Entity-Relationship-Modell-Clustering

Peter Jaeschke

PROMATIS Informatik
Descostraße 10
D-76307 Karlsbad

Zusammenfassung
Die Entity-Relationship-Modellierung hat sich als Methode zur konzeptuellen Datenmodellierung etabliert. Trotz der verfügbaren Tool-Unterstützung treten verschiedene Probleme auf, sobald sehr große Datenschemata erstellt werden. Es zeigt sich immer wieder die Notwendigkeit, sowohl einzelne Teilbereiche im Detail als auch den Zusammenhang dieser Teilbereiche mit dem Gesamtdatenschema darzustellen. Um diesen Überblick zu ermöglichen, kann die Technik des Entity-Clustering eingesetzt werden. Teildiagramme lassen sich in sogenannten Entity-Clusters zusammenfassen, die dann als Entity-Typen in einem Entity-Relationship-Diagramm einer höheren Abstraktionsebene dargestellt werden.
Der hier vorgestellte Ansatz geht über die bereits bestehenden Clustering-Ansätze hinaus, da er sich nicht nur zur Darstellung eines vergröberten Entity-Relationship-Diagramms, sondern auch in den folgenden Bereichen einsetzen läßt: *Top-Down-Datenbankentwurf, Datenbank-Reengineering, Standarddiagramme bzw. Referenzschemata, Modellierung zusätzlicher Integritätsbedingungen und Schemaintegration*
Die bereits existierenden Ansätze, die nur Entity-Clustering unterstützen, werden um die Technik des Relationship-Clustering erweitert. Diese Technik ermöglicht es, nicht nur Entity-Typen sondern auch Beziehungstypen zu verfeinern. Insgesamt werden drei verschiedene Clustering-Techniken unterstützt: *Entity-Clustering*, *Simple-Relationship-Clustering* und *Complex-Relationship-Clustering*

1 Einleitung

Die Entity-Relationship-Modellierung [Chen76] hat sich als Methode zur konzeptuellen Datenmodellierung etabliert. Trotz der verfügbaren Tool-Unterstützung treten verschiedene Probleme auf, sobald sehr große Datenschemata erstellt werden. Bei der Abstimmung mit dem Fachbereich, der Einarbeitung neuer Projektmitarbeiter oder in der späteren Wartungsphase zeigt sich immer wieder die Notwendigkeit, sowohl einzelne Teilbereiche im Detail als auch den Zusammenhang dieser Teilbereiche mit dem Gesamtdatenschema darzustellen.
[FeM86, TWB89, Mis91, RaS92] schlagen die Technik des Entity-Clustering vor, um diesen Überblick zu ermöglichen. Sie fassen ganze Teildiagramme in sogenannten Entity-Clusters zusammen, die dann als (komplexe) Entity-Typen in einem Entity-Relationship-Diagramm einer höheren Abstraktionsebene dargestellt werden. Diese Technik läßt sich wiederholt anwenden, so daß verschiedene Entity-Relationship-Diagramme mit unterschiedlichem Abstraktionsgrad entstehen.

Der hier vorgestellte Ansatz präzisiert und erweitert die bereits bestehenden Clustering-Ansätze. Er geht über die zuvor genannten Ansätze hinaus, da er sich nicht nur zur Darstellung eines vergröberten Entity-Relationship-Diagramms, sondern auch in den folgenden Bereichen einsetzen läßt:

- Er kann insbesondere im Rahmen eines *Top-Down-Datenbankentwurfs* eingesetzt werden, um die schrittweise Analyse des Anwendungsbereichs oder des Unternehmens und damit den schrittweisen Entwurf des zugehörigen Informationsschemas zu unterstützen.
- Er läßt sich für das *Datenbank-Reengineering* einsetzen. Die unterschiedlichen Abstraktionsebenen werden während eines eventuell notwendigen Reverse Engineering wie bisher bottom-up erstellt. Im Anschluß daran erfolgt dann das Redesign im Gegensatz zu bestehenden Ansätzen top-down.
- Es lassen sich *Standarddiagramme* bzw. *Referenzschemata* für spezielle Branchen entwerfen, die für bestimmte Unternehmen individuell verfeinert und angepaßt werden können.
- Eine spezielle Clustering-Technik, das Simple-Relationship-Clustering, läßt sich zur Modellierung *zusätzlicher Integritätsbedingungen* verwenden. Die Darstellung dieser Integritätsbedingungen läßt sich jedoch ausblenden, wenn sie zum allgemeinen Verständnis nicht erforderlich ist.
- Clustering-Ansätze lassen sich für die Schemaintegration einsetzen. Ein Ansatz hierzu wird in [SHO95] vorgestellt und legt den Clustering-Algorithmus von [RaS92] zugrunde. Die Überlegungen aus [SHO95] lassen sich auf den in dieser Arbeit vorgestellten Clustering-Ansatz übertragen.

