

Null-Werte in relationalen Datenbanken

Bearbeiter: Thomas Bierhance
Betreuer: Philipp Bender

Einleitung

Null-Werte sind sowohl im SQL-Standard als auch in einer Vielzahl höherer Programmiersprachen zu finden und stellen zweifellos ein wichtiges Instrument zur Behandlung von Imperfektion dar. Sie finden grundsätzlich Anwendung bei zwei Typen von Imperfektion:

1. *Wert existent aber unbekannt*

Werte können vorübergehend oder dauerhaft unbekannt sein, obwohl man sich ihrer Existenz sicher ist. Die Imperfektion liegt im Wissensstand über den Wert.

Ein Beispiel für einen existenten aber unbekanntem Wert ist das fehlende Geburtsdatum eines Kunden. Ein Kunde hat sicher ein Geburtsdatum, der tatsächliche Wert kann aber dem System, seinen Benutzern oder sogar aber allgemein unbekannt sein.

2. *Eigenschaft nicht anwendbar*

Werte können in einem gegebenen Kontext nicht existent sein. Die Imperfektion liegt in der verwendeten Datenstruktur und nicht im Wissensstand (denn es ist bekannt, dass der Wert nicht existiert).

Ein Beispiel für eine nicht anwendbare Eigenschaft ist das Datum der Führerscheinprüfung eines Angestellten. Es ist unsinnig nach dem Datum der Führerscheinprüfung eines Angestellten zu fragen, der keinen Führerschein hat. In diesem Fall ist die Datenstruktur deshalb imperfekt, weil sie die Frage überhaupt zulässt.

Eng mit dieser Art der Null-Werte sind offenbar Fragen der Normalisierung verbunden.

Gegenstand der Forschung war und ist, die Besonderheiten bei der Verarbeitung von Daten mit Null-Werten zu klären.

Das vorherrschende Modell für die Speicherung von grossen Datenmengen ist das relationale Modell. Es ist deshalb von besonderem Interesse wie die Verarbeitung von Null-

Werten in der relationalen Algebra gelingen kann, bzw. wie dies auch auf die weitverbreitete Abfragesprache SQL übertragen werden kann.

Im folgenden werden eine paar grundlegende Problem bei der Verarbeitung von Null-Werten mit der relationalen Algebra sowie zwei mögliche Lösungsansätze vorgestellt.

Probleme und Motivierung

Exemplarisch seien einige Fragen vorgestellt, die sich bei der Auswertung von relationalen Ausdrücken ergeben, falls die beteiligten Relationen Null-Werte enthalten.

Es dienen zwei Relationen als Basis dieser Überlegungen:

| <i>R</i> | <i>A</i> | <i>B</i> | <i>S</i> | <i>A</i> | <i>C</i> |
|----------|----------|----------|----------|----------|----------|
| | 1 | b | | 2 | c |
| | 2 | Null | | 1 | Null |
| | Null | b' | | Null | Null |

- ***vergleichende Operatoren und Null-Werte***

Wird auf die Relation R eine Abfrage der Form $\sigma_{A>1}(R)$ gestellt, ist in der Ergebnismenge sicherlich das Tupel $\{1, b\}$ nicht ($1>1=F$) enthalten, das Tupel $\{2, \text{Null}\}$ dagegen sicher ja ($2>1=T$). Fraglich ist, ob das Tupel $\{\text{Null}, b'\}$ in der Ergebnismenge enthalten sein soll, da der Wert für das Attribut A unbekannt ist, also echt grösser 1 oder auch nicht sein könnte. Die Eigenschaft könnte sogar überhaupt nicht anwendbar sein.

- ***Verknüpfung von boolschen Ausdrücken mit Null-Werten***

Enthält ein boolscher Ausdruck Null-Werten kann der Wahrheitswert bestimmter Teile des Ausdrucks unbestimmt sein, anderer dagegen klar bestimmbar. Fraglich ist dann, wie diese Teile verknüpft werden können. Wird beispielsweise die obige Selektion nach $\sigma_{A>1 \wedge B=b}(R)$ erweitert, ist das Tupel $\{1, b\}$ sicherlich nicht enthalten. Für die anderen zwei Tupel bleibt es aber ungewiss, ob sie in der Ergebnismenge enthalten sein sollen, da jeweils ein Teil des Selektionskriteriums unentschieden bleibt ($\text{Null}=b'$ bzw. $\text{Null}>1$).

