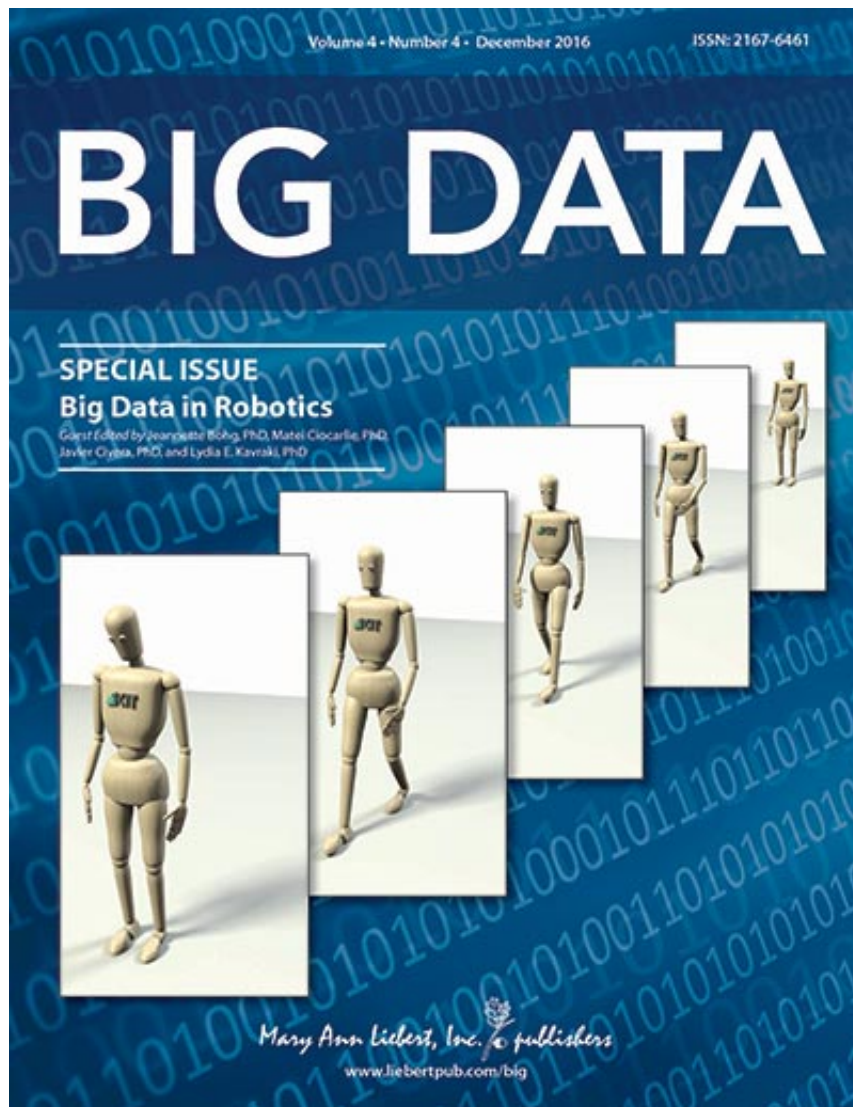


A new open source dataset links human motion and language

February 10 2017



Credit: Mary Ann Liebert, Inc., publishers

Researchers have created a large, open source database to support the development of robot activities based on natural language input. The new KIT Motion-Language Dataset will help to unify and standardize research linking human motion and natural language, as presented in an article in *Big Data*.

In the article "The KIT Motion-Language Dataset," Matthias Plappert, Christian Mandery, and Tamim Asfour, Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology (KIT), Germany, describe a novel crowd-sourcing approach and purpose-built web-based tool they used to develop their publicly available dataset that annotates [motions](#). Their approach relies on a unified representation that is independent of the capture system or marker set to be able to merge data from different existing motion capture databases into the KIT Motion-Language Dataset. It currently includes about 4,000 motions and more than 6,200 annotations in [natural language](#) that contain nearly 53,000 words.

The article is part of a special issue of *Big Data* on "*Big Data* in Robotics" led by Guest Editors Jeannette Bohg, PhD, Matei Ciocarlie, PhD, Jaview Civera, PhD, and Lydia Kavraki, PhD.

"Human motion is complex and nuanced in terms of how it can be described, and it is surprisingly difficult to even retrieve motions from databases corresponding to natural language descriptions. There is a great need to describe robotic systems in natural language that captures the richness associated with motion, but doing this accurately is an extremely challenging problem," says *Big Data* Editor-in-Chief Vasant Dhar, Professor at the Stern School of Business and the Center for Data Science at New York University. "Plappert and his colleagues do a wonderful job using a novel crowd-sourcing approach and a tool to document the annotation process itself along with methods for obtaining high quality inputs and selecting motions that require further annotation

automatically. They have constructed an impressive database of motions and annotations that can serve as a test-bed for research in this area. It is a great service to the research community."

More information: Matthias Plappert et al, The KIT Motion-Language Dataset, *Big Data* (2016). [DOI: 10.1089/big.2016.0028](https://doi.org/10.1089/big.2016.0028)

Provided by Mary Ann Liebert, Inc

Citation: A new open source dataset links human motion and language (2017, February 10) retrieved 25 April 2024 from <https://phys.org/news/2017-02-source-dataset-links-human-motion.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.