

Übung Data Warehousing, WS 2007/08		Übung 1
Abgabeschluss: Mo., 29. Okt. 2007, 18:00 Uhr		12 Punkte
Matrikel-Nr.:	Name:	
Matrikel-Nr.:	Name:	

Auszuarbeitende Aufgaben:

Nr.	Punkte	Beispiel
1	12	Wohnbau-Genossenschaft: Konzeptueller DW-Entwurf

Konzeptueller Data-Warehouse-Entwurf

Hinweise für die Übungs-Ausarbeitung:

- Erstellen Sie ein Dokument mit dem konzeptuellen Entwurf des Data Warehouse in DFM-Notation nach Golfarelli. Sie können wahlweise ein beliebiges grafisches Werkzeug (z.B. Visio) verwenden oder das DFM-Schema auf Papier entwerfen.
- Dokumentieren Sie Ihren Entwurf ausreichend, insbesondere falls Sie ergänzende Annahmen zur Aufgabenstellung treffen.
- Verwenden Sie diese Seite als Deckblatt für den Ausdruck Ihrer Übungs-Ausarbeitung. Geben Sie das ausgedruckte Dokument im Briefkasten vor dem DKE-Institut im Management-Zentrum, 3. Stock Trakt B, ab (Abgabeschluss beachten).

Gegeben ist das konzeptuelle Schema der operativen Datenbank einer Wohnbau-Genossenschaft zur Verwaltung von Mietverträgen (siehe Abbildung 1). Das in einer UML-Klasse durch {id} gekennzeichnete Attribut bildet den Schlüssel; zusammen gesetzte Schlüssel werden in einem {id}-Constraint neben der UML-Klasse angegeben.

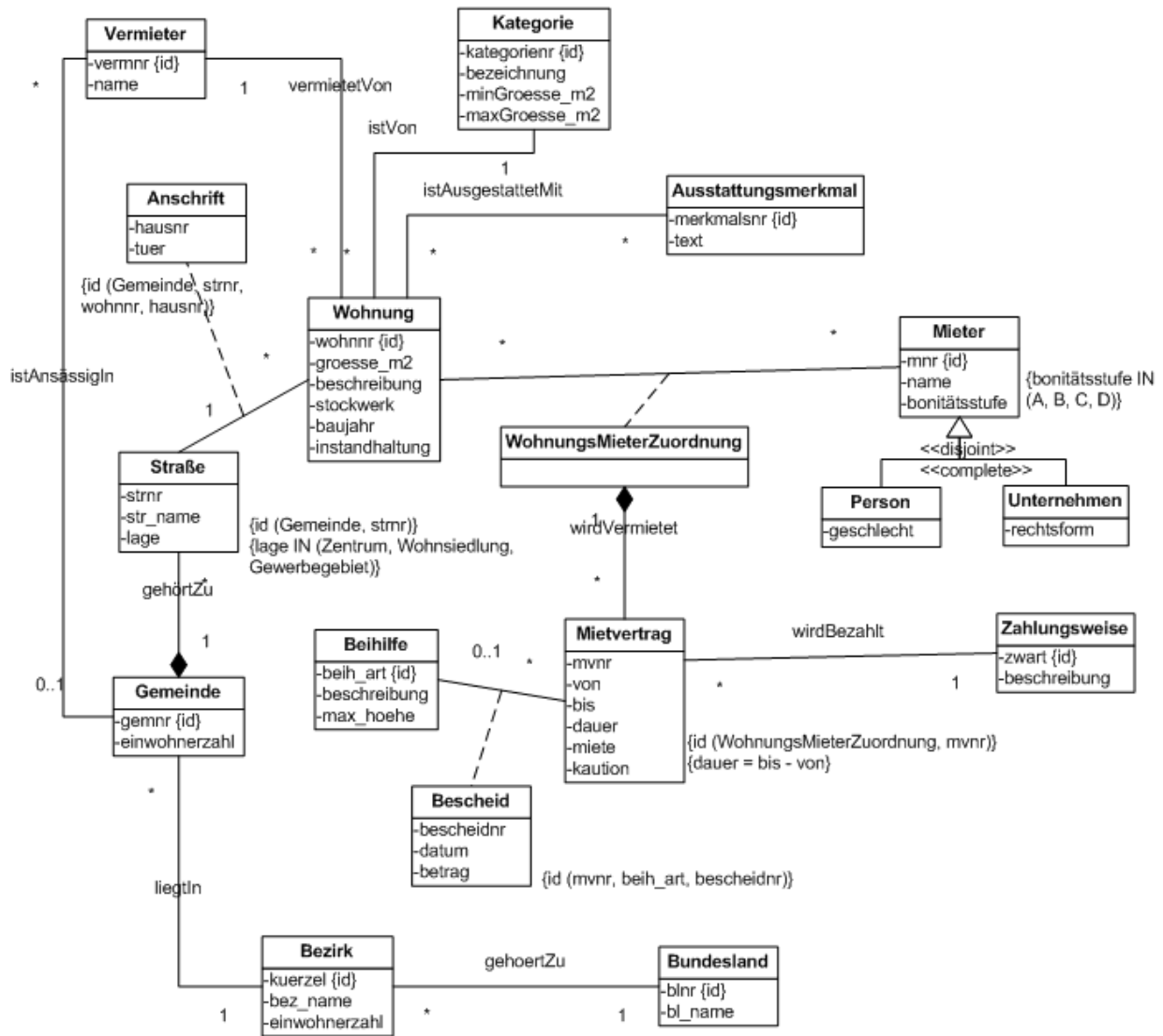


Abbildung 1. UML-Diagramm der operativen Datenbank der Wohnbau-Genossenschaft

Die operative Datenbank hat folgende Eigenschaften:

