



**PETER L.
REICHERTZ INSTITUT
FÜR MEDIZINISCHE
INFORMATIK**



Ringvorlesung Medizinische Informatik

Data Mining in Data Warehouses

Dr. rer. nat., Dipl.-Inform. Matthias Gietzelt
Universitätsklinikum Heidelberg

Vorlesung: 28.11.2017, 13:15 – 14:45
Nachbesprechung: 14.12.2017, 08:45 – 09:30
Ort: IZ 404

Data Warehouses zeichnen sich durch spezielle Konzepte in der Art der Datenhaltung in Datenbanken aus. Dabei werden die Daten meist in einem sogenannten Sternschema gespeichert. Das Sternschema gibt vor, in welcher Form die festgelegten Tabellen und deren Schlüssel miteinander verknüpft sein müssen, um Daten wiederauffindbar zu speichern. Die zu speichernden Daten werden unstrukturiert in einer sogenannten Faktentabelle abgelegt, wobei jeweils ein Fakt (ein einzelnes Datum) zusammen mit Zeitstempeln und erklärenden Schlüsseln als einzelne Zeile gespeichert wird.

Die entstehende Faktentabelle hat somit eine faktenbasierte (vertikale) Form. Methoden des Data Mining andererseits basieren auf einer instanzbasierten (horizontalen) Form der Speicherung. D.h., jeder Fall (Instanz) wird durcheinere Reihe von Attributen beschrieben, deren Ausprägungen eine einzelne Zeile bilden. Um Methoden des Data Mining auf Daten in einem Data Warehouse anwenden zu können, müssten die Daten zunächst in eine instanzbasierte Repräsentation überführt werden. Da dieser Umweg umständlich und zeitintensiv ist, war es unser Ziel, ohne einen solchen Transformationsschritt auszukommen und direkt die Datenstruktur des Data Warehouses für das Data Mining zu nutzen.



Dr. rer. Nat., Dipl.-Inform.
M. Gietzelt