Der in der vorliegenden Arbeit vorgestellte Ansatz erweitert die zuvor genannten Ansätze, die nur Entity-Clustering unterstützen, um die Technik des Relationship-Clustering. Diese Technik ermöglicht es, nicht nur Entity-Typen sondern auch Beziehungstypen zu verfeinern. Die Regeln zur Cluster-Bildung sind in [JOS93] beschrieben. Insgesamt werden drei verschiedene Clustering-Techniken unterstützt:

① *Entity-Clustering*, das in [FeM86, TWB89, Mis91, RaS92] eingeführt bzw. verwendet wird.
② *Simple-Relationship-Clustering*, das dazu verwendet werden kann, um Beziehungstypen durch mehrere semantisch ähnliche Beziehungstypen zu verfeinern oder um zusätzliche Integritätsbedingungen darzustellen.
③ *Complex-Relationship-Clustering*, das sich zur Verfeinerung von Beziehungstypen durch ganze Entity-Relationship-Diagramme einsetzen läßt.

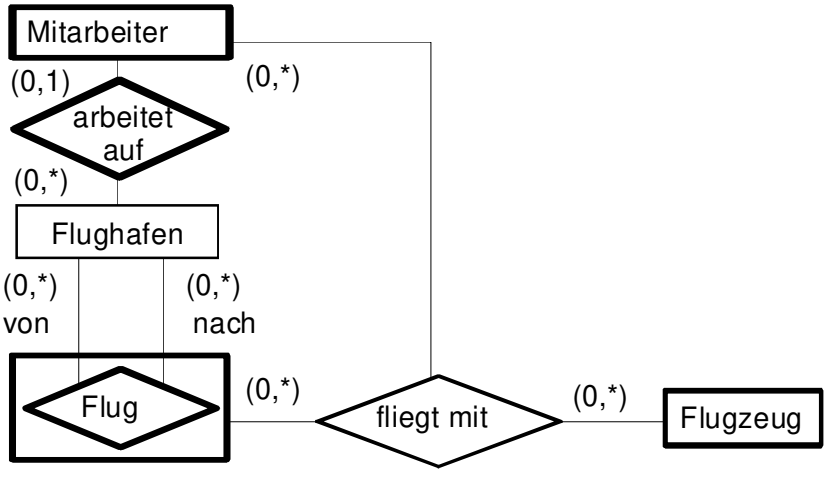

Abb. 1: Übersichtsdiagramm einer Fluggesellschaft

Im folgenden werden die verschiedenen Techniken und die durch diesen Ansatz ermöglichte Top-Down-Vorgehensweise am Entwurf des konzeptuellen Schemas einer Fluggesellschaft exemplarisch vorgestellt. Das in [RaS92] verwendete Beispiel wird hierzu in leicht modifizierter Version

genutzt. Das Übersichtsdiagramm des Beispiels ist in Abb. 1 dargestellt; fett umrandete Symbole stellen Clusters dar. Abschließend folgt eine Zusammenfassung und ein kurzer Ausblick.

2 Entity-Relationship-Modell-Clustering

2.1 Entity-Clustering

Für die Bildung von Entity-Clusters wird auf den Ansatz von [TWB89] zurückgegriffen. Von den dort angegebenen Grouping-Operationen werden Dominance-Grouping und Abstraction-Grouping für Generalisierung verwendet. Das im Entity-Relationship-Modell-Clustering verwendete Entity-Clustering zielt darauf ab, Informationen, die in einem engen Bezug zu einem Entity-Typ, der Cluster-Basis, stehen, mit diesem auf einer höheren Abstraktionsebene zusammenzufassen. Dabei handelt es sich typischerweise entweder um Informationen, die klassifizierend und damit beschreibend sind, oder um die Modellierung von Komponenten. Ebenso läßt sich auch die Generalisierung zur Cluster-Bildung auswerten. Aufgrund dieser Einschränkung ist die Bedeutung des Clusters bezogen auf die Realwelt identisch mit oder zumindest sehr ähnlich zu der Bedeutung der Cluster-Basis bezogen auf die Realwelt. Es soll vermieden werden, daß Entity- oder Beziehungstypen in einen Entity-Cluster aufgenommen werden, die nicht Aspekte der Cluster-Basis beschreiben, sondern vielmehr den Bezug zu einem anderen Entity- oder Beziehungstyp darstellen. Um ein Diagramm noch weiter zu abstrahieren, wird anstelle der anderen Grouping-Operationen aus [TWB89] das flexiblere Konzept des Relationship-Clustering eingesetzt.

