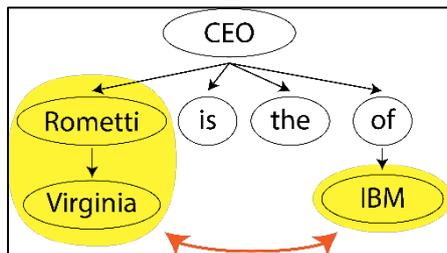


Der Shortest Path Subsequence Kernel

Ein Tree Kernel für ein Relation Extraction System



Student	Norman Süssstrunk
Referent	Prof. Martin Studer
Korreferent	Prof. Albert Weichselbraun
Master Research Unit	Software and Systems, HTW Chur



Annotierte Relation in einem Dependency Tree

In dieser Arbeit wurde ein Relation Extraction System als Komponente für ein umfangreiches NLP- und Information Retrieval System entwickelt. Relation Extraction bezeichnet die Erkennung von semantischen Beziehungen zwischen Named Entities innerhalb von natürlichsprachigem Text. Eine Named Entity beschreibt die textliche Nennung eines realen Objektes. Die aktuelle Forschung auf dem Gebiet des Information Retrievals zeigt, dass die Verwendung von Support Vector Machines für unterschiedliche Klassifikationsprobleme sehr gute Ergebnisse liefert. Eine Support Vector Machine löst ein Klassifikationsproblem auf der Grundlage eines Kernels, welcher die Ähnlichkeit zwischen den zu klassifizierenden Instanzen beschreibt. Ein solcher Kernel kann für den Vergleich von unterschiedlichen Strukturen verwendet werden.

Diese Eigenschaft hat zu der Entwicklung und Erforschung von unterschiedlichen Tree Kernel geführt. Die Relation Extraction kann als Klassifikationsproblem aufgefasst werden, welches mittels einer Support Vector Machine gelöst werden kann. Die syntaktische Struktur eines Satzes, welcher eine Relation beschreibt, wird mittels eines Dependency Tree beschrieben. Für die SVM wird dann ein Tree Kernel definiert, welcher solche Baumstrukturen in der Form des Dependency Tree verwendet, um die Ähnlichkeit zwischen Relation zu berechnen.

In dieser Arbeit wurden die nötigen Grundlagen erarbeitet, welche zu der Entwicklung des **Shortest Path Subsequence Kernel** geführt haben. In Java wurde anschliessend eine Komponente entwickelt, welche den Kernel für eine bestehende SVM-Software implementiert. Eine Evaluation des entwickelten Systemes mit einem Korpus zeigt, dass der entwickelte Kernel für die Relation Extraction vielversprechende Resultate liefert.