

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6400588号
(P6400588)

(45) 発行日 平成30年10月3日(2018.10.3)

(24) 登録日 平成30年9月14日(2018.9.14)

(51) Int. Cl. F I
G06F 12/00 (2006.01) G O 6 F 12/00 5 1 3 Z
G06F 17/30 (2006.01) G O 6 F 17/30 2 3 0 Z

請求項の数 19 (全 19 頁)

(21) 出願番号	特願2015-537766 (P2015-537766)	(73) 特許権者	509123208
(86) (22) 出願日	平成25年10月15日(2013.10.15)		アビニシオ テクノロジー エルエルシー
(65) 公表番号	特表2016-504642 (P2016-504642A)		アメリカ合衆国 02421 マサチュー
(43) 公表日	平成28年2月12日(2016.2.12)		セッツ州 レキシントン スプリング ス
(86) 国際出願番号	PCT/US2013/064979		トリート 201
(87) 国際公開番号	W02014/062638	(74) 代理人	100107984
(87) 国際公開日	平成26年4月24日(2014.4.24)		弁理士 廣田 雅紀
審査請求日	平成28年10月11日(2016.10.11)	(74) 代理人	100102255
(31) 優先権主張番号	13/653,995		弁理士 小澤 誠次
(32) 優先日	平成24年10月17日(2012.10.17)	(74) 代理人	100096482
(33) 優先権主張国	米国 (US)		弁理士 東海 裕作
		(74) 代理人	100188352
			弁理士 松田 一弘
		(74) 代理人	100131093
			弁理士 堀内 真

最終頁に続く

(54) 【発明の名称】 データに対する規則の指定及び適用

(57) 【特許請求の範囲】

【請求項1】

データセットの複数の要素の各要素の1又は2以上のフィールドに含まれるデータの妥当性を検証するための1又は2以上の妥当性検証規則を指定するためのユーザーインターフェイスを表示可能なコンピューティングシステムであって、

前記コンピューティングシステムが、

(1) 前記妥当性検証規則が適用される前記データセットの複数の要素のうちの1つの要素のそれぞれのフィールドにおけるデータ値と、(2) 第1の軸及び第2の軸を有する2次元グリッドに配列された複数のセルとを備える前記ユーザーインターフェイスを描画するように構成されたユーザーインターフェイスモジュールであって、前記2次元グリッドが

10

前記2次元グリッドの前記第1の軸に沿った方向に延伸した前記セルの1又は2以上のサブセットであって、それぞれのサブセットが前記データセットの前記複数の要素のうちの1つの要素のそれぞれのフィールドに関連付けられる、1又は2以上のサブセット、及び

前記2次元グリッドの前記第2の軸に沿った方向に延伸した前記セルの複数のサブセットであって、その1又は2以上がそれぞれの妥当性検証規則に関連付けられる、複数のサブセットを含む、前記ユーザーインターフェイスモジュールと、

前記セルの少なくともいくつかのセルから受け取られたユーザ入力に基づいて前記データセットの少なくとも1つの要素に妥当性検証規則を適用するように構成された処理モジ

20

ルールとを含み、

フィールド及び妥当性検証規則に関連する少なくともいくつかのセルが、それぞれ、
関連する妥当性検証規則が関連するフィールドに適用されるか否かを判定する入力を受け取るための入力要素、及び

前記要素の前記関連するフィールドに含まれるデータに前記関連する妥当性検証規則を適用することに基づく妥当性検証結果に関連するフィードバックを示すためのインジケータを含む、前記コンピューティングシステム。

【請求項 2】

第 1 の要素の第 1 のフィールドに含まれるデータに妥当性検証規則を適用することが、セルの入力要素において受け取られた任意の入力に基づいて、前記第 1 の要素の前記第 1 のフィールドに関連する第 1 の軸に沿った方向に延伸したセルのサブセットからの前記セルに関連する任意の選択された妥当性検証規則を決定することと、

前記選択された妥当性検証規則に基づいて前記第 1 の要素の前記第 1 のフィールドに含まれる前記データに関する妥当性検証結果を決定することを含む、請求項 1 に記載のシステム。

【請求項 3】

第 1 の軸に沿った方向に延伸したセルの 1 又は 2 以上のサブセットが、セルの行である、請求項 1 に記載のシステム。

【請求項 4】

第 2 の軸に沿った方向に延伸したセルの複数のサブセットが、セルの列である、請求項 1 に記載のシステム。

【請求項 5】

入力要素が、1 又は 2 以上の妥当性検証規則パラメータを指定する入力を受け取るように構成される、請求項 1 に記載のシステム。

【請求項 6】

妥当性検証規則のうちの 1 又は 2 以上が、評価されるときに、少なくとも 2 つの妥当性検証結果のセットの妥当性検証結果を生じ、前記妥当性検証結果が、妥当な結果及び妥当でない結果を含む、請求項 1 に記載のシステム。

【請求項 7】

セルの少なくともいくつかのセルに含まれるフィードバックを示すためのインジケータが、妥当性検証結果が妥当でない結果である場合にセルに網かけを適用するように構成される、請求項 6 に記載のシステム。

【請求項 8】

入力要素が、妥当性検証規則パラメータのそれぞれの正しさを判定するようにさらに構成される、請求項 5 に記載のシステム。

【請求項 9】

フィールド及び妥当性検証規則に関連する少なくともいくつかのセルが、それぞれ、前記セルに関連する妥当性検証規則パラメータの正しさを判定した結果を表示するための第 2 のインジケータを含む、請求項 8 に記載のシステム。

【請求項 10】

フィードバックを示すためのインジケータが、妥当でない結果の数を表示するように構成される数値インジケータを含み、妥当でない結果の前記数が、データセットの要素のすべてに関する関連するフィールドに含まれるデータに関連する妥当性検証規則を適用することによって決定される、請求項 1 に記載のシステム。

【請求項 11】

データセットが、データベースの 1 又は 2 以上のテーブルを含み、前記データセットの要素が、データベースのレコードを含む、請求項 1 に記載のシステム。

【請求項 12】

妥当性検証規則のうちの 1 又は 2 以上が、ユーザによって定義される、請求項 1 に記載のシステム。

10

20

30

40

50

【請求項 1 3】

妥当性検証規則のうちの 1 又は 2 以上が、予め定義される、請求項 1 に記載のシステム。

【請求項 1 4】

2 次元グリッドの第 2 の軸に沿った方向に延伸したセルの複数のサブセットのうちの 1 又は 2 以上が、第 1 の妥当性検証規則に関連する第 1 のセル及び第 2 の妥当性検証規則に関連する第 2 のセルを含み、前記第 2 の妥当性検証規則が、前記第 1 の妥当性検証規則とは異なる、請求項 1 に記載のシステム。

【請求項 1 5】

2 次元グリッドの第 2 の軸に沿った方向に延伸したセルの複数のサブセットのうちの 1 又は 2 以上が、既存の値に適用された妥当性検証規則のうちの 1 つに関する妥当でない結果に応じて対応するフィールドの前記既存の値を置き換えるための値を受け取るための入力要素を含むセルのサブセットを含む、請求項 1 に記載のシステム。

10

【請求項 1 6】

2 次元グリッドの第 2 の軸に沿った方向に延伸したセルの複数のサブセットのうちの 1 又は 2 以上が、除外される値が対応するフィールドに表示されることが妥当性検証規則が既存の値に適用されることを防止する結果となるように、前記除外される値を受け取るための入力要素を含むセルのサブセットを含む、請求項 1 に記載のシステム。

【請求項 1 7】

データセットの複数の要素の各要素の 1 又は 2 以上のフィールドに含まれるデータの妥当性を検証するための 1 又は 2 以上の妥当性検証規則を指定するための ユーザーインターフェイスを表示可能なコンピューティングシステムであって、