Ein Entity-Typ *dominiert* einen anderen, wenn beide durch einen binären Beziehungsytp verbunden sind und die Maximumkardinalität des anderen Entity-Typs hinsichtlich des verbindenden Beziehungstyp 1 ist. [TWB89] fassen den dominierenden Entity-Typ mit den von ihm dominierten Entity-Typen zusammen und legen den dominierenden Entity-Typ als Cluster-Basis (Abb. 2a) fest. Ein Entity-Cluster wird nach seiner Basis benannt.

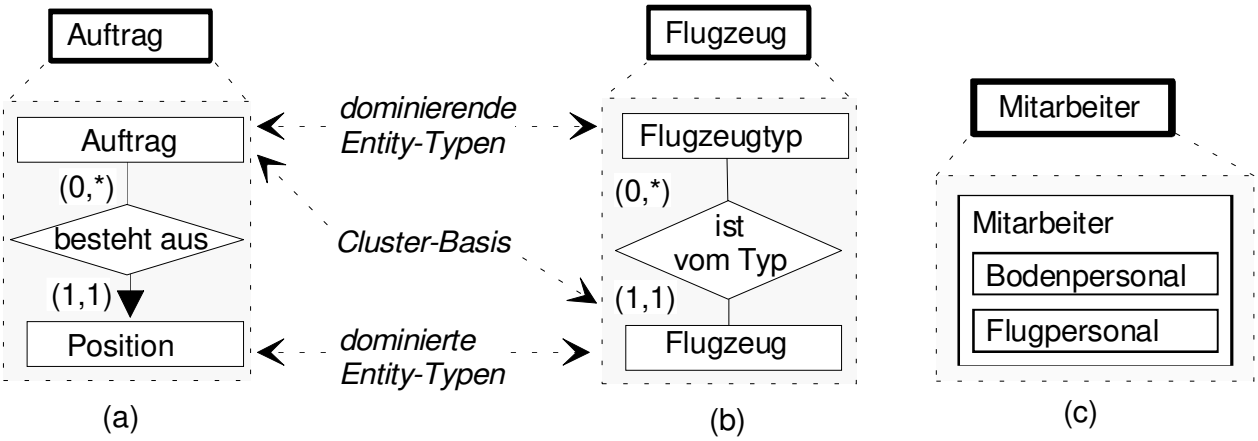

Abb. 2: Dominance-Grouping Dominance-Grouping Abstraction-Grouping

Es ist nicht immer sinnvoll, den *dominierenden* Entity-Typ als Cluster-Basis zu wählen, vielmehr kann es auch sinnvoll sein, *dominierte* Entity-Typen als Cluster-Basis zu wählen. Wenn die dominierenden Entities zur Klassifikation oder Beschreibung der dominierten verwendet werden, ist der dominierte Entity-Typ als Cluster-Basis zu wählen (Abb. 2b). Der dominierende Entity-Typ ist als Cluster-Basis festzulegen, wenn der dominierte Entity-Typ Details oder Komponenten des dominierenden Entity-Typs beschreibt (Abb. 2a).

Abstraction-Grouping wird in diesem Ansatz ausschließlich dazu verwendet, Subtypen mit ihrem Supertyp in einem Cluster zusammenzufassen. Im Beispiel (Abb. 2c) werden die Subtypen `Flug-personal` und `Bodenpersonal` mit ihrem Supertyp `Mitarbeiter` zusammengefaßt. Die Clusters werden gemäß dem Supertyp benannt.

2.2 Simple-Relationship-Clustering

Simple-Relationship-Clustering läßt sich dazu verwenden, Integritätsbedingungen zu präzisieren und generische Beziehungstypen zu modellieren. Dies wird an einem anderen Realweltausschnitt (Abb. 3) dargestellt. Häufig sind verschiedene Generalisierungshierarchien auf höchster Ebene durch einen Beziehungstyp verknüpft. In einem solchen Fall existieren oft Abhängigkeiten zwischen den unterschiedlichen Subtypen hinsichtlich ihrer gegenseitigen Zuordnungsmöglichkeiten. In Abb. 3 wird der Ausschnitt eines existierenden mittelständischen Unternehmens modelliert.

