

## Aufgabe 1: Fragen zu Informationssuche & Data Mining (1 P.)

Um diese Aufgabe als gelöst zu markieren, müssen Sie 7 aus 9 Teilaufgaben lösen.

1. Widerlegen oder begründen Sie folgende Aussage: Die längste gemeinsame Teilsequenz zweier Worte ist mindestens so lang wie das kleinere Wort.
2. Geben Sie zwei Szenarien an, in denen die normalisierten Termfrequenzen  $rtf_{t,d}$  und  $ntf_{t,d}$  gegenüber  $tf_{t,d}$  keine Vorteile haben.
3. Wie muss die Ergebnisliste der Länge 20 bzgl. relevanter bzw. irrelevanter Dokumente aussehen, damit die Average-Precision gleich dem Wert für Precision@20 ist.
4. Ist folgende Matrix zeilenstochastisch oder spaltenstochastisch oder weder noch?

$$\begin{pmatrix} 0.1 & 0.5 & 0.4 \\ 0.4 & 0.3 & 0.3 \\ 0.2 & 0.2 & 0.4 \end{pmatrix}$$

5. Bei der Erstellung der Page-Rank-Matrix  $P$  müssen Knoten (Dokumente), die keine ausgehenden Links (sogenannte “dangling nodes”) haben besonders gehandhabt werden. In welchem Teil der Definition von  $P$  passiert dies und was geschieht, wenn man dies nicht tut?
6. Begründen Sie, dass für die in der Vorlesung verwendete Definition der PageRank-Matrix  $P$  der gesuchte PageRank-Vektor ein linker Eigenvektor ist und nicht ein rechter. Beziehen Sie sich dabei direkt auf die ursprüngliche PageRank Definition  $PR(p) = (1 - \varepsilon) \times \sum_{q \rightarrow p} \frac{PR(q)}{out(q)} + \varepsilon \times \frac{1}{N}$
7. Für ein häufiges Itemset  $Z$ , wieso kann man alle Regeln  $W \rightarrow Z \setminus W$  ignorieren, wenn man weiß, dass  $X \rightarrow Z \setminus W$  nicht confident ist, für  $W \subseteq X$ ?
8. Geben Sie ein Beispiel an und begründen Sie, wieso das Support-Maß nicht monoton ist.
9. Wieso gibt es für  $k$  Cluster und  $n$  Objekte  $k^n$  Möglichkeiten die  $k$  Cluster (darunter auch leere) zu erzeugen?

## Aufgabe 2: ER-Modellierung (1 P.)

- (a) Ein Spielehersteller will ein neues Multiplayer-Rollenspiel entwickeln und sendet Ihnen für den Entwurf des Datenschemas den folgenden Text für den Sie ein ER-Modell erstellen sollen:

Wir haben Spieler, welche sich mit Nickname, Emailadresse und Passwort registrieren. Die Nicknames können sich beliebig ändern, aber Emailadressen müssen eindeutig sein und können sich nicht ändern. Jeder Spieler kann bis zu zehn Charaktere erstellen, welche auch wieder gelöscht werden, sobald ein Spieler seinen Account löscht. Diese Charaktere haben einen Namen und ein Level. Das Spiel soll regelmäßig und automatisch den Spielstand speichern. Um einen Charakter wieder zu laden, sind x,y,z-Koordinaten und die Blickrichtung nötig, außerdem Erfahrungs- und Lebenspunkte. Jeder Charakter verfügt über ein unbegrenzt großes Inventar, in dem Gegenstände abgelegt werden können. Die Gegenstände haben einen Namen, einen Rüstungs- und einen Kampfwert. Im Spielverlauf soll es die Möglichkeit geben, diese Werte zu verändern (zum Beispiel als Belohnung für das Lösen von Quests). Wir wollen uns außerdem merken, wie lange am Stück ein Gegenstand im Inventar eines Spielers ist.

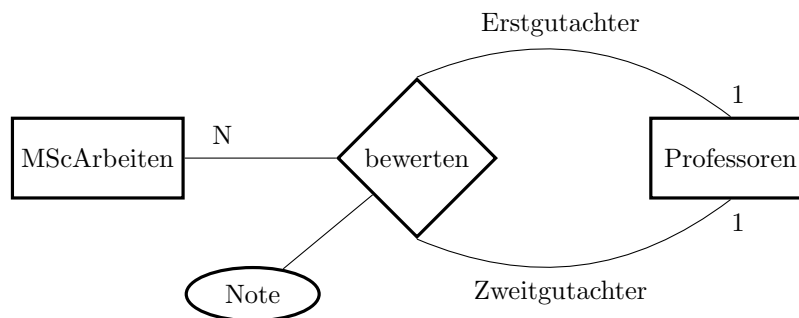
- (b) Erstellen Sie für folgende Beispieldaten ein mögliches ER-Diagramm inkl. Angabe der Kardinalitäten nach Chen, wo dies aus den Daten ersichtlich ist. Diskutieren Sie ferner kurz ob die Zusammenfassung all dieser Spalten in einer Tabelle sinnvoll ist.

Kundennummer	Datum	Artikelnummer	Anzahl	Rechnungsnummer
1	02.05.2016	1711	4	1000
2	17.12.2015	1711	1	780
1	02.05.2016	3014	3	1002
2	04.01.2016	9012	9	800
1	17.12.2015	4287	3	780

### Aufgabe 3: ER-Modellierung: Grundlagen

(1 P.)

- (a) Finden Sie ein Beispiel für eine dreistellige 1:1:1-Beziehung (z.B. aus dem Kontext Universitätsverwaltung). Es sollte eine dreistellige Beziehung sein, die nicht durch (mehrere) zweistellige Beziehungen dargestellt werden kann. Unter welchen Bedingungen ist dies der Fall?
- (b) Die Beziehung zwischen einem schwachen Entitytyp und dem starken Entitytyp kann keine N:M-Beziehung sein. Warum?
- (c) Gegeben folgendes ER-Diagramm der dreistelligen Beziehung “bewerten”:



Man kann die Beziehung “bewerten” als partielle Funktion wie folgt auffassen:

$$\text{bewerten} : \text{MScArbeiten} \times \text{Erstgutachter} \rightarrow \text{Zweitgutachter}$$

$$\text{bewerten} : \text{MScArbeiten} \times \text{Zweitgutachter} \rightarrow \text{Erstgutachter}$$

Diskutieren Sie, ob man diese Beziehung auch durch (mehrere) zweistellige Beziehungen modellieren kann.

Achtung: Für dieses und die folgenden Aufgabenblätter ist die Abgabedeadline Montags um 15:00 Uhr.