

Pressemitteilung**Universität des Saarlandes****Friederike Meyer zu Tittingdorf**

30.03.2023

<http://idw-online.de/de/news811829>Forschungsprojekte
Informationstechnik, Medien- und Kommunikationswissenschaften, Wirtschaft
überregional**UNIVERSITÄT
DES
SAARLANDES****Turbo boost for Big Data: Computer scientist wants to 'breed' superfast and efficient databases**

Computer science professor Jens Dittrich and his team at Saarland University are developing a new approach to optimizing databases that looks set to turn previously used methodologies on their head – away from hand-crafted search methods and towards automatically generated ones. The team uses what are known as evolutionary algorithms to 'breed' the best possible results. The paper in which they present the concept has been published in the proceedings of one of the world's largest specialist conferences on databases. The German Research Foundation (DFG) is now funding the team to develop their method further.

The team led by Jens Dittrich, Professor of Big Data Analytics at Saarland University, is completely rethinking some of the established practices in data science. Their focus is on two of the today's most important resources: data and databases. Whether it's internet search queries, researching events that have global impact like climate change, or developing and using technologies such as AI chatbots – data is the fuel that is powering modern digital technology, and databases are the tools that help us organize, store and analyse that data. And we come into contact with these two resources every day, whether we're aware of it or not.

The area of data processing that the Saarbrücken computer scientists are working on involves 'index structures', which determine the way in which a database is accessed. 'Index structures are essential for working with databases, because they allow stored data to be found quickly and efficiently. Indexes speed up the search enormously, as it's no longer necessary to search the entire database,' explained Professor Dittrich, who conducts research at the Saarland Informatics Campus. Most people know how library catalogues are sorted. 'In computer science, however, the data we work with is often very complex and is present in immense quantities, so more sophisticated methods are needed for indexing,' said Dittrich.

Index structures are a well-researched field in computer science. For decades, organizational methods have been developed 'by hand' that work comparatively well in a wide range of applications, but are not really optimized for any one of them. Dittrich and his team have come up with an approach that automatically generates index structures that are matched to any database and any application.

They call the method 'Genetic Generic Generation of Index Structures (GENE)'. It is based on 'evolutionary algorithms', which represent a special subset of optimization algorithms. These algorithms emulate the natural process of evolution. 'The starting point is a "normal", non-optimized index. Random mutations are then allowed to evolve from this starting index. These mutations are sorted according to their performance and only the best are carried forward into the next generation. These steps are then repeated until there are no more significant improvements between generations,' explained Dittrich.

Previously, index structures were treated as closed systems. 'It was common to think, for example, that tree-based index structures were the best organizational method for some problems, while for others it was better to use hash tables. But in our approach, we look at all previous index structures as if they had a common "ancestor",' said Professor Dittrich. This allows the researchers to mutate this 'meta-index structure' and to combine the best properties of today's common index structures so that the result is individually optimized for each database and each application. 'Our aim is to "breed" an index structure that is the perfect match for each database, and that outperforms all previous ones,' explained Dittrich. Their preliminary work has already demonstrated the huge potential of their approach. Common index structures that were developed manually in previous decades can be 'rediscovered' with this approach, i.e. automatically generated or replicated.

The research work conducted by Jens Dittrich's team is basic research that has yet to be applied. But as with much basic research, this could change radically in the future, given the incredibly rapid growth in the amount of data being produced. In future, highly optimized methods like Dittrich's that can quickly search and process vast quantities of data could become increasingly important.

The paper 'The next 50 Years in Database Indexing or: The Case for Automatically Generated Index Structures' was published in the Proceedings of the International Conference on Very Large Data Bases (VLDB), one of the two leading conferences on databases worldwide. In addition, the DFG has been funding the project 'GENE: Genetic Generic Generation of Index Structures' since January 2023, providing around €300,000 over three years.

Background information on the Saarland Informatics Campus:

900 scientists (including 400 PhD students) and about 2500 students from more than 80 nations make the Saarland Informatics Campus (SIC) one of the leading locations for computer science in Germany and Europe. Four world-renowned research institutes, namely the German Research Center for Artificial Intelligence (DFKI), the Max Planck Institute for Informatics, the Max Planck Institute for Software Systems, the Center for Bioinformatics as well as Saarland University with three closely aligned departments and 24 degree programmes cover the entire spectrum of computer science.