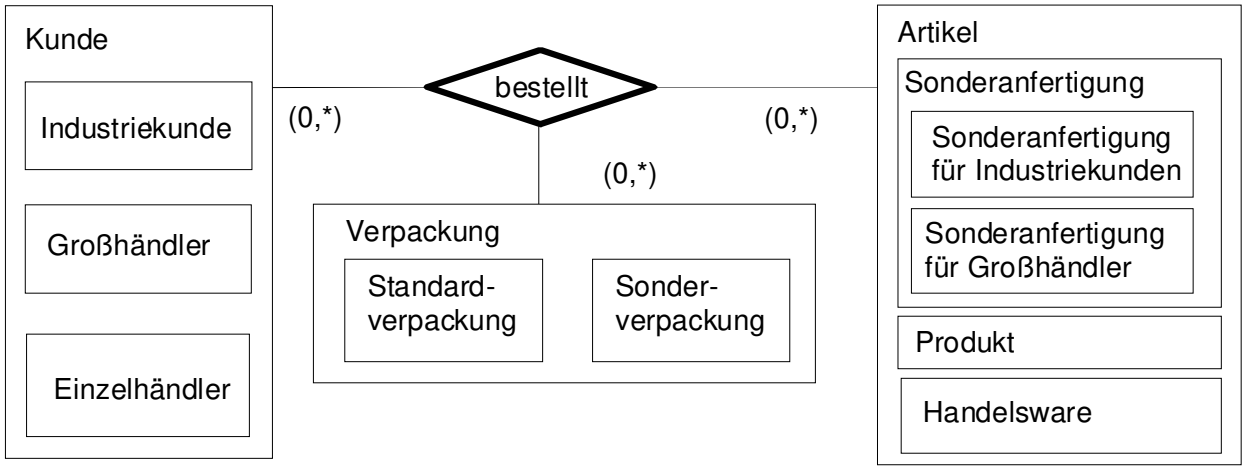

Abb. 3: Ausschnitt eines Entity-Relationship-Diagramms eines Industrieunternehmens: Kunden bestellen Produkte in bestimmten Verpackungen bzw. Packungsgrößen

Zusätzlich wurden die folgenden Zusammenhänge und Bedingungen identifiziert:

- Nur Einzelhändler können Handelsware bestellen.
- Nur Industriekunden und Großhändler können Sonderanfertigungen bestellen. Sonderanfertigungen sind entweder für Großhändler oder für Industriekunden.
- Für jede Sonderanfertigung muß mindestens eine Bestellung existieren.
- Einzelhändler können nur Standardverpackungen bestellen, Industriekunden und Großhändler können auch Sonderverpackungen bestellen.
- Produkte können mit Sonder- oder Standardverpackung bestellt werden.
- Sonderanfertigungen können nur mit Sonderverpackung bestellt werden, Handelsware nur mit Standardverpackung.

Die Entity-Typen werden mit ihren Subtypen in Entity-Clusters auf Basis von Abstraction-Grouping zusammengefaßt. Simple-Relationship-Clustering wird eingesetzt, um die oben formulierten zusätzlichen Integritätsbedingungen in der Verfeinerung zu modellieren. Im oberen Teil von Abb. 4. wird das Übersichtsdiagramm gezeigt, darunter werden die Verfeinerungen der Clusters dargestellt.