1. *Konsistenz der Kategorie einer Wohnung:* wie in Abbildung 1 ersichtlich, ist jeder Wohnung eine Kategorie zugeordnet, die Aufschluss über die bewohnbare Fläche (Wohnungsgröße in m^2) gibt. Das Attribut `groesse_m2` der Klasse `Wohnung` liegt immer zwischen den Werten `minGroesse_m2` und `maxGroesse_m2` der Klasse `Kategorie` (d.h. $\text{minGroesse_m2} \leq \text{groesse_m2} \leq \text{maxGroesse_m2}$).
2. *Beschreibung der Ausstattung einer Wohnung:* zusätzlich zur Wohnfläche (die in der Klasse `Kategorie` abgebildet ist, siehe oben) wird in der Datenbank die Ausstattung einer Wohnung erfasst. Zu diesem Zweck ist die Klasse `Ausstattungsmerkmal` vorgesehen, mit der z.B. eine eventuell bereitgestellte Möblierung, die Art der Heizung oder die vorhandenen Räume gespeichert werden.
3. *Veränderlichkeit des Bonitäts-Status:* je nach Zuverlässigkeit bei der Bezahlung der Miete kann sich die Bonitätsstufe eines Mieters vierteljährlich ändern. In der operativen Datenbank wird immer die aktuell gültige Bonitätsstufe gespeichert.
4. *Speicherung von Wohnungs-Adressen:* in der operativen Datenbank der Wohnbau-Genossenschaft sind die Adressen in einer eigenen Klasse `Straße` gespeichert. Jede Straße liegt in einer Gemeinde (Hinweis: das Attribut `gemnr` gibt die Postleitzahl der Gemeinde an). Diese Design-Entscheidung wurde getroffen, um die Lage einer Straße in Bezug auf die zu erwartende Wohnqualität zu speichern. Zudem ist dann die Datenbank besser änderbar, z.B. wenn Straßen umbenannt werden.
5. *Gewährung von Beihilfen für Mietverträge:* Personen, welche Mietverträge abgeschlossen haben, können bei der Wohnbau-Genossenschaft die Gewährung einer Beihilfe beantragen. Diese wird abhängig vom Einkommen genehmigt, das in der Datenbank nicht gespeichert ist. Die Bescheide zur Genehmigung der Wohnbeihilfe werden als Instanzen der Assoziations-Klasse `Beihilfe` erfasst und jeweils einem Mietvertrag zugeordnet.
6. *Abänderung von Beihilfen-Bescheiden:* fallweise kann es vorkommen, dass Beihilfen-Bescheide abzuändern sind (z.B. wenn der Mieter erfolgreich Berufung eingelegt hat). Gibt es zu einem Mietvertrag (`mvnr`) und einer Beihilfen-Art (`beih_art`) zwei oder mehrere Bescheide (`bescheidnr + datum`), so ist die semantische Einschränkung zu beachten, dass ein neuerer Bescheid alle älteren Bescheide ungültig werden lässt. *Hinweis:* die Bescheid-Nummer ist immer eindeutig, auch wenn ein Bescheid zur Abänderung eines älteren Bescheids erlassen wird.
7. *Einwohner-Zahlen von Gemeinden, Bezirken und Bundesländern:* in der Datenbank werden die Einwohnerzahlen von Gemeinden und Bezirken nach jeder Volkszählung aktualisiert. Zur Vereinfachung können Sie jedoch annehmen, dass die `einwohnerzahl`-Attribute unveränderlich sind.

