

Digital Curation Technologies (DKT)

Georg Rehm, Felix Sasaki

DFKI GmbH, Language Technology Lab, Alt-Moabit 91c, 10559 Berlin, Germany

georg.rehm@dfki.de, felix.sasaki@dfki.de

Abstract. Digital Curation Technologies (“Digitale Kuratierungstechnologien”, DKT) is a project that involves four Berlin-based SMEs (ART+COM AG, Condat AG, 3pc GmbH and Kreuzwerker GmbH) and DFKI GmbH (Language Technology Lab). The two-year action started in Sept. 2015 and aims at supporting digital curation processes, carried out by knowledge workers, through robust, precise and modular language and knowledge technologies. We combine these into workflows for the efficient processing, creation and dissemination of digital content. DFKI contributes language and knowledge technology components, including MT, and develops them further. Together with our partners, DFKI develops a platform for digital curation technologies, which offers services such as, e.g., translation, search, analytics, re-combination and summarisation. DKT is supported by the German Federal Ministry of Education and Research, Wachstumskern-Potenzial (no. 03WKP45). Details: <http://www.digitale-kuratierung.de>.

Description

The curation of digital content is a complex, knowledge- and time-intensive process in which authors or editors work, analyse, process and integrate highly heterogeneous data, content and media modules from disparate sources into a new content product that has a specific focus and communicative purpose. Among the processes involved are the selection, summarisation, classification, internationalisation, analysis, reordering, restructuring, and visualisation of various types of content. At the same time we need to bear in mind the continuously growing speed at which new information is coming in, the quantity and number of sources as well as of information pieces to be processed. The DKT SME partners contribute their respective sector-specific expertise (e.g., interactive showrooms, museums, television stations, newsrooms). In DKT we build showcases that demonstrate how language and knowledge technologies can be used in these typical digital curation workflows. The technologies contributed to DKT by DFKI belong to three areas: *semantic analysis* (e.g., information extraction), *semantic generation* (e.g., semantic story telling) and *multilingual technologies*, i.e., robust, adaptable MT components and integration of different monolingual or multilingual data and knowledge sources (including linked open data and standardised workflow data) for the sector-specific curation workflows. MT is used both for inbound and outbound translation (i.e., documents to be published). We also plan to incorporate MT engine improvement from user feedback. For the platform, we focus in particular the following features: fully integrated, robust, precise and scalable components with open interfaces in order to be efficiently embedded in sector-specific curation workflows; easy access to the cloud platform; application-oriented sector-technologies with high usability.