィールドに関連付けられた前記1つ以上の属性は、前記フィールド内の前記値の平均、前記値の最大長、前記値の最小長、前記値のデータタイプ、及び前記値のフォーマットのうちの少なくとも1つを表す、請求項23に記載の方法。

20

【請求項 28】

前記方法は、別のデータ記録内の別のフィールドに前記データ品質ルールを適用することを含む、請求項23に記載の方法。

【請求項 29】

データ記録のセット内のデータ記録のフィールド内の値に対するデータ品質ルールを、
前記フィールドに関連付けられたラベルに基づいて決定するためのデータ処理システムであって、前記ラベルは、前記フィールドの前記値の特性を示し、前記データ処理システムは、

30

少なくとも1つの処理デバイスと、
前記少なくとも1つの処理デバイスと通信する少なくとも1つのメモリであって、前記少なくとも1つのメモリは、命令を記憶し、前記命令は、前記少なくとも1つの処理デバイスによって実行されると、前記少なくとも1つの処理デバイスに、

ラベルをデータ記録内の1つ以上のフィールドのセットに関連付けるラベルインデックスを取り出すことであって、前記ラベルは、前記1つ以上のフィールドのセットのそれぞれのフィールドで予想される情報のタイプを識別する、ことと、

前記ラベルによって示される前記情報のタイプを、前記ラベルに関連付けられた前記1つ以上のフィールドの値に対する要件を表す属性値のセットに関連付けるデータ辞書にアクセスすることであって、前記要件は、前記1つ以上のフィールドの前記値の論理的又は構文的特性を含む、ことと、

40

特定のデータ記録のフィールドについて、

前記ラベルインデックスにアクセスすることによって、前記特定のデータ記録の前記フィールドに関連付けられた特定のラベルを識別することと、

前記データ辞書から、前記特定のラベルの属性値を取り出すことであって、前記属性値は、前記フィールドに対する特定の要件を指定する、ことと、

データ品質ルールを生成することであって、前記データ品質ルールは、実行されると、
前記フィールドの値が前記属性値によって表される前記特定の要件を満たすかどうかを検証し、

50

前記特定の要件が満たされているかどうかを示す出力データを生成するように構成されている、ことと、含む動作を実行させる、少なくとも1つのメモリと、を含む、データ処理システム。

【請求項30】

前記データ品質ルールは、
前記要件からの前記フィールド内の前記値の許容偏差、
前記フィールド内の1つ以上の許容値、及び
前記フィールド内の1つ以上の禁止値のうちの1つ以上を示す、請求項29に記載のデータ処理システム。

【請求項31】

前記フィールドは第1のフィールドであり、前記データ記録は第2のフィールドを含み、前記データ品質ルールは第1のデータ品質ルールであり、前記動作は、
前記特定のラベルの前記属性値に基づいて、前記第1のフィールドと前記第2のフィールドとの間に関係が存在すると決定することと、
前記第2のフィールドと前記第1のフィールドとの間の前記関係に基づいて、前記第2のフィールドに対する第2のデータ品質ルールを生成することと、を含む、請求項29に記載のデータ処理システム。

【請求項32】

前記動作は、
前記特定のラベルに関連付けられた前記属性値が、前記フィールドが前記特定のデータ記録の一次キー値を含むことを示すと決定することと、
前記一次キー値が前記フィールド内でそれぞれ固有であることを要求するようにデータ品質ルールを構成することと、を含む、請求項29に記載のデータ処理システム。

【請求項33】

前記特定のラベルの前記属性値は、前記フィールド内の前記値の平均、前記値の最大長、前記値の最小長、前記値のデータタイプ、及び前記値のフォーマットのうちの少なくとも1つを表す、請求項29に記載のデータ処理システム。

【請求項34】

前記特定のデータ記録は第1のデータ記録であり、前記動作は、前記第1のデータ記録とは異なる第2のデータ記録内の前記特定のラベルに関連付けられた別のフィールドに前記データ品質ルールを適用することを含む、請求項29に記載のデータ処理システム。

【請求項35】

データ記録のセット内のデータ記録のフィールド内の値に対するデータ品質ルールを、前記フィールドに関連付けられたラベルに基づいて決定するための命令を記憶する1つ以上の非一時的コンピュータ可読媒体であって、前記ラベルは、前記フィールドの前記値の特性を示し、前記命令は、少なくとも1つの処理デバイスによって実行されると、前記少なくとも1つの処理デバイスに、

ラベルをデータ記録内の1つ以上のフィールドのセットに関連付けるラベルインデックスを取り出すことであって、前記ラベルは、前記1つ以上のフィールドのセットのそれぞれのフィールドで予想される情報のタイプを識別する、ことと、

前記ラベルによって示される前記情報のタイプを、前記ラベルに関連付けられた前記1つ以上のフィールドの値に対する要件を表す属性値のセットに関連付けるデータ辞書にアクセスすることであって、前記要件は、前記1つ以上のフィールドの前記値の論理的又は構文的特性を含む、ことと、

特定のデータ記録のフィールドについて、

前記ラベルインデックスにアクセスすることによって、前記特定のデータ記録の前記フィールドに関連付けられた特定のラベルを識別することと、

前記データ辞書から、前記特定のラベルの属性値を取り出すことであって、前記属性値は、前記フィールドに対する特定の要件を指定する、ことと、

データ品質ルールを生成することであって、前記データ品質ルールは、実行されると、

10

20

30

40

50

前記フィールドの値が前記属性値によって表される前記特定の要件を満たすかどうかを検証し、

前記特定の要件が満たされているかどうかを示す出力データを生成するように構成されている、ことと、を含む動作を実行させるように構成されている、1つ以上の非一時的コンピュータ可読媒体。

【請求項36】

前記データ品質ルールは、

前記要件からの前記フィールド内の前記値の許容偏差、

前記フィールド内の1つ以上の許容値、及び

前記フィールド内の1つ以上の禁止値のうちの1つ以上を示す、請求項35に記載の1つ以上の非一時的コンピュータ可読媒体。

10

【請求項37】

前記フィールドは第1のフィールドであり、前記データ記録は第2のフィールドを含み、前記データ品質ルールは第1のデータ品質ルールであり、前記動作は、

前記特定のラベルの前記属性値に基づいて、前記第1のフィールドと前記第2のフィールドとの間に関係が存在すると決定することと、

前記第2のフィールドと前記第1のフィールドとの間の前記関係に基づいて、前記第2のフィールドに対する第2のデータ品質ルールを生成することと、を含む、請求項35に記載の1つ以上の非一時的コンピュータ可読媒体。

【請求項38】

前記動作は、

前記特定のラベルに関連付けられた前記属性値が、前記フィールドが前記特定のデータ記録の一次キー値を含むことを示すと決定することと、

前記一次キー値が前記フィールド内でそれぞれ固有であることを要求するようにデータ品質ルールを構成することと、を含む、請求項35に記載の1つ以上の非一時的コンピュータ可読媒体。

20

【請求項39】

前記特定のラベルの前記属性値は、前記フィールド内の前記値の平均、前記値の最大長、前記値の最小長、前記値のデータタイプ、及び前記値のフォーマットのうちの少なくとも1つを表す、請求項35に記載の1つ以上の非一時的コンピュータ可読媒体。

30

【請求項40】

前記特定のデータ記録は第1のデータ記録であり、前記動作は、前記第1のデータ記録とは異なる第2のデータ記録内の前記特定のラベルに関連付けられた別のフィールドに前記データ品質ルールを適用することを含む、請求項35に記載の1つ以上の非一時的コンピュータ可読媒体。

40

50