# Transformation von Integritätsbedingungen bei der Schemaintegration*

Stefan Conrad  Ingo Schmitt  Can Türker

Institut für Technische Informationssysteme
Otto-von-Guericke-Universität Magdeburg
Postfach 4120, D–39016 Magdeburg, Germany

E-mail: {conrad|schmitt|tuerker}@iti.cs.uni-magdeburg.de

— Extended Abstract —

## 1 Einleitung

Im Bereich der Schemaintegration gibt es zwar vielfältige Integrationsmethoden (etwa [SPD92, RPRG94, SS96a]), doch werden, wenn überhaupt, nur einzelne Arten von einfachen Integritätsbedingungen von diesen berücksichtigt (vgl. z.B. [RPG95]). In [VA96] wird die Bedeutung von Integritätsbedingungen für die Interoperation von Datenbanksystemen deutlich hervorgehoben und einige zentrale Probleme der Integration von Integritätsbedingungen rudimentär diskutiert. In [CHJ+96] haben wir anhand einer Fallstudie bei der Integration verschiedener Datenbestände die Behandlung von Integritätsbedingungen untersucht. Daraus lassen sich einige allgemeine Gesetzmäßigkeiten ableiten, die nicht nur für die Schemaintegration gelten, sondern sich auch auf Sichtenbildung oder andere Schemaveränderungen wie Transformation, Restrukturierung und teilweise auch Schemaevolution übertragen lassen.

In diesem Beitrag zeigen wir, welche Regeln für die Behandlung von Integritätsbedingungen bei der Schemaintegration gelten. Dazu betrachten wir elementare Operationen auf Klassenextensionen, die als Basis für alle Schemaintegrationsansätze dienen. Allerdings lassen sich nicht alle Arten von Integritätsbedingungen dabei gleich behandeln. Daher unterscheiden wir verschiedene Arten von Integritätsbedingungen und zeigen, inwieweit die Anwendbarkeit der Regeln eingeschränkt werden muß. Wir beschränken uns dabei aus Platzgründen auf einige Arten von Integritätsbedingungen, die in der Praxis häufig vorkommen.

Im nächsten Abschnitt führen wir zunächst die Basisoperationen ein, die für die Integration von Schemata benötigt werden. Für diese Basisoperationen stellen wir in Abschnitt 3 Regeln für die Behandlung von Integritätsbedingungen auf. Ferner gehen wir kurz darauf ein, inwieweit diese Regeln auf verschiedene Klassen von Integritätsbedingungen anwendbar sind. Anhand eines Beispieles zeigen wir dann die Anwendung der von uns aufgestellten Regeln (Abschnitt 4) und diskutieren an einem konkreten Problem, wie der Verlust oder die unnötige Abschwächung von bestimmten Integritätsbedingungen vermieden werden kann. Abschließend fassen wir unsere Ergebnisse zusammen und geben einen Ausblick auf weitere Arbeiten.

## 2 Basisoperationen

In diesem Abschnitt sollen kurz die Operationen vorgestellt werden, die bei der Integration von Schemata verwendet werden können. Dabei beschränken wir uns hier auf die extensionale Behandlung von Klassen. Mit einer Extension einer Klasse ist nicht die Menge ihrer aktuellen Objekte gemeint. Statt dessen verwenden wir hier diesen Begriff, um die Menge von möglichen Datenbankobjekten einer Klasse zu bezeichnen. Die Menge der aktuell existierenden Datenbankobjekte stellt damit immer eine Untermenge der Klassenextension dar.

Bei der Schemaintegration kommt es darauf an, die verschiedenen Beziehungen zu entdecken, die zwischen Klassenextensionen verschiedener Schemata bestehen, und bei der Integration korrekt zu behandeln. Typischerweise unterscheidet man dabei *disjunkte*, *gleiche*, *sich einschließende* und *sich überlap-*

*Diese Arbeit wurde gefördert vom Land Sachsen-Anhalt unter FKZ 1987A/0025 (Projekt: Föderierung heterogener Datenbanksysteme und lokaler Datenhaltungskomponenten zur systemübergreifenden Integritätssicherung).

*pende* Klassenextensionen (vgl. auch [SPD92]). Der paarweise Vergleich von Klassenextensionen liefert aber nicht immer alle gewünschten und für eine korrekte Integration notwendigen Informationen. Statt dessen müssen alle Klassenextensionen gegenseitig in Beziehung gesetzt werden. Das Ergebnis einer solchen Analyse kann, wie in Abbildung 1 beispielhaft dargestellt, grafisch präsentiert werden (vgl. auch [SS96a, SS96b]. Die horizontalen Striche der vorgegebenen Klassen A, B, C, D und E stellen deren Extensionen dar. Die Klassenextensionen B und C überlappen sich gegenseitig. Die Klasse E ist extensional in Klasse D eingeschlossen, weil Klasse E eine Subklasse der Klasse D ist. Auf der Grundlage der Information über diese Überlappungen ist nun eine Zerlegung in Basisextensionen möglich.

