

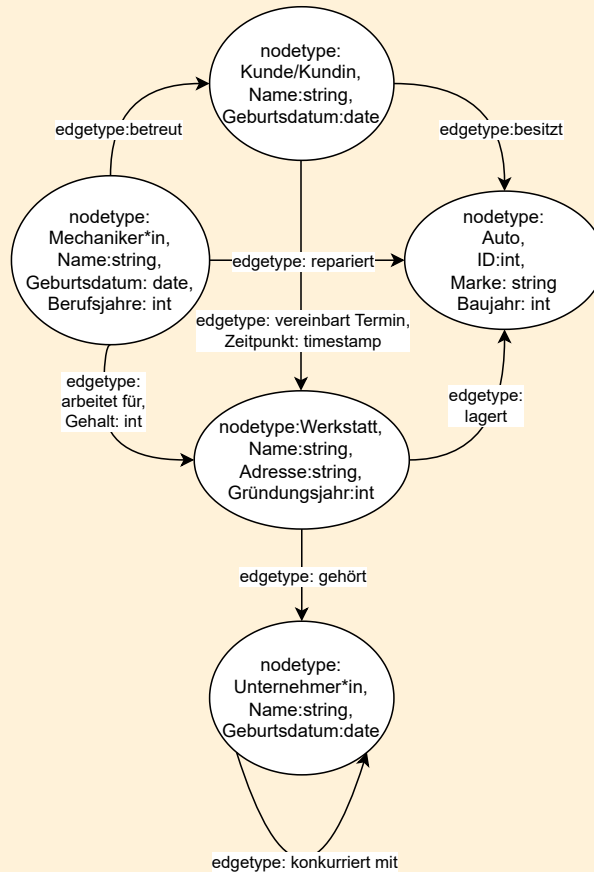
1 Graphmodellierung (5 Punkte)

Die folgende Spezifikation modelliert eine Werkstatt. Erstellen Sie hierzu das entsprechende Graphmodell. Orientieren Sie sich dabei am rechten Beispielgraph auf Folie 27 aus der Vorlesung Datenjournalismus (Teil 1).

- Kunden/Kundinnen und Unternehmer*innen haben einen eindeutigen Namen sowie ein Geburtsdatum.
- Mechaniker*innen haben einen eindeutigen Namen, ein Geburtsdatum sowie Berufsjahre.
- Eine Werkstatt hat einen eindeutigen Namen, eine Adresse sowie ein Gründungsjahr.
- Autos haben eine eindeutige ID, eine Marke sowie ein Baujahr.
- Kunden/Kundinnen besitzen Autos.
- Mechaniker*innen betreuen Kunden/Kundinnen und reparieren Autos.
- Mechaniker*innen arbeiten für Werkstätten und verdienen dabei ein gewisses Gehalt.
- Kunden/Kundinnen vereinbaren Termine mit Werkstätten an gewissen Zeitpunkten
- Werkstätten lagern Autos und gehören einer/einem Unternehmer*in.
- Unternehmer*innen konkurrieren miteinander.

