Editorial

Jens Dittrich¹ · Felix Naumann² · Norbert Ritter³ · Theo Härder⁴

Online publiziert: 14. Juni 2019

© Gesellschaft für Informatik e.V. and Springer-Verlag GmbH Germany, part of Springer Nature 2019

1 Schwerpunktthema: Data and Repeatability

Die Nachvollziehbarkeit (repeatability) und Wiederholbarkeit (reproducibility) wissenschaftlicher Erkenntnisse ist ein essentiell wichtiges Thema nicht nur in der Informatik, sondern in nahezu allen wissenschaftlichen Disziplinen. Wie kann ich mir sicher sein, dass ein in einem begutachteten Fachartikel beschriebenes wissenschaftliches Experiment tatsächlich die angeführten Ergebnisse erbracht hat? Wenn jemand Unabhängiges dasselbe Experiment wiederholte, würden dieselben Ergebnisse beobachtet und dieselben Schlussfolgerungen gezogen werden?

Diese Überprüfung ist immer dann schwierig, wenn nicht alle Informationen vollständig im Fachartikel dargelegt oder referenziert sind, die für eine Wiederholung eines Experiments notwendig sind. In der Informatik gehören hierzu: die verwendeten Algorithmen, Anfragen und Daten, Vorverarbeitungs- und Reinigungsschritte, die Parameter der Algorithmen, eine klare Beschreibung der verwendeten Hardware, die Versionen der verwendeten Software-Bibliotheken, ein annotierter Gold-Standard usw. Fehlt nur eine die-

ser Informationen, kann eine Wiederholung der Experimente bereits stark erschwert sein.

Die internationale Datenbankgemeinde hat sich schon früh über dieses Problem Gedanken gemacht, sei es durch einschlägige Forschung in diesem Bereich, aber auch z. B. durch Repeatability-Komitees, die zumindest Blackbox-Tests der Software angenommener Beiträge durchführen, und den sehr erfolgreichen Experiments & Analysis-Track auf (P)VLDB, der bei Autoren keinen unnötigen Druck aufbaut, Neuheit („...our approach is novel?“) zu behaupten und/oder künstlich herbeizuschreiben („...our algorithm is very different from related work?“).

Diese Ausgabe des Datenbank-Spektrums ist ein weiterer Versuch, ein wenig Licht in diese noch zu dunkle Ecke der Wissenschaft zu bringen. Hierzu tragen sechs ausgewählte Fachbeiträge bei, die die Themen Nachvollziehbarkeit und Wiederholbarkeit aus ganz unterschiedlichen Blickwinkeln beleuchten. So schlagen Chamanara et al. im Hinblick auf eine nationale Forschungsdateninfrastruktur ein semantisch verknüpft Netzwerk von Forschungsdatensystemen vor. Auch Vidal et al. beschäftigen sich mit der Nutzarmachung heterogener Daten, hier jedoch mit dem Domänen-spezifischen Ziel der Verbesserung von medizinischen Behandlungen. Pawlik et al. identifizieren fünf Ebenen der Reproduzierbarkeit von Daten und beleuchten damit die möglichen Auswirkungen von Vorverarbeitungsschritten. Risch et al. führen als neues Konzept die „partial data repeatability“ zur besseren Nachvollziehbarkeit von Web-Science-Experimenten ein. Ebenfalls mit Web-Daten beschäftigt sich der Beitrag von Bizer et al., wobei hier die Nutzung von semantischen Annotationen vorgeschlagen wird. Und der von Kiefer et al. beschriebene Ansatz versucht schließlich, auch ohne (Gold-)Annotationen die Qualität von (Text-)Daten abschätzen und entsprechend verbessern zu können.

Wir hoffen, Ihr Interesse geweckt zu haben und wünschen eine anregende Lektüre.

Ihre Gasteditoren

Jens Dittrich
jens.dittrich@bigdata.uni-saarland.de

Felix Naumann
felix.naumann@hpi.de

Norbert Ritter
norbert.ritter@min.uni-hamburg.de

✉ Theo Härder
haerder@cs.uni-kl.de

¹ Fakultät für Mathematik und Informatik, Universität des Saarlandes, 66123 Saarbrücken, Deutschland

² Hasso-Plattner-Institut für Digital Engineering gGmbH, Universität Potsdam, 14482 Potsdam, Deutschland

³ MIN-Fakultät, Universität Hamburg, 20354 Hamburg, Deutschland

⁴ AG Datenbanken und Informationssysteme, TU Kaiserslautern, 67663 Kaiserslautern, Deutschland