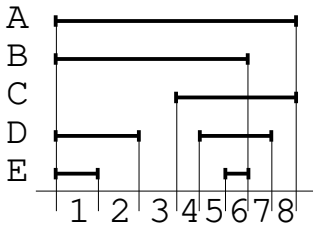


Abbildung 1: Extensionale Überlappungen

Durch die Mengenoperationen *Schnitt*, *Differenz* und *Selektion*[1] können die Basisextensionen aus den Ursprungsextensionen abgeleitet werden. Um bei der Zerlegung einen Informationsverlust zu vermeiden, muß darauf geachtet werden, daß alle möglichen Basisextensionen gefunden werden. Nach der Zerlegung können dann die Basisextensionen zu neuen Klassenextensionen zusammengesetzt werden (durch *Vereinigung*).

Diese vier Mengenoperationen sind die Basisoperationen, die von allen Ansätzen zur Schemaintegration verwendet werden und mit denen der extensionale Teil der Integration ausgedrückt werden kann. Im nächsten Abschnitt geben wir Regeln für die Behandlung von Integritätsbedingungen für die Operationen an.

# 3 Regeln für die Behandlung von Integritätsbedingungen

Im folgenden führen wir zunächst eine spezielle Notation ein, die wir bei der Formulierung der Integrationsregeln für Integritätsbedingungen nutzen werden:

$\Theta_\phi(E)$ bedeutet, daß die Integritätsbedingung $\phi$ für die Menge $E$ von Objekten gilt. Für alle nicht explizit anders gebundenen Variablen in $\phi$ ist damit $E$ der Definitionsbereich.

Neben dieser Notation benötigen wir noch eine einfache Klassifikation von Integritätsbedingungen für die Integrationsregeln. Wir unterscheiden im folgenden zwischen *Objektbedingungen* und *Klassenbedingungen*:

**Objektbedingungen:** Diese Art von Integritätsbedingungen zeichnet sich dadurch aus, daß ihre Erfüllung jeweils durch eine Überprüfung an den einzelnen Objekten einer Extension unabhängig voneinander festgestellt werden kann. Beim Einfügen oder Ändern kann eine solche Integritätsbedingung damit nur durch Betrachtung der neuen Objekte bzw. der geänderten Objekte überprüft werden.

**Klassenbedingungen:** Im Gegensatz zu den Objektbedingungen sind Klassenbedingungen dadurch charakterisiert, daß für ihre Überwachung immer mehrere oder alle Objekte einer Klasse überprüft werden müssen. Für diese Bedingungen reicht es bei Änderungsoperationen nicht aus, nur das neue, gelöschte oder veränderte Objekt zu betrachten.

Im folgenden geben wir für die vier Basisoperationen, die im vorherigen Abschnitt vorgestellt wurden, Regeln an, die die Gesetzmäßigkeiten beschreiben, die bei diesen Operationen für Integritätsbedingungen gelten. Dabei beschränken wir uns zunächst auf Objektbedingungen. Inwieweit diese Regeln auch auf Klassenbedingungen übertragbar sind, diskutieren wir später.

[1] Die Selektion ist aufgrund der beiden erstgenannten Operationen für die Erzeugung von Basisextensionen nicht unbedingt nötig. Jedoch spielt diese Operation bei der Ableitung von Sichten eine wichtige Rolle. Daher schließen wir diese Operation in unsere Betrachtungen ein.

**Selektion.** Die Anwendung einer Selektion $\sigma_\psi$ bedeutet vom logischen Standpunkt, daß wir genau die Objekte einer Extension $E$ in einer Teilmenge ($\sigma_\psi(E)$) zusammenfassen, die die vorgegebenen Selektionsbedingung $\psi$ erfüllen. Eine Selektionsbedingung ist dabei immer eine Objektbedingung, die für die Ergebnismenge $\sigma_\psi(E)$ auch als Integritätsbedingung aufgefaßt werden kann. Gilt für die Ursprungsextension $E$ bereits eine Objektbedingung $\phi$, so folgt, daß für $\sigma_\psi(E)$ die logische Konjunktion $\phi \wedge \psi$ gelten muß:

$$\Theta_\phi(E) \rightarrow \Theta_{\phi\wedge\psi}(\sigma_\psi(E)).$$

**Differenz/Teilmenge.** Die Operationen zur Bildung einer Mengendifferenz und zur Bildung einer Teilmenge unterliegen hinsichtlich der Behandlung von Integritätsbedingungen denselben Gesetzmäßigkeiten, da das Ergebnis einer Mengendifferenz zweier Extensionen immer Teilmenge einer dieser Extensionen ist. Wenn nun eine Objektbedingung $\phi$ für eine Extension gilt, gilt sie natürlich auch für jede Teilmenge dieser Extension. Dies liegt daran, daß die Bedingung für jedes einzelne Objekt der Extension unabhängig von anderen Objekten gilt. Als Regel können wir dies nun wie folgt ausdrücken (dabei nutzen wir in der Formulierung der Regel die implizite Eigenschaft aus, daß $E_1$ immer Teilmenge von $E_1 \cup E_2$ ist):

$$\Theta_\phi(E_1 \cup E_2) \rightarrow \Theta_\phi(E_1).$$

Für die Anwendung dieser Regel ist wichtig, daß die Ursprungsextension, für die $\phi$ gilt, sich als Vereinigung zweier Extensionen darstellen läßt. Wenn wir eine Ursprungsextension $E$ mit einer Teilmenge $E_1$ haben, dann können wir $E$ auch als $E_1 \cup (E - E_1)$ darstellen und die Regel anwenden.