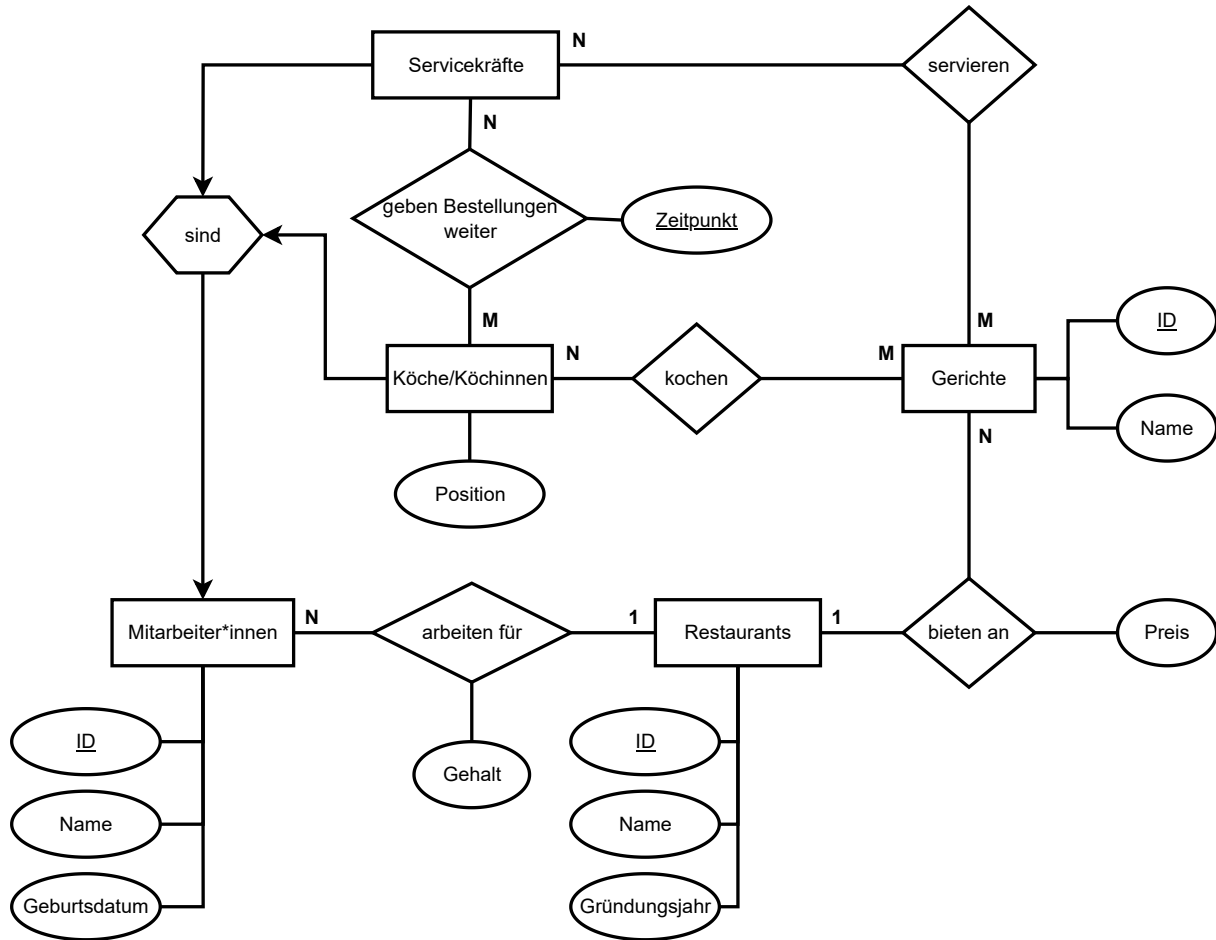
Lösung:

Vorgesehen sind 0,5 Punkte pro Stichpunkt der Spezifikation (keine Punkte bei fehlenden Attributen, unsinnigen Typen, usw.). Dabei ist es dem Studierenden überlassen, ob die Knoten (und damit die Kanten) im Singular oder Plural gehalten werden, es sollte lediglich konsistent gehandhabt werden. Die Verwendung von Schlüsseln ist ebenfalls freiwillig, aber konsistent anzuwenden.



