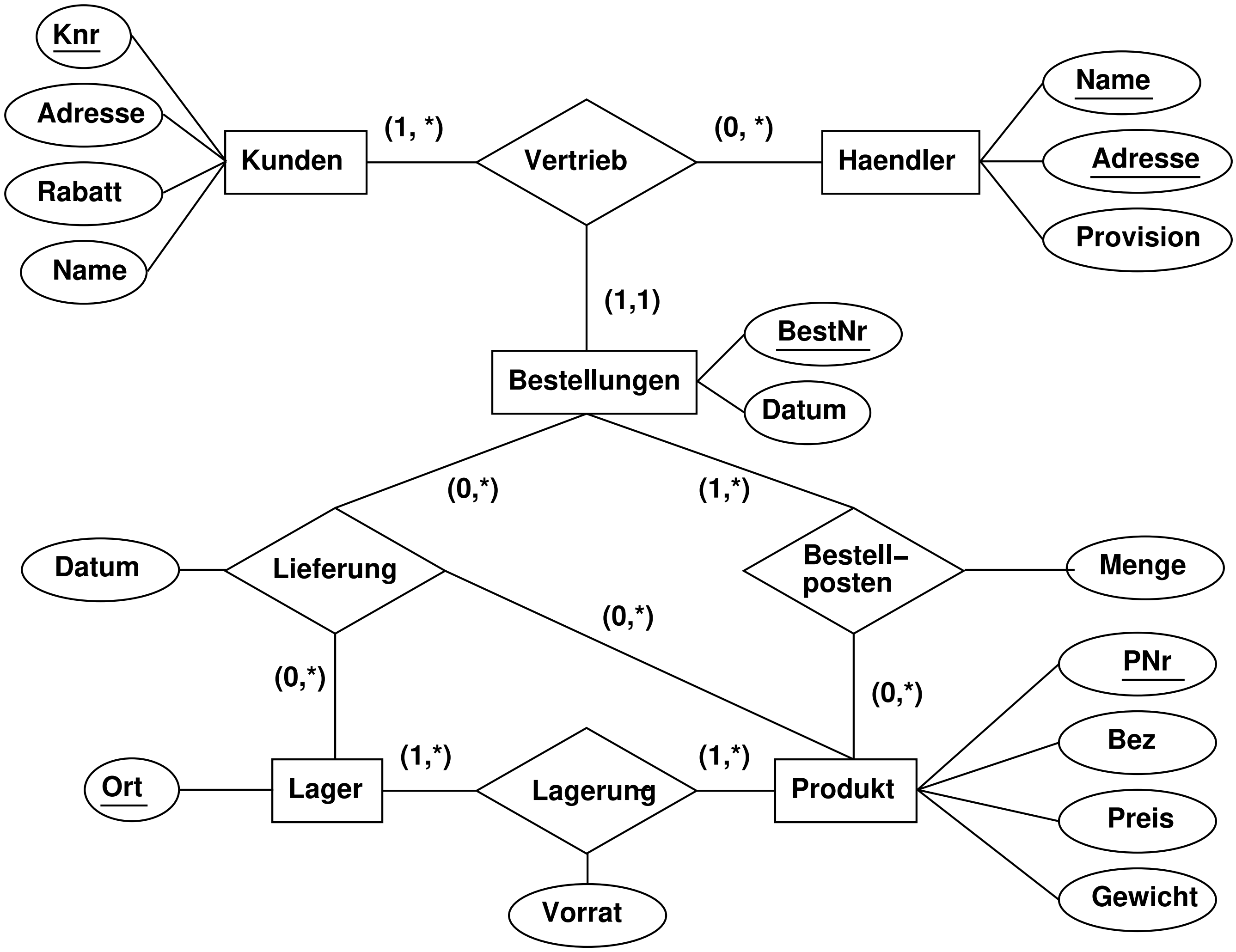
UNIVERSITÄT DES SAARLANDES
Fachbereich 14 – Informatik
Lehrstuhl Prof. Weikum

Datenbanksysteme WS 1999/2000

Übung 13 Abgabe: 14.2.2000

Aufgabe 1: Transformation ERM nach Relationenmodell

Transformieren Sie das nachfolgende ERM–Diagramm einer Vertriebskette in ein relationales Schema.