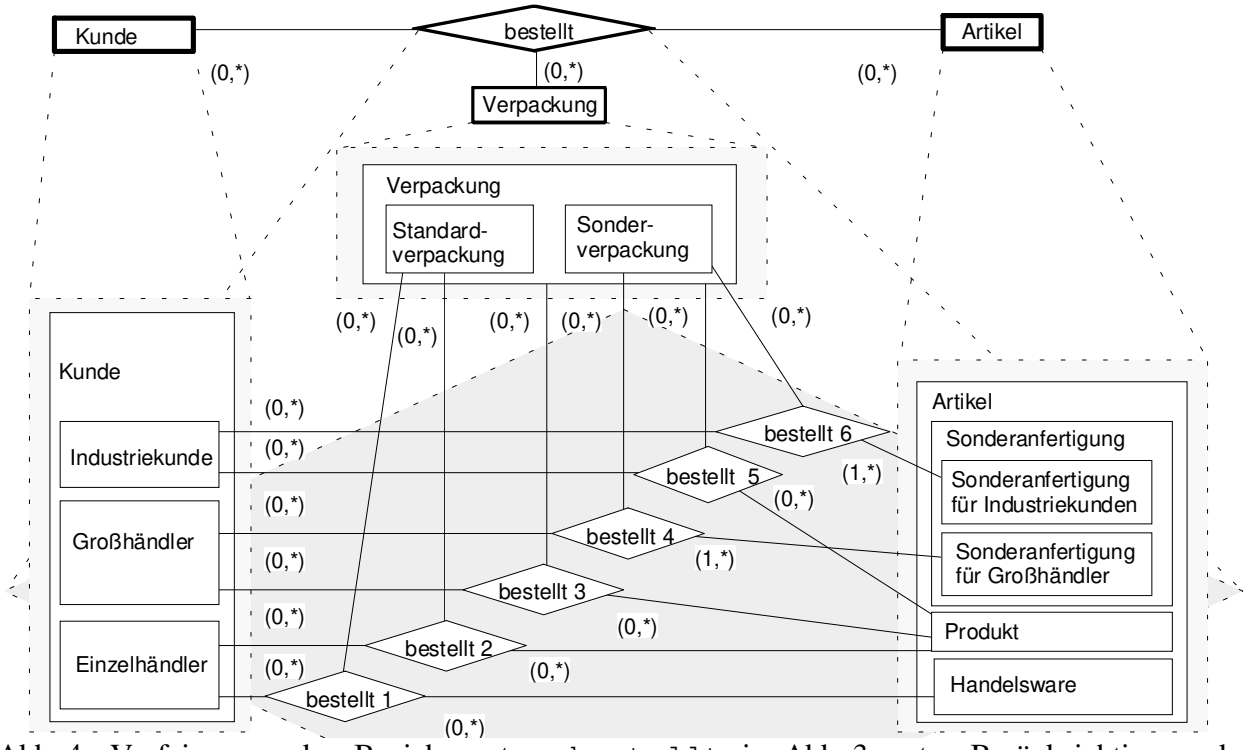

Abb. 4.: Verfeinerung des Beziehungstyps `bestellt` in Abb. 3 unter Berücksichtigung der zusätzlichen Integritätsbedingungen

3.3 Complex-Relationship-Clustering

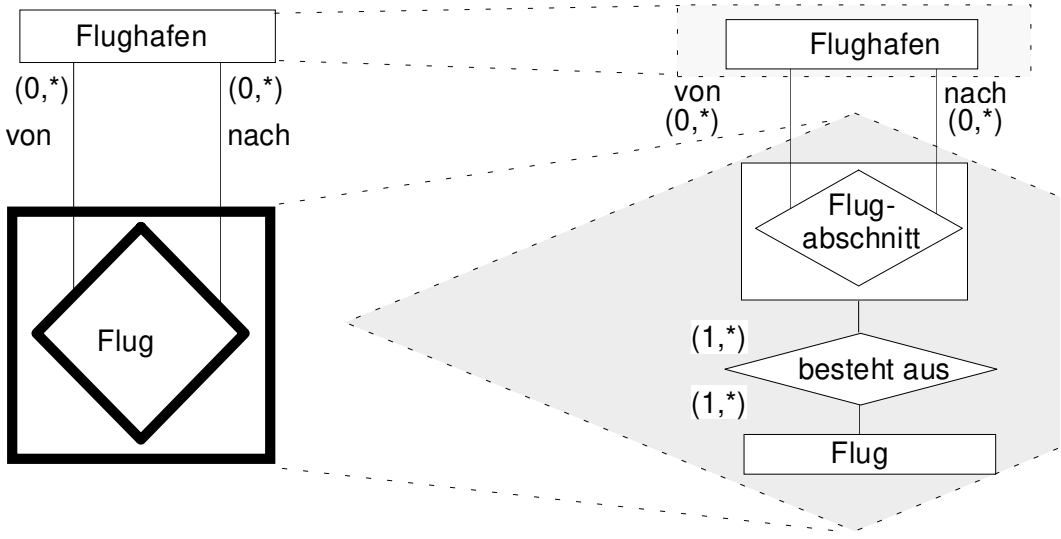

Abb. 5: Verfeinerung von `Flug` mittels Complex-Relationship-Clustering

Bei großen Entity-Relationship-Schemata bestehen häufig komplexe Beziehungen zwischen den wichtigen Entity-Typen. Die detaillierte Struktur dieser komplexen Beziehungen wird durch die Verkettung unterschiedlicher Beziehungs- und Entity-Typen modelliert. Für den Entwurf komplexer Entity-Relationship-Diagramme wird daher das Konzept des *Complex-Relationship-Clustering* unterstützt, das die Zusammenfassung solcher verketteter Entity- und Beziehungstypen