2 ER-Modell zu Graphmodell (3 Punkte)

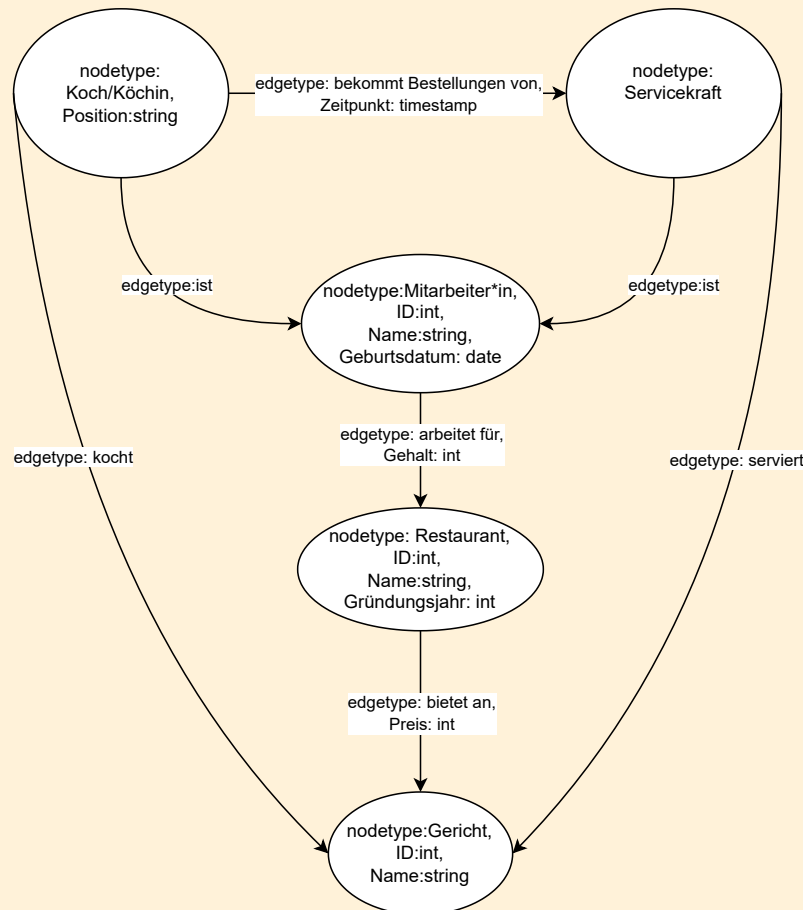
Betrachten Sie das folgende ER-Modell.



- Überführen Sie das ER-Modell in ein Graphmodell.
- Welche Schwierigkeiten treten bei der Umsetzung auf? Welche Informationen aus dem ER-Modell können im Graphmodell nicht umgesetzt werden?

Lösung:

- (a) Vorgesehen sind 2 Punkte (je 0,5 Punkte Abzug für fehlende Knoten/Kanten, fehlende Attribute, unsinnige Typen, usw.). Dabei ist es dem Studierenden überlassen, ob die Knoten (und damit die Kanten) im Singular oder Plural gehalten werden, es sollte lediglich konsistent gehandhabt werden. Die Verwendung von Schlüsseln ist ebenfalls freiwillig, aber konsistent anzuwenden. Ebenso sind auch andere, sinnvolle Modellierungen der Vererbungen erlaubt.

