

Oldenburg Computer Science Series

**Univ.-Prof. Dr. Dr. h.c. Hans-Jürgen Appelrath,
Univ.-Prof. Dr. Susanne Boll (Hrsg.)**

Stefan Brüggemann

Konsistenzsicherung im Datenqualitätsmanagement

Die Betrachtung der Datenqualitätsdimension Konsistenz gewinnt im Datenqualitätsmanagement (DQM) zunehmend an Bedeutung. Der Dimension Konsistenz wurde bislang meist durch das im Datenbankentwurf weit verbreitete Konzept der Funktionalen Abhängigkeiten (FA) begegnet, welches der Prüfung der externen Konsistenz einer Datenhaltung dient. FA sind durch Bedingte FA (BFA) erweitert worden, um FA nur unter bestimmten Bedingungen gelten zu lassen. FA und BFA eignen sich zur Sicherstellung der externen Konsistenz hinsichtlich einer Datenhaltung, lassen sich aber nicht zur Prüfung der internen Konsistenz eines einzelnen Tupels verwenden.

In dieser Arbeit wird das Konzept BFA durch Konsistenzbedingungen (KB) und Konsistenzinstanzen (KI) dahingehend erweitert, dass diese auch zur Prüfung der Datenqualität (DQ) hinsichtlich der internen Konsistenz eingesetzt werden. Mit der Sicherung der internen Konsistenz verbunden sind steigende Anforderungen an die Erstellung und Wartung von KB und KI, damit diese in realen Datenmanagementszenarien sinnvoll erstellt und verwaltet werden können.

Die vorliegende Arbeit hat zur Lösung dieser Probleme und Herausforderungen als Ziel ein Vorgehensmodell zur Etablierung pro- und reaktiver Konsistenzsicherung (VEK) im DQM, in das insbesondere Domänenexperten einbezogen werden. VEK bündelt die relevanten Schritte, die notwendig sind, um Konsistenzsicherung im DQM zu etablieren. Dabei werden sowohl Aktivitäten zur Erstellung von KB und KI als auch Möglichkeiten zur pro- und reaktiven Nutzung dieser definiert. Im Rahmen einer Fallstudie wird VEK in CLiQ (Data Cleansing mit intelligentem Qualitätsmanagement), einem Vorgehensmodell zum DQM in Data Warehouse Systemen, angewendet.

Die zentralen Beiträge dieser Arbeit sind eine domänenspezifische Sprache zur Definition von KB und KI, welche es Domänenexperten erlaubt, DQ-Anforderungen zu definieren. Diese Sprache wird mit Procon (Proaktives Management von KB) in eine Methode zum proaktiven Management von KB und KI integriert. Weiter wird mit Reacon (Reaktives Management von KB) eine Methode zur reaktiven Lösung des constraint repair problems angegeben. Dabei werden für erkannte Inkonsistenzen Ersetzungsvorschläge angegeben, damit in Interaktion mit Domänenexperten Daten semiautomatisch und assistiert bereinigt werden können.

Bd. 14, XIV, 187 S., Edeweicht 2011, € 39,80
ISBN-13 978-3-939704-59-1