20

前記コンピューティングシステムが、

(1) 前記妥当性検証規則が適用される前記データセットの複数の要素のうちの 1 つの要素のそれぞれのフィールドにおけるデータ値と、(2) 第 1 の軸及び第 2 の軸を有する 2 次元グリッドに配列された複数のセルとを備える前記ユーザーインターフェイスを描画するための手段であって、前記 2 次元グリッドが、

前記 2 次元グリッドの前記第 1 の軸に沿った方向に延伸した前記セルの 1 又は 2 以上のサブセットであって、それぞれのサブセットが前記データセットの前記複数の要素のうちの 1 つの要素のそれぞれのフィールドに関連付けられる、1 又は 2 以上のサブセット、及び

30

前記 2 次元グリッドの前記第 2 の軸に沿った方向に延伸した前記セルの複数のサブセットであって、その 1 又は 2 以上がそれぞれの妥当性検証規則に関連付けられる、複数のサブセットを含む、前記描画するための手段と、

前記セルの少なくともいくつかのセルから受け取られたユーザ入力に基づいて前記データセットの少なくとも 1 つの要素に妥当性検証規則を適用するための手段とを含み、

フィールド及び妥当性検証規則に関連する少なくともいくつかのセルが、それぞれ、

関連する妥当性検証規則が関連するフィールドに適用されるか否かを判定する入力を受け取るための入力要素、及び

前記要素の前記関連するフィールドに含まれるデータに前記関連する妥当性検証規則を適用することに基づく妥当性検証結果に関連するフィードバックを示すためのインジケータを含む、前記コンピューティングシステム。

40

【請求項 1 8】

データセットの複数の要素の各要素の 1 又は 2 以上のフィールドに含まれるデータの妥当性を検証するための 1 又は 2 以上の妥当性検証規則を指定するための方法であって、

ユーザーインターフェイスモジュールによって、(1) 前記妥当性検証規則が適用される前記データセットの複数の要素のうちの 1 つの要素のそれぞれのフィールドにおけるデータ値と、(2) 第 1 の軸及び第 2 の軸を有する 2 次元グリッドに配列された複数のセルとを備えるユーザーインターフェイスを描画するステップであって、前記 2 次元グリッドが

50

前記 2 次元グリッドの前記第 1 の軸に沿った方向に延伸した前記セルの 1 又は 2 以上のサブセットであって、それぞれのサブセットが前記データセットの前記複数の要素のうちの 1 つの要素のそれぞれのフィールドに関連付けられる、1 又は 2 以上のサブセット、及び

前記 2 次元グリッドの前記第 2 の軸に沿った方向に延伸した前記セルの複数のサブセットであって、その 1 又は 2 以上がそれぞれの妥当性検証規則に関連付けられる、複数のサブセットを含む、前記描画するステップと、

少なくとも 1 つのプロセッサによって、前記セルの少なくともいくつかのセルから受け取られたユーザ入力に基づいて前記データセットの少なくとも 1 つの要素に妥当性検証規則を適用するステップとを含み、

フィールド及び妥当性検証規則に関連する少なくともいくつかのセルが、それぞれ、関連する妥当性検証規則が関連するフィールドに適用されるか否かを判定する入力を受け取るための入力要素、及び

前記要素の前記関連するフィールドに含まれるデータに前記関連する妥当性検証規則を適用することに基づく妥当性検証結果に関連するフィードバックを示すためのインジケータを含む、前記方法。

【請求項 19】

データセットの複数の要素の各要素の 1 又は 2 以上のフィールドに含まれるデータの妥当性を検証するための 1 又は 2 以上の妥当性検証規則を指定するための、コンピュータ可読ストレージ媒体に記憶されたコンピュータプログラムであって、ユーザーインターフェイスを表示可能なコンピュータシステムに、

(1) 前記妥当性検証規則が適用される前記データセットの複数の要素のうちの 1 つの要素のそれぞれのフィールドにおけるデータ値と、(2) 第 1 の軸及び第 2 の軸を有する 2 次元グリッドに配置された複数のセルとを備える前記ユーザーインターフェイスを描画することであって、前記 2 次元グリッドが、

前記 2 次元グリッドの前記第 1 の軸に沿った方向に延伸した前記セルの 1 又は 2 以上のサブセットであって、それぞれのサブセットが前記データセットの前記複数の要素のうちの 1 つの要素のそれぞれのフィールドに関連付けられる、1 又は 2 以上のサブセット、及び

前記 2 次元グリッドの前記第 2 の軸に沿った方向に延伸した前記セルの複数のサブセットであって、その 1 又は 2 以上がそれぞれの妥当性検証規則に関連付けられる、複数のサブセットを含む、前記描画することと、

前記セルの少なくともいくつかのセルから受け取られたユーザ入力に基づいて前記データセットの少なくとも 1 つの要素に妥当性検証規則を適用することとを実行させるための命令を含み、

フィールド及び妥当性検証規則に関連する少なくともいくつかのセルが、それぞれ、関連する妥当性検証規則が関連するフィールドに適用されるか否かを判定する入力を受け取るための入力要素、及び

前記要素の前記関連するフィールドに含まれるデータに前記関連する妥当性検証規則を適用することに基づく妥当性検証結果に関連するフィードバックを示すためのインジケータを含む、前記コンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

関連出願の相互参照

本出願は、参照により本明細書に組み込まれている、2012年10月17日に出願した米国特許出願第 13 / 653 , 995 号の優先権を主張するものである。

【0002】

本明細書は、データに対する規則の指定及び適用に関する。

【背景技術】

10

20

30

40

50

【0003】

ビジネスアプリケーションを含む多くの最近のアプリケーションは、さまざまなソースから集められる可能性があるデータの大きな組（すなわち、「データセット」）を処理する。データセットにデータを提供するさまざまなソースは、異なるレベルのデータ品質を有する可能性がある。アプリケーションが適切に機能することを保証するために、データセットの適切なレベルのデータ品質が維持されるべきである。適切なレベルのデータ品質を維持するために、データセットを、データ妥当性検証システムによって処理することができる。そのようなシステムは、データセットがアプリケーションに提供される前にそのデータセットに妥当性検証規則を適用する。一部の例において、データ妥当性検証システムは、妥当性検証規則の結果を用いてデータ品質の測定値を計算し、データ品質の測定値が予め決められた閾値未満になる場合、アプリケーションの管理者に警告する。その他の例において、データ妥当性検証システムは、妥当性検証規則のうちの1又は2以上に反するデータを取り扱うためのモジュールを含む。例えば、データ妥当性検証システムは、妥当性検証規則のうちの1又は2以上に反するデータを破棄又は修復することができる。

10

【0004】

概して、データ妥当性検証システムによって適用される妥当性検証規則は、データ妥当性検証システムの管理者によって定義される。

【発明の概要】

【課題を解決するための手段】

【0005】

一態様においては、概して、コンピューティングシステムは、データセットの複数の要素の各要素の1又は2以上のフィールドに含まれるデータの妥当性を検証するための1又は2以上の妥当性検証規則を指定する。コンピューティングシステムは、第1の軸及び第2の軸を有する2次元グリッドに配置された複数のセルを描画するように構成されたユーザインターフェースモジュールを含む。2次元グリッドは、2次元グリッドの第1の軸に沿った方向に延伸したセルの1又は2以上のサブセットであって、1又は2以上のサブセットのそれぞれのサブセットがデータセットの複数の要素のうちの1つの要素のそれぞれのフィールドに関連付けられる、1又は2以上のサブセット、及び2次元グリッドの第2の軸に沿った方向に延伸したセルの複数のサブセットであって、複数のサブセットのうちの1又は2以上がそれぞれの妥当性検証規則に関連付けられる、複数のサブセットを含む。コンピューティングシステムは、セルの少なくともいくつかのセルから受け取られたユーザ入力に基づいてデータセットの少なくとも1つの要素に妥当性検証規則を適用するように構成された処理モジュールも含む。一部の実施形態においては、フィールド及び妥当性検証規則に関連する少なくともいくつかのセルが、それぞれ、関連する妥当性検証規則が関連するフィールドに適用されるか否かを判定する入力を受け取るための入力要素を含む。一部の実施形態においては、フィールド及び妥当性検証規則に関連する少なくともいくつかのセルが、それぞれ、要素の関連するフィールドに含まれるデータに関連する妥当性検証規則を適用することに基づく妥当性検証結果に関連するフィードバックを示すためのインジケータを含む。

