Extracting Notional Machines for Databases

Miedema, Daphne; Fletcher, George; Aivaloglou, Fenia; Busuttil, Leonard; Farinetti, Laura; Goodfellow, Martin; Guerrini, Giovanna; Haldeman, Georgiana; Pan, Yuhan; More Authors

DOI
10.1145/3724389.3731277

Publication date
2025

Document Version
Final published version

Published in
ITiCSE 2025: Proceedings of the 30th ACM Conference on Innovation and Technology in Computer Science Education V. 2

Citation (APA)
Miedema, D., Fletcher, G., Aivaloglou, F., Busuttil, L., Farinetti, L., Goodfellow, M., Guerrini, G., Haldeman, G., Pan, Y., & More Authors (2025). Extracting Notional Machines for Databases. In *ITiCSE 2025: Proceedings of the 30th ACM Conference on Innovation and Technology in Computer Science Education V. 2* (pp. 693-694). ACM. https://doi.org/10.1145/3724389.3731277

Extracting Notional Machines for Databases

Daphne Miedema*
d.e.miedema@uva.nl
University of Amsterdam
Amsterdam, the Netherlands

George Fletcher*
g.h.l.fletcher@tue.nl
Eindhoven University of Technology
Eindhoven, the Netherlands

Fenia Aivaloglou*
e.aivaloglou@tudelft.nl
Delft University of Technology
Delft, the Netherlands

Leonard Busuttil
leonard.busuttil@um.edu.mt
University of Malta
Msida, Malta

Laura Farinetti
laura.farinetti@polito.it
Politecnico di Torino
Turin, Italy

Martin Goodfellow
martin.h.goodfellow@strath.ac.uk
University of Strathclyde
Glasgow, Scotland

Giovanna Guerrini
giovanna.guerrini@unige.it
University of Genoa
Genoa, Italy

Georgiana Haldeman
ghaldeman@colgate.edu
Colgate University
Hamilton, New York, USA

Yuhan Pan
evelyn.pan@mail.utoronto.ca
University of Toronto
Toronto, Canada

Sujeeth Goud Ramagoni
sujeethgoud.ramagoni@marquette.edu
Marquette University
Milwaukee, Wisconsin, USA

Chandrika Satyavolu
jsatyav@ncsu.edu
North Carolina State University
Raleigh, North Carolina, USA

Raja Sooriamurthi
raja@cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Xiaoying Tu
xiaoying@cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Liviana Tudor
liviana.tudor@upb.ro
Petroleum-Gas University of Ploiesti
Ploiesti, Romania

Abstract

Database education is a cornerstone under many of the more popular topics in computer science such as machine learning and visualization. Although, in recent years, more fundamental research into database education has come out, there are many more ways in which it can be extended. Research on the practice of teaching databases, namely on the educational materials and explanations of teachers, can help us create new building blocks for fundamental research. This working group aims to collect and present notional machines of different types, for a wide range of database subtopics. These materials offer and updated context for database educators to design their courses from, as well as open up pathways of further research into database education.

ACM Reference Format:
Daphne Miedema, George Fletcher, Fenia Aivaloglou, Leonard Busuttil, Laura Farinetti, Martin Goodfellow, Giovanna Guerrini, Georgiana Haldeman, Yuhan Pan, Sujeeth Goud Ramagoni, Chandrika Satyavolu, Raja Sooriamurthi, Xiaoying Tu, and Liviana Tudor. 2025. Extracting Notional Machines for Databases. In *Proceedings of the 30th ACM Conference on Innovation and Technology in Computer Science Education V. 2 (ITiCSE 2025), June 27-July 2, 2025, Nijmegen, Netherlands.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3724389.3731277

*Working group leader.

ITiCSE 2025, June 27-July 2, 2025, Nijmegen, Netherlands

ACM ISBN 979-8-4007-1569-3/2025/06
https://doi.org/10.1145/3724389.3731277

1 Background

There is a growing demand for computer and data scientists, statisticians, and other STEM professionals who can manipulate data. Data processing, and especially data transformation, are at the core of data-driven projects. Much of the work of data scientists is inherently data processing work; this accounts for more than 70% of their time, which is significantly more than the time spent on other data pipeline tasks such as model selection, training, and deployment [2]. This also applies to AI projects, with enterprises spending the majority of their time and resources wrangling data [1]. Furthermore, these competencies do not apply only to data scientists; many STEM workers are also required to be proficient in their knowledge of computing for data processing [6].

Databases courses are the primary place in computer and data science curricula in teaching about data processing. There, students are taught about data modeling and data operations such as projections, aggregations, filtering, grouping, joining data collections, etc. From computing education research focusing on SQL, and specifically from studies on student errors [10] and misconceptions [7], we know that these topics are challenging for students to master.

To support knowledge transfer about challenging concepts, educators often employ notional machines. Notional machines are pedagogic devices that support explaining and understanding complex concepts through representations (such as visualizations or tool-supported representations of program executions) or analogies (such as concept metaphors, for example, discussing a programming variable as a label or a box) [3]. Notional machines have been researched for CT concepts such as sequences, repetition, and conditionals [5, 8], but not yet for database concepts.

2 Goals of the proposal

The primary objectives of the working group are *(i)* to develop a taxonomy/categorization of the notional machines identified through empirical methods and *(ii)* to evaluate to what extent the concept of a notional machine as currently defined for programming education can be applied to database education contexts. In our working group, we build on the work of Fincher et al. [3] and the categories of applied notional machines they defined: machine-generated representations, handmade representations, and analogies.

