

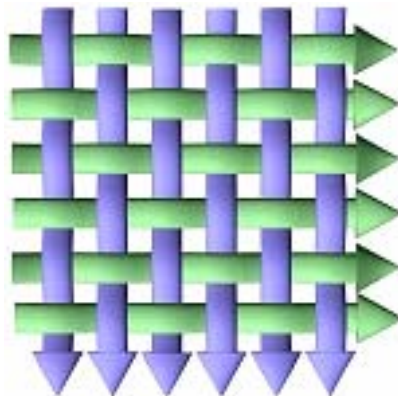


Technical Report 07006

ISSN 1612-1376

Bewertung der Informationsgüte in der Informationsgewinnung für die modellgestützte Analyse großer Netze der Logistik

Jochen Bernhard, Miroslaw Dragan



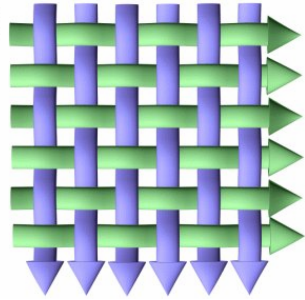
Sonderforschungsbereich 559
Modellierung großer Netze in der Logistik

Universität Dortmund
44221 Dortmund



Sonderforschungsbereich 559

**Modellierung großer
Netze in der Logistik**



Technical Report 07006

ISSN 1612-1376

**Bewertung der Informationsgüte in der
Informationsgewinnung für die modellgestützte
Analyse großer Netze der Logistik**

Teilprojekt M9:

Jochen Bernhard,
Miroslaw Dragan

Fraunhofer IML
Joseph-von-Fraunhofer-Straße 2-4
44227 Dortmund

Sigrid Wenzel

Universität Kassel
Fachbereich Maschinenbau Fachgebiet PFP
Kurt-Wolters-Straße 3
34125 Kassel

Dortmund, den 31. Oktober 2007

Inhalt

1	Einleitung und Motivation	3
2	Der Begriff „Informationsqualität“ in der Informationsgewinnung für die Analyse von GNL.....	3
3	Taxonomie zur Informationsqualität	3
	3.1 Anforderungen an die Taxonomie	3
	3.2 Informationsqualität in der Literatur	3
4	Taxonomie zur Bewertung der Informationsqualität	3
5	Nutzung der Taxonomie im Rahmen des Vorgehensmodells zur Informationsgewinnung.....	3
6	Zusammenfassung und Ausblick.....	3
7	Anhang.....	3
8	Literatur.....	3

1 Einleitung und Motivation

Valide Eingangsdaten in der richtigen Qualität, Quantität und Granularität sind Grundlagen jeder effizienten Analyse und Modellierung von großen Netzen der Logistik (GNL). Die über die Modellierung und anschließende modellgestützte Analyse erzielbare Ergebnisqualität hängt direkt von der Güte der zu Grunde liegenden Information und den daraus extrahierten Eingangsdaten ab. Vor diesem Hintergrund wurde im Rahmen der Arbeiten durch das Teilprojekt M9 aus dem Sonderforschungsbereich 559 „Modellierung großer Netze in der Logistik“ ein Vorgehen und die Methodiken für eine zielgerichtete und der Aufgabe angemessenen Informationsgewinnung mit Fokus auf die Methoden zur Datenerhebung, Statistik und Visualisierung entwickelt [BJH+06].

Unter Verwendung realisierter Methoden- und Informationsklassifikationen [BDW05], [JMa05], [FKK05] sowie eines Vorgehensmodells zur Informationsgewinnung [BJH+06] wird ein Methodennutzungsmodell entwickelt, das es dem Anwender ermöglicht, aufgrund einer gegebenen Problemstellung und einer vorgegebenen Analysemethode eine zielgerichtete Informationsgewinnung durchzuführen. Zu diesem Zweck ist die Verknüpfung der Methoden, einschließlich der enthaltenen Methoden- und Informationskategorien, mit dem Vorgehensmodell zwingend erforderlich. Zur Schaffung dieser Verknüpfung wird auf die Grundelemente Methoden, Taxonomien und Vorgehensmodell eine steuernde Metainformationsschicht gesetzt, die dem Anwender als Schnittstelle zur operativen Methodennutzung dient. Diese Schicht dient der Formulierung der Nutzeranforderungen an das Methodennutzungsmodell aus einer konkreten Anwendung heraus, beispielsweise durch Angaben bzgl. des Projektziels oder der benötigten Eingangsdatenqualität. Der Nutzer erhält hierdurch über alle Schritte des Vorgehensmodells eine Anleitung zur synergetischen Nutzung der Methoden unter Berücksichtigung der von ihm identifizierten Informationskategorien sowie bestehender Querverweise, Abhängigkeiten und Kausalzusammenhänge der Methoden- und Informationskategorien untereinander.

Zur Vervollständigung des Qualitätsmanagements über den gesamten Informationsgewinnungsprozess ist es darüber hinaus notwendig, zusätzlich zu den qualitätsbewertenden Maßnahmen beim Einsatz einzelner Methoden auch zu Aussagen bzgl. der Informationsqualität während der einzelnen Prozessschritte des Vorgehensmodells beginnend bei der Informationsidentifikation und -erhebung zu gelangen. Die Informationsqualität muss dazu anhand zu spezifizierender Kriterien – hierzu zählen beispielsweise die Bewertung der Informationsquelle hinsichtlich Aktualität und Zuverlässigkeit, aber auch qualitätsverändernde Einflüsse aufgrund der Anwendung der Methoden Datenerhebung, Statistik und Visualisierung – sukzessive innerhalb des Vorgehensmodells mitgeführt und im Sinne der erzielbaren Ergebnisse aus dem Informationsgewinnungsprozess präzisiert werden. Erst die damit realisierbare durchgängige Berücksichtigung des Faktors Qualität ermöglicht eindeutige Aussagen bezüglich der Güte nutzbarer Eingangsdaten, welche wiederum einen maßgeblichen Einfluss auf die Qualität der modellgestützten Analyseergebnisse haben. Aus diesem Grund wird auf der Basis einer Taxonomie ein Bewertungs- und Handlungsschema erarbeitet, das sowohl die verschiedenen Kriterien und Attribute, die für die Bewertung der Informationsqualität herangezogen werden, als auch deren Systematik zur Bestimmung und Nutzung der Informationsgüte im konkreten Anwendungsfall umfasst.

2 Der Begriff „Informationsqualität“ in der Informationsgewinnung für die Analyse von GNL

Der Begriff „Qualität“ wird in der Literatur häufig mit wechselnden Definitionen verwendet. Häufig wird Qualität wie in [ELi99], S. 15: „Quality is meeting or exceeding customer expectation“ oder wie in [Epp06], S. 20 „Quality is the totality of features and characteristics of a product or service that bears on its ability to satisfy given needs.“ beschrieben. Im

Kontext der Informationsgewinnung für die Analyse von GNL wird die Informationsqualität aufbauend auf [DIN04] folgendermaßen definiert:

„Informationsqualität ist die Gesamtheit von Qualitätsattributen einer Information oder eines Informationsprozesses, die sich auf deren Eignung und Güte zur Erfüllung gegebener Erfordernisse bezieht“.

