Reality is not a game!

Extracting Semantics from Unconstrained Navigation on Wikipedia

Thomas Niebler, Daniel Schlör, Martin Becker, and Andreas Hotho

University of Wuerzburg
Tel.: +49-931-31-89094
{niebler,schloer,becker,hotho}@informatik.uni-wuerzburg.de

Abstract. Semantic relatedness between words has been successfully extracted from navigation on Wikipedia pages. However, the navigational data used in the corresponding works are sparse and expected to be biased since they have been collected in the context of games. In this paper, we raise this limitation and explore if semantic relatedness can also be extracted from unconstrained navigation. To this end, we first highlight structural differences between unconstrained navigation and game data. Then, we adapt a state of the art approach to extract semantic relatedness on Wikipedia paths. We apply this approach to transitions derived from two unconstrained navigation datasets as well as transitions from WikiGame and compare the results based on two common gold standards. We confirm expected structural differences when comparing unconstrained navigation with the paths collected by WikiGame. In line with this result, the mentioned state of the art approach for semantic extraction on navigation data does not yield good results for unconstrained navigation. Yet, we are able to derive a relatedness measure that performs well on both, unconstrained navigation data as well as game data. Overall, we show that unconstrained navigation data on Wikipedia is suited for extracting semantics.
The original paper is currently under review [Niebler et al(2015)].

References

Niebler et al(2015). Niebler T, Schlör D, Becker M, Hotho A (2015) Extracting semantics from unconstrained navigation on wikipedia. Künstliche Intelligenz - Special Issue: Semantic Web *currently under review*