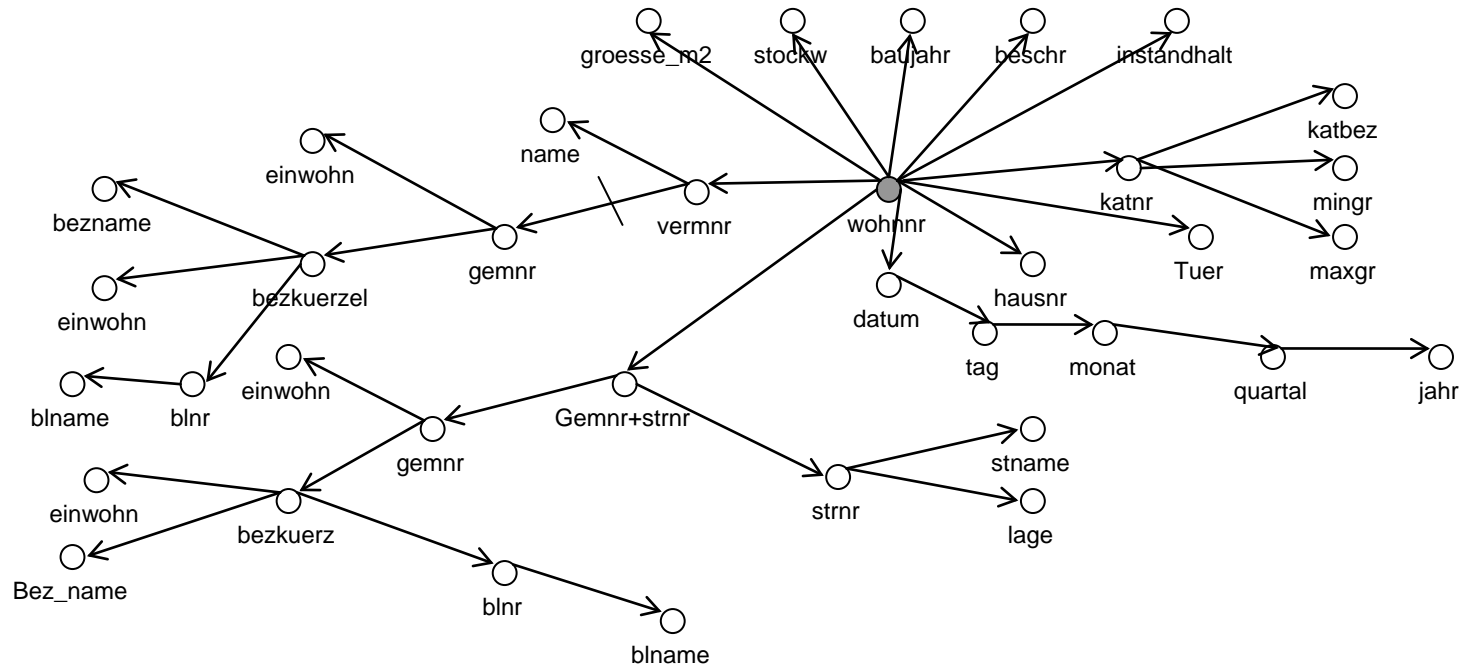
Auf Basis des operativen Schemas soll ein Data Warehouse modelliert werden, das folgende Anforderungen erfüllt:

- Im Data Warehouse sind Analysen über Mietverträge und Instandhaltungskosten durchzuführen. Die Kennzahl **miete** kann in allen Dimensionen sowohl durch Summen- als auch Mittelwertbildung aggregiert werden.
- Bei Datums-Werten sollen neben Analysen über einzelne Tage auch Aggregationen über Monat, Quartal und Jahr möglich sein.
- Im Data Warehouse muss ersichtlich sein, wann ein Mieter welche Bonitätsstufe hatte. Beim Abschluss von Mietverträgen wird im Data Warehouse immer die zuletzt gültige Bonität gespeichert.
- Schließlich soll im Data Warehouse gewährleistet sein, dass nur gültige Beihilfen-Bescheide gespeichert sind. Veraltete (d.h. durch Bescheide neueren Datums aufgehobene) Bescheide sollen im Data Warehouse nicht aufscheinen.

Ihre Aufgabe:

1. Erstellen Sie ein konzeptuelles Data-Warehouse-Schema für die Mietverträge-Datenbank. Verwenden Sie dabei die Notation und Vorgehensweise des Dimensional Fact Model (DFM) nach Golfarelli. Treffen Sie vernünftige Annahmen, welche Attribute dimensional und welche nicht-dimensional sind. Erläutern Sie, wie Sie die obigen Anforderungen (Verwaltung der historischen Bonitätsstufe etc.) umgesetzt haben.
2. Geben Sie sowohl Ihren Attribut-Baum vor der Bereinigung durch Grafting und Pruning sowie das fertige konzeptuelle DFM-Schema ab. Dokumentieren Sie Ihre Entscheidungen bei der Bereinigung des Attributbaums.