Aufgabe 2: Modellierung mit ERM

Betrachten wir die folgenden Fakten in einer Firma:

Jeder Angestellte hat einen Nachnamen und ist eindeutig identifizierbar durch eine Personalnummer. Jeder Angestellte hat einen oder mehrere Arbeitsplätze, die er alleine benutzt. Jeder Arbeitsplatz befindet sich in einem Büro und hat eine Briefkastennummer. Büro und Briefkastennummer identifizeren einen Arbeitsplatz.

Jedes Büro hat eine Zimmernummer. Jede Etage hat eine eindeutige Nummer. Jedes Büro kann firmenweit eindeutig identifiziert werden durch die Etage und die Zimmernummer.

Es sind verschiedene Aufgaben vorhanden. Jeder Angestellte kann mehrere Aufgaben zugeteilt bekommen (mindestens aber eine) und eine Aufgabe kann von mehreren Angestellten erledigt werden. Aufgaben sind eindeutig identifizierbar durch Aufgabennamen.

Jeder Angestellte bekommt eine Summe Geld als Gehalt und muß einer Gehaltsgruppe angehören. Die Gehaltsgruppen sind durch einen Code identifizierbar, wobei eine Gehaltsgruppe angibt, in welchem *Bereich* das Gehalt liegt. Die Höhe des Gehalts muß zunächst nicht festgelegt sein.

Jede Abteilung ist eindeutig identifizierbar durch den Abteilungsnamen. Jede Abteilung wird durch genau einen Angestellten geleitet und kann mehrere Angehörige haben. Ein Angestellter muß in genau einer Abteilung entweder Manager oder einfacher Angehöriger sein.

a) Modellieren Sie diese Informationszusammenhänge in Form eines ERM–Diagramms. Geben Sie insbesondere die Kardinalitätsbedingungen für Relationships an.

b) Transformieren Sie Ihren ERM–Entwurf in das Relationenmodell.

Aufgabe 3: Ausbeute (Recall) und Präzision (Precision)

a) Es seien $C_1 > 0$ die Kosten, welche entstehen, wenn ein Benutzer ein vom Suchsystem gefundenes Dokument inspiziert, das irrelevant ist. Außerdem seien $C_2 > 0$ die Kosten, die dadurch entstehen, wenn das Suchsystem ein relevantes Dokument nicht unter den ersten r Dokumenten rangiert, die der Benutzer inspiziert. Ermitteln Sie, ob in folgenden Situationen gilt $C_1 \ll C_2$, $C_1 \approx C_2$ oder $C_1 \gg C_2$.

- Ein Patentanwalt überprüft, ob eine Erfindung bereits patentiert ist.
- Ein Journalist möchte herausfinden, wann die Swissair das erste Mal über den Atlantik flog.

b) Eine Dokumentenkollektion enthalte $n = 1$ Mio. Dokumente, wovon der Benutzer $r = 10$ nach einer Anfrage inspiziert. Dabei wird eine Ausbeute von 20% und eine Präzision von 50% erzielt. Der „Fallout“einer Anfrage ist definiert als das Verhältnis

$$\frac{\text{Anzahl inspizierter irrelevanter Dokumente}}{\text{Anzahl aller irrelevanten Dokumente}}.$$

- Wie hoch ist der Fallout?
- Wie verhält sich der Fallout für $r \to n$?

Aufgabe 4: Vektorraum-Modell

Gegeben seien die Features (bzw. Terme) $F = \{a, b, c, d, e\}$ und die Dokumente

$$\begin{array}{ll} d_1 : a\ b\ b\ e\ a\ a & d_2 : c\ b\ b\ c\ a\ c\ d \\ d_3 : b\ b\ b\ b\ e\ b & d_4 : c\ b\ e \\ d_5 : b\ a\ c\ b & d_6 : e\ e\ e\ c\ e \end{array}$$

.

Berechnen Sie die ff*idf-Gewichte der Indexfeatures sowie die Ergebnisse der Anfragen unter Verwendung des Cosinusmaßes und der ff*idf-Gewichtung.

$$\begin{array}{ll} q_1 : d & q_2 : c \\ q_3 : c\ d & q_4 : a\ b\ e \\ q_5 : a\ b\ c\ d\ e & \end{array}$$