2 Community-Beiträge

Pünktlich zu einem Multi-Jubiläum besuchte die BTW 2019 zum ersten Mal in ihrer Geschichte die Ostseeküste: Stadt und Universität Rostock feiern ein Doppeljubiläum (800 Jahre Stadt Rostock in 2018, 600 Jahre Universität Rostock in 2019). Daneben feiert auch der Rostocker Lehrstuhl Datenbank- und Informationssysteme ein Jubiläum in 2019: er wurde vor 25 Jahren eingerichtet. In einem mit vielen Bildern illustrierten Beitrag *BTW 2019 – Datenbanksysteme im Zeitalter der Künstlichen Intelligenz, Data Science und neuen Hardware* berichten Andreas Heuer, Meike Klettke und Holger Meyer ausführlich über die 18. GI-Fachtagung „Datenbanksysteme für Business, Technologie und Web“, die vom 4. bis 8. März 2019 an der Universität Rostock stattfand. 215 Teilnehmer besuchten drei Workshops, fünf Tutorien, ein Studierendenprogramm sowie zehn Sessions mit wissenschaftlichen und Industrie-Beiträgen. Dazu gab es ein reichhaltiges Demo-Programm mit Vorstellungen von Forschungsprototypen.

Die Rubrik „Dissertationen“ enthält in diesem Heft 5 Kurzfassungen von Dissertationen aus der deutschsprachigen DBIS-Community.

Schließlich berichtet die Rubrik „Community“ unter *News* über weitere aktuelle Informationen, welche die DBIS-Gemeinde betreffen.

3 Künftige Schwerpunktthemen

3.1 Best Workshop Papers of BTW 2019

This special issue of the “Datenbank-Spektrum” is dedicated to the Best Papers of the Workshops running at the BTW 2019 at the University of Rostock. The selected Workshop contributions should be extended to match the format of regular DASP papers.

Paper format: 8–10 pages, double-column

Selection of the Best Papers by the Workshop chairs and the guest editor: April 15th, 2019

Deadline for submissions: June 10th, 2019

Issue delivery: DASP-3-2019 (November 2019)

Guest editor:

Theo Härder, University of Kaiserslautern
haerder@cs.uni-kl.de

3.2 Trends in Information Retrieval Evaluation

Evaluation is a central aspect in the research and development of information retrieval systems. In academia, the quantitative evaluation of such systems is mostly known under the term Cranfield paradigm. This research method has been established for more than 25 years in international evaluation campaigns such as the Text Retrieval Conference (TREC) or the Conference and Labs of the Evaluation Forum (CLEF). Meanwhile industrial research has taken a completely different approach. Many companies are able to access a large number of users and their interactions, which can be recorded and evaluated. These infrastructures allow alternative evaluations like large-scale A/B experiments or other online methods. In the last years, different approaches to go beyond TREC-style evaluations emerged to close the gap and to bring together academic and industrial evaluation.

We are calling for articles that report on novel evaluation efforts, like:

- Living labs
- Evaluation as a service
- Large-scale A/B tests
- Interactive retrieval evaluation
- Session-based evaluation
- User-centered evaluation
- Counterfactual evaluation
- Novel evaluations in application domains such as cultural heritage, digital libraries, social media, expert search, health information, etc.
- Other evaluations that go beyond TREC

Expected size of the paper: 8–10 pages, double-column (cf. the author guidelines at www.springer.com/13222). Contributions either in German or in English are welcome.

Deadline for submissions: Oct. 1st, 2019

Issue delivery: DASP-1-2020 (March 2020)

Guest editors:

Philipp Schaer, Technische Hochschule Köln

philipp.schaer@th-koeln.de

Klaus Berberich, Hochschule für Technik und Wirtschaft des Saarlandes

klaus.berberich@htwsaar.de

3.3 Argumentative Intelligence (Arguing)

This special issue of the “Datenbank-Spektrum” is dedicated to the area of “argumentative intelligence” concerned with the question of how to design intelligent systems that reach a deep understanding of states of affairs and domains in order to be able to analyze and generate arguments to

support users in the activity of rationalizing as a basis for informed decision making. We will invite submissions from different research fields including semantics, natural language processing, information retrieval, knowledge representation and artificial intelligence in general to contribute to this new field of research. The special issue will feature key results and advancements achieved as a result of the DFG-funded priority program “Robust Argumentation Machines”.

Expected size of the paper: 8–10 pages, double-column (cf. the author guidelines at www.springer.com/13222).

Contributions either in German or in English are welcome.

Deadline for submissions: Feb. 1st, 2020

Issue delivery: DASP-2-2020 (July 2020)

Guest editors:

Philipp Cimiano (Universität Bielefeld)

Gerhard Heyer (Universität Leipzig)

Michael Kohlhase (FAU Erlangen-Nürnberg)

Benno Stein (Bauhaus-Universität Weimar)

Jürgen Ziegler (Universität Duisburg-Essen)