20

30

【0006】

態様は、以下の特徴のうちの1又は2以上を含み得る。

40

【0007】

第1の要素の第1のフィールドに含まれるデータに妥当性検証規則を適用することは、セルの入力要素において受け取られた任意の入力に基づいて、第1の要素の第1のフィールドに関連する第1の軸に沿った方向に延伸したセルのサブセットからのセルに関連する任意の選択された妥当性検証規則を決定することと、選択された妥当性検証規則に基づいて第1の要素の第1のフィールドに含まれるデータに関する妥当性検証結果を決定することを含む。

【0008】

第1の軸に沿った方向に延伸したセルの1又は2以上のサブセットは、セルの行である

50

。

【0009】

第2の軸に沿った方向に延伸したセルの複数のサブセットは、セルの列である。

【0010】

入力要素は、1又は2以上の妥当性検証規則パラメータを指定する入力を受け取るように構成される。

【0011】

妥当性検証規則のうちの1又は2以上は、評価されるときに、少なくとも2つの妥当性検証結果の組のうちの妥当性検証結果を生じ、妥当性検証結果は、妥当な結果及び妥当でない結果を含む。

10

【0012】

セルの少なくともいくつかのセルに含まれるフィードバックを示すためのインジケータは、妥当性検証結果が妥当でない結果である場合にセルに網かけ (shade) を適用するように構成される。

【0013】

入力要素は、妥当性検証規則パラメータのそれぞれの正しさを判定するようにさらに構成される。

【0014】

フィールド及び妥当性検証規則に関連する少なくともいくつかのセルは、それぞれ、セルに関連する妥当性検証規則パラメータの正しさを判定した結果を表示するための第2のインジケータを含む。

20

【0015】

フィードバックを示すためのインジケータは、妥当でない結果の数を表示するように構成される数値インジケータを含み、妥当でない結果の数は、データセットの要素のすべてに関する関連するフィールドに含まれるデータに関連する妥当性検証規則を適用することによって決定される。

【0016】

データセットは、データベースの1又は2以上のテーブルを含み、データセットの要素は、データベースのレコードを含む。

【0017】

妥当性検証規則のうちの1又は2以上は、ユーザによって定義される。

30

【0018】

妥当性検証規則のうちの1又は2以上は、予め定義される。

【0019】

2次元グリッドの第2の軸に沿った方向に延伸したセルの複数のサブセットのうちの1又は2以上は、第1の妥当性検証規則に関連する第1のセル及び第2の妥当性検証規則に関連する第2のセルを含み、第2の妥当性検証規則は、第1の妥当性検証規則とは異なる。

。

【0020】

2次元グリッドの第2の軸に沿った方向に延伸したセルの複数のサブセットのうちの1又は2以上は、既存の値に適用された妥当性検証規則のうちの1つに関する妥当でない結果に応じて対応するフィールドの既存の値を置き換えるための値を受け取るための入力要素を含むセルのサブセットを含む。

40

【0021】

2次元グリッドの第2の軸に沿った方向に延伸したセルの複数のサブセットのうちの1又は2以上は、除外される値が対応するフィールドに表示されることが妥当性検証規則が既存の値に適用されることを防止する結果となるように、除外される値を受け取るための入力要素を含むセルのサブセットを含む。

【0022】

別の態様においては、概して、コンピューティングシステムが、データセットの複数の

50

要素の各要素の1又は2以上のフィールドに含まれるデータの妥当性を検証するための1又は2以上の妥当性検証規則を指定する。コンピューティングシステムは、第1の軸及び第2の軸を有する2次元グリッドに配置された複数のセルを描画するための手段を含む。2次元グリッドは、2次元グリッドの第1の軸に沿った方向に延伸したセルの1又は2以上のサブセットであって、1又は2以上のサブセットのそれぞれのサブセットがデータセットの複数の要素のうちの1つの要素のそれぞれのフィールドに関連付けられる、1又は2以上のサブセット、及び2次元グリッドの第2の軸に沿った方向に延伸したセルの複数のサブセットであって、複数のサブセットのうちの1又は2以上がそれぞれの妥当性検証規則に関連付けられる、複数のサブセットを含む。コンピューティングシステムは、セルの少なくともいくつかのセルから受け取られたユーザ入力に基づいてデータセットの少なくとも1つの要素に妥当性検証規則を適用するための手段も含む。一部の実施形態においては、フィールド及び妥当性検証規則に関連する少なくともいくつかのセルが、それぞれ、関連する妥当性検証規則が関連するフィールドに適用されるか否かを判定する入力を受け取るための入力要素を含む。一部の実施形態においては、フィールド及び妥当性検証規則に関連する少なくともいくつかのセルが、それぞれ、要素の関連するフィールドに含まれるデータに関連する妥当性検証規則を適用することに基づく妥当性検証結果に関連するフィードバックを示すためのインジケータを含む。

10

【0023】

別の態様においては、方法は、データセットの複数の要素の各要素の1又は2以上のフィールドに含まれるデータの妥当性を検証するための1又は2以上の妥当性検証規則を指定する。方法は、ユーザインターフェースモジュールによって、第1の軸及び第2の軸を有する2次元グリッドに配置された複数のセルを描画するステップを含む。2次元グリッドは、2次元グリッドの第1の軸に沿った方向に延伸したセルの1又は2以上のサブセットであって、1又は2以上のサブセットのそれぞれのサブセットがデータセットの複数の要素のうちの1つの要素のそれぞれのフィールドに関連付けられる、1又は2以上のサブセット、及び2次元グリッドの第2の軸に沿った方向に延伸したセルの複数のサブセットであって、複数のサブセットのうちの1又は2以上がそれぞれの妥当性検証規則に関連付けられる、複数のサブセットを含む。方法は、少なくとも1つのプロセッサによって、セルの少なくともいくつかのセルから受け取られたユーザ入力に基づいてデータセットの少なくとも1つの要素に妥当性検証規則を適用するステップも含む。一部の実施形態においては、フィールド及び妥当性検証規則に関連する少なくともいくつかのセルが、それぞれ、関連する妥当性検証規則が関連するフィールドに適用されるか否かを判定する入力を受け取るための入力要素を含む。一部の実施形態においては、フィールド及び妥当性検証規則に関連する少なくともいくつかのセルが、それぞれ、要素の関連するフィールドに含まれるデータに関連する妥当性検証規則を適用することに基づく妥当性検証結果に関連するフィードバックを示すためのインジケータを含む。