Wollen wir diese Regel für die Mengendifferenz $E' - E''$ anwenden, so müssen wir $E_1$ durch $E' - E''$ und $E_2$ durch $E' \cap E''$ ersetzen. Da immer $E' = (E' - E'') \cup (E' \cap E'')$ gilt, erhalten wir:

$$\Theta_\phi(E') \rightarrow \Theta_\phi(E' - E'').$$

**Schnittmenge.** Für die Schnittmenge von Extensionen kann die Konjunktion der jeweiligen Integritätsbedingungen gebildet werden. Da jedes Objekt in $E_1 \cap E_2$ sowohl in $E_1$ als auch in $E_2$ enthalten ist, erfüllt es natürlich gleichzeitig die Integritätsbedingung $\phi$ der Extension $E_1$ und $\psi$ der Extension $E_2$:

$$\Theta_\phi(E_1), \Theta_\psi(E_2) \rightarrow \Theta_{\phi\wedge\psi}(E_1 \cap E_2).$$

**Vereinigung.** Bei der Vereinigung von Extensionen $E_1$ und $E_2$ können wir im allgemeinen nur das Wissen ausnutzen, daß jedes Objekt der Vereinigung $E_1 \cup E_2$ auch Objekt der Extension $E_1$ *oder* der Extension $E_2$ ist (wobei dieses *Oder* ein logisches Oder und kein Entweder-Oder ist). Wenn also für die Objekte in $E_1$ eine Objektbedingung $\phi$ gilt und $\psi$ für die Objekte in $E_2$, dann wissen wir von jedem einzelnen Objekt in $E_1 \cup E_2$ nur, daß es mindestens eine dieser Objektbedingungen erfüllt:

$$\Theta_\phi(E_1), \Theta_\psi(E_2) \rightarrow \Theta_{\phi\vee\psi}(E_1 \cup E_2).$$

Seien $\phi$ und $\psi$ Objektbedingungen, dann gilt:

$$\Theta_\phi(E) \rightarrow \Theta_{\phi\wedge\psi}(\sigma_\psi(E)) \quad (1)$$

$$\Theta_\phi(E_1 \cup E_2) \rightarrow \Theta_\phi(E_1) \quad (2)$$

$$\Theta_\phi(E_1), \Theta_\psi(E_2) \rightarrow \Theta_{\phi\wedge\psi}(E_1 \cap E_2) \quad (3)$$

$$\Theta_\phi(E_1), \Theta_\psi(E_2) \rightarrow \Theta_{\phi\vee\psi}(E_1 \cup E_2) \quad (4)$$

$$\Theta_{\psi\wedge\phi}(E) \rightarrow \Theta_\psi(E) \quad (5)$$

$$\Theta_\psi(E) \rightarrow \Theta_{\psi\vee\phi}(E) \quad (6)$$

$$\Theta_\phi(E), \Theta_\psi(E) \rightarrow \Theta_{\phi\wedge\psi}(E) \quad (7)$$

$$\rightarrow \Theta_{\mathsf{true}}(E) \quad (8)$$

$$E \equiv E_1 \cup E_2 : \ \Theta_\phi(E) \leftrightarrow \Theta_\phi(E_1 \cup E_2) \quad (9)$$

Abbildung 2: Zusammenstellung der Regeln für *Objektbedingungen*.

In Abbildung 2 haben wir die zuvor besprochenen Regeln zusammengefaßt (Regel (1) bis (4)). Die Regeln (5) und (6) behandeln noch einige elementare logische Eigenschaften, die durch die vorhergehenden Regeln nicht erfaßt werden. Dabei handelt es darum, daß jeder einzelne Teil einer Konjunktion für

sich alleine auch als Integritätsbedingung aufgefaßt werden kann (Regel (5)) und daß jede Integritätsbedingung durch disjunktives Hinzufügen einer beliebigen anderen Bedingung abgeschwächt werden kann (Regel (6)). Regel (7) drückt aus, daß zwei Integritätsbedingungen, die für dieselbe Extension gelten, auch als Konjunktion zusammengefaßt werden dürfen. Die Regel (8) gehört aus Vollständigkeitsgründen auch dazu, da trivialerweise für jede beliebige Extension die Bedingung, die immer erfüllt ist, gilt. Eine weitere allgemeingültige Regel ist (9). Sie besagt, daß bei einer (vollständigen) Zerlegung einer Menge in Teilmengen, alle für die Ausgangsmenge geltenden Integritätsbedingungen auch für die Vereinigung der Teilmengen gilt, da diese zusammen die Ausgangsmenge bilden.