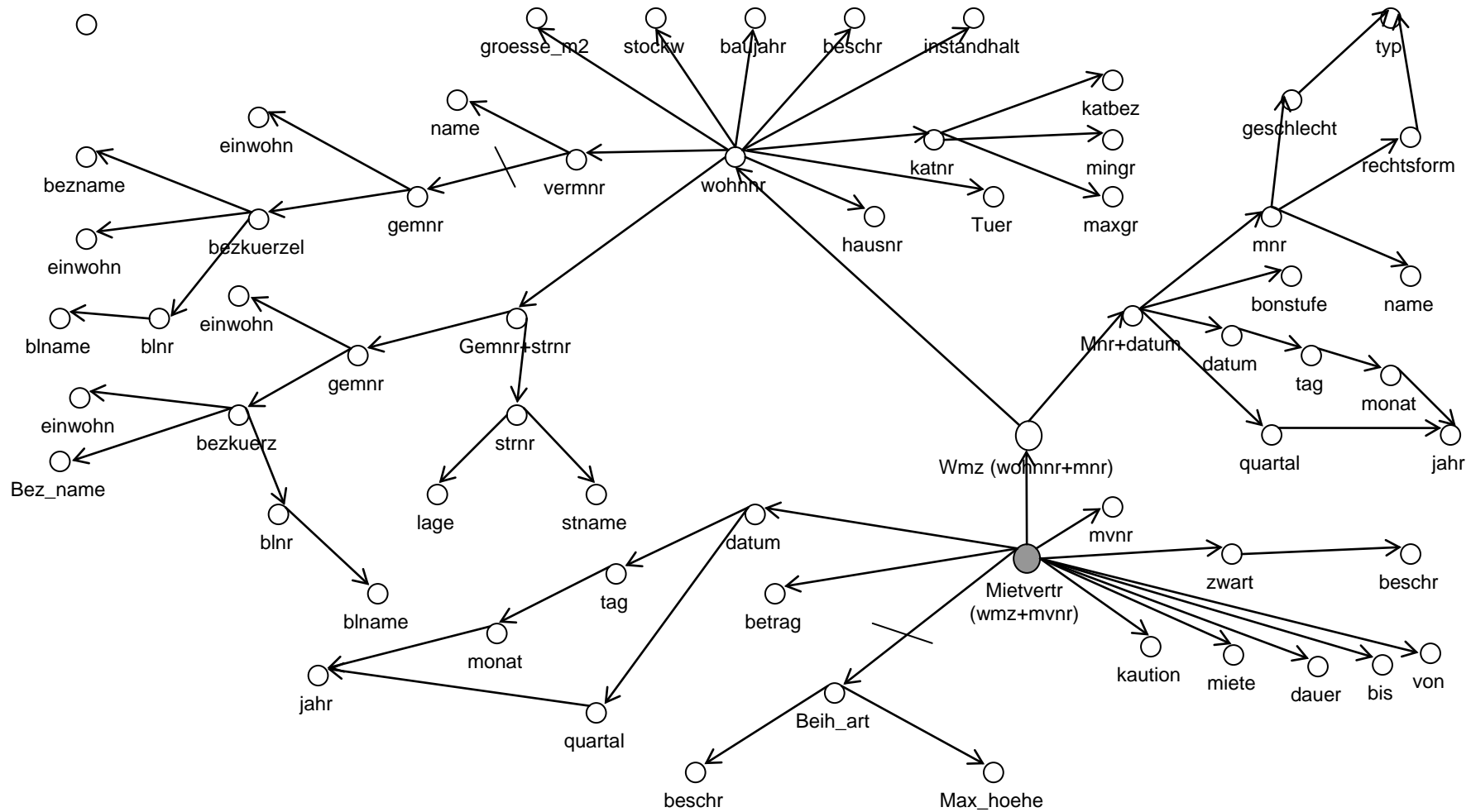
Vor Bereinigung durch Pruning und Grafting: Faktklasse Wohnung



Anmerkung zur Erstellung des Attributbaums

- Aus dem UML Modell der Angabe geht nicht eindeutig hervor, ob eine Wohnung mehrere Anschriften besitzen kann. Aus der Sicht einer Wohnungsgenossenschaft erscheint es nicht notwendig zu sein in einem DW die Historie der Haus- bzw. Türnummer einer Wohnung zu speichern. Aus diesem Grund wurde die Annahme getroffen, dass im DW ein Haus immer nur eine Anschrift hat. Deshalb wurde die Assoziationsklasse des operativen Datenbanks gemäß dem Algorithmus von Golfarelli in ein DFM umgewandelt. Die gleiche Argumentation gilt auch bei der Modellierung der Assoziationsklasse Bescheid (für die Faktklasse Mietvertrag).

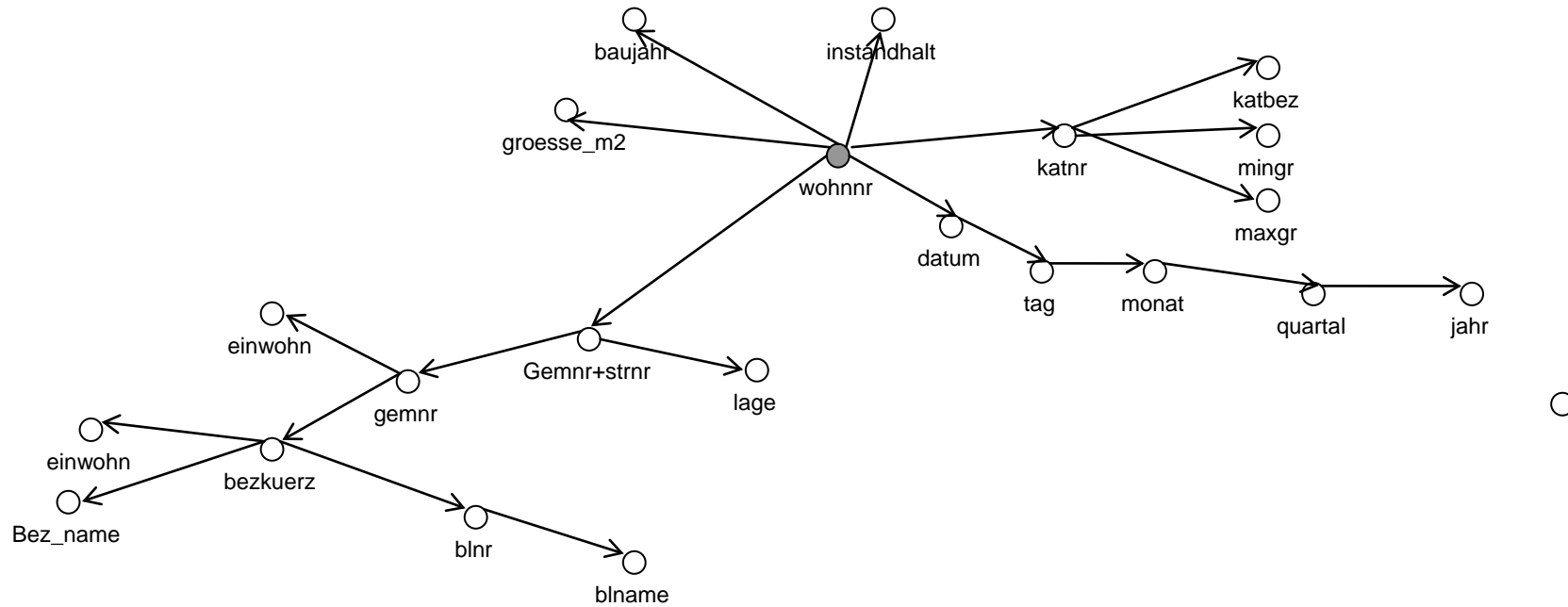
Vor Bereinigung durch Pruning und Grafting: Faktklasse Mietvertrag



