

# データ仕様復元に向けたパターン検出技術の適用検討

(株)日立製作所

橋本 康範

yasunori.hashimoto.uh@hitachi.com

## 開発における問題点

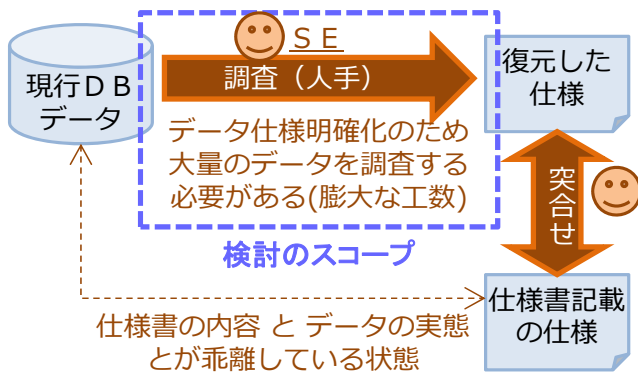
エンタープライズシステムの再構築においては、現行機能を踏襲する範囲も広いいため、現行仕様を十分に調査し明確化する必要がある。この調査対象のひとつに「現行データ」があるが、大規模システムを対象とした場合、大量のデータの調査が必要になるため、膨大な工数がかかってしまうことが問題であった。

## 手法・ツールの適用による解決

エンタープライズシステムのデータ仕様の多くは、比較的一般性が高いものであると仮説立てた。  
 このようなデータ仕様を予めパターン化しておき、パターン検出技術の利用により仕様調査を支援することで、調査作業の効率化が見込める。  
 本検討では、実例を用いた評価実験により、提案した3つのパターンについて有効性を確認した。

## 背景と課題

**背景** 現行データ仕様を明確化する際に、大量のデータを調査する必要があるため、膨大な工数がかかる

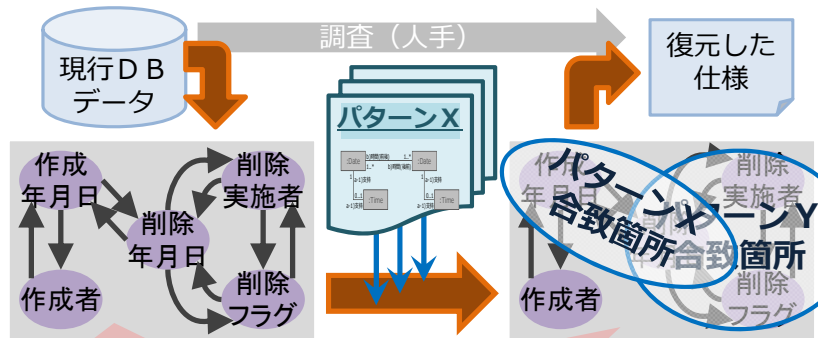


**課題** 現行データ仕様の調査作業の支援

特に、RDBカラム間の“データ依存関係”にフォーカス

## 提案手法の概要

頻出するデータ仕様(データ依存関係)を予めパターン化  
 ⇒ パターンに従う部分を自動検出することで仕様調査を支援



**STEP1:** データ仕様をグラフ構造化  
 ⇒ データを抽象化して扱うことで  
 1件1件調査する手間を削減

**STEP2:** 既知のデータ仕様をパターン検出  
 ⇒ 一般的なデータ仕様を自動で  
 検出し、人手調査範囲を削減

## 評価実験

パターンに基づくデータ仕様自動検出の有効性について評価

- 予め3種類のデータ仕様のパターン(例: 時系列の関係)を規定し、2件の実例からパターンを自動検出する実験を実施
- 以下RQの検証を通し、パターン検出技術の有効性を評価
  - RQ1:《実現性》想定したパターンを検出することは可能か?  
 ⇒ 規定した3パターン全てについて検出できることを確認
  - RQ2:《再現率》パターンは十分な範囲の仕様をカバーするか?  
 ⇒ スcope内のデータ仕様の約75.2%をパターンでカバー
  - RQ3:《適合率》検出結果を用いて仕様調査を効率化できるか?  
 ⇒ パターン検出結果の約85.4%について妥当性を確認

データ仕様調査の作業支援について一定の見通しを得た

## 考察と今後の展望

- 考察(妥当性への脅威)
  - ✓ 限られた事例を利用して評価実験を実施したため、調査対象により、有効性に差が出る可能性がある
  - ✓ 主観でパターン検出結果の妥当性を判断したため、評価者により、有効性に差が出る可能性がある
- 今後の展望
  - ✓ データ仕様のパターンの拡充
  - ✓ データ仕様パターン検出機構の汎用化
  - ⇒ データ仕様調査におけるパターン検出技術の適用効果拡大が期待できる