

30.10.2014

# Master-/Diplomarbeit

## „Finde den Nachfolger.“

### Erlernen von neuen Entitäten in einer Reihe von Entitäten“ zu vergeben!

#### Um was geht es?

Ziel ist es, Entitäten, die noch nicht in Wikipedia sind, aber für das Textdokument zentral sind, in einem gegebenen Text erstmalig zu erkennen. Als zentral werden dabei diejenigen Entitäten betrachtet, die als Nachfolger von bekannten Entitäten stehen (Entitätenreihe), etwa bei Produkten das *Windows 10* in der Reihe der Windows-Betriebssysteme (z.B. *Windows 95*,..., *Windows 8.1*).

Microsoft hat den Nachfolger von Windows 8.1 vorgestellt und richtet sich dabei vor allem an Großkunden. Der Name lautet überraschenderweise "Windows 10".

Windows 8.1



WIKIPEDIA



Der Fokus der Arbeit liegt darin, zunächst die „Entitätenketten“ automatisch zu erlernen und/oder aus einer Wissensbasis zu extrahieren. In einem zweiten Schritt kann dann ein Ansatz entwickelt werden, der die Nachfolger in diesen „Entitätenketten“ in einem gegebenen Text findet, ohne den Namen bzw. das Label der Entität vorher zu kennen (z.B. „Windows 10“). Die Erkennung soll dabei nicht auf Versionszahlen beschränkt sein, sondern es soll ein generischer Ansatz entwickelt werden. Hierzu können zwei Ansätze verglichen werden: Das Ausnutzen gleicher oder ähnlicher Attribute (also einer Wissensbasis) und das Ausnutzen des gleichen historischen Hintergrundes (also des Kontextes in den Dokumenten mittels Annotationen). Details des Ansatzes werden im persönlichen Gespräch erläutert.

Microsoft hat den Nachfolger von **Windows 8.1** vorgestellt und richtet sich dabei vor allem an Großkunden. Der Name lautet überraschenderweise "**Windows 10**".

neu in der Reihe der Microsoft-Betriebssysteme

Windows 10



WIKIPEDIA

Als Wissensbasis kann Wikipedia bzw. DBpedia verwendet werden. Zur Textannotation werden adaptierte Varianten des Wikipedia Miner zur Verfügung gestellt.

#### Was sollten Sie mitbringen?

- Interesse an der Entwicklung neuer Algorithmen in Zusammenarbeit mit dem Betreuer.
- Gute Programmierkenntnisse in Java.

Kontaktperson:  
**Michael Färber**  
michael.faeerber@kit.edu  
Tel.: 0721/608 479 46