Nach [BDW05] setzt sich der Begriff „Information“ wie folgt zusammen:

1. Die Inhalte einer Information bestehen aus *Informationsobjekten*, die in Beziehung zueinander stehen können.
2. Informationsobjekte sind Instanzen einer *Informationskomponente* (Klasse) im Sinne der objektorientierten Programmierung. Die Informationskomponente beinhaltet die gemeinsamen Eigenschaften/Attribute der Objekte.
3. Die Art der Beziehungen zwischen den Informationsobjekten wird als *Informationsstruktur* bezeichnet. Bestehen Beziehungen zwischen Objekten unterschiedlicher Informationskomponenten, kann dies gleichzeitig auch eine grundsätzliche Beziehung zwischen den betroffenen Informationskomponenten darstellen.

Wie schon in [BDW05] differenziert, wird im Folgenden zwischen Daten als analoge oder digitale Signale und Informationen, die Aussagen umfassen und durch Daten codiert werden, unterschieden. Zur Bildung einer Taxonomie von Qualitätskriterien und -attributen wurde des Weiteren eine Eingrenzung auf spezielle Anwendungsgebiete wie z. B. Datenbanktechnologie oder Internet möglichst vermieden. Beide hier aufgeführten Anwendungsbeispiele wurden in der hier vorgestellten Taxonomie als mögliche Informationsquellen betrachtet, womit z. B. die Homogenität von Informationsquellen als Qualitätsattribut der Informationsquellen betrachtet wurde. Anhand der im folgenden Kapitel vorgestellten Taxonomie soll Information bezüglich ihrer Informationsqualität beurteilt werden können, um damit eine Aussage zur Informationsgüte bzgl. einer Anwendung, der modellbasierten Analyse von GNL, zu erhalten.

3 Taxonomie zur Informationsqualität

Eine Taxonomie zur Bewertung der Informationsqualität für die Informationsgewinnung umfasst eine hierarchische Klassifikation von Qualitätsattributen, die eine bewertende Aussage über die Informationsqualität ermöglichen. Eine solche Taxonomie soll innerhalb des Informationsgewinnungsprozesses [BJH+06] begleitend zum Vorgehensmodell eingesetzt werden, um die Informationsqualität zu bewerten und die bestmögliche Informationsgüte bzgl. der Anwendung zu erreichen.

Im folgenden Unterkapitel werden zunächst die Anforderungen an die Taxonomie präzisiert. Anschließend folgt eine Übersicht bisheriger Taxonomiemodelle aus der Fachliteratur. Im Anschluss stellen die Autoren die entwickelte eigene Taxonomie für die Informationsqualität einschließlich der Beschreibung der Qualitätsattribute vor.

3.1 Anforderungen an die Taxonomie

Für einen adäquaten Informationsgewinnungsprozess ist neben dem Einsatz von Methoden- und Informationskategorien eine qualitative Bewertung der Informationsquellen, der Information selbst sowie der erhobenen Daten zur Bestimmung der Informationsgüte der Eingangsdaten für die Modellbildung notwendig. Für die Anwendung muss gewährleistet sein, dass die Güte der gewonnenen Eingangsdaten mit der durch die Anwendung geforderten Güte übereinstimmt, da die Güte der Eingangsdaten maßgeblichen Einfluss auf die Ergebnisse der Untersuchung hat. Hierzu wird ein Nutzungs- und Bewertungsmodell als Leitfaden für den Anwender entwickelt, das zum einen die Vorgehensweise und zum anderen die Kriterien selbst für eine solche Beurteilung beinhaltet, womit eine korrekte Bewertung der Informationsqualität selbst sowie der Güte der Eingangsdaten für die Modellbildung im Anwendungskontext sichergestellt ist.

Die primäre Aufgabe der Taxonomie für die Informationsqualitätsbewertung ist die Bereitstellung von Qualitätsattributen mit dem Ziel der Zusicherung, dass die erhobene Information eine adäquate Güte bezüglich der Anwendung aufweist. Eine Taxonomie für die Informationsqualität sollte daher folgende Aufgaben erfüllen [Epp06]:

- Informationsqualität sollte analysiert und gemessen werden können.
- Informationsqualität sowie ihre Handhabung sollten durch Anwendung der Taxonomie verbessert werden können.
- Die Taxonomie sollte als Instrument zur Bildungszwecken genutzt werden können (um z. B. die Mitarbeiter zu schulen).

3.2 Informationsqualität in der Literatur

Die Analyse des Standes der Forschung zum Thema „Informationsqualität“ legt ihren Schwerpunkt auf die Ermittlung der in der Literatur bereits bekannten Taxonomien, die zur Entwicklung einer eigenen Taxonomie beitragen können. Besonderes Augenmerk wird auf möglichst allgemeine Ansätze zur Klassifizierung der Informationsqualität gerichtet, da diese den größten Beitrag für die Anwendung innerhalb des Informationsgewinnungsprozesses leisten. Neben allgemeinen Ansätzen zu entsprechenden Taxonomien für Informationsqualität (vgl. Tabelle 1) existieren auch eine Reihe spezifischer Taxonomien wie z. B. die im Bereich des World Wide Web ([Ata99], [KSi99]), die jedoch in der folgenden Betrachtung vernachlässigt werden, da sie zu speziell für das eine Anwendungsgebiet entwickelt und somit für den behandelten Kontext nicht geeignet sind. Die allgemeinen Ansätze unterscheiden sich sowohl in der Bezeichnung der Kategorien und Attribute als auch in ihren Ordnungsschemata d. h. der Aufteilung in Kategorien und Attribute. Es ist allerdings erkennbar, dass viele Attribute in mehreren Taxonomien vorkommen, ihre Zuordnung zu einer übergeordneten Kategorie jedoch unterschiedlich ausfällt. Zu den fast durchgängig in allen Taxonomien vorkommenden Qualitätsattributen zählen „Accuracy“, „Completeness“, „Relevancy“, „Availability“. Die genauen Definitionen der Attribute werden allerdings nicht in der nachfolgenden Tabelle dargestellt, da diese immer dem Anwendungsbereich zutreffend definiert werden müssen. Des Weiteren werden die Begriffe in der jeweiligen Originalsprache der Veröffentlichung beibehalten, um evtl. Missinterpretationen durch die Übersetzung zu vermeiden.