20

30

【0024】

別の態様においては、概して、コンピュータ可読ストレージ媒体に記憶されたコンピュータプログラムが、データセットの複数の要素の各要素の1又は2以上のフィールドに含まれるデータの妥当性を検証するための1又は2以上の妥当性検証規則を指定する。コンピュータプログラムは、コンピュータシステムに、第1の軸及び第2の軸を有する2次元グリッドに配置された複数のセルを描画させるための命令を含む。2次元グリッドは、2次元グリッドの第1の軸に沿った方向に延伸したセルの1又は2以上のサブセットであって、1又は2以上のサブセットのそれぞれのサブセットがデータセットの複数の要素のうちの1つの要素のそれぞれのフィールドに関連付けられる、1又は2以上のサブセット、及び2次元グリッドの第2の軸に沿った方向に延伸したセルの複数のサブセットであって、複数のサブセットのうちの1又は2以上がそれぞれの妥当性検証規則に関連付けられる、複数のサブセットを含む。コンピュータプログラムは、コンピュータシステムに、セルの少なくともいくつかのセルから受け取られたユーザ入力に基づいてデータセットの少なくとも1つの要素に妥当性検証規則を適用させるための命令も含む。一部の実施形態にお

40

50

いては、フィールド及び妥当性検証規則に関連する少なくともいくつかのセルが、それぞれ、関連する妥当性検証規則が関連するフィールドに適用されるか否かを判定する入力を受け取るための入力要素を含む。一部の実施形態においては、フィールド及び妥当性検証規則に関連する少なくともいくつかのセルが、それぞれ、要素の関連するフィールドに含まれるデータに関連する妥当性検証規則を適用することに基づく妥当性検証結果に関連するフィードバックを示すためのインジケータを含む。

【0025】

態様は、以下の利点のうちの1又は2以上を有する場合がある。

【0026】

利点の中でもとりわけ、ユーザインターフェースは、規則が入力される時、データセットの単一のデータ要素に規則を適用した結果のライブフィードバックを提供することができる。このようにして、ユーザは、データセット全体に規則を適用する（潜在的に時間のかかるプロセス）必要なしにそれらの規則の有効性をテストすることができる。

10

【0027】

ユーザインターフェースは、ユーザがデータセットに対して指定された規則を実行し、データセット全体に対する指定された規則のそれぞれの実行に関するフィードバックを受け取れることを可能にする。そして、ユーザは、ユーザの期待にそぐわない指定された規則のいずれかを修正する機会を得る。

【0028】

ユーザインターフェースは、ユーザが規則を迅速且つ直感的に指定及び修正することを可能にし、時間及びリソースを節約する。

20

【0029】

本発明のその他の特徴及び利点は、以下の説明及び請求項から明らかになるであろう。

【図面の簡単な説明】

【0030】

【図1】データの妥当性を検証するための妥当性検証規則を指定するためのシステムのブロック図である。

【図2】データの妥当性を検証するための妥当性検証規則を指定するためのユーザインターフェースの図である。

【図3】妥当性検証規則を指定するためのユーザインターフェースのスクリーンキャプチャの図である。

30

【発明を実施するための形態】

【0031】

図1は、妥当性検証方法が使用され得る例示的なデータ処理システム100を示す。データ処理システム100は、ストレージデバイス、又はオンラインデータストリームへの接続などのデータの1又は2以上のソースを含み得るデータソース102を含み、それらの1又は2以上のソースのそれぞれは、さまざまな記憶フォーマット（例えば、データベーステーブル、スプレッドシートファイル、フラットテキストファイル、又はメインフレームによって使用されるネイティブフォーマット）のいずれかでデータ（「データセット」と呼ばれる場合がある）を記憶し得る。実行環境104は、ユーザインターフェース（UI, user interface）モジュール106及び処理モジュール108を含む。UIモジュール106は、データソース102からのデータを処理するために処理モジュール108によって使用される妥当性検証規則を指定するための、ユーザインターフェース112（例えば、ディスプレイスクリーン上のグラフィカルな表示）を介してユーザ110から受け取られる入力を管理する。

40

【0032】

実行環境104は、UNIXオペレーティングシステムなどの好適なオペレーティングシステムの制御の下で1又は2以上の多目的コンピュータでホストされてもよい。例えば、実行環境104は、ローカルの（例えば、SMPコンピュータなどのマルチプロセッサシステム）、又はローカルに分散された（例えば、クラスタ若しくはMPPとして接続さ

50

れた複数のプロセッサ)、又はリモートの、又はリモートに分散された(例えば、ローカルエリアネットワーク(LAN, local area network)及び/若しくは広域ネットワーク(WAN, wide-area network)を介して接続された複数のプロセッサ)、或いはそれらの任意の組み合わせのいずれかの複数の中央演算処理装置(CPU, central processing unit)を用いるコンピュータシステムの構成を含むマルチノード並列コンピューティング環境を含み得る。

【0033】

処理モジュール108は、データソース102からデータを読み取り、UIモジュール106によって取得された妥当性検証情報に基づいて妥当性検証手順を実行する。データソース102を提供するストレージデバイスは、実行環境104のローカルにあってもよく、例えば、実行環境104を実行するコンピュータに接続されたストレージ媒体(例えば、ハードドライブ114)に記憶され、又は実行環境104のリモートにあってもよく、例えば、リモート接続を介して実行環境104を実行するコンピュータと通信するリモートシステム(例えば、メインフレーム116)でホストされる。

10

【0034】

概して、データソース102からアクセスされるデータセットは、いくつかのデータ要素(例えば、予め決められたレコード構造にしたがってフォーマットされたレコード、又はデータベーステーブルの行)を含む。いくつかのデータ要素のそれぞれの要素は、場合によってはヌル(ゼロ)又は空値を含むいくつかのフィールド(例えば、レコード構造内で定義された属性、又はデータベーステーブルの列)(例えば、「名前」、「名字」、「電子メールアドレス」など)に関する値を含む場合がある。(例えば、内容若しくはデータ型に関連する)フィールドの値のさまざまな特徴、又は特定のフィールドの値の有無が、妥当である又は妥当でないと思われ得る場合がある。例えば、文字列「Smith」を含む「名字」フィールドは、妥当であると考えられ得る一方、空白の「名字」フィールドは、妥当でないと考えられ得る。

20

【0035】

データソース102からのデータセットを利用するアプリケーションの性能は、データセットが1又は2以上の妥当でないフィールドを有する、かなり多くのデータ要素を含む場合、悪影響を受ける可能性がある。処理モジュール108は、データセットにデータ妥当性検証規則を適用することを含むデータ妥当性検証手順を実行して、データセットが妥当性検証規則によって定義された品質の制約を満たすことを保証する。データ処理システム100は、データセットの品質が品質の制約を満たさない場合、システム管理者に警告する。一部の例において、処理モジュール108は、可能であれば妥当でないデータを修復し、又はさまざまなデータクレンジング手順を実行してクレンジングされたデータ要素のデータセットを生成するように構成され得る。さらに、その他の例において、処理モジュール108は、妥当でないデータを含むフィールドのリストを生成するように構成されてもよく、このフィールドのリストから、レポートを生成することができる。一部の例において、レポートは、フィールドのリストの中のフィールドのうち1又は2以上に関して、妥当でないデータを含んでいたレコードの数を含む。その他の例においては、妥当でないフィールドの集約が、フィールドのリストから計算される。

30

40

【0036】

概して、異なるアプリケーションは、異なる種類のデータを処理する。したがって、アプリケーションに応じて、データセットの要素は、異なるフィールドを含んでもよい。UIモジュール106は、データセットの妥当性を検証するために、1組の妥当性検証規則が指定され、使用されることを可能にするユーザインターフェース112を提供する。ユーザインターフェース112は、特定のデータ要素の構造の複数のフィールド(一部の実施形態においては、すべての利用可能なフィールド)を含む単一の表示を提供することができる。したがって、所与のアプリケーションに関して、ユーザ110(例えば、システム管理者)は、データのための適切な妥当性検証規則を指定することができる。

