
Authorship Verification and Dating of Latin Sources

Dominik Both
Universität Heidelberg
both@cl.uni-heidelberg.de

Ines Rehbein
Leibniz ScienceCampus
rehbein@cl.uni-heidelberg.de

For historians, information on authorship and time of creation is of equal importance to the analysis of a source text as its content. This information, however, is not always available but often has to be inferred from the text. In our work, we present a system for the automatic detection of author and creation year of unknown Latin texts using stylometric cues. We will evaluate our system on data from the Corpus Corporum, a large corpus of Latin texts created between 50 BC and 1800 AD from more than 1,000 distinct authors and from different domains.

We approach the authorship assessment task as an authorship verification problem: Given a text by a known author we try to assess the probability that a new document is written by the same author using a similarity-based approach. Distortion (Stamatatos 2017) and eliding techniques (Hitschler et al. 2017) are tested that replace content words with different placeholders. For calculating the similarity between two documents we compare a simple geometric measure with a measure of similarity based on the *Impostors Method* (see Koppel and Winter, 2014).

The dating task is approached as a regression problem where different cues are used that indicate the language stage for a given document, from syntactic changes over changes of word usage to orthographic changes. We evaluate our system in a ranking task, i.e. if the system correctly sorts a list of random document samples by year.

References: Hitschler, J. et al. (2017). Robust authorship attribution with convolutional neural networks and pos-eliding. Workshop on Stylistic Variation, 53–85. • Koppel, M. and Winter, Y. (2014). Determining if two documents are written by the same author. ASIS&T 65(1), 178–187. • Niculae, V. et al. (2014). Temporal text ranking and automatic dating of texts. EACL 14, 17–21. • Stamatatos, E. (2017). Authorship attribution using text distortion. EACL 15, 1138–1149. • Tilahun, G. et al. (2012). Dating medieval english charters. AOAS 6(4), 1615–1640. • Zampieri, M. et al. (2016). Modeling language change in historical corpora: The case of portuguese. LREC, 4098–4104.