

Editorial

Paul Molitor* and Jörg Ritter

Digital methods for intertextuality studies

<https://doi.org/10.1515/itit-2020-0006>

It is undisputed that with the application of digital methods humanities issues can be addressed which could not be addressed so far. This is the third special issue on digital methods in the humanities that is published in *it – Information Technology*. The first one was published in 2009 by Thomas Burch, Claudine Moulin and Andrea Rapp, at that time all working at the Trier Center for Digital Humanities [1], the second one in 2016 by Manfred Thaller from University of Cologne [2]. Both issues presented humanities questions where the use of digital methods is useful or even necessary. As Manfred Thaller wrote in his editorial, many of them are *challenges definitely worthy of a computer scientist*. However, the situation has hardly changed in the last decade. Digital Humanities continue to be an issue only in the humanities and are largely ignored by computer scientists. In Germany in particular, there are only a few working groups in computer science dealing with the counterpart of Digital Humanities, the so called eHumanities which is concerned with the development of new and non-trivial information technology approaches to support humanities scholars in addressing their issues.

With this special issue we would like to draw our computer science colleagues' attention to the exciting topics and the associated challenges at the interface between humanities and computer science, once again. Even though the special issue only deals with digital methods for text-based studies, there are exciting issues for computer scientists in almost all humanities fields. Take art history or archaeology, for example, where computer vision or 3D-modelling and -reconstruction play central roles. Or take political and social sciences, where big data analysis and deep learning methods have become indispensable digital methods. To answer questions in the humanities, a sound knowledge of data structures and efficient algo-

rithms, modern information retrieval, parallelism, modern web technologies and software engineering is often essential.

It should be noted, however, that the active participation of the humanities scholars in the development of digital methods for humanities questions is important, if not crucial. The meaning, purpose and reliability of the methods must be discussed in constant exchange with the humanities scholars. Besides that, many humanities questions are complex and are controversially discussed among experts. These issues cannot be reliably answered by algorithms, not even by the 'panaceas' artificial intelligence or deep learning. Here, digital methods can only support the humanities scholars in their research. As formulated by Dennis Tenen (cf. [3]), *computation in the humanities does not substitute but supplements the traditional research toolkit of humanities scholars*.

This special issue focuses on digital methods that help the humanities scholar to reveal intertextuality in texts. Intertextuality is a central aspect in literary and ancient history. The challenge posed by this topic is that most aspects of intertextuality are difficult to grasp in mathematical terms. There is no mathematical definition of what intertextuality means and the concept of intertextuality is discussed very diversely in the humanities. Without wanting to anticipate the first article of this special issue, simply put, intertextuality refers to the relationships between texts, for example through explicit quotations, through plagiarism, through allusions or paraphrases, but also to the question of who is the author of a text or a specific text passage. Intertextuality is partly a question of the respective interpretation of a text by the individual humanities scholar. Thus, algorithmic approaches for intertextuality studies usually cannot provide answers, only suggestions. However, they have the enormous advantage that no prior subjective selection criteria are used in the calculation of the proposals.

The first article comes from a colleague from Ancient History. In her contribution "Intertextuality and Digital Humanities" Charlotte Schubert from the University of Leipzig introduces us to the topic of Digital Humanities and intertextuality. She first briefly reviews the debate on the Digital Humanities that has been going on for more than 20 years in the humanities and then introduces us to the topic of intertextuality. In her essay, she argues in

*Corresponding author: Paul Molitor, Martin Luther University Halle-Wittenberg, Institute of Computer Science, D-06099 Halle, Germany, e-mail: paul.molitor@informatik.uni-halle.de, ORCID: <https://orcid.org/0000-0003-1461-7936>

Jörg Ritter, Martin Luther University Halle-Wittenberg, Institute of Computer Science, D-06099 Halle, Germany, e-mail: joerg.ritter@informatik.uni-halle.de, ORCID: <https://orcid.org/0000-0003-3080-7801>

particular that with the use of algorithmic methods it will be or should be possible to make different interpretations comparable over a reproducible and transparent measure. She then discusses her thesis using the example of the algorithmic search for text passages in Ancient Greek literature which paraphrase Plato.

The word-by-word comparison between different versions of a text, i. e. finding differences (or similarities) between them on the word level, is a central step in the investigation of the origin of a text or the transmission of a text. The paper “From giant despair to a new heaven: The early years of automatic collation” by Elisa Nury and Elena Spadini, who are working in the field of Digital Philology at the University of Geneva and the University of Lausanne, respectively, presents a commented history of automatic collation. The article reports on the expectations of the computer in supporting the creation of critical editions, which the humanities scholars have had since the “invention of the computer”. The article concludes with a look at current developments in this field.