Tritt an die Stelle der und-Verknüpfung eine oder-Verknüpfung $\sigma_{A>1 \vee B=b}(R)$ stellt sich ein anderes Problem. In diesem Fall ist es für das Selektionskriterium intuitiv nicht entscheidend, wie die Vergleiche $\text{Null}=b'$ bzw. $\text{Null}>1$ ausgehen, da die Vergleiche $2>1$ und $b'=b'$ schon jeweils den Wahrheitswert T ergeben.

- ***Duplikate bei Projektion***

Da Relationen Mengen sind, und damit keine Duplikate enthalten dürfen, stellt sich die Frage wie mit Null-Werten bei der Projektion umgegangen wird. Der relationale Ausdruck $\pi_c(S)$ wird ohne Duplikatsentfernung die Ergebnisrelation $\{\{c\}, \{Null\}, \{Null\}\}$ ergeben. Wird von dem zweifach enthaltenen Tupel $\{Null\}$ eins entfernt, enthält das Ergebnis zwar keine Duplikate mehr, es wird aber die Information unterdrückt, dass die Null-Werte für unterschiedliche Werte stehen könnten (mehr dazu in den Ausführungen zu Codd's Arbeit).

- ***Null-Werte in Verbänden***

Werden die Relationen R und S miteinander verbunden ($R \bowtie S$) stellt sich die Frage, wie der Vergleich $Null=Null$ ausgeht. Die Null-Werte könnten für den gleichen Wert stehen, in vielen Fällen werden sie wohl aber für unterschiedliche Werte stehen.

Erweiterung von Codd

Codd schlug 1972 eine Erweiterung der relationalen Algebra zur Behandlung von Null-Werten vor. In seinem Vorschlag soll der Null-Wert für einen zur Zeit unbekanntem Wert stehen ("value at present unknown"); der Fall, dass ein Null-Wert für eine nicht anwendbare Eigenschaft ("property inapplicable") steht, wird nicht berücksichtigt.

Codd bezeichnet den Null-Wert mit ω .

Dreiwertige Logik

Der Kern der Erweiterung von Codd ist eine dreiwertige Logik, die neben den Werten "wahr" und "falsch" der klassischen Logik auch einen "unbekannten" Wahrheitswert kennt. Auch der unbekanntem Wahrheitswert wird mit ω bezeichnet.

Für die logischen Operatoren "und" bzw. "oder" ergeben sich folgende intuitive

Wahrheitstabellen:

Da ω für einen unbekanntem Wert steht, muss jeder Ausdruck der von den tatsächlichen Werten dieser

| \wedge | F | ω | T | \vee | F | ω | T |
|----------|----------|----------|----------|----------|----------|----------|----------|
| F | F | F | F | F | F | ω | T |
| ω | F | ω | ω | ω | ω | ω | T |
| T | F | ω | T | T | T | T | T |

Unbekanntem abhängt auch mit ω

bewertet werden. So wird z.B. der Ausdruck $T \wedge \omega$ mit ω bewertet, da je nachdem ob ω für den Wahrheitswert T oder F steht, eine andere Bewertung anzuführen wäre. Dagegen wird der Ausdruck $F \wedge \omega$ sicher mit F bewertet, da es für die Bewertung unerheblich ist, für welchen tatsächlichen Wahrheitswert ω steht.

Prinzip der Nullsubstitution

Eine allgemeine Regel für die Bewertung von booleschen Ausdrücken nach Codd's dreiwertiger Logik klingt in der obigen Beschreibung an und wurde von Codd als Prinzip der Nullsubstitution präzisiert.

Ein boolescher Ausdruck A wird genau dann mit dem Wahrheitswert ω bewertet, wenn...

- (1) jedes ω in A mit Werten ungleich ω so ersetzt werden kann, dass $A=T$ und gleichzeitig
- (2) jedes ω in A mit Werten ungleich ω so ersetzt werden kann, dass $A=F$

Aus diesem Prinzip lassen sich sowohl die oben aufgeführten Wahrheitstabellen, als auch

der Wahrheitswert für jeden anderen Ausdruck bestimmen. Es sei angemerkt, dass sich aus der Negation des Prinzips unmittelbar auch folgende Regel ergibt:

Ein boolescher Ausdruck A wird genau dann mit dem Wahrheitswert T (bzw. F) bewertet, wenn für jede mögliche Ersetzung von ω in A mit Werten ungleich ω gilt, dass $A=T$ (bzw. $A=F$)

Duplikatsentfernung und Konsequenzen

Da jede Relation eine Menge ist, gilt auch für Relationen, dass sie kein Element (Tupel) doppelt enthalten dürfen.

