

Web-based Information Systems

Exercise 7

Felix-Johannes Jendrusch, Robert Böhnke, Simon Philipp Hohberg
Group 4

January 31, 2011

1 Semantic Search Engine

- a. Bei dieser Ontology handelt es sich um eine “Upper Ontology” bzw. “Top-level Ontology”. Sie beschreibt nur grundlegende Konzepte und eine allgemeine hierarchische Ordnung über die Ähnlichkeit der Konzepte.
- b. Sei $G = (V, E)$ der Graph der beschriebenen Ontologie, wobei die Menge der Knoten V durch die Konzepte und die Menge der Kanten E durch ihre Beziehungen gegeben sind und Konzepte mit einer *sameAs* Beziehung durch einen einzigen Knoten repräsentiert werden; und $d : V \times V \rightarrow R_0^+$ eine Funktion, welche die Länge eines kürzesten Pfades in G zwischen zwei Knoten $v_i, v_j \in V$ — zwei Konzepten $c_i, c_j \in O$ — berechnet.

Die Ähnlichkeit zweier Konzepte $c_i, c_j \in O$ lässt sich beispielsweise wie folgt definieren:

$$s(c_i, c_j) = 1/(1 + d(v_i, v_j)).$$

2 RapidMiner

- a. Es wird das “Golf” Dataset analysiert, welches Aussagen darüber macht, wann im Zusammenhang mit der Wettervorhersage Golf gespielt werden soll.

Wenn wir den resultierenden “Decision Tree” von links nach rechts durchgehen, ergeben sich folgende Aussagen:

- Wird bedecktes Wetter vorhergesagt, wird gespielt.
 - Wird Regen und keinen Wind vorhergesagt, wird gespielt.
 - Wird Regen und Wind vorhergesagt, wird nicht gespielt.
 - Wird Sonne und eine Luftfeuchtigkeit vorhergesagt, die größer ist als 77.5, wird nicht gespielt.
 - Wird Sonne und eine Luftfeuchtigkeit vorhergesagt, die kleiner oder gleich 77.5 ist, wird gespielt.
- b. Die vertikalen Linien sind die Ähnlichkeit zwischen den einzelnen Clustern bzw. Items, wobei eine kurze vertikale Linie eine große Ähnlichkeit und eine lange Linie eine geringe Ähnlichkeit beschreibt.

Um mithilfe des Dendograms einzelne Cluster zu definieren, wird im Dendogram einfach in der Höhe der gewünschten Genauigkeit/Ähnlichkeit die dort vorherrschende Clustereinteilung verwendet.