Taxonomie	Klassifizierung	
Conceptual Framework of Data Quality ([WSt96])	Category	Dimension
	Intrinsic Data Quality	Accuracy, Objectivity, Believability, Reputation
	Contextual Data Quality	Value-Added, Relevancy, Timeliness, Completeness, Appropriate Amount of Data
	Representational Data Quality	Interpretability, Easy of Understanding, Representational Consistency, Concise Representation
	Accessibility Data Quality	Accessibility, Access security
Extended ISO Model ([ZHe96])	Characteristics	Sub-Characteristics
	Functionality	Suitability, Accuracy, Interoperability, Compliance, Security, Traceability
	Reliability	Maturity, Recoverability, Availability, Degradability, Fault Tolerance
	Efficiency	Time Behaviour, Resource Behaviour

Taxonomie	Klassifizierung		
	Usability	Understandability, Learnability, Operability, Luxury, Clarity, Helpfulness, Explicitness, Customisability, Userfriendliness,	
	Maintainability	Analysability, Changeability, Stability, Testability, Manageability, Reusability	
	Portability	Adaptability, Conformance, Replaceability, Installability.	
Informationsqualitätsmodell ([Wol99])	Qualitätsdimension	Qualitätsattribute	
	Entscheidungsrelevanz	Relevanz	
	Inhalt	Richtigkeit	
	Zeit	Terminierung, Aktualität	
	Ort	Verfügbarkeit	
	Menge	Vollständigkeit/Detaillierungsgrad, Häufigkeit	
	Form	Übersichtlichkeit, Weiterverwendbarkeit	
	Wirtschaftlichkeit	Aufwand/Nutzen	
Semiotic-based Framework for Data Quality ([SCo99])	Semiotic Level	Goal	Dimension
	Syntactic	Consistent	Well-defined Syntax
	Semantic	Complete and Accurate	Comprehensive, Unambiguous, Meaningful, Correct
	Pragmatic	Usable and Useful	Timely, Concise, Easily Accessed, Reputable
	Social	Shared Understanding of Meaning	Understood, Awareness of Bias
Conceptual Framework for measuring IS Quality ([Ded00])	Quality Category	Dimensions	
	Ergonomic Quality	Ease of Navigation, Comfortability, Learnability, Visual Signals, Audio Signals	
	Accessibility Quality	Technical Access, System Availability, Technical Security, Data Accessibility, Data Sharing, Data Convertibility	
	Transactional Quality	Controllability, Error Tolerance, Adaptability, System Feedback, Efficiency, Responsiveness	
	Contextual Quality	Value Added, Relevancy, Timeliness, Completeness, Appropriate Data	
	Representation Quality	Interpretability, Consistency, Conciseness, Structure, Readability, Contrast	

Taxonomie	Klassifizierung		
Classification based on Conflicts ([RDo00])	Data Quality Problems	Level	Main Problems
	Single-Source Problems	Schema Level	Uniqueness, Referential Integrity, ...
		Instance Level	Misspellings, Redundancy/Duplicates, Contradictory Values,...
	Multi-Source Problems	Schema Level	Naming Conflicts, Structural Conflicts,...
Instance Level		Inconsistent Aggregating, Inconsistent Timing,...	
Classification of IQ Metadata Criteria ([NRo00])	Assessment Class	IQ Criterion	
	Subject Criteria	Believability, Concise Representation, Interpretability, Relevancy, Reputation, Understandability, Value-Added	
	Object Criteria	Completeness, Customer Support, Documentation, Objectivity, Price, Reliability, Security, Timeliness, Verifiability	
	Process Criteria	Accuracy, Amount of Data, Availability, Consistent Representation, Latency, Response Time	
Mapping the IQ Dimensions into the PSP/IQ Model ([KSW02])	Quality Type	Classification	Dimension
	Product Quality	Sound Information	Free-of-Error, Concise, Representation, Completeness, Consistent Representation
		Useful Information	Appropriate Amount, Relevancy, Understandability, Interpretability, Objectivity
	Service Quality	Dependable Information	Timeliness, Security
Usable Information		Believability, Accessibility, Easy of Manipulation, Reputation, Value-Added	
A Framework for Information Quality Management([Epp06])	Information Quality Level	Information Quality Criteria	
	Community Level (Relevance)	Comprehensiveness, Accuracy, Clarity, Applicability	
	Product Level (Soundness)	Conciseness, Consistency, Correctness, Currency	
	Process Level	Convenience, Timeliness, Traceability, Interactivity	
	Infrastructure Level	Accessibility, Security, Maintainability, Speed	

Tabelle 1: Zusammenfassung bekannter Informationsqualitätstaxonomien

Die Ansätze für die Entwicklung einer für die Anwendung innerhalb des Informationsgewinnungsprozesses einsetzbaren Taxonomie für die Informationsqualität basiert maßgeblich auf [WSt96] und [Epp06]. Zunächst stellen die Autoren die Qualitätsattribute aus den in der Literatur bekannten Taxonomien zusammen. In dieser Aufstellung kommt es bewusst zu inhaltlichen Überlappungen bzw. Redundanzen, so dass eine gezielte anwendungsspezifische Selektion und Zusammenfassung der Attribute möglich wird. Zusätzlich werden eigene Attribute definiert bzw. die Definition der aus der Literatur bekannten Attribute dem eigenen Kontext angepasst. Ausgehend von der getroffenen Auswahl und unter Betrachtung einer Projektion auf das Vorgehensmodell zur Informationsgewinnung [BJH+06] verbleiben letztlich 14 Qualitätsattribute. Eine mögliche Gruppierung der Qualitätsattribute führt zu den insgesamt vier Qualitätskategorien Inhalt, Bedeutung, Herkunft und Verwendung, auf die in Kapitel 4 näher eingegangen wird. Auf eine zusätzliche Verwendung der in der Literatur häufig anzutreffenden Kriterien und Attribute, die die Informationsdarstellung betreffen, wird gezielt verzichtet, da diese Attribute bereits in der Methoden- und Informationskategorisierung enthalten sind (vgl. [BDW03] und [BDW05]).

4 Taxonomie zur Bewertung der Informationsqualität

Wie bereits im Kapitel 2 beschrieben stellt der Begriff der Informationsqualität die Gesamtheit der Qualitätsattribute der Information bzw. eines Informationsprozesses dar. Im Folgenden werden die Attribute einer Information definiert und je nach Zugehörigkeit in eine Qualitätskategorie eingeordnet (Abbildung 1). Das Ordnungsschema besteht aus den oben bereits genannten vier Kategorien. Jede Kategorie besteht wiederum aus mehreren Attributen. Zusätzlich zu der Definition der Attribute wird jeweils ein Maßstab zur Bewertung der Attributerfüllung vorgeschlagen. Die Kategorie „Inhalt“ bezieht sich in erster Linie auf das Vorhandensein und den Umfang der Daten. Die Kategorie „Bedeutung“ bezieht sich dagegen auf die in den Daten codierte Information, die durch semantische Attribute beschrieben wird. Über die Attribute der Kategorie „Herkunft“ werden insbesondere die Informationsquelle beurteilt und Aussagen darüber getroffen, inwieweit das Vorhandensein der Eingangsdaten für den Nutzer vertrauensvoll erscheint. Die Kategorie „Verwendung“ schließlich bezieht sich auf die Anwendung, in der die Information einzusetzen ist und beinhaltet Attribute, die die Qualität der Information bezüglich der Anwendung beschreiben. Zu jedem Attribut werden im Folgenden seine Definition sowie der attributspezifische Maßstab angegeben; darüber hinaus wird auf Interdependenzen zwischen den Qualitätsattributen eingegangen. Je nach Erfüllungsgrad eines Attributs, kann die Ausprägung anderer Attribute positiv (Korrelation) bzw. negativ (Gegenläufigkeit) beeinflusst werden.