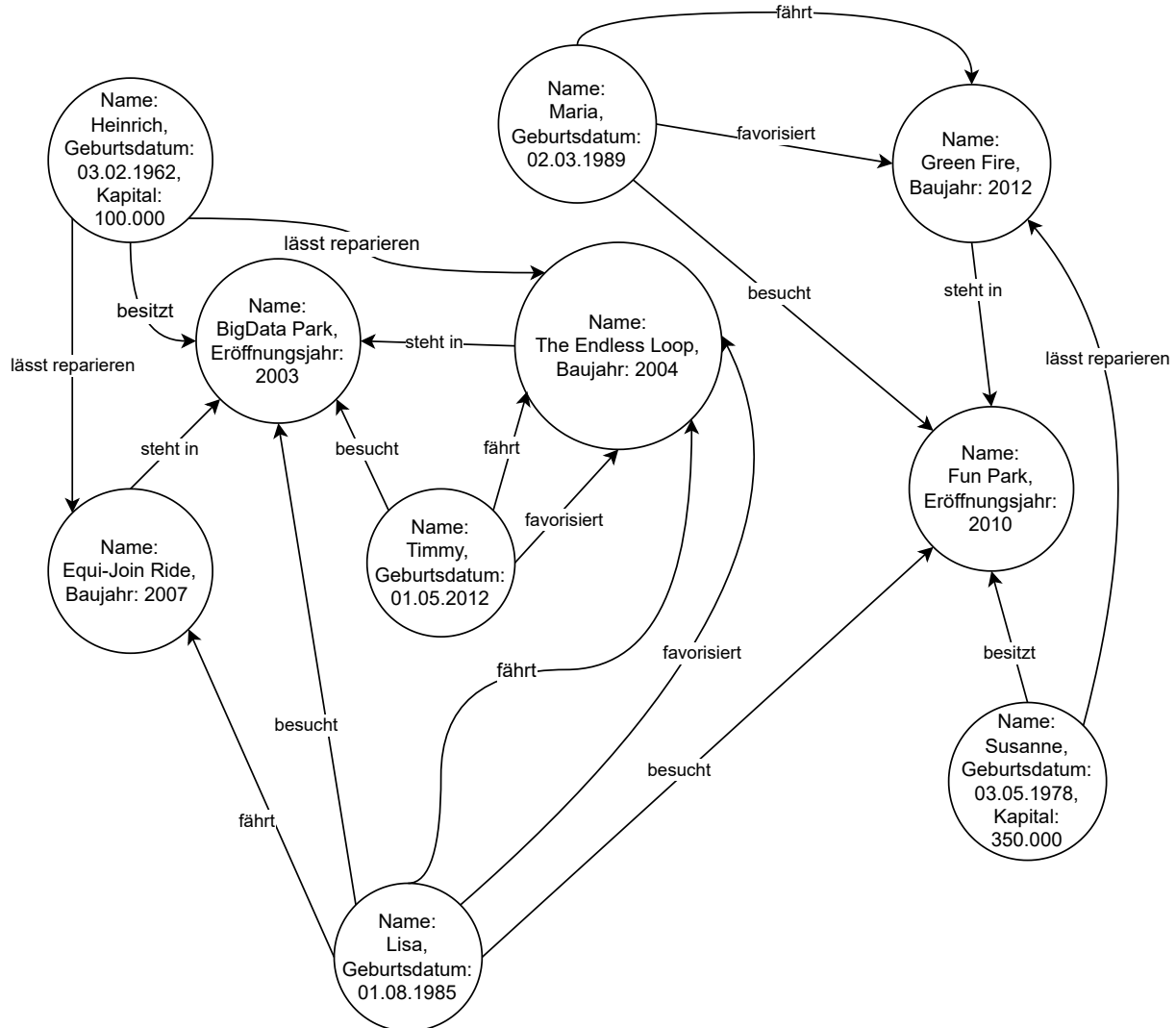


- (b) Vorgesehen ist insgesamt ein Punkt, davon 0,5 Punkte pro Schwierigkeit, d.h. es müssen nur 2 der 3 Schwierigkeiten erkannt werden.

- Funktionalitäten von Beziehungen können nicht dargestellt werden.
- Generalisierung kann nicht durch einen speziellen Beziehungstyp dargestellt werden, sondern kann nur simuliert werden.
- Im ER-Modell sind Beziehungen bzw. Kanten zwischen Beziehungstypen und Entitätstypen ungerichtet.

