

Flexible and Reliable Text Analytics in the Digital Humanities -- Some Methodological Considerations

Jonas Kuhn

Institute for Natural Language Processing
University of Stuttgart
jonas.kuhn@ims.uni-stuttgart.de

Abstract

The availability of Language Technology Resources and Tools generates a considerable methodological potential in the Digital Humanities: aspects of research questions from the Humanities and Social Sciences can be addressed on text collections in ways that were unavailable to traditional approaches. I start this talk by sketching some sample scenarios of Digital Humanities projects which involve various Humanities and Social Science disciplines, noting that the potential for a meaningful contribution to higher-level questions is highest when the employed language technological models are carefully tailored both (a) to characteristics of the given target corpus, and (b) to relevant analytical subtasks feeding the discipline-specific research questions.

Keeping up a multidisciplinary perspective, I then point out a recurrent dilemma in Digital Humanities projects that follow the conventional set-up of collaboration: to build high-quality computational models for the data, fixed analytical targets should be specified as early as possible -- but to be able to respond to Humanities questions as they evolve over the course of analysis, the analytical machinery should be kept maximally flexible. To reach both, I argue for a novel collaborative culture that rests on a more interleaved, continuous dialogue. (Re-)Specification of analytical targets should be an ongoing process in which the Humanities Scholars and Social Scientists play a role that is as important as the Computational Scientists' role. A promising approach lies in the identification of re-occurring types of analytical subtasks, beyond linguistic standard tasks, which can form building blocks for text analysis across disciplines, and for which corpus-based characterizations (viz. annotations) can be collected, compared and revised. On such grounds, computational modeling is more directly tied to the evolving research questions, and hence the seemingly opposing needs of reliable target specifications vs. "malleable" frameworks of analysis can be reconciled. Experimental work following this approach is under way in the Center for Reflected Text Analytics (CRETA) in Stuttgart.

This work is licenced under a Creative Commons Attribution 4.0 International Licence. Licence details:
<http://creativecommons.org/licenses/by/4.0/>