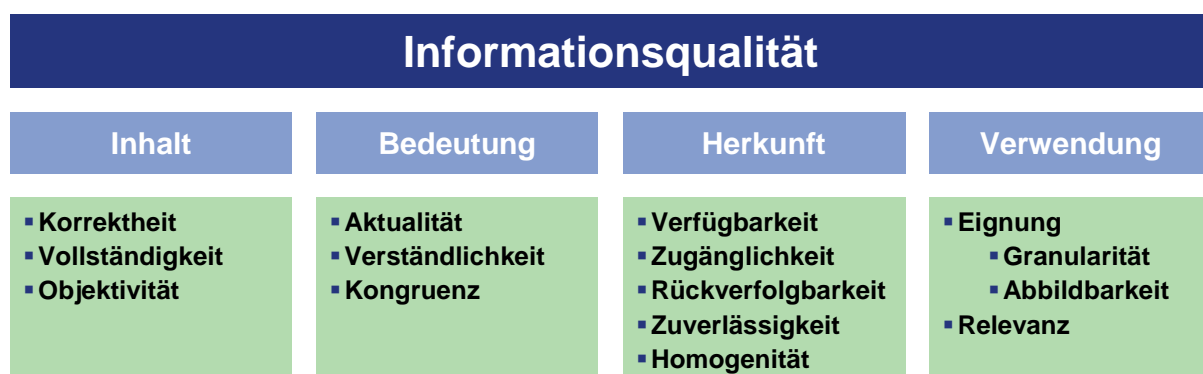


Abbildung 1: Zuordnung der Qualitätsattribute zu Qualitätskategorien

Inhalt

Mittels der Kategorie „Inhalt“ wird die technische Informationsgüte unabhängig von dem eigentlichen Informationsgehalt bewertet. Hierbei werden z. B. Aufzeichnungsfehler, Verfälschungen und Aufzeichnungslücken als negative Einflussfaktoren bzgl. der Inhaltsattribute bewertet. Die inhaltliche Informationsgüte hat einen direkten Einfluss auf die

Qualitätskriterien der Kategorien „Bedeutung“ und „Verwendung“ und wird wiederum selbst von denen der „Herkunft“ beeinflusst.

- *Korrektheit:*

Richtigkeit des Informationsinhalts. Daten und die damit verbundenen Informationen müssen das real verfügbare Informationsangebot inhaltlich korrekt wiedergeben. Mess- oder Aufzeichnungsfehler in den Datensätzen haben einen negativen Einfluss auf die Informationsqualität.

Maßstab: korrekt / korrigierbar / fehlerhaft

- *Vollständigkeit:*

Die Vollständigkeit ist ein Maß für die Abdeckung des Informationsbedarfs durch vorhandene Daten. Eine Übervollständigkeit liegt beim Auftreten redundanter Inhalte vor, Unvollständigkeit hingegen entsteht z. B. durch das Vorhandensein von Lücken im Datensatz. Dieses Attribut steht im engen Zusammenhang mit dem Qualitätsattribut Kongruenz der Qualitätskategorie „Bedeutung“.

Maßstab: übervollständig / vollständig / unvollständig

- *Objektivität:*

Objektivität beschreibt den Grad der Unabhängigkeit des Informationsinhaltes von den beteiligten Personen, Werkzeugen, Hilfsmitteln oder Institutionen während der Informationserhebung oder -dokumentation.

Maßstab: objektiv / teils / subjektiv

Bedeutung

Daten, die inhaltlich eine hohe Qualität aufweisen, implizieren nicht zwangsläufig eine bezüglich der Bedeutung hohe Qualitätsgüte der codierten Information. Die Information muss auch bezüglich ihrer semantischen Attribute hohe kontextabhängige Qualität aufweisen. Die Attribute der Kategorie „Bedeutung“ beschreibt die Informationsgüte bezüglich der Interpretierbarkeit der Information über die folgenden Attribute:

- *Aktualität:*

Die Aktualität ist ein Maß für die Zeitnähe des Informationsangebotes bezüglich des zu untersuchenden Zeitpunktes. Mögliche Einflussfaktoren sind Erhebungszeitpunkt, Aktualisierungszeitpunkt oder die Latenzzeit. Dieses Attribut steht bzgl. seiner Wichtigkeit im Zusammenhang mit dem Attribut „Relevanz“ der Information in Bezug auf seine Verwendung.

Maßstab: aktuell / eingeschränkt / unpassend

- *Verständlichkeit:*

Verständlichkeit bezeichnet den Grad der Anschaulichkeit und Eindeutigkeit des Informationsinhalts eines Informationsobjektes, so dass der Adressat diesen verstehen und nachvollziehen kann.

Maßstab: verständlich / eingeschränkt verständlich / unverständlich

- *Kongruenz:*

Kongruenz bezeichnet den Grad der Abdeckung des vorhandenen Informationsangebotes bestehend aus Informationsobjekten und deren Ausprägungen bzgl. des notwendigen Informationsbedarfs. Eine Überkongruenz kann dabei u. U. als Mehrwert verstanden werden. Dieses Attribut steht in engem Zusammenhang zur inhaltlichen Vollständigkeit sowie zur Relevanz in der Kategorie „Verwendung“.

Maßstab: überkongruent / kongruent / teilweise kongruent / disgruent

Herkunft

Qualitätsattribute dieser Kategorie bilden ein Maß für die Bewertung der Informationsquelle und der damit zu erwartenden Güte der darin enthaltenen Information. Sie sind besonders wichtig für die Vertrauensbildung bezüglich der Informationsquellen. Die die Kategorie „Herkunft“ beschreibenden Qualitätsattribute haben einen direkten Einfluss auf die Qualitätsattribute der Kategorie „Inhalt“, „Bedeutung“ sowie „Verwendung“.

- *Verfügbarkeit:*

Die Verfügbarkeit stellt ein Maß für das grundsätzliche Vorhandensein von Informationsquellen für einen notwendigen Informationsbedarf dar. Dieses Attribut bewertet jedoch weder die Zugänglichkeit zu dieser Information noch den dafür notwendigen Aufwand für die Erhebung. Das Nichtvorhandensein einer Information ist ein K.o.-Kriterium für die weitere Bewertung und führt zu einer inakzeptablen Informationsqualität.