Regel (7) ist eigentlich nur ein Spezialfall der Regel (3), der sich daraus ergibt, daß $E_1$ und $E_2$ gleich gesetzt werden, also dieselbe Menge von Objekten bezeichnen. Man kann in diesem Fall auch die Regel (4) anwenden, da die Vereinigung einer Menge mit sich selbst auch wieder dieselbe Menge ergibt. Die Regel (4) liefert uns aber eine Disjunktion der beiden Integritätsbedingungen zurück. Dieses Ergebnis ist formal korrekt, aber nicht das gewünschte. Ziel ist es, eine möglichst scharfe Integritätsbedingung zu erhalten, in diesem Fall also die Konjunktion von $\phi$ und $\psi$, wie sie Regel (7) liefert.

Die Regeln (5), (6) und (7) repräsentieren allgemeine logische Eigenschaften. Daher gelten sie nicht nur für Objektbedingungen sondern für beliebige Integritätsbedingungen. Dies trifft auf die Regeln (1) bis (4) nicht zu.

### Anwendbarkeit der Regeln auf Klassenbedingungen

Die im vorangegangenen Abschnitt diskutierten Regeln gelten für Objektbedingungen. Damit stellt sich natürlich die Frage, ob sie auf Klassenbedingungen übertragbar sind oder ob für diese Art von Integritätsbedingungen ganz andere Regeln gelten. Aus Platzgründen wollen wir hier nur einige erste Ergebnisse für spezielle Arten von Klassenbedingungen, die in der Praxis häufig auftreten (wie etwa *Eindeutigkeitsbedingungen/Schlüsselbedingungen*, *referentielle Integrität* sowie Bedingungen, die die Aggregatsfunktionen max und min verwenden) zusammenfassen, ohne die Details zu diskutieren.

| **Regel** | **unique** | **ref** | $\mathbf{max}_\theta$ | $\mathbf{min}_\theta$ |
|---|---|---|---|---|
| **(1)** *Selektion* | $\surd$ | — | $\theta \in \{\leq, <\}$ | $\theta \in \{\geq, >\}$ |
| **(2)** *Differenz* | $\surd$ | — | $\theta \in \{\leq, <\}$ | $\theta \in \{\geq, >\}$ |
| **(3)** *Schnitt* | $\surd$ | — | $\theta \in \{\leq, <\}$ | $\theta \in \{\geq, >\}$ |
| **(4)** *Vereinigung* | — | $\surd$ | $\theta \in \{\geq, >, <, \leq\}$ | $\theta \in \{\leq, <, \geq, >\}$ |

Abbildung 3: Anwendbarkeit der Regeln auf Klassenbedingungen

Abbildung 3 gibt an, welche der vier Grundregeln für die Basisoperationen auf welche Klassenbedingungen übertragbar sind. Die Regel (4) gilt nicht bei Eindeutigkeitsbedingungen auf Klassenextensionen, da bei der Vereinigung mit einer anderen Klassenextension Objekte hinzukommen können, die die Eindeutigkeit verletzen. Ferner gelten die Regeln (1), (2) und (3) nicht für referentielle Integritätsbedingungen, da bei den ihnen zugrundeliegenden Operationen (Selektion, Differenz/Teilmengenbildung, Schnitt) Objekte der Ausgangsextension wegfallen können, die für die referentielle Integrität erforderlich sind. Bei den Aggregationsfunktionen max und min hängt die Anwendbarkeit der Regeln davon ab, mit welchem Vergleichsoperator sind kombiniert werden. So können z.B. die Regeln (1), (2) und (3) nicht für eine Integritätsbedingung max(Alter) > 20 angewendet werden.

Zum Verständnis dieser kurzen Zusammenstellung muß noch angemerkt werden, daß z.B. bei Regel (4) zwei unterschiedliche Arten von Integritätsbedingungen für die zu vereinigenden Extensionen zusammentreffen können. In dem Fall muß die Anwendbarkeit der Regel für beide Integritätsbedingungen unabhängig voneinander gewährleistet sein.

## 4 Beispielhafte Anwendung der Regeln

Im Abschnitt 3 haben wir die einzelnen Regeln zur Behandlung von Integritätsbedingungen diskutiert. Zusammen bilden sie einen Regelsatz. Wir wollen an dieser Stelle einige Richtlinien zur Verwendung dieses Regelsatzes geben, damit mit möglichst wenig Aufwand eine Integration ohne Informationsverlust durchgeführt werden kann. Das folgende Beispiel soll dazu dienen, die Anwendung der Regeln entspre-

chend den Richtlinien zu veranschaulichen. Die Integration findet in zwei Etappen statt: (1) Zerlegung der Ursprungsextensionen in Basisextensionen und (2) Vereinigung der Basisextensionen zu abgeleiteten Extensionen.

**Zerlegung von Extensionen.** Die Zerlegung der Ursprungsextensionen basiert auf der extensionalen Analyse. Aufgrund dieser Analyse ist bekannt, mit welchen Mengenoperationen die Basisextensionen erzeugt werden können. In unserem Beispiel betrachten wir zwei Extensionen, die Mitarbeiter jeweils eines bestimmten Projektes beinhalten. Das Ergebnis der extensionalen Analyse ist in Abbildung 4 dargestellt. Es kann Mitarbeiter geben, die jeweils nur einem Projekt zugeordnet sind. Zusätzlich können aber auch Mitarbeiter existieren, die beiden Projekten angehören (horizontaler Überlappungsbereich).

