

Bachelorarbeit Proposal

Online Wissensgraphkonstruktion aus natürlicher Sprache

Clemens Damke

Matrikelnr. 7011488

23. März 2017

betreut von

Prof. Dr. Eyke Hüllermeier
Intelligente Systeme
Institut für Informatik
Universität Paderborn

1 Motivation und Hintergrund

In den letzten Jahren hat die Repräsentation von Wissensbasen durch Graphen immer mehr an Bedeutung gewonnen. Google benutzt solche Wissensgraphen z. B. zum Beantworten von komplexen Suchanfragen.

Die Grundidee dabei ist, Entitäten durch Knoten und Relationen durch Kanten abzubilden. Entitäten können konkrete Dinge, wie z. B. Personen, aber auch abstrakte Konzepte, wie z. B. historische Epochen, sein. Relationen beschreiben beliebige Beziehungen zwischen den Entitäten, z. B. $Person(\text{Da Vinci}) \xrightarrow{\text{lebte in}} Epoche(\text{Renaissance})$.

Da solche Graphen in zahlreichen Domänen einsetzbar sind, wird deren automatisierte Konstruktion bereits seit Jahren erforscht. Manuelles Konstruieren und vor allem anschließendes Warten und Aktualisieren von Wissensgraphen ist aufgrund der abzubildenden Datenmengen nicht praktikabel. Bei einer maschinellen automatisierten Konstruktion sind insbesondere zwei Anforderungen problematisch:

1. Das Verarbeiten von unstrukturierten Eingaben, wie z. B. natürlichsprachlichen Texten.
2. Effizientes Eingliedern neuer Informationen in einen bestehenden Wissensgraphen. Dieses Eingliedern von Informationen umfasst im Speziellen:
 - a) **Entity Resolution:** Hinzukommende Entitäten, die bereits im Graphen enthalten sind, müssen als Duplikate erkannt werden. Dies ist i. d. R. nicht trivial, da die selbe Entität durch viele verschiedene, oftmals vom Kontext abhängige, Token repräsentiert werden kann; z. B. *Bob* vs. *Robert* oder *Der Papst* vs. *Franziskus*.
 - b) **Link Prediction:** Hinzukommende Entitäten müssen mit bereits bestehenden Entitäten in Relation gesetzt werden. Hinzukommende Relationen können zudem benutzt werden um andere Relationen zu inferieren; z. B.

$$Weiblich(A) \wedge B \xrightarrow{\text{Sohn von}} A \implies A \xrightarrow{\text{Mutter von}} B$$

Die Kombination dieser beiden Anforderungen ist interessant, da das meiste verfügbare Wissen in natürlichsprachlicher Textform vorliegt und zudem permanent neues Wissen entsteht. Ein automatisiertes Wissensgraphkonstruktionsverfahren sollte daher beide Anforderungen berücksichtigen.

2 Ziele

3 Verwandte Arbeiten

4 Inhalte

5 Zeitplan