

07.06.2013

# Bachelor-/Master-/Diplomarbeit

## „Erkennung von neuartigen Entitäten in Text“ zu vergeben!

### Um was geht es?

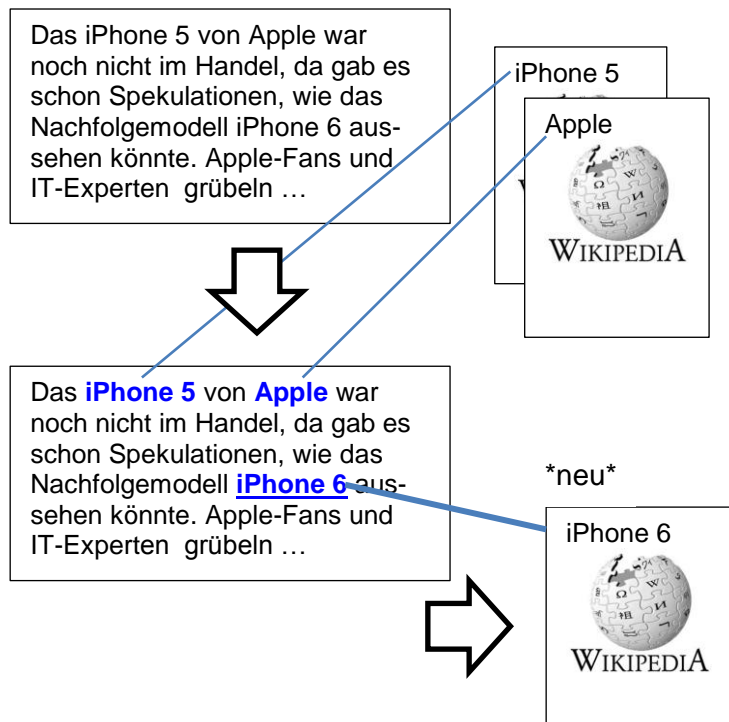
Unter der Voraussetzung, dass eine Wissensbasis wie Wikipedia gegeben sei, sollen bisher unbekannte Entitäten in einem Text erkannt werden. Unbekannt bedeutet dabei bspw., dass die Entität noch nicht durch einen Wikipedia-Eintrag in Wikipedia repräsentiert ist. Dies ist beispielsweise bei neuartigen Erfindungen wie dem „iPhone 6“ der Fall.

Für die Erkennung von Entitäten in einem Fließtext existieren bereits unzählige Tools (sog. NERC-Tools). Bisherige Verfahren schaffen es jedoch nur sehr begrenzt, „Neuheiten“ zu entdecken. Im konkreten Beispiel wird bspw. „iPhone“ als Entität und existierender Wikipedia-Eintrag erkannt, jedoch nicht, dass „iPhone 6“ gemeint ist, dies ein Nachfolger der iPhone-Serie ist (d.h. vom Entitätstyp „Smartphone“ oder „Produkt“) und der Text relevant für den noch fehlenden Eintrag „iPhone 6“ relevant sein könnte. Die Erkennung kann auf Versionszahlen basieren, soll aber nicht darauf beschränkt sein.

In der Abschlussarbeit soll zum einen ein (kurzer) Überblick über bestehende Verfahren/Tools geschaffen werden, die das beschriebene Problem bereits angehen. Ferner sollen geeignete Algorithmen zur Erkennung erdacht und implementiert werden. Dies kann in Form einer Erweiterung eines existierenden etablierten NERC-Tools wie dem *Wikipedia Miner* geschehen.

### Was sollten Sie mitbringen?

- Gute Programmierkenntnisse in Java
- Interesse an semantischen Technologien und der Verarbeitung natürlich-sprachiger Texte.



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