



Universität Augsburg
Prof. Dr. Hans Ulrich Buhl
Inhaber des Lehrstuhls für BWL, Wirtschaftsinformatik,
Informations- & Finanzmanagement
Wissenschaftlicher Leiter Kernkompetenzzentrum
Finanz- & Informationsmanagement
Vorsitzender des Boards Elitenetzwerk-Studiengang
Finanz- & Informationsmanagement
Geschäftsführender Herausgeber
Zeitschriften WIRTSCHAFTSINFORMATIK und
Business & Information Systems Engineering

Telefon
+49 821 598-4140

Telefax
+49 821 598-4225

E-Mail
hans-ulrich.buhl@wiwi.uni-augsburg.de

Datum
Augsburg, 17. August 2009

Datenqualitätsmetriken: Ein wahrscheinlichkeitstheoretischer Ansatz zur Quantifizierung der Aktualität von Daten

Verfasser: Quirin Görz - Erstprüfer: Prof. Dr. Hans Ulrich Buhl

Seit einigen Jahren gewinnt das Thema Datenqualität - unter anderem bedingt durch den verstärkten Einsatz von Data Warehouse-Systemen im Bereich des Customer Relationship Management - in Wissenschaft und Praxis immer mehr an Bedeutung. Die zunehmende Relevanz, die dieser Thematik beigemessen wird, überrascht nicht, da der Nutzen, welcher aus der Versorgung von Entscheidungsträgern und Mitarbeitern mit Daten resultiert, mit der Vollständigkeit, Korrektheit, Konsistenz und Aktualität der Daten steigt beziehungsweise fällt. Mit der steigenden Bedeutung der Datenqualität wird auch die Notwendigkeit geeigneter Mess- und Bewertungsverfahren deutlich, die für eine Planung, Steuerung und Kontrolle von Datenqualitätsmaßnahmen unabdingbar sind.

Vor diesem Hintergrund wird in der Diplomarbeit von Herrn Görz eine wahrscheinlichkeitstheoretisch fundierte Metrik zur Quantifizierung der Aktualität von Daten entwickelt, die eine Interpretierbarkeit der Metrikergebnisse gewährleistet und weitgehend automatisiert eingesetzt werden kann. Die grundsätzliche Idee besteht darin, dass Aktualität als Wahrscheinlichkeit dafür interpretiert wird, dass ein Datenattribut zum Bewertungszeitpunkt noch den aktuellen Gegebenheiten der Realwelt entspricht und nicht - bedingt durch zeitlichen Verfall - bereits veraltet ist. Die Interpretierbarkeit der Metrikergebnisse als Wahrscheinlichkeit hat mehrere Vorteile: Zum einen besitzen diese auch für Fachabteilungen Aussagekraft. Zum anderen können die Metrikergebnisse methodisch fundiert in Erwartungswertkalkülen berücksichtigt werden.

Darüber hinaus ist es im Gegensatz zu bestehenden Ansätzen möglich, relevante Zusatzdaten wie beispielsweise die Art der Hochschule bei der Quantifizierung der

Aktualität des Attributs Student, zu berücksichtigen. Dies erlaubt in der Regel eine verbesserte Quantifizierung der Aktualität eines Datenattributs. In diesem Fall wird methodisch mit bedingten Wahrscheinlichkeiten gearbeitet. Neben dem Grundmodell der Metrik wird zudem diskutiert, wie die Annahmen des Grundmodells entsprechend relaxiert werden können, um realen Gegebenheiten besser Rechnung tragen zu können und wie die Metrikergebnisse für mehrere Datenattribute (z. B. Tupel oder Relationen) aggregiert werden können.

Die entwickelte Metrik wird im Hinblick auf sechs Anforderungen evaluiert. Zudem wird deren Operationalisierung und Anwendung am Fallbeispiel eines Mobilfunkanbieters illustriert und unter anderem der Mehrwert für die Unternehmenspraxis dargestellt. Am Beispiel des Attributs Student wird so die konkrete Instanziierung der Metrik erläutert und deren konkrete Anwendung im Kampagnenmanagement aufgezeigt. So konnte bspw. die Erfolgsquote der betrachteten Kampagne bei Anwendung der hier entwickelten Metrik von 5,7% auf 11,5% erhöht werden. Zuletzt wird ein theoretischer Vergleich der entwickelten Metrik mit bestehenden Ansätzen auf Basis einer Simulationsanalyse vorgenommen.

Neben einer wissenschaftlich stringenten Modellentwicklung wurde unter anderem Wert auf eine anschauliche Kommunikation der konkreten Umsetzung der Metrik in der Unternehmenspraxis gelegt. Die Arbeit erfüllt somit neben dem Kriterium der wissenschaftlichen Rigorosität auch das vor allem in der Wirtschaftsinformatik geforderte Relevanz-Kriterium. Wesentliche Teile der Diplomarbeit sind Bestandteil eines wissenschaftlichen Beitrags. Dieser befindet sich derzeit in Begutachtung bei einer renommierten deutschsprachigen Zeitschrift.

