



Übung zur Vorlesung *Grundlagen: Datenbanken* im WS20/21
Christoph Anneser, Josef Schmeißer, Moritz Sichert, Lukas Vogel (gdb@in.tum.de)
<https://db.in.tum.de/teaching/ws2021/grundlagen/>

Blatt Nr. Z1

In den folgenden Zusatzblättern wird der gesamte Prozess von der Modellierung eines Datenbankschemas, der Extraktion und Bereinigung der Daten, sowie der Erstellung eines geeigneten Datenbankschemas mittels SQL-DDL Statements und dem anschließenden Import der Daten in die Datenbank anhand des Beispiels der Fußball-Weltmeisterschaften der Jahre 1930 bis einschließlich 2014 erklärt.

Alle in diesem Blatt behandelten datenbankspezifischen Aspekte sind selbstverständlicherweise auch *klausurrelevant*.

Hausaufgabe 1

Zusammen mit diesem Blatt erhalten Sie ein Archiv mit drei CSV-Dateien auf der Moodle-seite des Kurses, welches verschiedene Daten zu den Fußball-Weltmeisterschaften der Jahre 1930 bis einschließlich 2014 enthält.

Machen Sie sich nun im ersten Schritt mit den verfügbaren Daten vertraut. Fokussieren Sie sich hierbei auf die jeweils erste Zeile der Dateien, die teils deskriptive Namen der jeweiligen Spalten enthält. Erklärungen/Legenden zu den jeweiligen Spalten finden in der Datei *headers.txt*.

Hinweis: Die Daten enthalten zum Teil Duplikate und nicht bekannte Symbole. Wenn Sie beispielsweise unbekannte Symbole ersetzen oder die Spieler um Vornamen ergänzen, lassen Sie uns diese bitte zukommen und wir werden Ihre Änderungen in den Datensatz integrieren.

Hausaufgabe 2

In dieser Teilaufgabe möchten wir mit der Modellierung beginnen und verwenden hierfür ein Entity-Relationship-Modell.

Beachten Sie bei der Erstellung des Modells folgende Anforderungen:

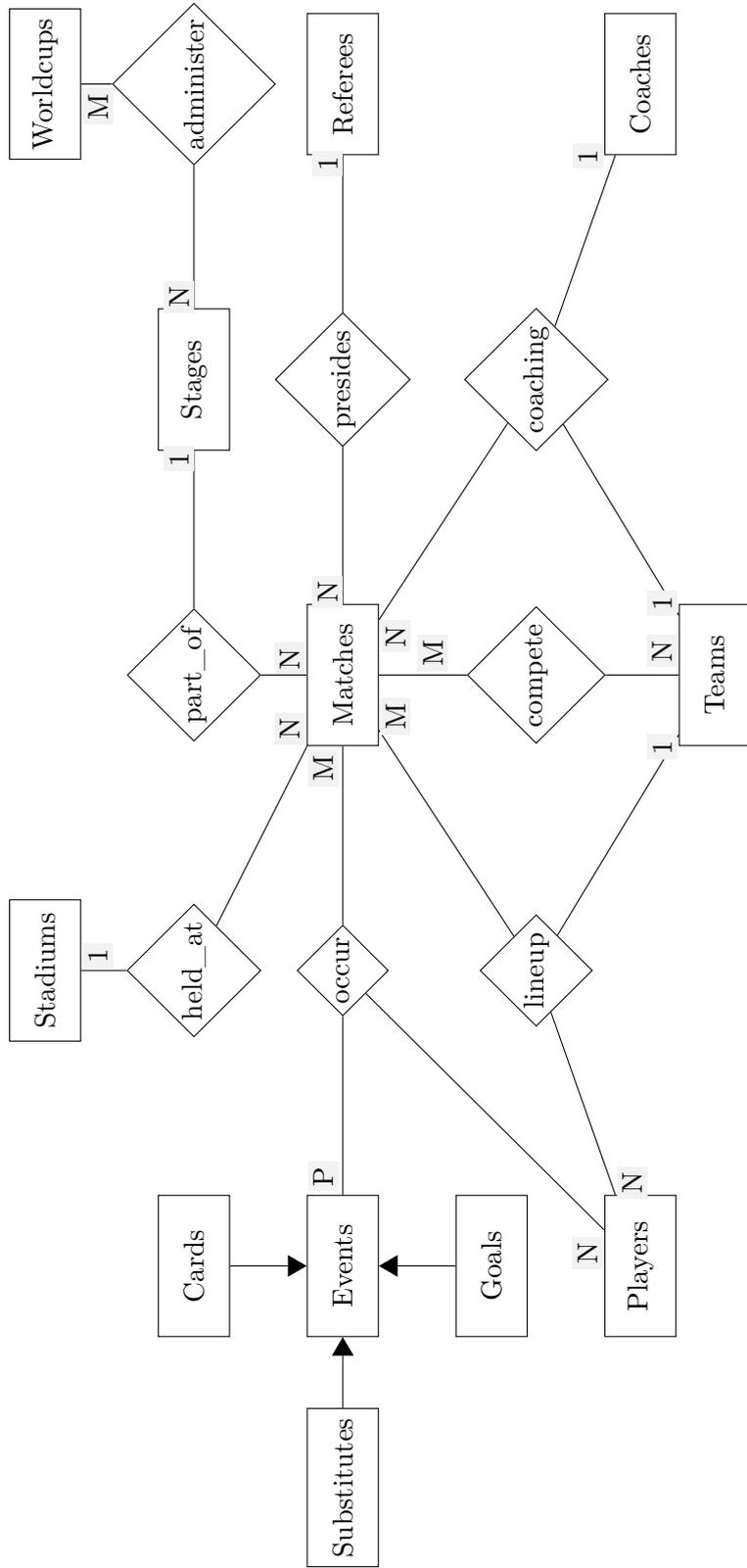
1. Spiele werden von Schiedsrichtern geleitet, die von jeweils zwei Assistenten unterstützt werden.
2. Schiedsrichter gehören anderen Nationen an als die sich in den jeweils geleiteten Spielen gegenüberstehenden Mannschaften.
3. In jedem Spiel treten zwei Mannschaften gegeneinander an - wobei zwischen Heim- und Gastmannschaft differenziert wird.
4. Jede Mannschaft hat in den von ihr bestrittenen Spielen einen festen Kader, bei dem genau elf Spieler in der Startaufstellung gesetzt sind: 10 Feldspieler und ein Torhüter.
5. Jede Mannschaft wird in den von ihr bestrittenen Spielen von genau einem Coach an der Seitenlinie gemanaged.