Editor:

Philipp Zapf-Schramm

Saarland Informatics Campus

Phone: +49 681 302-70741

Email: pzapf@cs.uni-saarland.de

wissenschaftliche Ansprechpartner:

Prof. Dr. Jens Dittrich

Big Data Analytics Group

Universität des Saarlandes

Saarland Informatics Campus

Tel.: + 49 681 302 70141

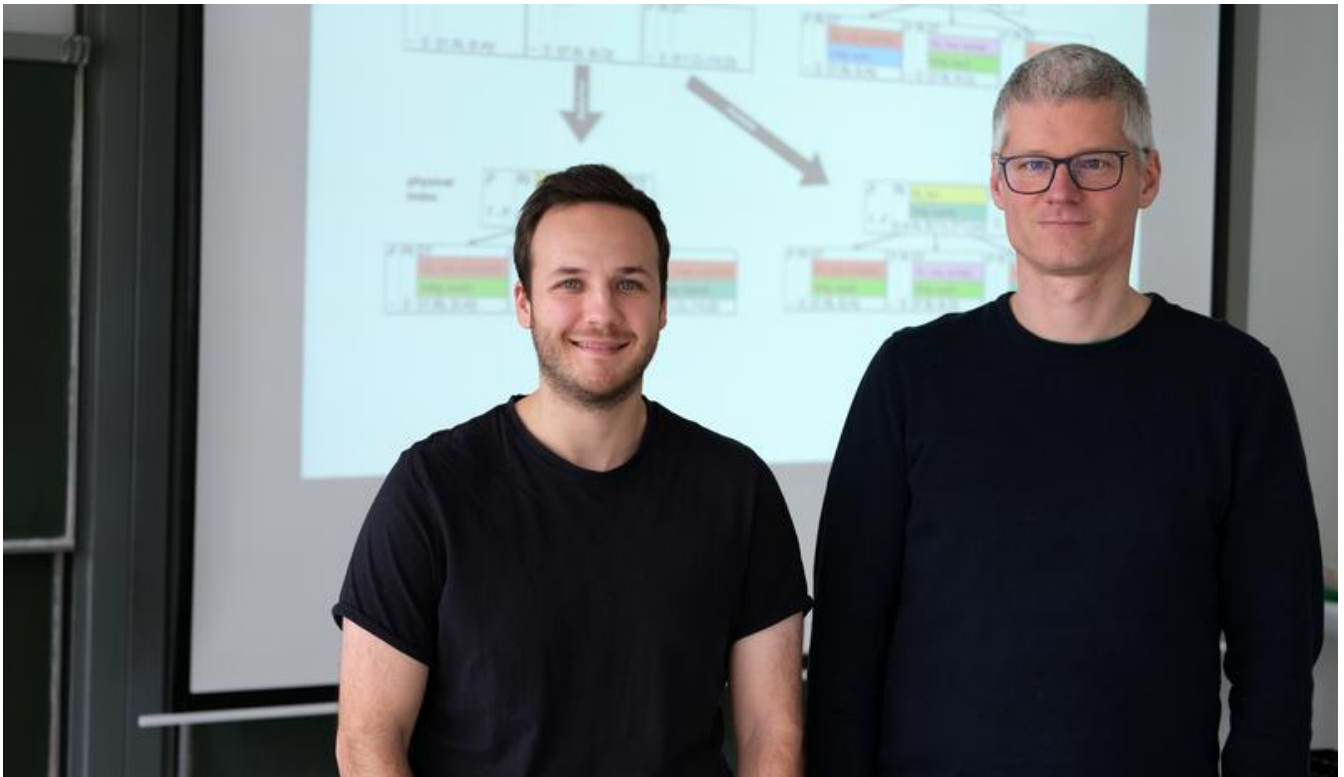
Email: dittrich@cs.uni-saarland.de

Originalpublikation:

Jens Dittrich, Joris Nix, Christian Schön: 'The next 50 Years in Database Indexing or: The Case for Automatically Generated Index Structures', Proceedings of the VLDB Endowment, Vol. 15(3): 527-540 (2021)

<https://www.vldb.org/pvldb/vol15/p527-dittrich.pdf>

URL zur Pressemitteilung: <https://gepris.dfg.de/gepris/projekt/513858547?language=en> - Information on DFG funding



Professor Jens Dittrich and PhD student Joris Nix want to "breed" highly efficient index structures.
Philipp Zapf-Schramm
SIC



Joris Nix is a PhD student in the Big Data Analytics group at Saarland University.
Philipp Zapf-Schramm
SIC