

Clustering

1. Die Dokumente, die für ein und denselben Konzept (z.B. Auto) unterschiedliche Terme benutzen (e.g. Pkw, VW Golf, etc.) werden beim Clustering mit einer hohen Wahrscheinlichkeit ein und demselben Cluster zugeordnet. Erklären Sie warum.

Gibt es Terme, die zwar zu unterschiedlichen Konzepten gehören, deren Dokumente aber aufgrund der oben genannten Eigenschaft einem gemeinsamen Cluster zugeordnet werden können? Nennen Sie Beispiele.

2. Erklären Sie die Gemeinsamkeiten und Unterschiede zwischen Klassifikation und Clustering. Achten Sie insbesondere auf die folgenden Eigenschaften:
 - a) Training des Systems
 - b) Anzahl der Cluster/Klassen
 - c) Erklärbarkeit und Darstellung des Ergebnisses