Maßstab: vollständig / eingeschränkt / nicht verfügbar

- *Zugänglichkeit:*

Die Zugänglichkeit einer Informationsquelle wird durch den zeitlichen Aufwand beschrieben, mit dem der der Nachfragende die Information identifizieren, lokalisieren und erheben kann.

Maßstab: vollständig / eingeschränkt / unzugänglich

- *Rückverfolgbarkeit:*

Die Rückverfolgbarkeit ist ein Maß für die Rekonstruktion des Prozesses der Informationserhebung bzw. -generierung bzgl. möglicher Einflussfaktoren, Restriktionen, beteiligter Personen oder auch Institutionen und deren Bewertbarkeit.

Maßstab: vollständig / eingeschränkt / unmöglich

- *Zuverlässigkeit:*

Die Zuverlässigkeit stellt ein Maß für die Vorabbewertung einer Informationsquelle bzgl. der Informationsqualität der darin enthaltenen Informationen (Inhalt und Bedeutung) dar. Sie basiert auf der Annahme, dass die Informationsquelle eine gemäß dem Informationsinhalt korrekte und vollständige Information liefert. Ein vorhandener Schutz vor Missbrauch sowie Verfälschung der Information durch unautorisierte Personen/Instanzen impliziert somit eine erhöhte Zuverlässigkeit.

Maßstab: hoch / eingeschränkt / unzuverlässig

- *Homogenität:*

Die Homogenität beschreibt den Grad der Gleichartigkeit der Informationsquellen aus struktureller (strukturell gleiche Abbildung von Informationsobjekten) und semantischer (keine Unterschiede in der Bedeutung enthaltener Informationsobjekte) Sicht. Im Falle einer einzigen Informationsquelle sollte die Homogenität in der Regel automatisch gegeben sein.

Maßstab: homogen / teilweise homogen / heterogen

Verwendung

Die Kategorie „Verwendung“ stellt ein Maß für die Übereinstimmung zwischen der erhobenen und aufbereiteten Information und den Anforderungen aus der Anwendung zur weiteren Verwendung der Information dar. Die Attribute der Kategorie „Verwendung“ beschreiben die Informationsgüte bezüglich der Zielsetzung sowie der verwendeten Methoden und Werkzeuge.

- *Relevanz:*

Relevanz bezeichnet den Grad für die Wichtigkeit bzw. Bedeutsamkeit einer Information bzgl. deren Verwendung. Dies ist einerseits direkt von der Aufgaben- und Zielstellung sowie von der für die Untersuchung gewählten Analysemethode abhängig und stellt auch einen maßgeblichen Validierungsfaktor dar. Die Relevanz stellt für einige der Attribute der Kriterien „Inhalt“ und „Bedeutung“ die Anforderungen für deren späteren Ist-Soll-Vergleich dar. Dies betrifft ebenso die Wichtung der einzelnen Attribute untereinander (siehe Abschnitt 5).

Maßstab: hoch / mittel / niedrig

- *Eignung:*

Das Attribut Eignung bezieht sich auf die Verwendung der Information im Anschluss an den Informationsgewinnungsprozess und damit auf die Verwendung in einer modellgestützten Analyse. Dieses setzt sich zusammen aus den Qualitätsattributen:

- *Granularität:*

Die Granularität bezeichnet den für die Nutzung der Information in der modellgestützten Analyse notwendigen Detaillierungsgrad.

Maßstab: passend / ausreichend / unpassend

- *Abbildbarkeit:*

Die Abbildbarkeit beschreibt, inwieweit sich die vorliegende Information, codiert durch Daten, in der verwendeten Methode bzw. in dem verwendeten Werkzeug auch in Bezug auf den Informationsbedarf nutzen lässt. Abbildbarkeit steht daher auch in engem Zusammenhang mit der inhaltlichen Vollständigkeit (Inhalt als Attribut der Kategorie „Vollständigkeit“).

Maßstab: vollständig / eingeschränkt / nicht abbildbar

5 Nutzung der Taxonomie im Rahmen des Vorgehensmodells zur Informationsgewinnung

Für die Bewertung der Informationsqualität im Rahmen eines Informationsgewinnungsprozesses für die Eingangsdatenermittlung in der modellgestützten Planung großer Netze der Logistik ist es sinnvoll, einen Bewertungsmaßstab im Hinblick auf die relative Wichtigkeit der Qualitätsattribute untereinander, die Wichtung eines Attributes, die Festlegung eines Soll-Wertes für das jeweilige Qualitätsattribut anhand der in Kapitel 4 definierten attributspezifischen Maßstäbe und die Soll-Bewertung festzulegen. Darüber hinaus muss auch eine Bewertung der erhobenen und verarbeiteten Information selbst, der Informationsquellen sowie der Verwendbarkeit der Information in der modellgestützten Analyse, d. h. die Ist-Bewertung, durchgeführt werden. Mit Hilfe der Soll-Bewertung wird ein Zielwert für die Gewinnung eines Informationsobjektes definiert. Die Ist-Bewertung hingegen spiegelt die tatsächlich erreichten Werte für die Qualitätsattribute wider. Beide Bewertungen beziehen sich auf einen attributspezifischen Maßstab, der durch einen vorher festgelegten, einheitlichen Zahlenbereich, z. B. von 0 bis 100, linear skaliert und damit durch ein einzelnes Datum auch bewertbar ist. Die Wichtung eines Attributes ermöglicht, ebenfalls unter Nutzung eines vorher festzulegenden einheitlichen, in der Berechnung automatisch normierten Zahlenbereiches, die relative Wichtung der Attribute untereinander. Eine Bewertung der Abweichung zwischen dem zu erreichenden Soll-Wert und dem erreichten Ist-Wert zusammen mit einer relativen Wichtung ermöglicht die Identifikation eines möglichen Handlungsbedarfs für die Identifikation weiterer Informationsquellen, der weiteren Aufbereitung einer erhobenen Information sowie den Qualitätseinfluss einer Information in der späteren modellgestützten Analyse.

Die Nutzung des oben beschriebenen Schemas für die Bewertung der Informationsqualität eines Informationsobjektes erfolgt entlang des in [BJH+06] beschriebenen Vorgehensmodells für die Informationsgewinnung. Für die Informationsqualitätsbewertung wird ein einzelnes Informationsobjekt (siehe auch Kapitel 2) zugrunde gelegt.