Die Auswertung von relationalen Ausdrücken kann in diesem Sinne in zwei Schritte aufgeteilt werden: Zum einen die Bewertung von Konditionen zur Datenextraktion aus den zu Grunde liegenden Relationen, zum anderen die Duplikatsentfernung aus den resultierenden "Roh-Ergebnissen", um sicherzustellen, dass jedes Ergebnis wieder eine Relation ist, die jedes Tupel nur einmal enthält.

Nun muss beachtet werden, dass bei der Entfernung von Duplikaten die Anwendung von Codd's dreiwertiger Logik zweifelhaft ist. Zum einen können zwei Vorkommnisse von ω für echt unterschiedliche Werte, zum anderen könnten sie auch für den selben unbekanntem Wert stehen. Im ersten Fall könnte die Entfernung eines ω aus der Ergebnismenge nicht wünschenswert sein, im anderen Fall dagegen ja.

Dieser Konflikt kann nicht befriedigend aufgelöst werden, da auch eine erweiterte Semantik allenfalls zwischen unbekanntem, aber bestimmt gleichen bzw. unbekanntem, aber bestimmt ungleichen Werten zu unterscheiden vermag (die später vorgestellte Erweiterung entspricht diesem Gedanken teilweise). Es kann aber eben auch unbekannt sein, ob zwei ω für den gleichen Wert stehen oder nicht.

Codd entschied für eine Lösung, die bei der Duplikatsentfernung von seiner dreiwertigen Logik abweicht, und ω bei der Identifikation von gleichen Tupeln während der Duplikatsentfernung als normalen Domänenwert behandelt.

Konsequenz dieses Vorgehens ist, dass ω in jeder Relation (zumindest in jedem Ergebnis eines relationalen Ausdrucks) für mehr als einen unbekanntem Wert stehen kann.

Änderungen an relationalen Operatoren

Für die Mehrheit der relationalen Operatoren müssen keine allgemeinen booleschen Ausdrücke ausgewertet werden, sondern allenfalls Tupel miteinander verglichen werden.

Dies ist der Fall für Vereinigung, Schnittmenge, Differenz, kartesischem Produkt und Projektion. Codd schlägt vor, dass hier bei der Identifikation von gleichen Tupeln auf die dreiwertige Logik verzichtet wird (wie oben beschrieben, Schnittmenge und Differenz benötigen aber nicht nur für die Duplikatsentfernung das Konzept der Tupelgleichheit).

Bei den relationalen Operatoren Selektion und Theta-Verbund sind dagegen arbiträre boolesche Ausdrücke zu bewerten, und es stellt sich somit die Frage, wie hier die dreiwertige Logik anzuwenden ist.

Codd entschied sich dafür zwei verschiedene Varianten dieser Operatoren einzuführen.

Für beide Operatoren existiert eine "wahre" und eine "mögliche" Variante, namentlich Wahre-Selektion/Wahrer-Theta-Verbund (True-Selection/True-Theta-Join) und Mögliche-Selektion/Möglicher-Theta-Verbund (Maybe-Selection/Maybe-Theta-Join).

Die wahre Variante enthält alle Tupel, für die das Selektionskriterium sicher wahr ist. In dem einleitend genannten Selektionsbeispiel wäre dies $\sigma_{A>1}(R) = \{\{2, \text{Null}\}\}$. Die mögliche Variante enthält dagegen alle Tupel, die das Selektionskriterium nicht sicher, sondern nur vielleicht erfüllen. Ein möglicher Vergleich wird mit ω gekennzeichnet, so dass sich für unser Beispiel $\sigma_{A>\omega 1}(R) = \{\{\text{Null}, b'\}\}$ ergibt.

Kritik

Duplikatsentfernung

Die ausführlicher erläuterten Probleme der Regel zur Duplikatsentfernung sind teilweise Quelle von Kritik an Codds Vorschlag. Die zweischrittige Aufteilung der Auswertung von relationalen Ausdrücken in Datenextraktion und Duplikatsentfernung und die damit einhergehende Anwendung der dreiwertigen Logik einerseits, der zweiwertigen Logik andererseits ist nicht intuitiv und wird von Codd auch nicht überzeugend begründet. Hier ist z.B. anzuführen, dass die Schnittmenge ein Spezialfall des natürlichen Verbundes ist. Nichtsdestotrotz führen beide Operatoren in Codds Vorschlag zu unterschiedlichen Ergebnissen. Das ist eine nicht zu übersehende Inkonsistenz.