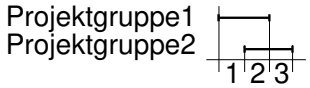


Abbildung 4: Extensionale Überlappungen

Entsprechend diesen Informationen kann die Zerlegung der Ursprungsextensionen in drei Basisextensionen durchgeführt werden. Das Ergebnis der Zerlegung ist im mittleren Teil der Abbildung 5 dargestellt.

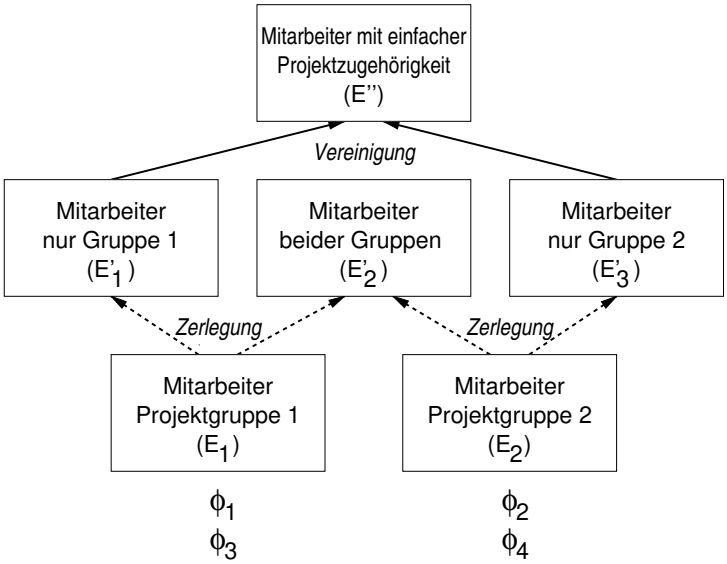


Abbildung 5: Zerlegung und Vereinigung von Extensionen am Beispiel

In unserem Beispiel gehen wir davon aus, daß den beiden Ursprungsextensionen verschiedene Integritätsbedingungen zugeordnet sind:

$$\Theta_{\phi_1}(E_1) \quad mit \quad \phi_1 \equiv \forall e : e.\text{qualifikation} > 4$$
$$\Theta_{\phi_3}(E_1) \quad mit \quad \phi_3 \equiv \forall e_1, e_2 : e_1.\text{pid} = e_2.\text{pid} \Longrightarrow e_1 = e_2$$
$$\Theta_{\phi_2}(E_2) \quad mit \quad \phi_2 \equiv \forall e : e.\text{qualifikation} > 3$$
$$\Theta_{\phi_4}(E_2) \quad mit \quad \phi_4 \equiv \exists e_1 : \forall e_2 : e_1.\text{gehaltsstufe} \geq e_2.\text{gehaltsstufe} \wedge e_1.\text{gehaltsstufe} > 3$$

Die Integritätsbedingungen $\phi_1$ und $\phi_3$ sind Objektbedingungen während $\phi_2$ und $\phi_4$ Klassenbedingungen darstellen; $\phi_2$ ist eine Eindeutigkeitsbedingung über das Attribut pid (in der Anwendung verwendete Mitarbeiteridentifikatoren) während $\phi_4$ eine Maximumbedingung (entspricht max(gehaltstufe)>3) darstellt. Die letzte Bedingung sagt aus, daß es in der Projektgruppe 2 mindestens einen Mitarbeiter g die Anwendungeben muß, dessen Gehaltsstufe die Stufe 3 übersteigt. Es gelten zwischen den Basisextensionen und den Ursprungsextensionen folgende Mengenbeziehungen:

$$E_1' = E_1 - E_2, \quad E_2' = E_1 \cap E_2, \quad E_3' = E_2 - E_1$$

Nachdem die Mengenbeziehungen zwischen den Basisextensionen und Ursprungsextensionen bekannt sind, können die Integritätsbedingungen der Basisextensionen entsprechend den im vorherigen Abschnitt genannten Regeln abgeleitet werden. Dazu werden die Integritätsbedingungen bei der Zerlegung einzeln betrachtet. Wenig Probleme bereiten die Objektbedingungen. Sie können durch Anwenden der Regel (2) und (9) abgeleitet werden:

$$\Theta_{\phi_1}(E_1) \overset{\text{Regel } 9}{\longrightarrow} \Theta_{\phi_1}(E_1' \cup E_2') \overset{\text{Regel } 2}{\longrightarrow} \Theta_{\phi_1}(E_1')$$
$$\Theta_{\phi_1}(E_1) \overset{\text{Regel } 9}{\longrightarrow} \Theta_{\phi_1}(E_2' \cup E_1') \overset{\text{Regel } 2}{\longrightarrow} \Theta_{\phi_1}(E_2')$$
$$\Theta_{\phi_2}(E_2) \overset{\text{Regel } 9}{\longrightarrow} \Theta_{\phi_2}(E_2' \cup E_3') \overset{\text{Regel } 2}{\longrightarrow} \Theta_{\phi_2}(E_2')$$
$$\Theta_{\phi_2}(E_2) \overset{\text{Regel } 9}{\longrightarrow} \Theta_{\phi_2}(E_3' \cup E_2') \overset{\text{Regel } 2}{\longrightarrow} \Theta_{\phi_2}(E_3')$$

Die Übertragung der Klassenbedingungen hingegen bereitet mehr Probleme. Die Regel (2) angewendet auf eine Eindeutigkeitsbedingung führt zu einer Abschwächung. Diese Abschwächung kann dadurch abgefangen werden, daß das Ergebnis der Regel (9) als eigenständige Bedingung betrachtet wird.