【0037】

50

1 妥当性検証ユーザインターフェース

図2を参照すると、ユーザインターフェース112の一例が、データセットの妥当性を検証するための1又は2以上の妥当性検証規則をユーザ110が指定(設定)し、確認することを容易にするように構成される。

【0038】

1.1 妥当性検証規則の指定

UIモジュール106は、第1の軸226及び第2の軸228を有する2次元グリッド(格子)225で配置された、いくつかのセル224を含むユーザインターフェース112を(例えば、コンピュータモニタ上に)描画する。セル224の1又は2以上のサブセット230(以後、発明を実施するための形態において行230という)は、2次元グリッド225の第1の軸226に沿った方向に延びる。行230のそれぞれは、フィールド218に関連付けられる。一部の例において、行230のそれぞれの最初のセル(すなわち、最も左のセル)は、行230に関連するフィールド218の名前を含む(この例において、フィールド名は「フィールド1」、「フィールド2」、...「フィールドM」である)。

10

【0039】

セル224の複数のサブセット232(以後、発明を実施するための形態において列232という)は、2次元グリッド225の第2の軸228に沿った方向に延びる。列232のうち1又は2以上は、それぞれの妥当性検証規則234に関連付けられる。一部の例において、列232のそれぞれの最初の(すなわち、最も上の)セルは、列232に関連する妥当性検証規則234の名前を含む(この例において、妥当性検証規則名は、「妥当性検証規則1」、「妥当性検証規則2」、...「妥当性検証規則N」である)。一部の例においては、第1の軸226及び第2の軸228の方向を入れ替え、フィールド218に関連する行230を列にし、妥当性検証規則234に関連する列232を行とすることができると留意されたい。

20

【0040】

一部の例において、ユーザインターフェース112は、予め定義された妥当性検証規則のリスト(図示せず)を含む(有する)。妥当性検証規則234は、例えば、ユーザ110が予め定義された妥当性検証規則のうち1若しくは2以上を2次元グリッド225内にドラッグするか、又は予め定義された妥当性検証規則のうち1つをダブルクリックし、その結果、1又は2以上の新しい列232が2次元グリッド225に追加されることによって2次元グリッド225に追加される。予め定義された妥当性検証規則は、対応するセル内で与えられ得る入力として予め定義された1組のパラメータを受け付け得る組み込み関数を有する。多くの状況に関して、妥当性検証規則の予め定義されたリストは、ユーザ110の必要を十分に満たす。しかし、一部の例においては、以下で説明されるように、ユーザ110は、2次元グリッド225に列232としてやはり追加され得るカスタムの妥当性検証規則を定義することができる。

30

【0041】

1又は2以上の妥当性検証規則の列232が2次元グリッド225に追加された後、ユーザ110は、どの妥当性検証規則234がどのフィールド218に適用されるべきかを指定することができる。所与の妥当性検証規則234が所与のフィールド218に適用されるべきであることを指定するために、ユーザ110は、まず、所与のフィールド218に関連する行230が所与の妥当性検証規則234に関連する列232と交差するセル224を選択する。それから、ユーザ110は、選択されたセル224の入力要素(例えば、テキストフィールド又はチェックボックス)に1又は2以上の妥当性検証規則パラメータ236を入力する。概して、セルに規則パラメータ236を含めることは、潜在的に、2つの目的に資する。第1の目的は、妥当性検証規則234を構成する「設定入力(configuration input)」を与えることであり、第2の目的は、所与の妥当性検証規則234が所与のフィールド218に適用されるべきであることを示すことである。したがって、セル224が妥当性検証規則パラメータ236を含まない(つまり、セルが空白のまま

40

50

ある)場合、処理モジュール108は、セル224に関連する妥当性検証規則234をセル224に関連するフィールド218に適用しないということになる。

【0042】

多くの異なる種類の規則パラメータ236が、セル224に入力され得る。場合によっては、設定入力規則を構成するために必要とされず、したがって、規則パラメータ236は、単に、対応する妥当性検証規則が適用されるべきであることを確定する「確定入力 (confirmation input)」規則パラメータである場合がある。例えば、確定入力規則パラメータを受け取るための入力要素の一例は、チェックされたときに、セル224に関連する妥当性検証規則234がセル224に関連するフィールド218に適用されるべきであることを示すチェックボックスである。さまざまな種類の妥当性検証規則の例が、妥当性

10

検証規則が設定入力によって構成されるのか否かを示す以下のリストに示される。

- ・整数 - フィールドが整数のみを含むことを検証する (設定入力は必要とされない)。
- ・妥当でない値 - フィールドが (設定入力として与えられる) ユーザ指定の妥当でない値を含まないことを検証する。

- ・最大精度 - フィールドが小数点以下の (設定入力として与えられる) ユーザ指定の数以下の桁数を有することを検証する。

- ・最大 - フィールドの値が (設定入力として与えられる) ユーザ指定の値よりも大きい場合、妥当でない。

- ・最大長 - フィールドが (設定入力として与えられる) ユーザ指定の数以下の文字数又はバイト数を有することを検証する。

20

- ・最小 - フィールドが (設定入力として与えられる) ユーザ指定の値未満である場合、妥当でない。

- ・空白でない - フィールドが空であるか又は空白のみを含む場合、妥当でない (設定入力は必要とされない)。

- ・ヌルでない - フィールドが (必要とされる設定入力として与えられる) ヌルである場合、妥当でない。

- ・パターン - 文字列フィールドが (設定入力として与えられる) 指定されたパターン通りであることを検証する。

- ・妥当な値 - フィールドが (設定入力として与えられる) ユーザ指定の妥当な値のみを含むことを検証する。

30

- ・型に対して妥当 - フィールドのデータがそのデータの型に対して妥当であることを検証する (設定入力は必要とされない)。

【0043】

妥当性検証規則の上記のリストは必ずしも包括的でないことが、留意される。

【0044】

1.2 妥当性検証規則の確認

一部の例において、UIモジュール106は、処理モジュール108がデータセットの要素の少なくとも一部にユーザ指定の妥当性検証規則234を適用した結果を表示することによってユーザインターフェース112を通じてユーザ110にフィードバックを提供する。

40

【0045】

図2に示されたユーザインターフェース112は、データセットの所与の要素244に関するフィールド218の値242を表示するように構成される。ユーザが妥当性検証規則234及びそれらの関連するパラメータ236を指定する (及び / 又は修正する) とき、処理モジュール108は、自動的に、指定された妥当性検証規則234を所与のデータ要素244のフィールド218の値242に適用し、妥当性検証規則234を適用した結果をUIモジュール106に与え、そして、UIモジュール106が、ユーザ110へのフィードバックとしてユーザインターフェース112に結果を表示する。概して、妥当性検証規則を適用した結果は、パス / フェイル (合 / 否) 結果である。そのようなパス / フェイル結果は、例えば、適切なセルを特定の色、パターン、又は網かけで満たすことによ

50

ってユーザ 110 に示され得る。図 2 において、フィールド 1 及び妥当性検証規則 1 に関連するセルは、フィールド 1 の値が妥当性検証規則 1 に反したことを示す灰色の網かけ 238 を含む。その他の例において、パス/フェイル結果は、適切なセルにインジケータアイコンを含むこと/含まないことによってユーザ 100 に示すことができる。例えば、フェイル(不合格)の結果を、セルに赤い感嘆符アイコンを含めることによって示すことができ、パス(合格)の結果を、赤い感嘆符アイコンがないことによって示すことができる。一部の例において、緑色の円などのアイコンを、パスの結果を示すためにセルに含めることができる。