Tautologien

Nicht wenige in der relationalen Algebra ohne Null-Werte bestehende Tautologien haben in Codds dreiwertiger Erweiterung keinen Bestand mehr. So wird z.B. aus

$\sigma_{A=a}(A) \cup \sigma_{A \neq a}(A) = A$ in Codds Modell $\sigma_{A=a}(A) \cup \sigma_{A \neq a}(A) = A - \sigma_{A=\omega}(A)$. Tupel, deren Wert für das Attribut A unbekannt ist, werden bei den Selektionen ausgeschlossen und sind

somit nicht im Endergebnis enthalten . Codd meint entgegen seinen Kritikern, dass es nicht unbedingt wünschenswert ist diese Tautologien beizubehalten.

Fehlende Bewertungsmöglichkeit

Grundlegend ist zu kritisieren, dass Codd keine Kriterien angab und zu dieser Zeit wohl auch keine Kriterien bekannt waren, die Erweiterungen der relationalen Algebra denn sinnvollerweise erfüllen sollten. Es war deshalb schwierig seine Erweiterung zu bewerten und auch nur schwer nachzuvollziehen warum bestimmte Entscheidungen so und nicht anders getroffen wurden.

Erweiterung von Imieliński und Lipski

Imieliński und Lipski stellten 1984 einen Ansatz vor, der statt einem einzigen Symbol für unbekannte Werte (wie das ω bei Codd) beliebige Variablen zulässt. Dadurch wird es möglich, den Umstand auszudrücken, dass zwei Symbole für den selben unbekanntem Wert stehen. Ein Beispiel zur Verdeutlichung:

| <i>Vorlesung</i> | <i>Dozent</i> | <i>Wochentag</i> |
|------------------|---------------|------------------|
| Datenbanken | x | Montag |
| Programmieren | y | Dienstag |
| Datenbanken | x | Donnerstag |
| Fortran | Schmidt | z |

Die Relation drückt in diesem Fall aus, dass die Vorlesung “Datenbanken” sowohl am Montag als auch am Donnerstag vom selben, unbekanntem Dozenten gehalten wird. Wer die Vorlesung “Programmieren” hält ist auch nicht bekannt. Es könnte der selbe Dozent sein, der auch die Vorlesung “Datenbanken” hält, muss es aber nicht zwangsweise sein. Über die Vorlesung “Fortran” ist bekannt, dass sie von Dozentin “Schmidt” gehalten wird, jedoch nicht an welchem Wochentag.

Relationale Operatoren und 2-wertige Logik

Die relationalen Operatoren behandeln die Variablen wie gewöhnliche Domänenwerte.

Erklärungsbedürftig ist, wie Variablen in booleschen Ausdrücken behandelt werden. Auch hier lässt sich ein Substitutions-Prinzip angeben:

Ein boolescher Ausdruck A wird genau dann mit dem Wahrheitswert T bewertet, wenn für jede Ersetzung der Variablen in A mit Domänen-Werten gilt, dass $A=T$.

Dementsprechend gilt dann auch die Negation:

Ein boolescher Ausdruck A wird genau dann mit dem Wahrheitswert F bewertet, wenn eine Ersetzung der Variablen in A mit Domänen-Werten existiert, so dass $A=F$.

Beispielsweise wird normalerweise gelten, dass $(x=y)=F$, $(x=x)=T$ und $(a=x)=F$.

Repräsentations Systeme

Ein Kritikpunkt an Codd's vorgestellter Arbeit war die fehlende Bewertungsmöglichkeit einer Erweiterung der relationalen Algebra. Imieliński und Lipski haben hier einen Vorschlag gemacht. Die grundlegende Idee ist, dass eine Erweiterung der relationalen

Algebra dann sinnvoll ist, wenn eine Abbildung zwischen der erweiterten und der regulären Algebra besteht, die bezüglich einer Menge von relationalen Operatoren bestimmten Anforderungen genügt. Diese Abbildung wird Repräsentationssystem genannt.