Nun könnte man fordern, daß immer nur die Regeln angewendet werden sollen, die zu keiner Abschwächung der Integritätsbedingung führen. Jedoch könnte es passieren, daß in selektiven Sichten solche umgewandelten starken Integritätsbedingungen im Gegensatz zu abgeschwächten Integritätsbedingungen nicht übernommen werden können. Um den Semantikverlust so gering wie möglich zu halten, empfehlen wir daher die gemeinsame Anwendung der Regeln, die zu keiner Abschwächung der Integritätsbedingung führen, und abschwächender Regeln. Daher wird bei der Behandlung der Eindeutigkeitsbedingung die Regel (2) und die Regel (9) angewendet.

$$\Theta_{\phi_3}(E_1) \overset{\text{Regel } 9}{\longrightarrow} \Theta_{\phi_3}(E_1' \cup E_2') \overset{\text{Regel } 2}{\longrightarrow} \Theta_{\phi_3}(E_1')$$
$$\Theta_{\phi_3}(E_1) \overset{\text{Regel } 9}{\longrightarrow} \Theta_{\phi_3}(E_2' \cup E_1') \overset{\text{Regel } 2}{\longrightarrow} \Theta_{\phi_3}(E_2')$$
$$\Theta_{\phi_3}(E_1) \overset{\text{Regel } 9}{\longrightarrow} \Theta_{\phi_3}(E_1' \cup E_2')$$

Im Gegensatz zu der Eindeutigkeitsbedingung können Bedingungen wie max(x)>y und min(x)<y nicht in abgeschwächter Form für die einzelnen entstehenden Basisextensionen hergeleitet werden. Daher kann hier nur Regel (9) genutzt werden:

$$\Theta_{\phi_4}(E_2) \overset{\text{Regel } 9}{\longrightarrow} \Theta_{\phi_4}(E_2' \cup E_3')$$

**Vereinigung von Extensionen.** Im Gegensatz zu den Zerlegungsregeln ist es nicht sinnvoll, die Vereinigungsregel auf einzelne Integritätsbedingungen anzuwenden. Um eine Abschwächung der Integritätsbedingung zu vermeiden, müßten nämlich alle Kombination von Integritätsbedingungen gebildet werden und die Regel (4) dementsprechend oft ausgeführt werden, um dann als letzte Aktion die entstehenden Integritätsregeln durch die Anwendung der Regel (7) zusammenzufassen. Statt dessen empfehlen wir vor der Vereinigung die Anwendung der Regel (7), die die einzelnen Integritätsbedingungen einer Basisextension konjunktiv verknüpft.

In unserem Beispiel wollen wir eine Extension konstruieren, die alle Mitarbeiter umfaßt, die jeweils nur einem der beiden Projekte zugeordnet sind. Wie aus der Abbildung 5 ersichtlich, wird diese Extension durch die Vereinigung der Basisextensionen $E_1'$ und $E_3'$ erzeugt: $E'' = E_1' \cup E_3'$.

Die konjunktive Verknüpfung durch Regel (7) wird in unserem Beispiel auf den Basisextensionen $E_1'$ und $E_2'$ durchgeführt, bevor Regel (4) angewendet wird. Auf $E_3'$ ist nur eine Bedingung definiert, so daß die konjunktive Verknüpfung nicht angewendet werden muß.

$$\Theta_{\phi_1}(E_1'), \Theta_{\phi_3}(E_1') \overset{\text{Regel } 7}{\longrightarrow} \Theta_{\phi_1 \wedge \phi_3}(E_1')$$
$$\Theta_{\phi_1}(E_2'), \Theta_{\phi_2}(E_2'), \Theta_{\phi_3}(E_2') \overset{\text{Regel } 7}{\longrightarrow} \Theta_{\phi_1 \wedge \phi_2 \wedge \phi_3}(E_2') = \Theta_{\phi_1 \wedge \phi_3}(E_2')$$

Da die Bedingung $\phi_2$ schwächer als die Bedingung $\phi_1$ ist, fällt sie bei der konjunktiven Verknüpfung weg. Nun wird zur Ableitung der Integritätsbedingungen für die Extension $E''$ die Regel (4) angewendet.

$$\Theta_{\phi_1 \wedge \phi_3}(E_1'), \Theta_{\phi_2}(E_3') \overset{\text{Regel } 4}{\longrightarrow} \Theta_{(\phi_1 \wedge \phi_3) \vee \phi_2}(E_1' \cup E_3') = \Theta_{\phi_2}(E'')$$