【0046】

妥当性検証規則 234 を指定するとき、ユーザ 110 がデータセットの異なる要素に対する妥当性検証規則の影響を評価するためにデータセット内をナビゲーションすることが有用である場合がある。したがって、ユーザインターフェース 112 は、ユーザが(この例において、シーケンス番号を入力することによって)データセットの異なる要素を選択することを可能にするコントロール 246 を含む。ユーザが 1 つの要素から次の要素へとナビゲーションするとき、処理モジュール 108 は、現在選択されている要素に妥当性検証規則 234 を自動的に適用する。

【0047】

一部の例において、ユーザインターフェース 112 は、処理モジュール 108 がデータセットの要素のすべてに指定された妥当性検証規則 234 を適用することを可能にする実行コントロール 248 を含む。データセットへの妥当性検証規則 234 の適用が完了すると、処理モジュール 108 は、データセットに妥当性検証規則 234 を適用した結果を UI モジュール 106 に与え、そして、UI モジュール 106 が、ユーザ 110 に対してユーザインターフェース 112 で結果を表示する。一部の例において、適用された妥当性検証規則 234 に関連する各セル 234 は、フェイル結果カウントインジケータ 240 を含む。フェイル結果カウントインジケータ 240 は、セル 224 によって指定された妥当性検証規則 234 に反したデータ要素の数を表示する。

【0048】

1.3 混合された列及びカスタムの妥当性検証規則

上で述べられたように、ユーザ 110 は、予め定義された妥当性検証規則のいずれにも含まれていない機能を有する妥当性検証規則を望む場合がある。一部の例において、ユーザインターフェース 112 は、1 又は 2 以上の混合された妥当性検証規則の列を 2 次元グリッド 225 に挿入するためのオプションを含む。混合された妥当性検証規則の列は、列に含まれる(所与のフィールド 218 に関連する)各セルに関する異なる妥当性検証規則をユーザ 110 が指定することを可能にする。例えば、混合された妥当性検証規則の列の 1 つのセルが「妥当な値」テストを含み得る一方、混合された妥当性検証規則の列の別のセルは「最大」テストを含み得る。概して、ユーザ 100 は、(テストが規則パラメータを受け付ける場合は)テストのための規則パラメータが後に続くテストの名前を入力することによって混合された妥当性検証規則の列の所与のセルに関する妥当性検証規則を指定する。概して、列として 2 次元グリッド 225 に追加され得る任意の妥当性検証規則が、混合された妥当性検証規則の列の単一のセルに入力され得る。混合された妥当性検証規則の列のセルの内容のいくつかの例は、「ヌルでない」、「最大(99)」、及び「妥当な値(VM, F)」である。

【0049】

混合された妥当性検証規則の列によってもたらされる 1 つの利点は、まれにしか使用されないテストをスクリーン上により効率的に示すことによってユーザインターフェース 112 の使い勝手が高められることである。特に、ユーザ 110 は、単一のフィールド 218 にのみ当てはまる妥当性検証規則に 2 次元グリッド 225 の列 232 全体を割り当てる必要はない。例えば、混合された妥当性検証規則の列は、「妥当な電子メール」テストが単一のフィールド 218 (例えば、「email__addr」フィールド)にのみ当てはまるが、2 次元グリッド 225 の列 232 全体を占有し、それによって、貴重なスクリー

10

20

30

40

50

ンの領域 (real estate) を無駄にする状況を避けることができる。

【 0 0 5 0 】

その他の例において、ユーザ 1 1 0 は、予め定義された妥当性検証規則のリストを新しい、再利用可能な、カスタムの妥当性検証規則 2 3 4 によって拡張することができる。ユーザインターフェース 1 1 2 は、ユーザ 1 1 0 が新しい妥当性検証規則 2 3 4 の機能を定義するためのテンプレートを提供する。ユーザ 1 1 0 は、例えば、プログラミング言語、又は例えば構造化されたコメントが施された D M L コードなどの式言語 (expression language) を用いて、テンプレートの範囲内で、所望のカスタムの機能を定義する。新しい妥当性検証規則 2 3 4 は、保存されると、予め定義された妥当性検証規則のリストに追加される。ユーザ 1 1 0 は、例えば、妥当性検証規則のリストから 2 次元グリッド 2 2 5 内に妥当性検証規則をドラッグするか、又は妥当性検証規則をダブルクリックすることによって、新しいカスタムの妥当性検証規則 2 3 4 を後で使用することができる。予め定義された妥当性検証規則と同様に、新しい妥当性検証規則を 2 次元グリッド 2 2 5 内にドラッグするか、又は新しい妥当性検証規則をダブルクリックすることは、新しい列 2 3 2 を 2 次元グリッド 2 2 5 に追加させ、新しい列 2 3 2 は、新しい妥当性検証規則に関連付けられる。

10

【 0 0 5 1 】

予め定義された妥当性検証規則か、又はカスタムの妥当性検証規則かに関わらず、妥当性検証規則は、規則がヌル値又は空白値に適用されるべきかどうかを示す属性を有する場合がある。規則がヌル値に適用されるべきでないとその規則が指定する場合、最初に、値がヌル値であるかに関してテストされ、それから、ヌルである場合は規則が適用されず、又はヌルでない場合は規則が適用される。規則が空白値に適用されるべきでないとその規則が指定する場合、最初に、値が空白であるかどうかを調べるためにその値がテストされ、それから、値が空白でないことが分かった場合にだけ規則が適用される。

20

【 0 0 5 2 】

予め定義されているか又はカスタムであるかに関わらず、妥当性検証規則は、セル 2 2 4 に入力された 1 組の規則パラメータ 2 3 6 が妥当性検証規則に関して妥当であるかどうかを判定するために使用され得る論理を示す属性を有する場合がある。例えば、ユーザインターフェース 1 1 2 は、この論理を用いて、セル 2 2 4 に入力されたそれぞれの 1 組の規則パラメータ 2 3 6 が正しいか否か (correctness) を判定し、規則パラメータが (例えば、構文 (syntax) エラーが原因で) 正しくないと判定される場合、インジケータ (例えば、赤い停止サイン) がセル内に表示され、論理によって決定されたエラーメッセージが (例えば、エラーのリストで、又はホバリングしているときにはホバーツールチップとして) 表示される。規則パラメータが正しいか否か (正しさ) を調べることの別の例は、指定されたルックアップファイルの識別子が処理モジュール 1 0 8 に実際に知らされていることを確認することなど、意味 (semantics) を調べることである。

30

【 0 0 5 3 】

1 . 4 前処理列又は後処理列

一部の例において、ユーザインターフェース 1 1 2 は、フィールドの値に任意の初期処理を適用するか、又はその他の列の妥当性検証規則によって異なるように取り扱われる任意の特定の値を指定するために使用され得る前処理列を含んでもよい。また、ユーザインターフェース 1 1 2 は、妥当性検証規則によって実行されたテストの結果に応じて任意の対処を適用するために使用され得る後処理列も含んでもよい。前処理列は、例えば、妥当性検証から除外される値と、フィールド 2 1 8 のうちの 1 又は 2 以上に関する妥当性検証のデータ型とをユーザ 1 1 0 が指定することを可能にするために使用され得る。後処理列は、例えば、(例えば、異なる種類の妥当でない値を適切な置換値で置き換えるために) 要素の既存の値を置き換える置換値をユーザ 1 1 0 が指定することを可能にするために使用され得る。