Anmerkungen zur Erstellung des Attributbaums

- Da die Generalisierung des Mieters komplett und disjunkt ist, wurde eine weitere Dimension „Typ“ eingeführt. Diese Dimension dient zur Unterscheidung ob der Mieter eine Person oder ein Unternehmen ist.
- Um temporale Analysen der Mieten durchführen zu können, wurde eine Datums-Dimension eingefügt.
- Vom Mietvertrag kommt der Algorithmus zum Erstellen des Attributbaums auch zur Entität der Wohnung. Von daher ist der Baum der vom Knoten wohnnr weggeht identisch zu dem Baum zur Faktklasse Wohnung.
- Im Data Warehouse wurde verlangt, dass eine Historie der Bonitätsstufen für jeden Mieter mitgespeichert wird. Deshalb wurde die Mieternummer um ein Datumsfeld ergänzt. Dies erlaubt die Unterscheidung zwischen den Bonitätsstufen entlang einer Zeitachse. Darüber hinaus bleibt eine :1 Beziehung erhalten, welche sich im DFM darstellen lässt.
- Die Assoziationsklasse WohnungsMieterZuordnung (WMZ) wurde wie eine eigene Entität behandelt. Sofern man die WMZ als Entität betrachtet, ergibt sich ausgehend von der WMZ eine n:1 Relation zur Entität Wohnung und eine n:1 Relation zur Entität Mieter.
- Die Modellierung von Bescheid-Beihilfe ist analog zur Anschrift-Straße von der Faktklasse Wohnung. Da laut Angabe im Data Warehouse nur der jeweils aktuelle Bescheid gespeichert werden soll, kann diese Assoziationsklasse als 1:1 Beziehung zwischen Bescheid und Mietvertrag gesehen werden. Von daher wurde diese Assoziationsklasse gemäß dem Algorithmus von Golfarelli umgewandelt.