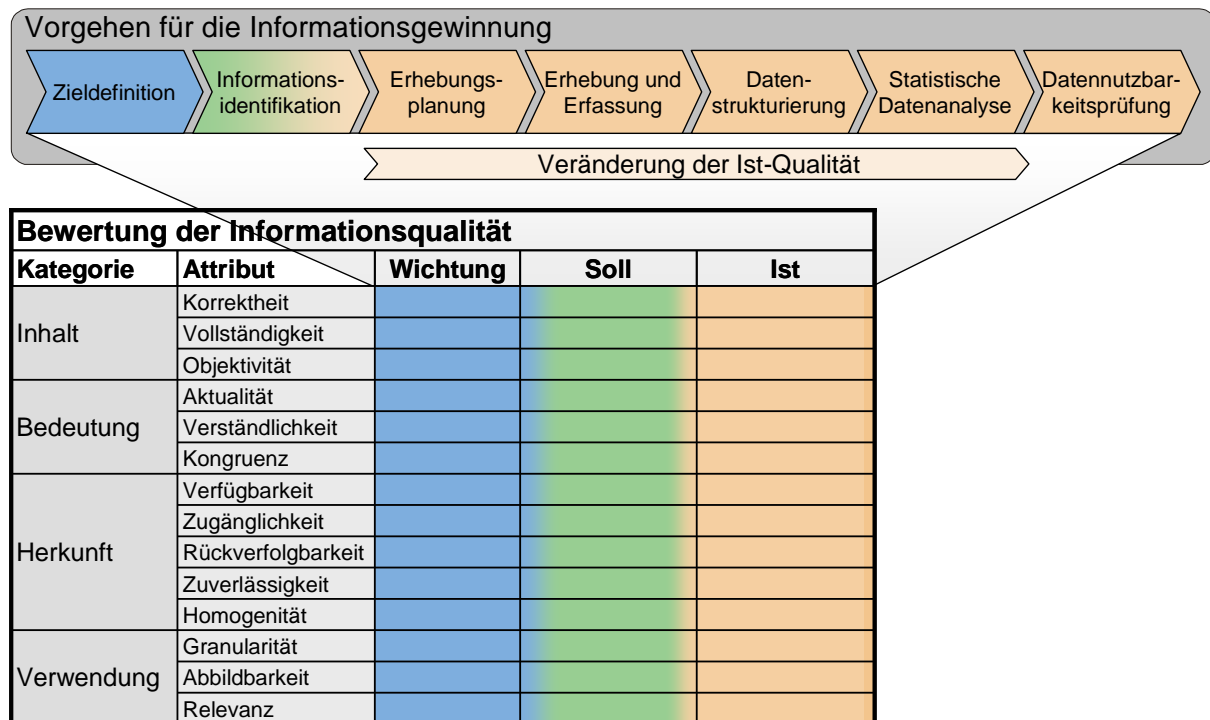


Abbildung 2: Bewertungsprinzip und Verknüpfung mit dem Informationsgewinnungsprozess.

Ausgangspunkt für den Prozess der Informationsgewinnung ist eine gegebene Aufgabenstellung zur modellgestützten Analyse mit der Festlegung des Untersuchungsgegenstandes, der Studienziele sowie der gewählten Analyse- und Modellierungsmethode. Auf dieser Basis und der Festlegung einer Zielhierarchie für die Informationsgewinnung sowie der Formulierung des objektiven, theoretischen Informationsbedarfs aus der *Zieldefinition* kann eine erste Wichtung der Qualitätsattribute untereinander erfolgen, welche im Rahmen der *Informationsidentifikation* vervollständigt wird. Maßgeblicher Prozess für die Festlegung der Soll-Bewertung ist ebenfalls die *Informationsidentifikation*. Insbesondere mit der Festlegung des subjektiven Informationsbedarfs in der Informationsbedarfsanalyse kann die Soll-Bewertung für die Qualitätsattribute für alle vier Qualitätskategorien erfolgen. Die Ist-Bewertung der Attribute der Kategorie „Herkunft“ beginnt mit der *Informationsangebotsanalyse*, in der ein möglichst umfassendes Screening aller verfügbaren Informationsquellen durchgeführt wird. Hierbei ist zu beachten, dass diese Attribute für jede Informationsquelle einzeln bewertet werden müssen, jedoch bei der Gesamtauswertung am Ende der Informationsgewinnungsprozesse nur die Bewertung der tatsächlich für ein Informationsobjekt genutzten Quellen zum Tragen kommt. Diese Auswahl findet bereits in der *Erhebungsplanung* bei der Auswahl adäquater Informationsquellen statt. Mit der Auswahl und Anwendung der Erhebungsmethode sowie der Datenerfassung im Prozessschritt der *Erhebung und Erfassung* können für die Kategorie „Inhalt“ und „Bedeutung“ erste Ist-Werte bestimmt werden. Die Weiterverarbeitung in der *Datenstrukturierung* und der *Statistischen Datenanalyse* bedingt einerseits detailliertere Aussagen bzgl. der Ist-Qualität aus „Inhalt“ und „Bedeutung“, andererseits wird mit Durchführung und Abschluss dieser Vorgehensschritte auch eine Qualitätsbewertung für die Verwendung der Daten in der späteren modellbasierten Analyse möglich. Durch das Vorhandensein eines jeweils abschließenden Validierungsprozesses in der *Erhebung und Erfassung*, der *Datenstrukturierung* und der *Statistischen Datenanalyse* sowie des expliziten Prozessschrittes der *Datennutzbarkeitsprüfung* erfolgt ein ständiger Soll-Ist-Vergleich der Informationsqualität vor dem Hintergrund einer relativen

Wichtung. Damit stellt das hier entwickelte Bewertungsschema für die Informationsqualität im Rahmen eines Informationsgewinnungsprozesses für die Eingangsdatenermittlung in der modellgestützten Planung großer Netze der Logistik ein maßgebliches Werkzeug für die Validierung der Informationsgewinnung in Bezug auf die erreichte Informationsgüte dar. In der abschließenden Datenvalidierung werden dazu hauptsächlich die Attribute der Kategorien „Inhalt“, „Bedeutung“ und „Herkunft“ verwendet, wohingegen die Attribute der Kategorie „Verwendung“ im Rahmen der Nutzbarkeitsprüfung eingesetzt werden.

6 Zusammenfassung und Ausblick

Für die Bewertung der Informationsgüte innerhalb der Informationsgewinnung in der modellgestützten Analyse großer Netze der Logistik wird in diesem Technical Report eine Taxonomie zur Bewertung der Informationsqualität entwickelt. Sie soll eine bewertende Aussage über die Informationsqualität während der einzelnen Prozessschritte des Vorgehensmodells zur Informationsgewinnung ermöglichen und so die Informationsgewinnung gezielt unterstützen. Um die Informationsqualität bewerten zu können, werden unterschiedliche Qualitätsattribute vier übergeordneten Kategorien zugeordnet. Für jedes Attribut werden ferner zwecks Bewertungsunterstützung ein attributspezifischer Maßstab vorgeschlagen sowie Korrelationsbeziehungen zu anderen Qualitätsattributen aufgestellt. Um die Anwendung der vorgestellten Taxonomie innerhalb der Informationsgewinnung zu verdeutlichen, wird in Kapitel 5 die Nutzung der Taxonomie, erweitert um die Wichtung der Attribute sowie quantitativer Soll- und Ist-Werte, anhand des Vorgehensmodells diskutiert.