40

【 0 0 5 4 】

概して、置換値は、後処理列の単一のセルに入力され、所与のフィールド 2 1 8 に関連

50

付けられる。置換値は、所与のフィールド 2 1 8 に関連する 1 又は 2 以上の妥当性検証規則 2 3 6 に違反するとき所与のフィールド 2 1 8 の値 2 4 2 を置き換える。例えば、「start_date」フィールドが 2 つの妥当性検証規則、最小 (1900 - 01 - 01) 及び最大 (2011 - 12 - 31) に関連付けられる場合、置換値の一例は、1970 - 01 - 01 である。したがって、所与のレコードに関する「start_date」フィールドの値が最小未満であるか (すなわち、1900 - 01 - 01 よりも前) 又は最大を超える (すなわち、2011 - 12 - 31 よりも後) 場合、値は、置換値 1970 - 01 - 01 で置き換えられる。文字列、日付 / 時間などのその他の種類の置換値も、後処理列で指定され得る。

【0055】

上述のように、ユーザ 1 1 0 は、除外される値の種類の前処理列で、妥当性検証から除外される 1 又は 2 以上の値を指定することもできる。例えば、「end_date」などのフィールドに関する妥当なデータは、概して、日付情報 (例えば、1900 - 01 - 01) のみを含む。しかし、一部のアプリケーションにおいては、「ACTIVE」などの別の値も「end_date」フィールドに関する妥当なデータであると指定することが望ましい場合もある。これは、除外される値の種類の前処理列に文字列「ACTIVE」を入力し、値「ACTIVE」が「start_date」フィールドに関して常に許容されることと、妥当性検証規則が指定された除外される値に適用される必要がないことを示すことによってなされ得る。

【0056】

前処理列は、フィールド 2 1 8 のうちの 1 又は 2 以上に関する妥当性検証のデータ型を指定する妥当性検証の型列も含むことができる。一部の例において、ユーザ 1 1 0 は、フィールドの妥当性を検証するために使用される DML の型宣言を入力することができる。例えば、フィールド 2 1 8 が日付を表す文字列値を含む場合、ユーザ 1 1 0 は、文字列値が実際は日付データ型を表し、したがって、そのように妥当性を検証されるべきであると指定するように DATE('YYYY-MM-DD') を入力することができる。同様に、10 進数として文字列の妥当性を検証するために、ユーザ 1 1 0 は、decimal(' ') を入力することができる。

【0057】

1.5 例示的なユーザインターフェース

図 3 を参照すると、スクリーンキャプチャが、図 2 のユーザインターフェース 1 1 2 の一実施形態を示す。ユーザインターフェース 1 1 2 は、ユーザ 1 1 0 が妥当性検証規則のフィードバックを受け取りながらデータセットに関する妥当性検証規則 2 3 4 を指定することを可能にするように構成される。

【0058】

上述のように、ユーザインターフェース 1 1 2 は、セル 2 2 4 の 2 次元グリッド 2 2 5 を含む。2 次元グリッド 2 2 5 は、データセットのデータ要素のフィールド 2 1 8 に関連するいくつかの行 2 3 0 を含む。行 2 3 0 のそれぞれの最初のセルは、行 2 3 0 に関連するフィールド 2 1 8 の名前を含み、データセットの現在選択されているデータ要素 2 4 4 に関するフィールド 2 1 8 の値 2 4 2 を括弧内に含む。フィールドについてのその他の情報も、ユーザが妥当性検証規則を指定するのを支援するために視覚的に表示され得る。この例において、第 1 のセルは、フィールド 2 1 8 の値のデータ型を視覚的に示すアイコン 2 2 0 も含む。

【0059】

図 3 において、ユーザ 1 1 0 は、2 次元グリッド 2 2 5 にいくつかの妥当性検証規則 2 3 4 を追加した。妥当性検証規則 2 3 4 は、いくつかの列 2 3 2 としてグリッドに表示される。それぞれの妥当性検証規則 2 3 4 の名前が、妥当性検証規則 2 3 4 に関連する列 2 3 2 の一番上に含まれる (例えば、「最大長」、「空白でない」、「パターン」など)。

【0060】

ユーザ 1 1 0 は、選択された妥当性検証規則 2 3 4 がデータセットの要素の 1 又は 2 以

10

20

30

40

50

上のフィールド 218 に適用されるべきであると指定している。そのようにするために、適用されるそれぞれの妥当性検証規則 234 に関して、ユーザ 110 は、妥当性検証規則 234 に関連する列 232 と妥当性検証規則 234 が適用されるべき (1 又は 2 以上の) フィールド 218 に関連する (1 又は 2 以上の) 行 230 との交差部分に規則パラメータ 236 を入力している。例えば、ユーザ 110 は、「パターン」妥当性検証規則と「zip code」フィールドとの交差部分に規則パラメータ S “99999” を入力している。入力された規則パラメータは、データセットの各要素の「zip code」フィールドを評価して要素のそれぞれの「zip code」フィールドの値が 5 つの連続する数字文字のパターンを有する文字列であるかどうかを判定するように「パターン」妥当性検証規則を構成する。同様に、「パターン」妥当性検証規則は、データセットの各要素の「pho 10
nenum」フィールドを評価して、各要素の「phonenumber」フィールドの値 242 が S “999-999-9999” (すなわち、3 つの数字文字、ダッシュ、3 つの数字文字、ダッシュ、及び 4 つの数字文字) のパターンを有する文字列であるかどうかを判定するように構成される。

【0061】

その他の種類の妥当性検証規則 234 及び規則パラメータも、図 3 に示されている。例えば、「妥当な値」妥当性検証規則が、「statename」フィールドに関する妥当な値を米国の州名の組として特定する規則パラメータ M “StateNames” を有する「statename」フィールドに適用される。上記の規則パラメータの「StateNames」の前の「M」は、実行環境 104 内でアクセス可能なメタデータ参照システムに記憶される (コードセットと呼ばれることもある) 別個のデータセットとして (例 20
えば、ユーザ 110 又はシステム管理者によって) 州名の組が定義されることを示す。この例において、州名を含むデータセットは、変数名「StateNames」によって参照される。

【0062】

一部の例において、コードセットは、ルックアップテーブルに記憶される。ルックアップテーブルのコードセットにアクセスするために、規則パラメータは、例えば、名前「StateNames」によってシステムに対して特定されたルックアップファイルが妥当な「statename」の値のソースであることを示す L “StateNames” 30
として入力される。さらにその他の例において、ユーザ 110 は、1 組の妥当な値を直接入力することができる。例えば、妥当な 1 組の性別コードは、V “M, F, U” として入力され得る。

【0063】

別の「空白でない」妥当性検証規則が、いくつかのフィールドに適用される。例えば、「空白でない」妥当性検証規則は、「空白でない」規則パラメータの列と「street」フィールドの行との交差部分のセルにチェックマーク規則パラメータが存在するので「street」フィールドに適用される。

【0064】

上述のように、ユーザインターフェース 112 は、所与の要素 244 に関するフィールド 218 の値 242 のすべてをユーザ 110 に対して表示することができる。また、UI 40
モジュール 106 は、要素 244 のフィールド 218 に関連する妥当性検証規則 234 の一部又はすべてを処理モジュール 108 に実行させる入力をユーザインターフェース 112 から受け取る。処理モジュール 108 によって生成された (1 又は 2 以上の) 結果が、UI モジュール 106 に与えられ、そして、UI モジュール 106 が、(1 又は 2 以上の) 結果に基づくフィードバックをユーザインターフェース 112 でユーザ 110 に対して表示する。図 3 においては、「statename」フィールドの値が 1 組の州名の構成要素であるかどうかをテストするために、「妥当な値」妥当性検証規則が「statename」フィールドに適用される。検査から、「statename」フィールドの値が州名「Pennsylvania」の綴りの誤り「Pennsylvannia」であることを知り得る。したがって、「妥当な値」妥当性検証規則は、所与の要素 244 に関す 50

