Karlsruhe Institute of Technology
Institute of Applied Informatics and Formal Description Methods (AIFB)
Knowledge Management Group
Michael Färber
michael.faerber@kit.edu

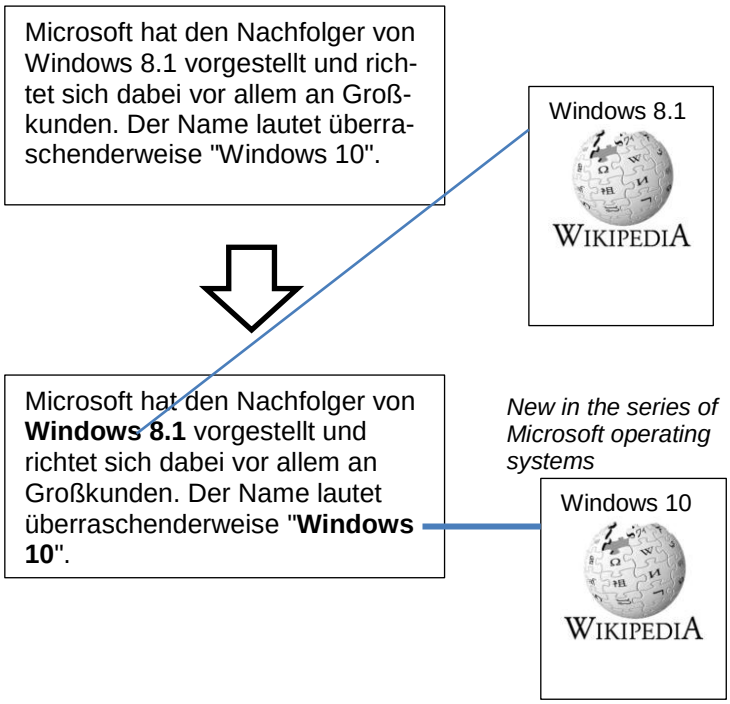
Nov. 14, 2014

Call for Master thesis

"Find the successor. Learning new entities in a sequence of entities"

What's it all about?

The aim is to identify for the first time entities that are not yet in Wikipedia, but central to the given text document. Entities are hereby regarded as central if they are successors of known entities (entity series). For instance, in case of products *Windows 10* stands in the sequence of Microsoft operating systems (e.g. *Windows 95*, ..., *Windows 8.1*).



The focus of the thesis is to learn first the "entity chains" automatically and/or to extract them from a knowledge base. In a second step, an approach is to be developed that finds successors of these "entities chains" in a given text, whereas the name or the label of the entity is not known so far (e.g. *Windows 10*). The recognition should not be limited to version numbers; instead, a generic approach is to be developed. For this purpose, two approaches can be compared: Exploiting the same or similar attributes (i.e. in the context of a knowledge base) and the exploitation of the same historical background (i.e. in the context of the documents using annotations). Details of the approach are explained to the student in person.
As a knowledge base Wikipedia or DBpedia can be used. For text annotation adapted versions of the Wikipedia Miner will be made available.

What you should bring along?

- Interest in the development of new algorithms in coordination with the supervisor.
- Good programming skills in Java.

Contact person:
Michael Färber
michael.faerber@kit.edu
Tel.: 0721/608 479 46