In diesem Fall konnte die Bedingung $(\phi_1 \wedge \phi_3) \vee \phi_2$ vereinfacht werden, da $\phi_1 \Rightarrow \phi_2$ gilt. Auf der Extension $E''$ ist nun weder eine Eindeutigkeitsbedingung noch eine Maximumbedingung definiert. Es kann ein

Objekt in $E''$ eingefügt werden, daß weder in $E_1$ noch in $E_2$ eingefügt werden kann, da es zum Beispiel sowohl $\phi_3$ als auch $\phi_4$ verletzt. Die Bedingungen $\phi_3$ und $\phi_4$ können jedoch nicht auf der Extension $E''$ überprüft werden, da sie das Vorhandensein weiterer Extensionen voraussetzen. Das bedeutet konkret, daß ein Scheitern einer Einfügeoperation eines Objektes in $E''$ nicht mit dem Verletzen einer auf dem abgeleiteten Schema sichtbaren Bedingung erklärt werden kann. Dieses Problem ist eng verwandt mit dem *View-Update*-Problem bei selektiven Sichten.

Interessant an dieser Vereinigung ist weiterhin die Tatsache, daß durch die logische Disjunktion sogar die abgeschwächte Eindeutigkeitsregel verschwindet. Dies kann verhindert werden, wenn bei der Erzeugung der Basisextensionen ein künstliches Diskriminantenattribut eingefügt wird. In unserem Beispiel würde dieses Attribut (zum Beispiel gruppe) die Information beinhalten, ob ein Mitarbeiter der Projektgruppe 1 oder 2 oder beiden Gruppen zugeordnet ist. Diese Information kann nun innerhalb der Integritätsbedingungen der Basisextensionen in der Form ausgenutzt werden, daß die ursprünglichen Bedingungen als rechter Teil einer Implikation auftreten und im linken Teil jeweils der Diskriminantentest durchgeführt wird.

$$\begin{aligned} \Theta_{\phi'_1 \wedge \phi'_3}(E'_1) \quad mit \quad & \phi'_1 \equiv \forall e : e.\text{gruppe} = 1 \Longrightarrow e.\text{qualifikation} > 4 \\ & \phi'_3 \equiv \forall e_1, e_2 : (e_1.\text{gruppe} = 1 \wedge e_2.\text{gruppe} = 1) \Longrightarrow (e_1.\text{pid} = e_2.\text{pid} \Longrightarrow e_1 = e_2) \\ \Theta_{\phi'_2}(E'_3) \quad mit \quad & \phi'_2 \equiv \forall e : e.\text{gruppe} = 2 \Longrightarrow e.\text{qualifikation} > 3 \end{aligned}$$

Solche umgewandelten Integritätsbedingungen können nun entgegen der Regel (4) bei der Vereinigung konjunktiv verknüpft werden.

$$\Theta_{\phi'_1 \wedge \phi'_3}(E'_1), \Theta_{\phi'_2}(E'_3) \longrightarrow \Theta_{\phi'_1 \wedge \phi'_2 \wedge \phi'_3}(E'_1 \cup E'_3) = \Theta_{\phi'_1 \wedge \phi'_2 \wedge \phi'_3}(E'')$$

Durch diese Umwandlung kann die Abschwächung von Integritätsbedingungen bei der Integration (oder allgemeiner: bei der Restrukturierung) reduziert werden.

Grundsätzlich ist bei der Ableitung der externen Integritätsbedingungen darauf zu achten, daß diese durch Anwenden der Vereinigungsregeln auf den Basisextensionen abgeleitet werden müssen, da bei einer Ableitung auf der Basis anderer schon abgeleiteter Klassenextensionen Integritätsbedingungen verloren gehen können. Damit man in solchen Fällen weiß, auf welche Basisextensionen zurückgegriffen werden muß, ist es sinnvoll, einer jeden abgeleiteten Klassenextension deren Ableitungsvorschrift auf der Grundlage der Basisextensionen zuzuordnen.

# 5 Zusammenfassung

Wir haben in diesem Beitrag einen Ansatz skizziert, der die umfassende Berücksichtigung von Integritätsbedingungen bei der Schemaintegration ermöglichen soll. Durch die Aufstellung elementarer Regeln für die Behandlung von Integritätsbedingungen steht ein Instrumentarium zur Verfügung, das in alle Schemaintegrationsansätze eingebunden werden kann. Die vorgestellten Grundprinzipien müssen allerdings noch weiter untersucht werden. Insbesondere die Anwendbarkeit der Regeln auf weitere relevante Arten von Integritätsbedingungen sowie das Finden ergänzender Regeln für die Fälle, in denen die vorgestellten Regeln nicht angewendet werden können, sind Fragestellungen für die Fortsetzung der hier begonnenen Arbeit.

