

# *Welt der Kinder*: Knowledge and Interpretation of the World as Portrayed in Textbooks and Children Books between 1850 and 1918

Carsten Schnober<sup>†‡</sup>, Ben Heuwing\*, Andreas Weiß<sup>‡</sup>,  
Richard Eckart de Castilho<sup>‡</sup>, Robert Strötgen<sup>‡</sup>,  
Iryna Gurevych<sup>†‡</sup>, Simone Lässig<sup>‡</sup>, and Christa Womser-Hacker\*

<sup>†</sup>Ubiquitous Knowledge Processing Lab (UKP-TUDA)  
Department of Computer Science, Technische Universität Darmstadt

<sup>‡</sup>Ubiquitous Knowledge Processing Lab (UKP-DIPF)  
German Institute for Educational Research and Educational Information  
[www.ukp.tu-darmstadt.de](http://www.ukp.tu-darmstadt.de)

\*Institut für Informationswissenschaft und Sprachtechnologie  
Universität Hildesheim  
[www.uni-hildesheim.de/fb3/institute/iwist](http://www.uni-hildesheim.de/fb3/institute/iwist)

<sup>‡</sup>Georg-Eckert Institute for International Textbook Research  
[www.gei.de](http://www.gei.de)

The digital humanities project *Welt der Kinder* (“Children and their world“) started in May 2014 and is designed to serve as a template for similar projects in the future. By fostering close cooperation between historians, information scientists and computer scientists it aims to gain new insights into the period from 1850 to 1918; a time in which the accelerated production of knowledge was dominated by both globalization and nationalisation simultaneously. The project is a cooperative venture between two institutions of the Leibniz Association – the Georg Eckert Institute for International Textbook Research in Braunschweig (GEI) and the German Institute for International Educational Research (DIPF) located in Frankfurt and Berlin –, four universities (Darmstadt, Hildesheim, Göttingen, and Zurich), and the Bavarian State Library (BSB) in Munich.

The research will provide access to German-language *mass sources* that originated between 1850 and 1918. This material reflected contemporary world interpretation patterns and elements of cultural memory yet, equally, helped form the same. Due to their sheer size, however, such sources cannot be penetrated by hermeneutic methods alone. In interdisciplinary and exploratory research, tools are therefore being developed that facilitate the analysis of large (digital) corpora. Diachronic and synchronic analyses of the (trans)formation of knowledge inventories are employed, as are meta analyses, to consolidate the potential

provided by insights gained through digital methods.

The development process implements user-centred methods in order to support the research goals of historians. The tools developed help to discover semantic structures and patterns in a variety of nineteenth century educational media. This enables us, as historians, to take an entirely new approach to the analysis of digitised material, which had previously been limited to full-text searches.

As its basic research resource, the project uses a continuously expanding corpus of several thousand books, which have been scanned and digitised using OCR technology capable of reading Gothic typefaces. The collection comprises more than 600,000 pages so far. *Welt der Kinder* has four main goals:

1. Historical research about representations and interpretations of the world in the given period, during which, knowledge about the world could generally not be gathered through one's own experiences or travels, nor through audiovisual media. That is why textbooks and other printed media formed the main information source for young adults other than oral traditions, which cannot be used as scientific sources.
2. Exploration of a specific media type (textbooks and juvenile literature) that subsequently influenced millions of citizens, but which has never been investigated in an integrated approach considering media type, circulation, and transformation of collective knowledge.
3. Combining established hermeneutic methodology with innovative methods and technologies for the exploration and scientific annotation of large amounts of textual data.
4. Fundamental research in computational linguistics; developing and adapting methods of semantic analysis and opinion mining for the challenges presented by the language of the nineteenth century and by this specific media type.

We shall present the resources and methodologies applied in the project as well as its progress so far. We will discuss challenges, experiences, and problems encountered in the early stages of the project.