**Vortrag von Mai Linh Pham, BA Betreuerin Yadollah Yaghoobzadeh**
**Thema: Analysis of NIL Results in an entity Linking System**

Als erstes erklärt die Vortragende drei wichtige Definitionen zum Thema. Unter „Entity Linking (EL)“ versteht man den Vorgang eine Erwähnung in einem Text mit einer entsprechenden Entität in einer Wissensdatenbank zu verbinden. „NIL Results“ sind Entitäten die nicht mit der Wissensdatenbank verbunden sind. „Fine-Grained Entity Annotation“ ist ein detailliertes Tag Set, das mehr Tags als die Standard NER Tags (LOC, ORG, PER, MISC) beinhaltet.
Motivation für die Arbeit ist u.a., das EL nicht alle Entitäten verbinden kann, da einige Entitäten in der Wissensdatenbank fehlen. Das EL allgemein verbessern durch analysieren der „NIL Results“. Eine neue Cluster-Methode einführen um NIL Erwähnungen für die Analyse zu nutzen.
Ziel der Arbeit ist es herauszufinden ob detaillierte „entity tags“ hilfreich sind um NIL Erwähnungen zu Clustern und zu Analysieren.
Um dieses Ziel zu erreichen werden zuerst eine Ausgabe von einem „entity annotation tool“ mit der Ausgabe eines „entity linking system“ kombiniert. Dann wird die NIL Ausgabe extrahiert und geclustert. Zuletzt werden detaillierte „entity tags“ zum Clustern benutzt.
Es werden zwei verschiedene Systeme vor dem Clustern verwendet, FIGER und WAT. FIGER ist ein detailliertes Entitäten-Annotierungssystem und bietet 112 verschiedene Tags und lässt auch Überlappungen zu. Des Weiteren ist das System gut darin, ungewöhnliche Entitäten zu erkennen. Die FIGER Ausgabe besteht aus Standard BIO (Begin-Inside-Outside) Tags. WAT (ein „Entity-Linking-System) arbeitet mit folgenden Komponenten: „Spotter“, Disambiguierer und einem „Pruner“. Der „Spotter“ gibt eine Liste möglicher Entitäten zurück. der Disambiguierer stuft mögliche Entitäten mit verschiedenen Disabmiguierungsalgortihmen ein. Der „Pruner“ entfernt unnütze Annotationen um die „Precision“ zu erhöhen. Die WAT Ausgabe muss dasselbe Format haben wie die FIGER Ausgabe.
Im nächsten Schritt werden NIL Erwähnungen extrahiert. Erwähnungen die FIGER annotiert hat, aber nicht von WAT verlinkt wurden heißen unverlinkte Erwähnungen. Um diese Erwähnungen zu taggen erstellt man eine Liste mit Entitäten Namen in der Wissensdatenbank. Es werden nur die unverlinkten Erwähnungen betrachtet die einen entsprechenden Wissensdatenbank-Eintrag haben.
Zum Clustern der NIL Erwähnungen gibt es drei Typen von Herangehensweisen, den „coarse-grained-type“, „fine-grained-type“ und den „top-level-type“. Beim „fine-grained-type“ werden Typen semantisch geclustert, z.B. art -> film -> play -> music. Beim „top-level-clustering“ werden Tags in einem Oberbegriff zusammengefasst, z.B. product -> game, weapon, computer.
Die Schlussfolgerung lautet: „Fine-grained entity types“ können verwendet werden um zusammenhängende NIL Erwähnungen semantisch zu Clustern. Information die in den Tags verankert ist, kann zur Analyse verwendet werden. „Fine-grained entity types“ sind informative als „coarse-grained types“.

**Verfasser: Thomas Ebert**