3 Graphmodell zu ER-Modell (4 Punkte)

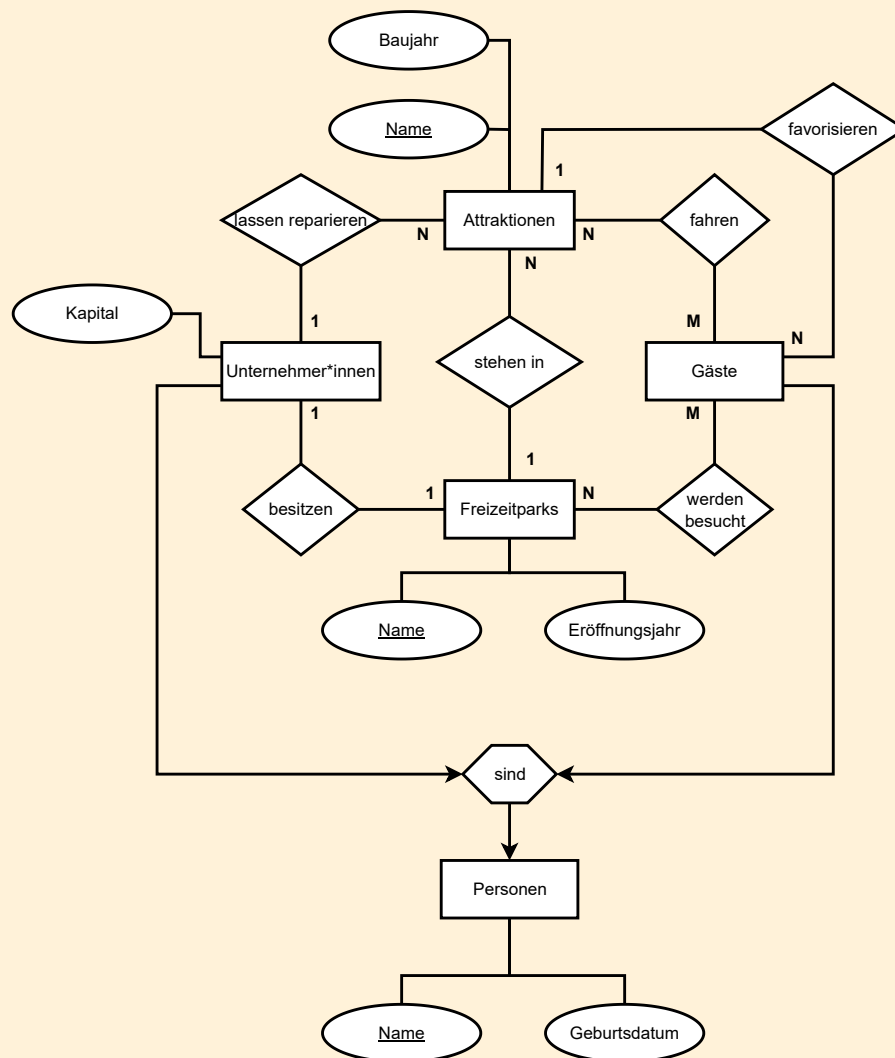
Betrachten Sie folgenden Graphen (Instanz).



Extrahieren Sie aus diesem Graphen ein ER-Modell. Gehen Sie dabei davon aus, dass das jeweils erste Attribut eines Knotens eindeutig ist. Verwenden Sie die Chen-Notation und wählen Sie Funktionalitäten so, dass sie genau den Daten der Instanz entsprechen. Beispielsweise sollten Sie den Beziehungstyp für die Kante „steht in“ als 1:N-Beziehung modellieren. Leiten Sie zudem Vererbungen ab, sofern dies möglich ist.

Lösung:

Vorgesehen sind 4 Punkte (je 0,5 Punkte Abzug für fehlende Entitäten/Beziehungen, fehlende Attribute, falsche Schlüsselattribute, falsche Funktionalitäten, fehlende Vererbung, usw.).



4 Cypher (8 Punkte)

Betrachten Sie die folgenden Anfragen in natürlicher Sprache, die sich auf eine Teilmenge des IMDb Datensatzes beziehen. Der zugehörige Graph besteht dabei aus den folgenden Knoten- und Beziehungstypen:

- ACTOR-Knoten mit den Attributen `node_id`, `first_name`, `last_name` und `gender`.
- DIRECTOR-Knoten mit den Attributen `node_id`, `first_name` und `last_name`.
- MOVIE-Knoten mit den Attributen `node_id`, `name`, `year` und `rank`.
- GENRE-Knoten mit den Attributen `node_id` und `name`.
- PLAYED_IN-Beziehung zwischen MOVIE und ACTOR.
- DIRECTED-Beziehung zwischen DIRECTOR und MOVIE.
- BELONGS_TO_GENRE-Beziehung zwischen MOVIE und GENRE.

