*Implementing an application for knowledge acquisition that uses Amazon’s Mechanical Turk and that is based on various methods for semantic content creation*

Throughout the last few years, semantic technologies became more and more prominent in the World Wide Web as well as within enterprises. Despite the fact that there is lots of research going on in this field, there is a lack of semantic content and semantic data which to have is a prerequisite for every application that is based on semantics.

Due to the fact that such data cannot be produced fully automatically, human contribution is required on a large scale, however, producing labeled data can be a cumbersome and difficult task.

Amazon’s Mechanical Turk (<https://www.mturk.com/mturk/welcome>) proved to be a convenient means to encourage people to create semantic data on a large scale resulting in benefits for the “workers” who earn money and for the community, being able to use the created knowledge in semantic applications.

The goal of this work aims at reviewing the start of the art in terms of knowledge acquisition and at creating an application that uses Amazon’s Mechanical Turk and that is based on the methods described in <http://ai.stanford.edu/~dvickrey/game.pdf> for creating semantic data.

This includes for example methods

 for extracting hypernym/hyponym pairs,

 for capturing selectional preferences of verbs and

 for retrieving attributes/noun pairs.

In den letzten Jahren entwickelten sich das World Wide Web und der Datenaustausch innerhalb von Unternehmen dahingehend, dass immer mehr auf semantische Technologien gesetzt wird. Trotz der regen Forschungsaktivität auf diesem Gebiet, fehlt es an semantischem Inhalt und an semantischen Daten, die jedoch die Grundlage jedes semantischen Programms sind.

Diese Daten können nicht rein automatisch, ohne den Nutzer erzeugt werden. Da der Prozess der Datengenerierung für Menschen aber oft schwer und ohne Anreiz ist, bedarf es neuer Wege den Nutzer zu motivieren diese Lücke in der Entwicklung zu schließen.

Es hat sich gezeigt, dass Amazon’s Mechanical Turk (<https://www.mturk.com/mturk/welcome>) ein bewährtes Mittel ist, um Nutzer dazu zu bewegen, große Mengen an semantischen Daten zu produzieren. Zum einen verdienen die Nutzer Geld beim erzeugen der Daten, zum anderen werden große Datenmengen für die Community bereitgestellt und so profitieren beide Seiten.

Ziel dieser Arbeit soll es sein, den gegenwärtigen Stand der Forschung in Bezug auf Methoden zur Datengenerierung zu erfassen und weiters ein Programm zu schreiben, welches Amazon’s Mechanical Turk mit den Methoden in <http://ai.stanford.edu/~dvickrey/game.pdf>, die der Datengenerierung dienen, zu verknüpfen.

Solche Methoden sind zum Beispiel

 Methoden zur Generierung von Oberbegriff/Unterbegriff – Paaren,

 Methoden zur Bestimmung von Abhängigkeiten zwischen Nomen und Verb und

 Methoden zur Bestimmung von Attribut/Nomen – Paaren.