Exploring Change

Tobias Bleifuß[1], Leon Bornemann[1], Theodore Johnson[2], Dmitri V. Kalashnikov[2], Felix Naumann[1], and Divesh Srivastava[2]

[1] Hasso Plattner Institute, University of Potsdam, Germany
{tobias.bleifuss,leon.bornemann,felix.naumann}@hpi.de
[2] AT&T Labs–Research, USA
{johnsont,dvk,divesh}@research.att.com

Abstract. Data and schema in datasets experience many different kinds of change: values are inserted, deleted or updated; rows appear and disappear; columns are added or repurposed, and so on. In such a dynamic situation, users might wonder: How many changes have there been in the recent minutes, days or years? What kind of changes were made at which points of time? How volatile are the values of a specific property or entity? How dirty is the data? The fact that data changed can hint at different hidden processes: a frequently crowd-updated city name may be controversial; a person whose name has been recently changed may be the target of vandalism; and so on. To interactively explore such changes, we present our DBChEx (Database Change Explorer) prototype system [2]. Using two real-world datasets, IMDB and Wikipedia infoboxes, we illustrate how users can gain valuable insights into data generation processes and data or schema evolution over time by a mix of serendipity and guided investigation using DBChEx. Finally, we identify a range of technical challenges that need to be addressed to fully realize our vision of change exploration [1].

References

1. Bleifuß, T., Bornemann, L., Johnson, T., Kalashnikov, D.V., Naumann, F., Srivastava, D.: Exploring change - A new dimension of data analytics. PVLDB **12**(2), 85–98 (2018)
2. Bleifuß, T., Bornemann, L., Kalashnikov, D.V., Naumann, F., Srivastava, D.: DBChEx: Interactive exploration of data and schema change. In: *Proceedings of the Conference on Innovative Data Systems Research* (CIDR) (2019)