Mit dem ursprünglichen Ziel des Teilprojektes M9, dem Anwender eine methodenintegrierte und disziplinübergreifende Arbeitsumgebung, bestehend aus Werkzeugen, Hilfsmitteln und Handlungsanleitungen zur Gewinnung von Eingangsdaten im Rahmen der Modellierung von GNL bereitzustellen, wurde die Entwicklung eines allgemeingültigen Methodennutzungsmodells zur zielgerichteten Informationsgewinnung auf Basis eines Methodenbaukastens (vgl. auch [HJL04], [BDW05], [FKK05]) und eines Vorgehensmodells [BJH+06] verfolgt. Die Verknüpfung des Methodenbaukastens einschließlich der enthaltenen Methoden- und Informationskategorien mit dem Vorgehensmodell sollte durch den Aufbau einer steuernden Metainformationsschicht gelöst werden. Ein wesentlicher Beitrag zur Formulierung dieser Schicht stellt die hier dargestellte Taxonomie zur Bewertung der Informationsqualität über den gesamten Informationsgewinnungsprozess dar. Insbesondere für die am Ende jedes Einzelprozesses stattfindenden Validierungsschritte liegt mit dem Bewertungsschema eine ausführliche Anleitung zur Beurteilung der Informationsgüte vor, die einen Maßstab für die zielgerichtete Gewinnung von valider Information und Daten für die Analyse von GNL zur Verfügung stellt.

7 Anhang

Bewertung der Informationsgüte

Informationsobjekt: **Systemlast**

Kategorie	Gewichtung	Soll	IST	Delta	Maßstab	Soll	Ist	
Inhalt	1,00	10,00	8,81	-1,19	hoch mittel gering	61 - 90 30 - 60 10 - 29	70,00	61,67
Korrektheit	1,00	26,67	16,67	-10,00	richtig korrigierbar fehlerhaft	61 - 90 30 - 60 10 - 29	80	50
Vollständigkeit	1,00	16,67	20,00	3,33	übervollständig vollständig unvollständig	61 - 90 30 - 60 10 - 29	50	60
Objektivität	1,00	26,67	25,00	-1,67	objektiv teils subjektiv	61 - 90 30 - 60 10 - 29	80	75
Teilwertung Inhalt	3,00	70,00	61,67	-8,33				
Bedeutung	2,00	18,10	20,00	1,90	hoch mittel gering	61 - 90 30 - 60 10 - 29	63,33	70,00
Aktualität	1,00	13,33	16,67	3,33	aktuell eingeschränkt unpassend	61 - 90 30 - 60 10 - 29	40	50
Verständlichkeit	1,00	30,00	30,00	0,00	verständlich eingeschränkt unverständlich	61 - 90 30 - 60 10 - 29	90	90
Kongruenz	1,00	20,00	23,33	3,33	kongruent teils disgruent	61 - 90 30 - 60 10 - 29	60	70
Teilwertung Bedeutung	3,00	63,33	70,00	6,67				
Herkunft	1,00	9,14	10,29	1,14	hoch mittel gering	61 - 90 30 - 60 10 - 29	64,00	72,00
Verfügbarkeit	1,00	12,00	18,00	6,00	vollständig eingeschränkt nicht verfügbar	61 - 90 30 - 60 10 - 29	60	90
Zugänglichkeit	1,00	18,00	14,00	-4,00	vollständig eingeschränkt unzugänglich	61 - 90 30 - 60 10 - 29	90	70
Rückverfolgbarkeit	1,00	6,00	9,00	3,00	vollständig eingeschränkt unmöglich	61 - 90 30 - 60 10 - 29	30	45
Zuverlässigkeit	1,00	18,00	17,00	-1,00	hoch eingeschränkt unzuverlässig	61 - 90 30 - 60 10 - 29	90	85
Homogenität	1,00	10,00	14,00	4,00	homogen teils heterogen	61 - 90 30 - 60 10 - 29	50	70
Teilwertung Herkunft	5,00	64,00	72,00	8,00				
Verwendung	3,00	32,86	30,00	-2,86	hoch mittel gering	61 - 90 30 - 60 10 - 29	76,67	70,00
Relevanz	1,00	26,67	26,67	0,00	hoch mittel gering	61 - 90 30 - 60 10 - 29	80	80
Granularität	1,00	23,33	20,00	-3,33	passend ausreichend unpassend	61 - 90 30 - 60 10 - 29	70	60
Abbildbarkeit	1,00	26,67	23,33	-3,33	vollständig eingeschränkt nicht abbildbar	61 - 90 30 - 60 10 - 29	80	70
Teilwertung Verwendung	3,00	76,67	70,00	-6,67				
Gesamteindruck	7,00	70,10	69,10	-1,00				

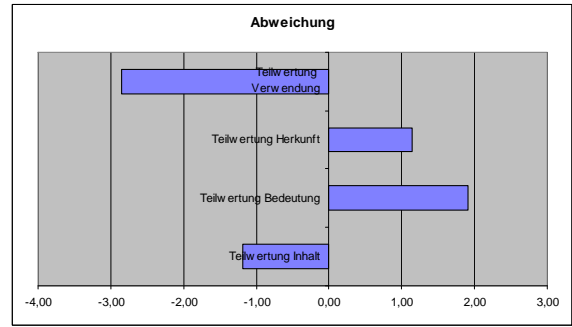
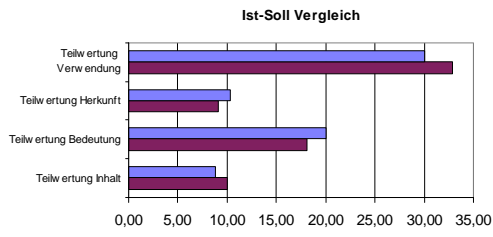