in einem Complex-Relationship-Cluster erlaubt. Ein `Flug` (Abb. 5) besteht aus verschiedenen `Flugabschnitt`en, um auch Flüge mit Zwischenlandungen zu ermöglichen.

Im Gegensatz zu einem Simple-Relationship-Cluster wird ein Complex-Relationship-Cluster nicht nur durch semantisch ähnliche Beziehungstypen verfeinert, sondern es werden zusätzliche Entity-Typen und Beziehungstypen eingeführt bzw. aufgenommen.

In Abb. 6 wird auch der Aggregationstyp `fliegt mit` verfeinert. Aus Gründen der Übersichtlichkeit ist die Verfeinerung von `fliegt mit` nicht grau unterlegt. Nicht nur Entity-Clusters sondern auch Complex-Relationship-Clusters können an einem anderen Relationship-Cluster teilnehmen; beispielsweise nimmt auch `Flug` an `Durchführung` teil. Für einen Flug gibt es zu unterschiedlichen Zeitpunkten Flugdurchführungen. Dementsprechend gibt es für die zugeordneten Flugabschnitte Flugabschnittsdurchführungen, für die jeweils ein bestimmtes Flugzeug benutzt wird.

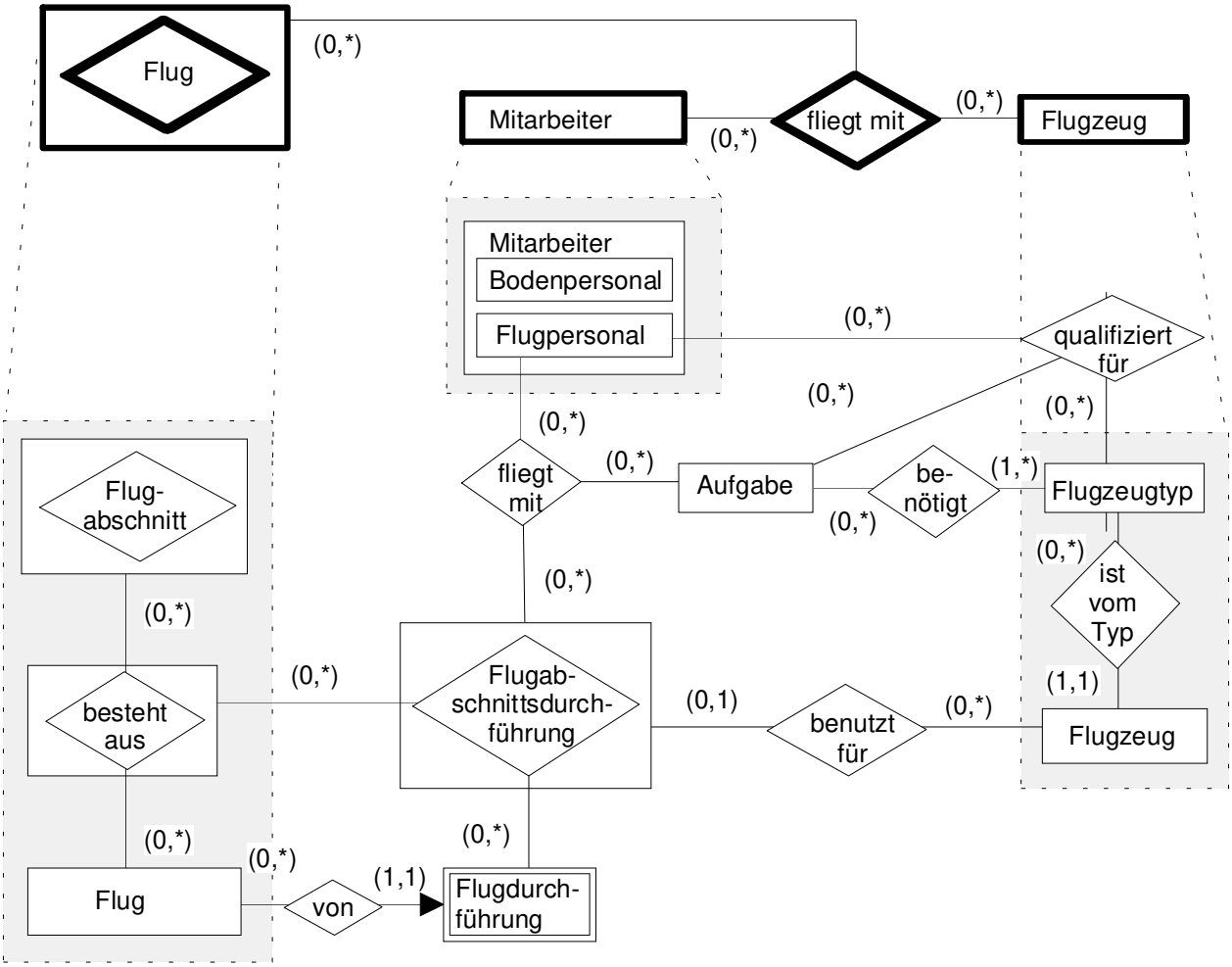

Abb. 6: Verfeinerung von `fliegt mit`

Abb. 7 zeigt das Detaildiagramm des Beispiels. Dieses Detaildiagramm liegt beiden vorgestellten Cluster-Hierarchien zugrunde.

