

Text mining, computational linguistics and the history of mathematics

Gerhard Brey

Research Fellow

Centre for Computing in the Humanities

King's College London

This talk will focus on potential uses of text mining and computational linguistics in various areas of the history of mathematics. The areas I will touch upon will range from the more tried and tested applications of these methods to those that venture into less established territory. The former could comprise for example, the application of text mining methods to incipits of medieval mathematical and astronomical manuscripts to extract and infer names, titles, etc. or their use to categorise and classify medieval and early modern mathematical text books. An example for the latter would be the potential use of these methods to contribute to and facilitate the creation of a database of mathematical problem texts - and my talk will concentrate mainly on this area.

Building on Warren Van Egmond's [1] suggestions as to how to approach such an undertaking, I will extend his recommendations by presenting some ideas about how text mining and computational linguistics could be used to alleviate this enormous task.

Mathematical problem texts have come down to us in large numbers; they can be found in all "languages, cultures and peoples who have left written records from the beginning of recorded history to the present day" (Van Egmond, p. 379). Chronologically they range from the Babylonian and Egyptian eras to the present, and culturally from the Chinese, Indian and Muslim cultures to the western world.

The same types of mathematical problems appear in all of these texts, each culture and time embedding them in a situation or context ("Einkleidung") to make it appropriate and meaningful for their readers. These texts therefore form an immensely rich source for the study of culture and history in general and of practical applications of mathematics in particular.

In this talk I will discuss the potential benefits and challenges that present themselves if text mining methods are applied to this kind of text.

[1] Warren Van Egmond, "Types and Traditions of Mathematical Problems: A Challenge for Historians of Mathematics", in: *Mathematische Probleme im Mittelalter. Der lateinische und arabische Sprachbereich*, ed.: Menso Folkerts, Wiesbaden (Harrassowitz) 1996 (*Wolfenbütteler Mittelalter-Studien*, 10), pp. 379-428.