Interactive Visualization Methods for Sketch Engine

Motivation
- Automatic analysis of large text corpora produces numerous useful figures which help to understand the underlying corpus data
- The results are usually presented in the form of tables containing raw data
- The representation is therefore crucial for correct understanding and interpretation

Sketch Engine
- Leading corpus management system for discovering how words behave in different contexts
- It has a wide range of analytic functions dealing with billion-word corpora

The thesis was focused on the following features of Sketch Engine:

- Corpora overview: all language corpora that are available in the system
- Thesaurus: words that occur in the same grammatical context
- Word Sketch: word’s collocational behavior according to grammar relations
- Word Sketch Difference: comparison of two words according to their context

Visualizations available online
- Features have interactive interfaces
- Implemented in JavaScript (D3, jQuery)
- Easy deployment on any website
- Language independent

Novel visualizations for Sketch Engine allow comprehensible presentation of query results
- Deeper exploration is possible via interactions
- Usability confirmed by user testing
- Presented at QueryVis at NoDaLiDa 2015 [1]