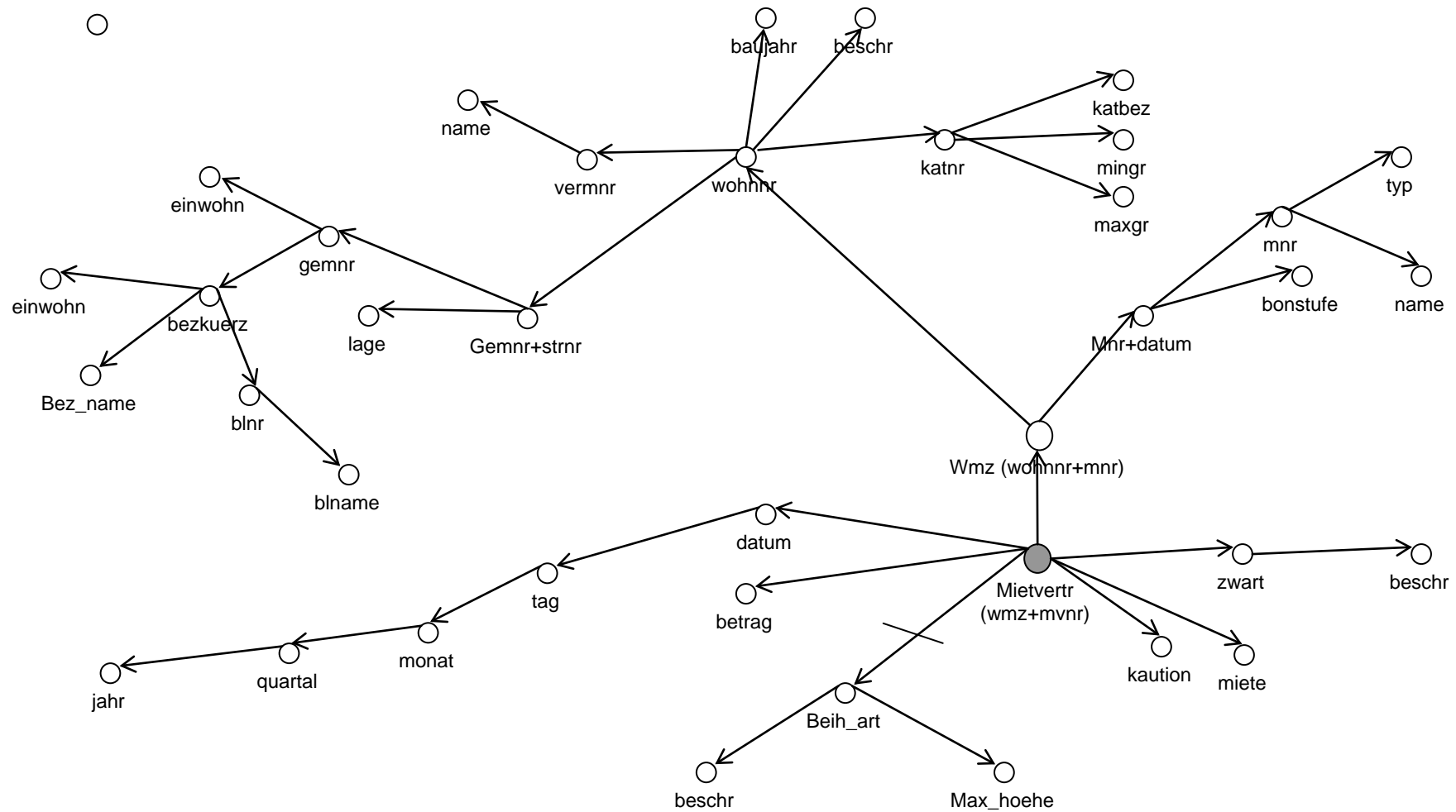
Nach Bereinigung durch Pruning und Grafting: Faktklasse Wohnung



Annahmen und Begründung für die Bereinigung:

- Der Baum beschreibt das DFM nach Pruning und Grafting. Eine Einteilung der Dimensionen bzw. der dimensional und nicht-dimensionalen Attribute sowie der Kennzahlen wird erst später getroffen.
- Instandhaltung: Angenommen wurde, dass es sich hierbei um periodisch (zum Beispiel monatlich) anfallende Kosten handelt, welche zur Erhaltung des bestimmungsmäßigen Gebrauchs aufgewendet werden müssen. Dazu gehören Kosten, um die durch Abnutzung, Alterung und Witterungseinwirkung entstehenden Mängel zu beseitigen. Hier handelt es sich also um einen Geldbetrag. Von daher sind folgende Aggregationsfunktionen prinzipiell sinnvoll: MAX, MIN, AVG, SUM.
- Die Instandhaltungskosten sind aller Voraussicht nach nicht vom Vermieter abhängig. Aus der Angabe geht auch nicht hervor, dass aus irgendwelchen anderen Gründen die Dimension des Vermieters nötig ist. Aus diesem Grund wurde der Knoten, der den Vermieter repräsentiert, entfernt. Es wurde der gesamte Tree entfernt, da eine Aggregation über Gemeinde, Bezirk oder Bundesland aus dem der Vermieter stammt ebenfalls nicht sinnvoll ist.
- Zur Identifikation einer Straße wird sowohl die Straßenummer, als auch die Gemeindenummer benötigt. Eine Analyse über alle Straßen mit einer bestimmten Nr. bzw. Namen ist nur unter Berücksichtigung der Gemeinde sinnvoll. Aus diesem Grund wurden die Knoten Straßenummer und Straßename entfernt.
- Eine Analyse über Hausnummer und Haustür ist ebenfalls nicht sinnvoll. Diese Attribute wurden ebenfalls entfernt. Gleiches gilt für Stockwerk und Beschreibung.

