

Ähnlichkeitssuche und Datenexploration über Entitäts-Ranglisten

[Abstract]

Sebastian Michel
TU Kaiserslautern
Kaiserslautern, Germany
smichel@cs.uni-kl.de

ABSTRACT

Tabellen oder geordnete Listen sind generische und weit verbreitete Mittel, um Informationen über Entitäten in einer prägnanten und strukturierten Form darzustellen. Ranglisten, die etwa die höchsten Gebäude der Welt, die reichsten Personen Deutschlands oder die leistungsstärksten Autos enthalten, sind allgegenwärtige Beispiele. Mit der Verfügbarkeit großer Mengen solcher Ranglisten, die Entitäten kategorisiert nach speziellen Eigenschaften und geordnet anhand unterschiedlichster Metriken in Relation zueinander setzen, können interessante Erkenntnisse hergeleitet werden, indem berechnet wird, welche Ranglisten ähnliche Entitäten ähnlich ordnen, wobei Rangkriterium und Eigenschaften der Entitäten (stark) unterschiedlich sein können. Dieser Berechnung liegt das Problem der Ähnlichkeitssuche zugrunde, für das wir zwei verschiedene – auf die spezielle Ähnlichkeitsfunktion zugeschnittene – Ansätze entwickelt haben. Desweiteren betrachten wir das Problem der Erkennung von Attributen zur sinnvollen Kategorisierung von Mengen von Entitäten und präsentieren ein Verfahren, das vollautomatisch entscheiden kann, ob ein Attribut geeignet oder ungeeignet ist, Entitäten zu kategorisieren. Darüber hinaus geben wir einen Überblick über PALEO und COMPETE, zwei Ansätze zur Ranglisten-basierten Datenexploration.