Übersetzen Sie diese Anfragen im beigefügten Notebook in passende Cypher Anfragen. Beachten Sie auch die Hinweise am Anfang der Aufgabe im Notebook. Bei Fragen bezüglich Cypher können Sie auf das Notebook Graphs in Cypher.ipynb sowie auf die offizielle Cypher Dokumentation zurückgreifen.

Sie können Ihre Anfragen durch Einfügen an die vorgesehene Stelle im Notebook testen. Reichen Sie Ihre Abgabe für diese Aufgabe bitte als `jupyter.txt`, die alle vier Anfragen klar voneinander getrennt enthält, ein.

1. Berechnen Sie die Anzahl unterschiedlicher “Sci-Fi” Filme, bei denen Arnold Schwarzenegger mitgespielt hat. Benennen Sie die Ausgabe in `numMovies` um.
2. Bestimmen sie die Anzahl unterschiedlicher Pfade mit der minimalen Länge von 3 und einer maximalen Länge von 6 zwischen dem Regisseur Quentin Tarantino und dem Schauspieler Tim Roth. Benennen Sie Ihre Ausgabe in `pathCount` um.
3. Berechnen Sie den maximalen Rang der Filme die nach 1990 erschienen sind und mindestens 3 Genren angehören. Benennen Sie die Ausgabe in `maxRank` um.
4. Ermitteln Sie die Länge des kürzesten Pfades zwischen einem der ältesten und einem der neuesten Filme. Beachten Sie, dass es sich hierbei um unterschiedliche Filme handeln muss. Benennen Sie die Ausgabe in `shortestPath` um.

Lösung:

Vorgesehen sind 0,5 Punkte pro Klausel (MATCH, WHERE, RETURN, WITH), d.h. jeweils 1,5 Punkte für die ersten beiden Anfragen und jeweils 2,5 Punkte für die beiden letzten.

1.

```
MATCH (a:ACTOR) -- (m:MOVIE) -- (g:GENRE)
WHERE a.first_name = 'Arnold'
      AND a.last_name = 'Schwarzenegger' AND g.name = 'Sci-Fi'
RETURN COUNT(DISTINCT m) AS numMovies;
```

2.

```
MATCH p = (d:DIRECTOR) -[*3..6]- (a:ACTOR)
WHERE d.first_name = 'Quentin' AND d.last_name = 'Tarantino'
      AND a.first_name = 'Tim' AND a.last_name = 'Roth'
RETURN COUNT(DISTINCT p) AS pathCount;
```

3.

```
MATCH (m:MOVIE) -- (g:GENRE)
WHERE m.year > 1990
WITH m, COUNT(DISTINCT g) AS numGenres
WHERE numGenres >= 3
RETURN MAX(m.rank) AS maxRank;
```

4.

```
MATCH (m:MOVIE)
WITH MIN(m.year) AS oldest, MAX(m.year) AS newest
MATCH (m1:MOVIE), (m2:MOVIE)
WHERE m1.year = oldest AND m2.year = newest
MATCH p = SHORTESTPATH((m1) -[*]- (m2))
RETURN MIN(LENGTH(p)) AS shortestPath
```


Abgabe

Lösungen sind in Teams von 2 bis 3 Studierenden bis zum 07. Juli 2022, 10:15 Uhr über Ihre persönlichen Statusseite im CMS einzureichen. Nutzen Sie hierfür die Team Groupings Funktionalität im CMS.

Ihre Abgabe muss dem folgenden Format entsprechen:

```
abgabe.zip
├── abgabe.pdf
└── jupyter.txt
```

Hierbei enthält `abgabe.pdf` Ihre Lösungen zu Aufgabe 1, 2 und 3 und `jupyter.txt` Ihre Lösung zu Aufgabe 4. Achten Sie darauf, dass Sie nur die von Ihnen zu ergänzenden Jupyter Zellen so kopieren, dass Einrückung und Formatierung korrekt sind.

Abgaben, die nicht den oben angegebenen Vorgaben entsprechen, führen zu Punktabzug. Einzelabgaben werden nicht mehr korrigiert und mit 0 Punkten bewertet.