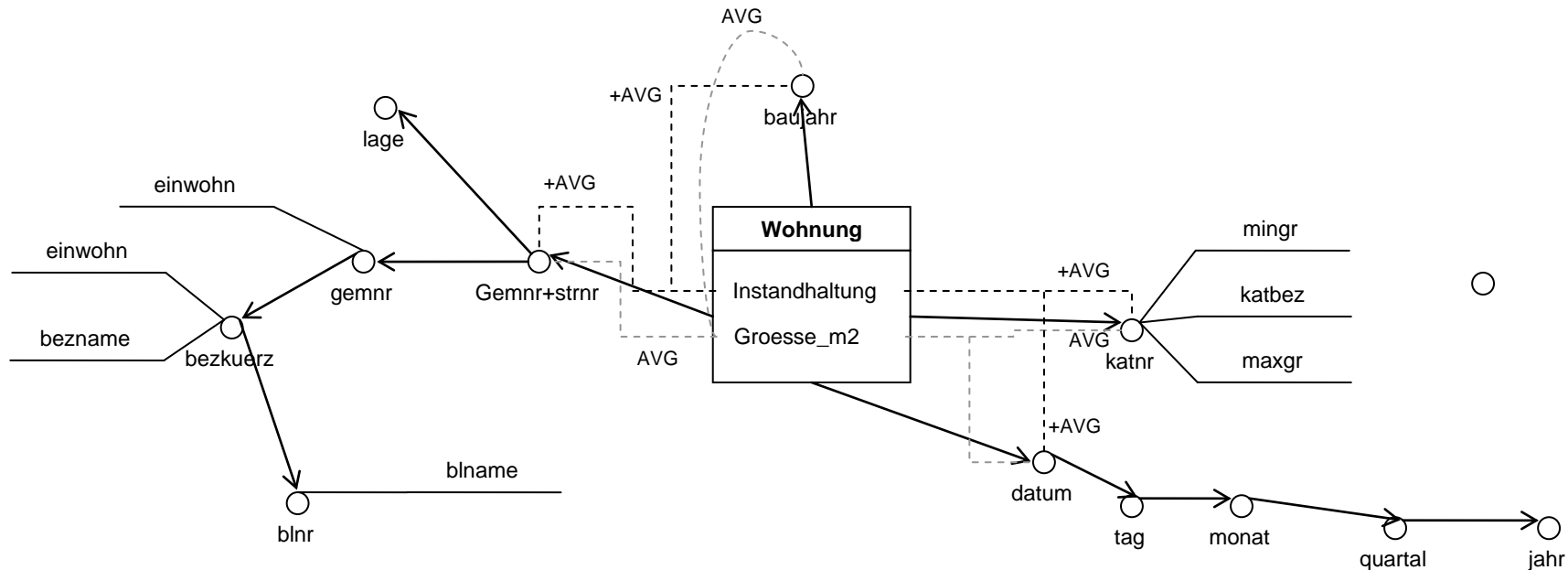
Nach Bereinigung durch Pruning und Grafting: Faktklasse Mietvertrag



Annahmen und Begründung für die Bereinigung:

- Der Baum beschreibt das DFM nach Pruning und Grafting. Eine Einteilung der Dimensionen bzw. der dimensional und nicht-dimensionalen Attribute sowie der Kennzahlen wird erst später getroffen.
- Die Mietvertragsnummer allein, sowie die Bescheidnummer sind für die Analyse uninteressant und wurden deshalb entfernt.
- Dauer ist ein redundantes Attribut welches sich aus den Attributen von und bis ergibt und wurde deshalb entfernt. Es wurde die Annahme getroffen, dass die Wohnbau-Genossenschaft keine Analysen über zukünftige Mieten durchführen will, von daher wurden die Attribute „von“ und „bis“ ebenfalls entfernt. Falls Analysen der in Zukunft zu erwartenden Mieten von Interesse sind, wäre es möglich für „von“ und „bis“ eine Zeitkomponente einzuführen. Damit könnte man zum Beispiel folgende Frage beantworten: Welcher Mietumsatz wird aller Voraussicht nach mit Wohnung die noch bis mindestens ins Jahr 2010 vermietet sind gemacht?
- Es wurde die Annahme getroffen, dass eine Analyse der Mieten über die Dimension Geschlecht und Rechtsform uninteressant ist. Deshalb wurden diese Dimensionen gelöscht. Es bleibt lediglich die Dimension Typ über um die Mieten von Personen und Unternehmen getrennt analysieren zu können.
- Obwohl eine Historie der Bonitätsstufen mitgespeichert werden muss, ist eine Analyse der Miete über dieses Datum (welches im Prinzip das Datum eines Bonitätsstufen-Updates darstellt) uninteressant und dieser Subtree wird daher entfernt.
- Die Größe einer Wohnung ist implizit durch die Kategorie festgelegt. Da ein Data Warehouse zur Analyse von Fragestellungen auf eher strategischer Ebene verwendet wird, wurde entschieden, dass die Wohnungsgröße im Data Warehouse nicht notwendig ist.
- Aus der Angabe ist nicht ersichtlich, ob die Analyse der Mieten auf allen örtlichen Ebenen (Straße+Gem, Gem, Bezirk und Bundesland) notwendig ist oder ob eine Ebene gestrichen werden kann. Aus diesem Grund wurden alle Ebenen im Data Warehouse behalten.

