

Bachelor-/Masterarbeit

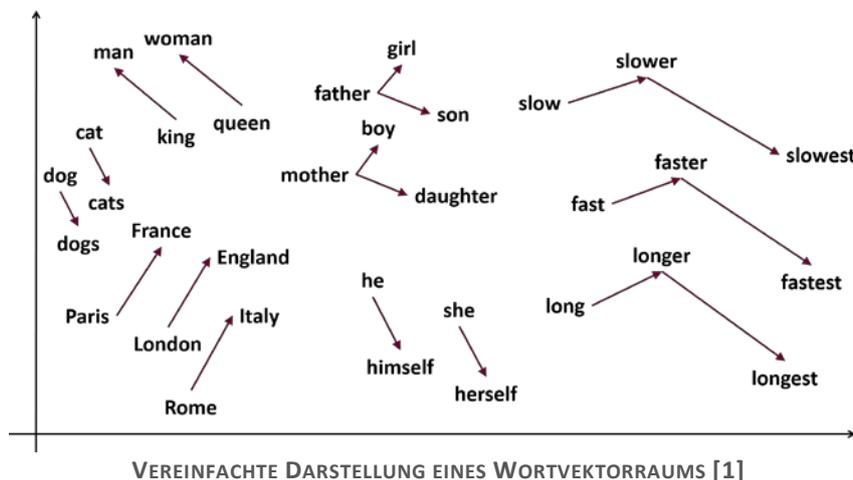
Word Embedding für semantische Textvergleiche in Gefährdungsanalysen

In Kooperation mit dem Ford Research and Innovation Center Aachen

Problemstellung

Sicherheitsrelevante Systeme müssen in der Automobilindustrie nach einem standardisierten Vorgehensmodell entwickelt werden. Eine Aktivität in diesem Modell ist die Gefährdungsanalyse und Risikoabschätzung (engl. Hazard Analysis and Risk Assessment – HARA). Üblicherweise, werden Daten, die während der HARA gesammelt werden, in tabellarischer Form mit natürlicher Sprache festgehalten. Dabei können leicht Inkonsistenzen und Mehrdeutigkeiten entstehen.

Im Rahmen eines Projekts am Lehrstuhl Informatik 11 wird ein Tool zur Unterstützung bei der Durchführung der oben genannten Aktivität entwickelt. In diesem Tool werden kontrollierte Sprachen (engl.: Controlled Natural Language – CNL) zur Dokumentation der Gefährdungsanalyse verwendet. Eine kontrollierte Sprache ist eine natürliche Sprache mit grammatischen als auch lexikalischen Einschränkungen.



In dieser Abschlussarbeit soll eine existierende semantische Textanalyse basierend auf den kontrollierten Sprachen um eine Verfahren für Word Embedding erweitert werden. Word Embedding ist ein Sammelbegriff für Methoden zur Sprachmodellierung in dem Wörter in einem Vektorraum angeordnet werden. Dabei soll die Anordnung der semantischen Ähnlichkeit der Wörter entsprechen. Ein Beispiel für ein solches Verfahren ist word2vec von Google.

Aufgabenstellung

- ▶ Literaturrecherche zum aktuellen Stand der Forschung von Word Embedding und verwandten maschinellen Lernverfahren
- ▶ Recherche geeigneter Frameworks zur Umsetzung
- ▶ Implementierung eines Word Embedding Verfahrens und Anbindung an die semantische Textanalyse
- ▶ Evaluation mit Hilfe geeigneter Methoden und industriellen Fallbeispiele

Ansprechpartner

Stefan Rakel, M.Sc. RWTH
rakel@embedded.rwth-aachen.de

Paul Chomicz, M.Sc. RWTH
chomicz@embedded.rwth-aachen.de

[1] <http://www.samyzaf.com/ML/nlp/nlp.html>