Aus unserem Ansatz lassen sich allgemeine Gesetzmäßigkeiten ableiten, die nicht nur für die Schemaintegration gelten, sondern sich auch auf Sichtenbildung oder andere Schemaveränderungen wie Transformation, Restrukturierung und teilweise auch Schemaevolution übertragen lassen. Wir gehen daher davon aus, daß der hier vorgestellte Ansatz auch in diesen Bereichen eingesetzt werden kann. Vielleicht lassen sich durch die allgemeinere Betrachtung von Schemaänderungen in den verschiedenen Bereichen neue Anregungen und zusätzliche Gesetzmäßigkeiten entdecken, die zur Weiterentwicklung und Vervollständigung unseres Ansatzes beitragen können.

Die umfassende Behandlung von Integritätsbedingungen bei der Schemaintegration ist auch eine notwendige Voraussetzung für die Einführung einer globalen Integritätssicherung in föderierten Datenbanksystemen [CT97]. Darüber hinaus kann die weitgehende Berücksichtigung von Integritätsbedingungen auch einen erheblichen Beitrag in Anwendungsgebieten leisten, bei denen Sichtenmaterialisierung [SJ96, KR95] eine wichtige Rolle spielt, so z.B. in mobilen Datenbanken [WSD+95].

Derzeit entwickeln wir einen Prototyp eines Entwurfs- und Integrationswerkzeuges, der die Berücksichtigung von Integritätsbedingungen bei Schematransformation und -integration sowie Sichtenbildung unterstützen soll. Hierzu werden wir auch die in der Diskussion des Beispiels angesprochene Methodik der Regelanwendung weiter verfeinern, um optimale Ergebnisse möglichst effizient zu erhalten. Für die ebenfalls schon angesprochene Möglichkeit, Diskriminanten einzuführen, um in bestimmten Fällen eine Abschwächung der Integritätsbedingungen zu vermeiden, arbeiten wir gegenwärtig daran, die Einführung von Diskriminanten in Integritätsbedingungen formal korrekt und vollständig zu definieren.

# Literatur


[CHJ+96] S. Conrad, M. Höding, S. Janssen, G. Saake, I. Schmitt und C. Türker. Integrity Constraints in Federated Database Design. Preprint Nr. 2, Fakultät für Informatik, Universität Magdeburg, April 1996.

[CT97] S. Conrad und C. Türker. Unternehmensweite Datenkonsistenz durch Föderierung bestehender Informationssysteme. In *Proc. Wirtschaftsinformatik (WI'97)*, 1997. *To appear.*

[KR95] H.A. Kuno und E.A. Rundensteiner. Materialized Object-Oriented Views in MultiView. In O. Bukhres, T. Özsu und M.-C. Shan, Herausgeber, *Proc. of the 5th Int. Workshop on Research Issues in Data Engineering: Distributed Object Management (RIDE-DOM'95)*, S. 78–85. IEEE Computer Society Press, 1995.

[RPG95] M. P. Reddy, B. E. Prasad und A. Gupta. Formulating Global Integrity Constraints during Derivation of Global Schema. *Data & Knowledge Engineering*, 16(3):241–268, Juli 1995.

[RPRG94] M. P. Reddy, B. E. Prasad, P. G. Reddy und A. Gupta. A Methodology for Integration of Heterogeneous Databases. *IEEE Transactions on Knowledge and Data Engineering*, 6(6):920–933, Dezember 1994.

[SJ96] M. Staudt und M. Jarke. Incremental Maintenance of Externally Materialized Views. In T.M. Vijayaraman, A.P. Buchmann, C. Mohan und N.L. Sarda, Herausgeber, *Proc. of the 22nd Int. Conf. on Very Large Data Bases (VLDB'96)*, S. 75–86. Morgan Kaufmann Publishers, 1996.

[SPD92] S. Spaccapietra, C. Parent und Y. Dupont. Model Independent Assertions for Integration of Heterogeneous Schemas. *The VLDB Journal*, 1(1):81–126, Juli 1992.

[SS96a] I. Schmitt und G. Saake. Integration of Inheritance Trees as Part of View Generation for Database Federations. In B. Thalheim, Herausgeber, *Conceptual Modelling — ER'96, Proc. of the 15th Int. Conf., Cottbus, Germany, October 1996*, S. 195–210. LNCS 1157, Springer-Verlag, Berlin, 1996.

[SS96b] I. Schmitt und G. Saake. Schema Integration and View Generation by Resolving Intensional and Extensional Overlappings. In K. Yetongnon und S. Hariri, Herausgeber, *Proc. of the 9th ISCA Int. Conf. on Parallel and Distributed Computing Systems (PDCS'96), Dijon, France*, S. 751–758, September 1996.

[VA96] M. W. W. Vermeer und P. M. G. Apers. The Role of Integrity Constraints in Database Interoperation. In T. M. Vijayaraman, A. P. Buchmann, C. Mohan und N. L. Sarda, Herausgeber, *Proc. of the 22nd Int. Conf. on Very Large Data Bases (VLDB'96), Bombay, India*, S. 425–435. Morgan Kaufmann Publishers, September 1996.

[WSD+95] O. Wolfson, P. Sistla, S. Dao, K. Narayanan und R. Raj. View Maintenance in Mobile Computing. *ACM SIGMOD Record*, 24(4):22–27, 1995.