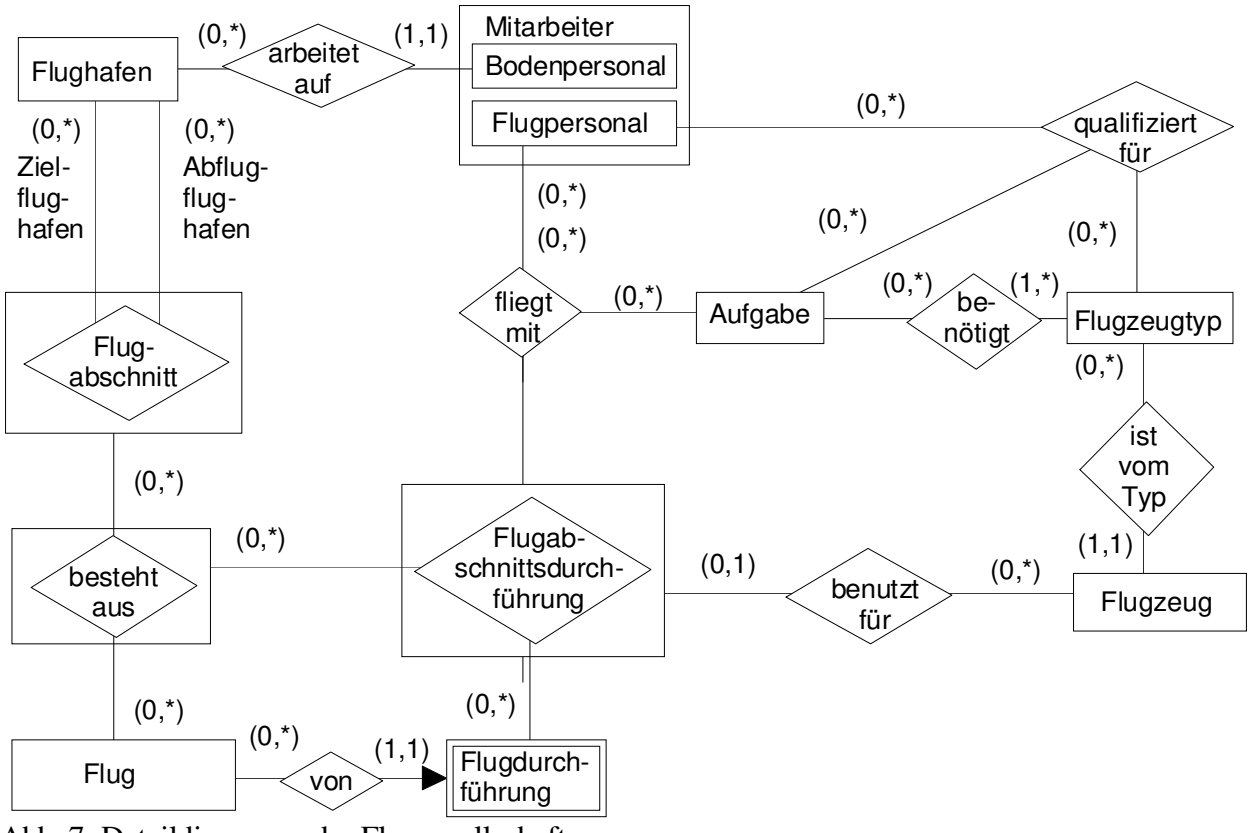

Abb. 7: Detaildiagramm der Fluggesellschaft

4. Zusammenfassung

Der Grundgedanke des in der vorliegenden Arbeit eingeführten Entity-Relationship-Modell-Clustering besteht darin, zunächst die wichtigsten und zentralen Entity-Typen und die groben Beziehungstypen zwischen ihnen zu bestimmen. Diese Beziehungstypen werden dann mittels Complex- und Simple-Relationship-Clustering verfeinert und präzisiert. Die Entity-Typen werden ebenfalls verfeinert. Entity-Clustering wird einerseits dazu verwendet Generalisierungsstrukturen festzulegen und andererseits Entity-Typen durch Dominance-Grouping exakter zu beschreiben. Wenn die Beziehungstypen des Übersichtsdiagramms geeignet gewählt werden, besteht die Möglichkeit, den Detailentwurf ihrer Verfeinerungen durch unterschiedliche Teams parallel durchführen zu lassen.

Die Relationship-Clusters sind *disjunkt*; die Verbindung zu anderen Relationship-Clusters kann nur durch Entity-Clusters oder Relationship-Clusters dargestellt werden, die an anderen Relationship-Clusters teilnehmen. Aufgrund dieser Einschränkung können Änderungen in einem Cluster keine unvorhersehbaren Auswirkungen in anderen Clusters bewirken, da sich solche Änderungen nur in der kontextsensitiven Darstellung durchführen lassen. Dies unterstützt die kooperative Erstellung von großen Entity-Relationship-Schemata.

In diesen Clustering-Ansatz sind nach der ersten Fassung aus [JOS93] noch zahlreiche Erfahrungen aus verschiedenen Praxisprojekten in den Bereichen öffentliche Verwaltung, Dienstleistung und Industrie eingeflossen. Da kein kommerzielles Tool zur Verfügung stand und steht, das die vorgestellte Clustering-Technik unmittelbar unterstützt, wurden geeignete Ausschnitte des Detail-