6. Jedes Spiel gehört zu einer bestimmten Turnierphase ({Gruppe A, ..., Gruppe F, Achtelfinale, ..., Finale})
7. Jede Turnierphase ist eindeutig einer Weltmeisterschaft zugeordnet.
8. Weltmeisterschaften finden in einem bestimmten Jahr und in einem bestimmten Gastland statt.
9. In einem Spiel treten Ereignisse wie beispielsweise Tore, Ein- und Auswechslungen sowie Sanktionen auf. Jedes dieser Ereignisse tritt in einer Spielminute auf und ist einem bestimmten Spieler zugeordnet.

Erstellen Sie anhand der oben genannten Anforderungen ein ER-Modell. Lassen Sie hierfür zunächst die jeweiligen Attribute aus Übersichtsgründen weg und konzentrieren Sie sich auf die korrekte Modellierung der Entitätstypen und Beziehungen.

Ergänzen Sie ihr Modell in einem weiteren Schritt um geeignete Funktionalitätsangaben.

Tipp: In diesem Sachverhalt kann es je nach Modellierung zu *einigen* ternären Beziehungen kommen. Manche oben genannten Anforderungen können nicht ohne Weiteres modelliert und daher vernachlässigt werden. Nutzen Sie Aggregationen für Entitätstypen, die ähnliche Eigenschaften teilen.



Nicht alle geforderten Bedingungen lassen sich gleichermaßen im Modell ausdrücken.

Beispielsweise können wir die Anforderungen 2 und 4 nicht ohne Weiteres auf direktem Weg umsetzen. Für Anforderungen dieser Art ist die Modellierung mittels ER Diagramm auch nicht geeignet – jedoch können diese Bedingungen später in der Datenbank durch **check-constraints** umgesetzt werden.

Anforderung 1 wurde hier ebenfalls nicht vollständig modelliert. Eine Möglichkeit wäre beispielsweise einen neuen Entitätstypen *RefereeGroup* bestehend aus einem Schiedsrichter und zwei Assistenten zu erstellen. Eine Alternative hierzu wäre die Modellierung mittels einer weiteren Relationship *assists*. Auch dies haben wir aus Gründen der Übersichtlichkeit vereinfacht bzw. weggelassen.

Die Anforderung 3 kann durch die Hinzunahme eines weiteren Attributs erfüllt werden, welches das jeweilige *home*-Team bestimmt. Jedoch kann auch hier die Modellierung mittels Funktionalitätsangaben nicht direkt ausdrücken, dass *genau zwei* Teams in einem Spiel aufeinandertreffen. Hierfür eignen sich (min, max)-Angaben. Später kann dies auch in der Datenbank durch Definition weiterer *check-constraints* festgelegt werden.

Für die in der letzten Anforderung 9 genannten Ein- und Auswechslungen wäre es selbstverständlich ideal, beide Informationen in einem Entitätstypen zu persistieren. Ein Blick in die Rohdaten lässt aber erkennen, dass insbesondere bei älteren Weltmeisterschaften konkretere Angaben zu paarweisen Ein- und Auswechslungen fehlen. Daher entscheiden wir uns hier für einen generischen Entitätstypen *Substitutes*, der unter Zuhilfenahme eines weiteren Attributs den Typ (Ein- oder Auswechslung) genau spezifiziert. Da nun die Entitätstypen *Cards*, *Goals* und *Substitutes* alle *genau* einem Spieler in *genau* einem Spiel zugeordnet sind, können wir hier eine Aggregation mittels des Entitätstypen *Events* vornehmen.

Des Weiteren haben wir in der Lösung für eine bessere Übersichtlichkeit die Attribute nicht modelliert.

Solche Vereinfachungen wie in diesem Beispiel treten in der Praxis häufig auf, um unnötige Komplexität in der frühen Entwicklung des Datenbankschemas zu vermeiden.

In der Klausur dürfen Vereinfachungen wie hier gezeigt jedoch **nicht ohne expliziten Hinweis** vorgenommen werden.