Einleitende Definitionen

Als (reguläre) Relation R wird eine Menge von Tupeln t bezeichnet, die auf einer bestimmten Attributmengemenge X definiert ist, die wiederum Untermenge der Menge aller Attribute A ist:

$$\alpha(t) = X \subseteq A$$

Eine Multi-Relation \mathbf{R} bezeichnet eine Menge von Relationen, steht also für einen konkreten Datenbankzustand zu einem gegebenen Schema. Eine Menge von homogenen Multi-Relationen wird mit \mathcal{R} bezeichnet. Homogen bedeutet hier, dass $\alpha(\mathbf{R}) = \alpha(\mathbf{S})$ für alle \mathbf{R}, \mathbf{S} aus \mathcal{R} . Somit ist \mathcal{R} eine Menge von Datenbankzuständen zu einem gegebenem Schema. Die Menge aller homogenen Mengen von Multi-Relationen wird mit \mathcal{P} bezeichnet, sie enthält damit alle vorstellbaren Datenbanken (alle Zustände zu jedem möglichen Schema).

Relationen in der erweiterten Algebra werden Tabellen genannt. Diese Tabellen enthalten im Gegensatz zu den regulären Relationen Null-Werte. In welcher Weise diese Null-Werte dargestellt werden, ob als Variablen oder mit einem einzigen Bezeichner wie ω , ist nicht relevant.

Eine Menge von Tabellen \mathbf{T} wird Multi-Tabelle genannt. Eine homogene Menge von Multi-Tabellen wird mit \mathcal{T} bezeichnet. \mathcal{T} entspricht damit einer Menge von Datenbankzuständen mit Null-Werten zu einem gegebenen Schema.

Eine Menge relationaler Operatoren wird mit Ω bezeichnet.

Definition eines Repräsentationssystems

Basis für eine Repräsentationssystem ist ein Tripel $(\mathcal{T}, Rep, \Omega)$. \mathcal{T} ist eine homogene Menge von Multi-Tabellen, Ω eine Menge relationaler Operatoren und Rep eine Abbildung $Rep: \mathcal{T} \rightarrow \mathcal{P}$.

Die Abbildung Rep entspricht damit einer Abbildung zwischen einem konkreten Datenbankzustand mit Null-Werten und einer Menge von möglichen Datenbankzuständen ohne Nullwerte, die sozusagen ersteren repräsentieren und die möglichen Konkretisierungen der unbekanntenen Werte aufzeigen.

Um von einem Repräsentationssystem sprechen zu können muss die Abbildung einer bestimmten Bedingung genügen. Wünschenswert wäre, wenn für jeden relationalen Ausdruck f mit Operatoren aus Ω gelten würde, dass die Komposition aus f und Rep kommutativ wäre: $f(Rep(\mathbf{T}))=Rep(f(\mathbf{T})) \forall \mathbf{T} \in \mathcal{T}, f \text{ aus } \Omega$

Bildlich gesprochen also das Ergebnis einer Abfrage auf die Repräsentation einer imperfekten Datenbank mit Null-Werten identisch mit der Repräsentation des Ergebnisses einer Abfrage auf diese Datenbank ist.

Imieliński und Lipski sind allerdings der Meinung, dass diese Bedingung normalerweise nicht erfüllt werden kann und suchten deshalb nach einer erfüllbaren Relaxierung.

Anstelle der Gleichheit tritt jetzt die sogenannte Ω -Äquivalenz. Zwei Mengen homogener Multi-Relationen \mathcal{X} und \mathcal{Y} sollen dann Ω -Äquivalent sein, wenn für jeden Ausdruck g aus Ω der Schnitt von $g(\mathcal{X})$ gleich dem Schnitt von $g(\mathcal{Y})$ (da \mathcal{X} eine Menge von Multi-Relationen ist, ist auch $g(\mathcal{X})$ eine Menge von Multi-Relationen und ein Schnitt kann gebildet werden):

$$\mathcal{X} \equiv_{\Omega} \mathcal{Y} \Leftrightarrow \bigcap g(\mathcal{X}) = \bigcap g(\mathcal{Y})$$

Bildlich gesprochen entsprechen \mathcal{X} und \mathcal{Y} Mengen von möglichen Datenbankzuständen. Für jede denkbare Abfrage g werden die möglichen Ergebnisse gebildet. Der Schnitt hieraus entspricht den Tupeln, die unabhängig davon, welcher der möglichen Datenbankzustände der "wahre" ist, auf jeden Fall im Ergebnis enthalten sind.