る「state name」フィールドに関して守られていない。妥当性検証規則が守られていないことをユーザ110に対して示すために、「妥当な値」妥当性検証規則及び「state name」フィールドに関連するセルが、網かけを付けられる。

【0065】

ユーザ110は、ナビゲーションコントロール246を用いてデータセットの要素内をナビゲーションすることができる。一部の例において、ナビゲーションコントロール246は、ユーザ110が1度に1つずつデータセットの要素内に入り込むことを可能にする矢印と、ユーザ110がそれらのユーザ110が見たいデータセットの要素の番号を入力することを可能にする数値フィールドとを含む。ユーザ110がナビゲーションコントロール246を用いて異なる要素にナビゲーションするたびに、処理モジュール108は、新しい要素の値に対して指定された妥当性検証規則を実行し、値242及び妥当性検証テストの結果を示すその他の視覚的なフィードバック（例えば、セルの網かけ）が、リフレッシュ/更新される。

10

【0066】

ユーザインターフェース112は、作動されるときに、処理モジュール108にデータセットの要素のすべてに関して指定された妥当性検証規則を実行させる「テスト」ボタン248も含む。上述のように、データセットの要素のすべてに関して指定された妥当性検証規則を実行した結果は、1又は2以上の要素が指定された妥当性検証規則に反した各セルにフェイル要素カウントインジケータ240を含めることによってユーザインターフェース112においてまとめられる。図3の実施形態において、フェイル要素カウントインジケータ240は、セルによって指定された妥当性検証規則に反したデータセットの要素の数を表す数である。例えば、「state name」フィールド及び「妥当な値」妥当性検証規則に関連するセルに関するフェイル要素カウントインジケータは、データセットの要素のうちの3886個が1組の妥当な州名の構成要素ではない州名を含むことを示す。ユーザは、そのセルをクリックして、フェイルであった要素についての情報を取得することができる。

20

【0067】

1又は2以上の妥当性検証規則のテスト結果に反した各要素に関して、妥当性検証の問題点について問題点の情報の集合を集約し、後で取得されるように記憶することができる。例えば、1又は2以上の妥当性検証規則が指定されたフィールドのリストを、そのフィールドに関して妥当性検証の問題点が存在しなかった場合のゼロ個の要素のカウントを含む、そのフィールドに関して妥当性検証の問題点を有していた要素の数のカウントとともに別のビューに表示することができる。これは、ユーザが、そのフィールドに関する妥当性検証規則が実際に実行されたことを確認しながらその特定の妥当性検証規則に反した要素がないと明白に判定することを可能にする。記憶された妥当性検証の問題点の情報は、さまざまな測定基準（例えば、特定の品質の問題点を有するレコードの割合）を計算するため、又は妥当性検証の問題点の情報によってデータ要素のデータセットを補足するためにも使用され得る。

30

【0068】

2 代替

一部の例において、フェイル結果カウントインジケータ240は、ユーザ110によってクリックされるときにUIモジュール106にフェイル要素のすべてをまとめるウィンドウをユーザ110に対して表示させるハイパーリンクである。

40

【0069】

一部の例において、データ妥当性検証規則を適用した結果は、データセットの測定基準を判定するために使用され得る。例えば、測定基準は、データ品質の問題点を有するデータセットのレコードの割合を含み得る。本明細書において説明されていないその他のユーザインターフェースを使用し、これらの測定基準を指定し、ユーザ110に対して示すことができる。

【0070】

50

上記の説明はセルに網かけを付けることによってユーザにフィードバックを与えることを示すが、その他の種類のフィードバックメカニズム（例えば、音声、ポップアップウィンドウ、特別なシンボルなど）を利用することができる。

【0071】

上記の説明は、完全なデータセットを処理しながら規則を指定することを示す。しかし、一部の例においては、削減されたより管理しやすいサイズを有し、完全なデータセットを代表するテストデータセットを使用することができる。

【0072】

上述の方法は、コンピュータで実行するためのソフトウェアを用いて実装され得る。例えば、ソフトウェアは、それぞれが少なくとも1つのプロセッサ、（揮発性及び不揮発性メモリ及び/又はストレージ要素を含む）少なくとも1つのデータストレージシステム、少なくとも1つの入力デバイス又はポート、並びに少なくとも1つの出力デバイス又はポートを含む（分散、クライアント/サーバ、又はグリッドなどのさまざまなアーキテクチャであってよい）1又は2以上のプログラムされた又はプログラム可能なコンピュータシステムで実行される1又は2以上のコンピュータプログラムで手順を形成する。ソフトウェアは、例えば、データフローグラフの設計及び構成に関連するその他のサービスを提供するより大きなプログラムの1又は2以上のモジュールを形成する場合がある。グラフのノード及び要素は、コンピュータ可読媒体に記憶されたデータ構造、又はデータリポジトリに記憶されたデータモデルに準拠するその他の編成されたデータとして実装され得る。

【0073】

ソフトウェアは、多目的又は専用のプログラム可能なコンピュータによって読み取り可能なCD-ROMなどのストレージ媒体で提供されるか、又はそのソフトウェアが実行されるコンピュータのストレージ媒体にネットワークの通信媒体を介して配信される（伝播信号に符号化される）場合がある。機能のすべてが、専用のコンピュータで、又はコプロセッサなどの専用のハードウェアを用いて実行される場合がある。ソフトウェアは、ソフトウェアによって指定された計算の異なる部分が異なるコンピュータによって実行される分散された方法で実装される場合がある。それぞれのそのようなコンピュータプログラムは、本明細書において説明された手順を実行するためにストレージ媒体又はデバイスがコンピュータシステムによって読まれるときにコンピュータを構成し、動作させるために、多目的又は専用のプログラム可能なコンピュータによって読み取り可能なストレージ媒体又はデバイス（例えば、ソリッドステートメモリ若しくは媒体、又は磁気若しくは光媒体）に記憶されるか又はダウンロードされることが好ましい。本発明のシステムは、コンピュータプログラムで構成されたコンピュータ可読ストレージ媒体として実装される場合もあり、そのように構成されたストレージ媒体は、本明細書において説明された機能を実行するために特定の予め定義された方法でコンピュータシステムを動作させる。

【0074】

本発明のいくつかの実施形態が、説明された。しかしながら、本発明の主旨及び範囲から逸脱することなしにさまざまな修正がなされ得ることが理解されるであろう。例えば、上述のステップの一部は、順番に依存しない場合があり、したがって、説明された順番とは異なる順番で実行され得る。

【0075】

上述の説明は、添付の請求項の範囲によって定義される本発明の範囲を例示するように意図されており、限定するように意図されていないことを理解されたい。例えば、上述のいくつかの機能のステップは、処理全体に実質的に影響を与えることなく異なる順番で実行され得る。その他の実施形態は、以下の請求項の範囲内にある。

10

20

30

40

フロントページの続き

- (74)代理人 100150902
弁理士 山内 正子
- (74)代理人 100141391
弁理士 園元 修一
- (74)代理人 100198074
弁理士 山村 昭裕
- (72)発明者 プロコップス ロイ レオナード
アメリカ国 マサチューセッツ01890 ウィンチェスター マッコールロード39
- (72)発明者 ゲールド ジョエル
アメリカ国 マサチューセッツ02474 アーリントン リーテラス27

審査官 福田 正悟

- (56)参考文献 特開平11-031042(JP,A)
特開2011-090598(JP,A)
特開平06-162075(JP,A)
国際公開第2011/035039(WO,A1)
国際公開第2012/097278(WO,A1)
国際公開第2011/060257(WO,A1)
特開2006-113810(JP,A)
米国特許出願公開第2007/0027858(US,A1)

- (58)調査した分野(Int.Cl., DB名)
G06F 12/00
G06F 17/30