Abbildung 3: Ist-Soll-Vergleich sowie gewichtete Abweichung der Informationsgüte

8 Literatur

- [ATa99] Alexander, J. E.; Tate, M. A.: Web wisdom: how to evaluate and create information quality on the web, Lawrence Erlbaum, New Jersey, 1999.
- [BDW03] Bernhard, J.; Dragan, M.; Wenzel, S.: Klassifikation von Visualisierungsverfahren für GNL. Technical Report – Sonderforschungsbereich 559 „Modellierung großer Netze in der Logistik“ 03005, 2003, ISSN 1612-1376.
- [BDW05] Bernhard, J.; Dragan, M.; Wenzel, S.: Evaluation und Erweiterung der Kriterien zur Klassifizierung von Visualisierungsverfahren für GNL. Technical Report – Sonderforschungsbereich 559 „Modellierung großer Netze in der Logistik“ 05001, 2005, ISSN 1612-1376.
- [BJH+06] Bernhard, J.; Jodin, D.; Hömberg, K.; Kuhnt, S.; Schürmann, C.; Wenzel, S.: Vorgehensmodell zur Informationsgewinnung – Prozessschritte und Methodennutzung. Technical Report – Sonderforschungsbereich 559 „Modellierung großer Netze in der Logistik“ 06008, 2006, ISSN 1612-1376.
- [Ded00] Dedeke, A.: A Conceptual Framework for Developing Quality Measures for Information Systems. In: Proceedings of the 2000 International Conference on Information Quality (MIT IQ Conference), MA, USA, 2000, S.126-128.
- [DIN04] DIN 55350-11 (Entwurf): Begriffe zu Qualitätsmanagement und Statistik - Teil 11. Beuth Verlag, Berlin, 2004.
- [ELI99] Evans, J.R.; Lindsay, W.M.: Management and Control of Quality. South-Western College Publishing, 1999.
- [Epp06] Eppler, M.: Managing Information Quality: Increasing the Value of Information in Knowledge-intensive Products and Processes. Springer, Heidelberg, 2006.
- [FKK05] Fender, T.; Kuhnt, S.; Krampe, A.: Kriterien für die Kategorisierung statistischer Methoden im Rahmen eines Methodennutzungsmodells zur Informationsgewinnung in GNL. Technical Report – Sonderforschungsbereich 559 „Modellierung großer Netze in der Logistik“ 05005, 2005, ISSN 1612-1376.
- [HJL04] Hömberg, K.; Jodin, D.; Leppin, M.: Methoden der Informations- und Datenerhebung. Technical Report – Sonderforschungsbereich 559 „Modellierung großer Netze in der Logistik“ 04002, 2004, ISSN 1612-1376.
- [JMa05] Jodin, D.; Mayer, A.: Automatisierte Methoden und Systeme der Datenerhebung. Technical Report – Sonderforschungsbereich 559 „Modellierung großer Netze in der Logistik“ 05004, 2005, ISSN 1612-1376.
- [KSi99] Katerattanakul, P.; Siau, K.: Measuring information quality of web sites: development of an instrument. In: Proceedings of the 20th International Conference on Information Systems, Charlotte, NC, 1999, S. 279-85.
- [KSW02] Kahn, B.; Strong, D.; Wang, R.: Product and Service Performance. In: Communications of the ACM, Vol. 45, No. 4, 2002, S. 184-192.
- [Nau07] Naumann, F.: Datenqualität. Informatik Spektrum. Vol. 30 No.1. Berlin, Heidelberg: Springer, Februar 2007, S. 27-31.
- [NRo00] Naumann, F.; Rolker, C.: Assessment Methods for information quality criteria. In: Proceedings of the International Conference on Information Quality (IQ), Cambridge, MA, 2000, S. 148-162.

- [RDo00] Rahm, E.; Do, H.: Data Cleaning: Problems and Current Approaches. In: IEEE Bulletin of the Technical Committee on Data Engineering, Vol. 23 No. 4, 2000, S. 3-13.
- [Sch05] Schmiedhauser, K.: Informationsqualität in Virtual Communities of Practice. Diplomarbeit, Universität Zürich, 2005.
- [SCo99] Shanks G.; Corbitt B.: Understanding Data Quality: Social and Cultural Aspects. In: Proceedings of the 10th Australasian Conference on Information Systems, Wellington, New Zealand, Victoria University of Wellington, 1999, S. 785-797.
- [SSN05] Saake, G., Sattler, K.-U., Naumann, F. (Eds.) Datenbankspektrum – Daten- und Informationsqualität, Volume 14, Heidelberg, dpunkt.verlag, 2005.
- [Wol99] Wolf, P.: Konzept eines TQM-basierten Regelkreismodells für ein „Information Quality Management“ (IQM). Kuhn, A. (Hrsg.), Verlag Praxiswissen, Dortmund, 1999.
- [WSt96] Wang, R.; Strong, D.: Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems, Bd. 12, Nr. 4, S. 5-33, 1996.
- [ZHe96] Zeist, R.H.J.; Hendriks, P.R.H.: Specifying software quality with the extended ISO model. In Software Quality Journal, Volume 5, Number 4, Springer, 1996, S. 273-284.

Sonderforschungsbereich 559

Bisher erschienene Technical Reports

- 06003 Mirko Eickhoff, Michael Hierweck, Mathias Schwenke: Hands On ProC/B-Tools – Eine beispielorientierte Einführung in die Anwendung der ProC/B-Tools
- 06004 Doris Blutner, Stephan Cramer, Tobias Haertel: Der Mensch in der Logistik: Planer, Operateur und Problemlöser
- 06005 Tobias Haertel: UsersAward: Ein Beitrag zur optimalen Gestaltung von Mensch-Maschine-Systemen in der Logistik
- 06006 Falko Bause, Tim Geißen, Anne Meinke, Veye Tatah, Marcus Völker: Performance Evaluation for Cost Calculation of Business Processes
- 06007 Peter Kemper, Carsten Tepper: Trace Analysis – Gain Insight through Modelchecking and Cycle Reduction
- 06008 Jochen Bernhard, Dirk Jodin, Kay Hömberg, Sonja Kuhnt, Christoph Schürmann, Sigrid Wenzel: Vorgehensmodell zur Informationsgewinnung – Prozessschritte und Methodennutzung
- 06009 Doris Blutner, Stephan Cramer, Sven Krause, Tycho Mönks, Lars Nagel, Andreas Reinholz, Markus Witthaut: Ergebnisbericht der Arbeitsgruppe 5 „Assistenzsysteme für die Entscheidungsunterstützung“
- 07001 Falko Bause, Tobias Hegmanns, Stefan Pietzarka, Veye Tatah, Markus Witthaut: Ergebnisbericht der Arbeitsgruppe Neues Problemverständnis: Ergänzung des Modellierungsparadigmas
- 07002 Arnd Bernsmann, Peter Buchholz, Stephan Kessler, Andreas Reinholz, Britta von Haaren, Markus Witthaut: Bewertungs- und Dimensionierungsmethoden im Sonderforschungsbereich 559
- 07003 Jochen Bernhard, Kay Hömberg, Lars Nagel, Iwo Riha, Christoph Schürmann, Harald Sieke, Marcus Völker: Standardisierte Modelle zur Systemlastbeschreibung
- 07004 Kay Hömberg, Jan Hustadt, Dirk Jodin, Joachim Kochsiek, Lars Nagel, Iwo Riha: Basisprozesse für die Modellierung in großen Netzen der Logistik
- 07005 Kay Hömberg, Dirk Jodin, Reineke: Bewertung und Kategorisierung der Methoden zur Datenerhebung
- 07006 Jochen Bernhard, Miroslaw Dragan: Bewertung der Informationsgüte in der Informationsgewinnung für die modellgestützte Analyse großer Netze der Logistik

Alle Technical Reports können im Internet unter
<http://www.sfb559.uni-dortmund.de/>
abgerufen werden. Für eine Druckversion wenden Sie
sich bitte an die SFB-Geschäftsstelle
e-mail: andrea.grossecapenberg@iml.fraunhofer.de