The third contribution is authored by Marcus Pöckelmann, Janis Dähne, Jörg Ritter and Paul Molitor. They are computer scientists at the University of Halle. Their article “Fast paraphrase extraction in Ancient Greek literature” addresses intertextuality in a more general sense. In contrast to word-by-word text comparison, it considers the problem of finding text passages that semantically resemble a given text passage, even if they have no words in common. The article focuses on ancient languages for which there are (so far) neither high-quality part of speech taggers nor lexical-semantic networks, so that the approaches to paraphrase extraction known from literature which all work on modern languages are not applicable. The authors present in their article a high-quality method for paraphrase extraction in Ancient Greek literature whose wall clock times are such that it is suitable for being used in an interactive application for intertextuality studies.

The fourth contribution is from Christopher Blackwell and Neel Smith who are professors of classics at Furman University and at the College of the Holy Cross, respectively. Their article “The CITE architecture (CTS/CITE) for analysis and alignment” deals with the problem of documenting text-reuse, especially how to cite results or findings in the field of intertextuality achieved by digital analytical tools or human experts. The authors first introduce special features that must be considered when documenting results of analysis and alignment. A documentation of these special features using XML is possible, but rather complicated. Using examples, the authors demonstrate the elegance with which such features can be documented with the new CITE architecture.

Another issue to be dealt with in intertextuality studies is the question of the authorship of a given text or a part of a text. If the experts know the author of a text, they may be able to draw conclusions about the origin of the ideas or opinions expressed in the texts. The article “On divergence-based author obfuscation: An attack on the state of the art in statistical authorship verification” by Janek Bevendorff, Tobias Wenzel, Martin Potthast, Matthias Hagen, and Benno Stein handles both authorship verification and author obfuscation, which are two sides of the same coin. It introduces new algorithms for authorship verification and author obfuscation and examines the extent to which current approaches to authorship verification can withstand obfuscation. The five authors are computer scientists from the Universities of Weimar, Leipzig and Halle, respectively, working in the fields of text mining, data- and knowledge-intensive information processing tasks and natural language processing.

We hope you enjoy reading this special issue.

References

1. Thomas Burch, Claudine Moulin, and Andrea Rapp (Eds.). *Informatik in den Geisteswissenschaften, it – Information Technology*, Vol. 51, No. 4, Oldenbourg Wissenschaftsverlag, Munich 2009.
2. Manfred Thaller (Ed.). *Computer Science in Humanities, it – Information Technology*, Vol. 58, No. 2, de Gruyter Oldenbourg, Munich 2016.
3. Dennis Tenen. Blunt Instrumentalism: On Tools and Methods. In: M. K. Gold and L. F. Klein (Eds.) *Debates in the Digital Humanities 2016*, University of Minnesota Press, 2016.

Bionotes



Prof. Dr. Paul Molitor

Martin Luther University Halle-Wittenberg,
Institute of Computer Science, D-06099
Halle, Germany
paul.molitor@informatik.uni-halle.de

Prof. Dr. Paul Molitor studied Computer Science and Mathematics at the University of Saarland (Diplom 1982, Promotion 1986, Habilitation 1992). He was member of the scientific staff of Prof. Dr. Günter Hotz (1982–1994) where he leads a project in the National Research Center 124 VLSI and Parallelism (1992–1994). In 1993, he was with the Humboldt-University of Berlin as Associate Professor for Circuit Technology. In 1994, at the age of 34, he became

a Full Professor for Technical Computer Science at Martin-Luther-University Halle-Wittenberg. Paul Molitor's interests lie in addition to technical computer science in combinatorial optimization and computational humanities/eHumanities. Together with co-editor Jörg Ritter and colleagues from the humanities, he has been leading several interdisciplinary third-party funded projects in the field of Digital Humanities, in particular with colleagues from the fields of German Studies, Romance studies, Jewish studies, Sanskrit studies, and ancient history / ancient Greek studies.



Dr. Jörg Ritter
Martin Luther University Halle-Wittenberg,
Institute of Computer Science, D-06099
Halle, Germany
joerg.ritter@informatik.uni-halle.de

Dr. Jörg Ritter studied Computer Science at the University of Saarland (Diplom 1997), he received his doctorate in technical computer

science with a thesis on a pipelined architecture for partitioned discrete wavelet transformation based lossy image compression using FPGA's. Since 1997, he is research associate in the group headed by Prof. Dr. Molitor at the Institute of Computer Science of Martin Luther University Halle-Wittenberg. For more than 10 years, Jörg Ritter has been working with colleagues from philology in interdisciplinary projects dealing with humanities issues, e. g., Digital Plato. Tradition and Reception (2016–2019), A New Supplement Dictionary of Sanskrit (2013–2016), Epistolary Networks. Visualizing multi-dimensional information structures in correspondence corpora (2013–2016), Semi-automatic Difference Analysis of Complex Text Variants (2012–2015). Jörg Ritter was and is significantly involved in the establishment of a research focus “eHumanities” at the Institute of Computer Science of the Martin Luther University Halle-Wittenberg.