diagramms als Sichten modelliert, die der kontextsensitiven Darstellung der jeweiligen Clusters entsprachen. Der Ansatz erwies sich als geeignet, schnell den Überblick zu erlangen, Schemata neu zu erstellen bzw. bereits bestehende Schemata zu restrukturieren.

A. Literatur

[Chen76] Chen, P. P.: The entity-relationship model: Toward a unified view of data, ACM TODS 1 (1976), No.1, 166 - 192

[FeMi86] Feldman, P.; Miller, D.: Entity model clustering: Structuring a data model by abstraction, The Computer Journal 29(1986), No. 4, 348-360

[JOS93] Jaeschke, P.;. Oberweis, A; Stucky, W.: Extending ER model clustering by relationship clustering, in Elmasri R., Kouramajian, V; .Thalheim, B. (Eds.): Proc. 12th International Conference on the Entity Relationship Approach, Arlington, TX, 1993, LNCS 823, Springer Verlag 1994, 451-462.

[Mis91] Mistelbauer, H.: Datenmodellverdichtung: Vom Projektdatenmodell zur Unternehmensarchitektur, Wirtschaftsinformatik 33(1991), No. 4, 289-299

[RaS92] Rauh, O.; Stickel, E.: Entity tree clustering - A method for simplifying ER design, in Pernul, G.; Tjoa A M. (Eds.): Proc. of the 11th Intern. Conf. on Entity-Relationship Approach, Karlsruhe, Germany 1992, Springer-Verlag 1992, 62-78

[SHO95] E. Stickel, J. Hunstock, A. Ortmann, J. Ortmann:
Verfahren zur werkzeuggestützten Integration von Datenbankschemata.
In: W. König (Hrsg.): *Wirtschaftsinformatik 95*, Frankfurt, Physica-Verlag, Heidelberg 1995, 205-222.

[TWB89] Teorey, T. J.; Guangping, W.; Bolton, D. L.; Koenig, J. A.: ER model clustering as an aid for user communication and documentation in database design, Communications of the ACM 32(1989), No. 8, 975-987