Von einem Repräsentationssystem wird dann gesprochen, wenn gilt:

$$f(Rep(\mathbf{T})) \equiv_{\Omega} Rep(f(\mathbf{T})) \forall \mathbf{T} \in \mathcal{T}, f \text{ aus } \Omega$$

Existiert ein solches Repräsentationssystem für eine bestimmte Erweiterung, kann diese relationale Ausdrücke aus Ω im Sinne dieses Kriteriums korrekt auswerten.

Untersuchungen und Ergebnisse

Imieliński und Lipski untersuchten sowohl Codds Vorschlag als auch ihren eigenen Vorschlag zur Behandlung von Null-Werten daraufhin, für welche Mengen von relationalen Operatoren die jeweilige Erweiterung korrekt sind (also Repräsentationssystem existieren).

Für Codds Erweiterung fanden sie ein Repräsentationssystem für $\Omega = \{\pi, \sigma\}$. Das heisst, das Codds Vorschlag beliebige relationale Ausdrücke, die nur Projektion und Selektion enthalten, in oben definiertem Sinne korrekt ausgewertet.

Dagegen zeigten sie für $\Omega = \{\pi, \sigma, \cup\}$ und $\Omega = \{\pi, |><|\}$ ein solches Repräsentationssystem nicht existiert. Codd's Erweiterung wertet dementsprechend relationale Ausdrücke, die die Operatoren Projektion und Verbund, bzw. Projektion, Selektion und Vereinigung enthalten, nicht korrekt aus.

Ihre eigene Erweiterung mit Variablen statt einfacher Null-Werte kann für $\Omega = \{\pi, \sigma\}$ im Gegensatz zu Codd keine korrekte Auswertung garantieren. Wird die Selektion eingeschränkt auf Selektionen, die im Selektionskriterium keine Negation enthalten (der Operator \neg ist nicht im Kriterium enthalten), und wird eine solche Selektion mit σ^+ bezeichnet, existiert für $\Omega = \{\pi, \sigma^+, \cup, |><|\}$ ein Repräsentationssystem.

Fazit

Durch die Definition einer Eigenschaft, die eine Erweiterung der relationalen Algebra erfüllen sollte, haben Imieliński und Lipski sicher ein Kriterium geschaffen, an dem sich andere Vorschläge messen können.

Ob eine ideale Erweiterung der relationalen Algebra zur Behandlung von Null-Werten jedoch gefunden werden kann, ist fraglich.

Die vorgestellte Erweiterung von Imieliński und Lipski mag in bestimmten Grenzen ausdrucksstärker als der ursprüngliche Vorschlag von Codd sein, bringt durch die Variablen aber auch eine nicht zu unterschätzende Komplexität in die Algebra ein.

Es ist eine weitergehende Erweiterung dahin vorstellbar, dass nicht nur die Gleichheit sondern auch die Ungleichheit zwischen zwei Variablen, bzw. auch noch die Unanwendbarkeit einer Eigenschaft zum Ausdruck gebracht wird. Eine solche Erweiterung würde allerdings sicherlich in der Praxis nur schwer Verbreitung finden. Es stellt sich nämlich die Frage, ob das Wissen über die Gleichheit bzw. Ungleichheit zwischen zwei unbekanntem Werten mit einer Häufigkeit in der Praxis auftritt, die diesen Aufwand rechtfertigen würde.

Nichtsdestotrotz kann festgestellt werden, dass eine Vielzahl von Abfragen, die heute mit SQL gestellt werden und sich damit weitestgehend dem Vorschlag von Codd entsprechend verhalten, im Sinne von Imieliński und Lipski nicht korrekt beantwortet werden und damit durchaus eine weitergehende Untersuchung der Problematik gerechtfertigt scheint.

Literatur

Codd, E.F. "Extending the database relational model to capture more meaning", ACM Transactions on Database Systems, Vol 4, No 4, Dec 1979, pp. 379-434

Imielinski, Lipski "Incomplete Information in Relational Databases", ACM Transactions on Database Systems, Vol 31, No 4, Oct 1984, pp. 761-791

Biskup, J. "A Foundation of Codd's Relational Maybe-Operations", ACM Transactions on Database Systems, Vol 8, No 4, Dec 1983, pp. 608-636

Grant, J. "Null values in a relational data base" Inform. Process. Lett. 5, 1977, pp. 156-157