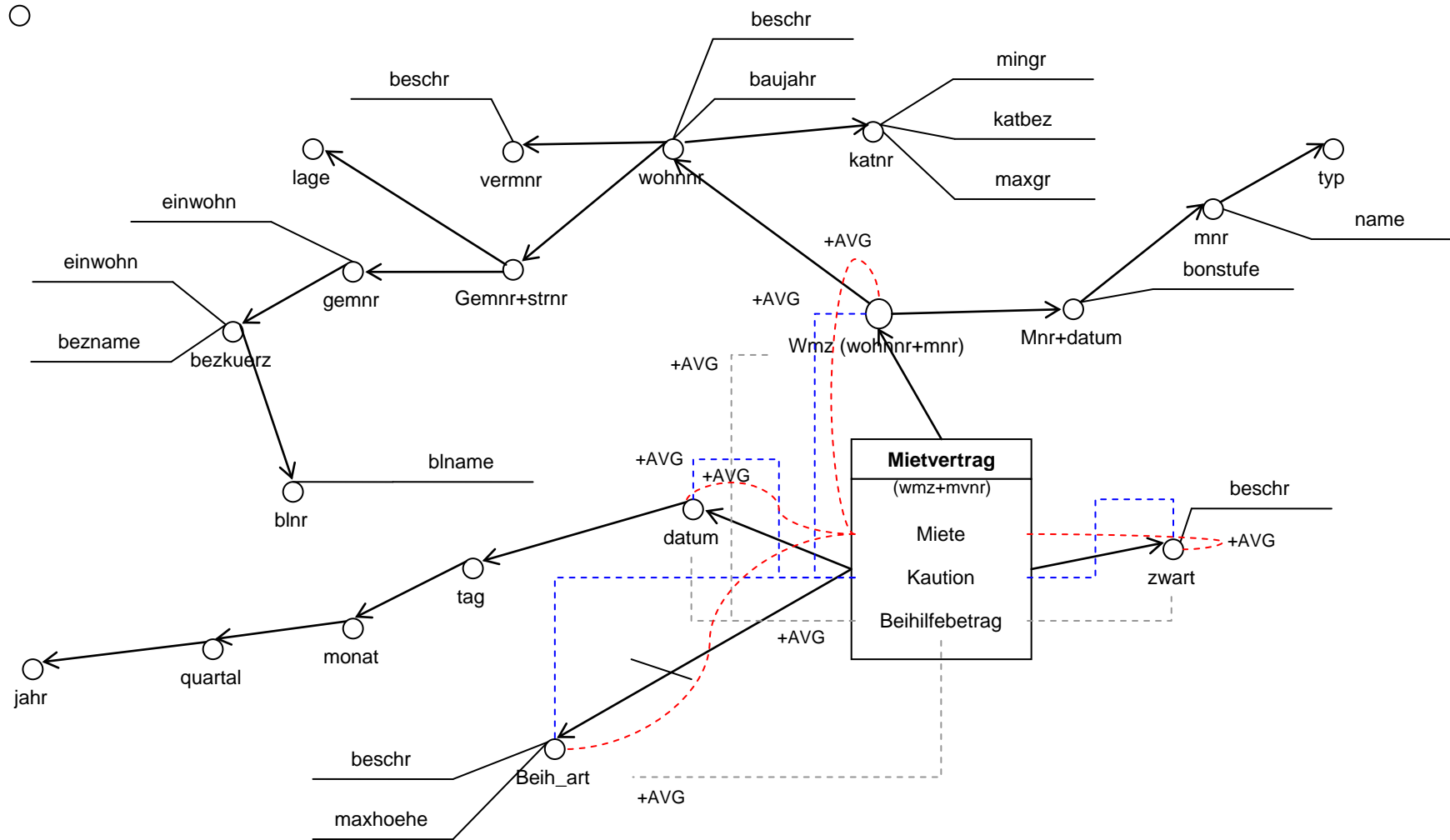
Finales DFM der Faktklasse Wohnung



Anmerkungen zum finalen DMF der Faktklasse Wohnung:

- Es können sowohl die Instandhaltungskosten, als auch die Wohnungsgrößen analysiert werden. Die Größe kann anhand der Dimensionen Kategorie, Gemeinde+Str und Baujahr aggregiert werden. Allerdings ist die Durchschnittsbildung die einzige erlaubte Aggregationsfunktion. So kann man zum Beispiel analysieren, ob die Wohnungen die 1985 gebaut wurden im Schnitt kleiner sind als jene 2005. Des Weiteren kann man so Änderungen in den Wohnungskategorien nachvollziehen, z.B. ob sich die durchschnittliche Größe von Wohnungen der Kategorie „klein“ verändert hat.
- Für die Analyse der Instandhaltungskosten stehen ebenfalls die oben genannten Dimensionen zur Verfügung. Die Genossenschaft kann zum Beispiel die durchschnittlichen Instandhaltungskosten in Oberösterreich im 4. Quartal im Jahr 2005 analysieren.

Finales DFM der Faktklasse Mietvertrag



Anmerkungen zum finalen DMF der Faktklasse Mietvertrag:

- Es können Mieten, Kautionen und auszuzahlende Beträge von Beihilfen analysiert werden.
- Die Zahlungsart ist nur für die Analyse der Miete relevant. Daher können die anderen Kennzahlen über diese Dimension nicht aggregiert werden. Ähnliches ist es bei der Dimension Beihilfen-Art, hier kann nur die Kennzahl Beihilfebetrag aggregiert werden.
- Es wurde eine Zeitdimension eingeführt um temporale Analysen zu ermöglichen.
- Die Bonitätsstufe wurde als nicht-dimensionales Attribut modelliert, mit der Begründung, dass eine Aggregation für die Wohnbau-Genossenschaft aus der Angabe in diesem Fall nicht sinnvoll erscheint.
- Die WohnungsMietZuordnung wurde als Dimension beibehalten, um die eine Miethistorie eines Mieters in einer Wohnung analysieren zu können.