Insights into notional machines for database education have two major benefits. First of all, research into misconceptions and errors becomes more straightforward as a taxonomy presents a starting point for investigating thought processes. Second, the notional machines can provide a rich context for database teachers to utilize in their course (re-)designs.

3 Methodology

We start this project by updating and scoping the literature review done by Fincher et al. [3] in their WG on notional machines, specifically focusing on the period from 2020 until now.

Then, for the uncovering of notional machines, we gather materials of different types:

• Textbooks. A preliminary study of literature on database textbooks led to the list of books in Table 1, ordered from most mentioned to least mentioned and then sorted by recency. One of our proposed WG leaders has students working on identifying analogies in some of these textbooks, which provides a good starting point for this research.
• Education material, such as video lectures, handouts and visualization tools, used in the (introductory and advanced) database courses of the participants and/or their institutions. We could also include some lecture observations here, where feasible.
• MOOCs. MOOC analysis of notional machines on variables has been done by before by Van Der Werf et al. [12], we could follow a similar process.
• Popular tutorials on YouTube. One of our WG leaders has a student creating a dataset of YouTube tutorials on databases, from which we could analyze the most popular videos per topic.

We organize the materials by topic, separating into partitions that will be analyzed by different groups. This helps us identify patterns in the use of notional machines. The topics are:

- Query languages
- Information Retrieval and Data Mining (as part of data systems resources).
- Conceptual modeling
- The query processing pipeline
- Data storage and indexing
- Design theory and normalization

Textbook	Author	Ed.	Pub	Mentioned
Database Systems: A Practical Approach to Design, Implementation, and Management	Connolly et al.	6	2014	[4, 9, 11]
Database Management Systems	Ramakrishnan et al.	3	2002	[4, 11]
Database System Concepts	Silberschatz et al.	6	2011	[4, 11]
Database Systems: The Complete Book	Garcia-Molina et al.	2	2008	[4, 11]
Fundamentals of Database Systems	Elmasri et al.	7	2015	[9, 11]
Modern Database Management	Hoffer et al.	12	2015	[4, 9]
Data Modeling Essentials	Simsion et al.	3	2004	[11]
Learning SQL	Beaulieu	2	2009	[11]
Introduction to Database Systems	Date	8	2003	[4]
Introduction to Database Systems	Bressan et al.	1	2005	[4]
A first course in database systems	Ullman et al.	3	2007	[4]
Readings in Database Systems (the red book)	Hellerstein et al.	4	2005	[4]
Essentials of Database Management	Hoffer et al.		2014	[9]
Database Processing: Fundamentals, Design, and Implementation	Kroenke et al.	14	2016	[9]

Table 1: Database textbooks mentioned in earlier research

References

[1] 2022. Artificial Intelligence (AI): 7 trends to watch for in 2022. *The Enterprisers Project* (January 2022). https://enterprisersproject.com/article/2022/1/artificial-intelligence-ai-7-trends-watch-2022.

[2] Anacoda. 2021. 2021 State of Data Science. https://www.anaconda.com/state-of-data-science-2021.

[3] Sally Fincher, Johan Jeuring, Craig S. Miller, Peter Donaldson, Benedict du Boulay, Matthias Hauswirth, Arto Hellas, Felienne Hermans, Colleen Lewis, Andreas Mühling, Janice L. Pearce, and Andrew Petersen. 2020. Notional Machines in Computing Education: The Education of Attention. In *Proceedings of the Working Group Reports on Innovation and Technology in Computer Science Education* (Trondheim, Norway) *(ITiCSE-WGR '20)*. Association for Computing Machinery, New York, NY, USA, 21–50. https://doi.org/10.1145/3437800.3439202

[4] Muhammad Ishaq, Adnan Abid, Muhammad Shoaib Farooq, Muhammad Faraz Manzoor, Uzma Farooq, Kamran Abid, and Mamoun Abu Helou. 2023. Advances in database systems education: Methods, tools, curricula, and way forward. *Education and Information Technologies* 28, 3 (2023), 2681–2725.

[5] Cruz Izu, Carsten Schulte, Ashish Aggarwal, Quintin Cutts, Rodrigo Duran, Mirela Gutica, Birte Heinemann, Eileen Kraemer, Violetta Lonati, Claudio Mirolo, and Renske Weeda. 2019. Fostering Program Comprehension in Novice Programmers - Learning Activities and Learning Trajectories. In *Proceedings of the Working Group Reports on Innovation and Technology in Computer Science Education* (Aberdeen, Scotland Uk) *(ITiCSE-WGR '19)*. Association for Computing Machinery, New York, NY, USA, 27–52. https://doi.org/10.1145/3344429.3372501

[6] Hyewon Jang. 2016. Identifying 21st century STEM competencies using workplace data. *Journal of science education and technology* 25, 2 (2016), 284–301.

[7] Daphne Miedema, Efthimia Aivaloglou, and George Fletcher. 2021. Identifying SQL Misconceptions of Novices: Findings from a Think-Aloud Study. In *Proceedings of the 17th ACM Conference on International Computing Education Research* (Virtual Event, USA) *(ICER 2021)*. Association for Computing Machinery, New York, NY, USA, 355–367. https://doi.org/10.1145/3446871.3469759

[8] Kathryn M. Rich, Carla Strickland, T. Andrew Binkowski, Cheryl Moran, and Diana Franklin. 2018. K–8 Learning Trajectories Derived from Research Literature: Sequence, Repetition, Conditionals. *ACM Inroads* 9, 1 (jan 2018), 46–55. https://doi.org/10.1145/3183508

[9] Toni Taipalus. 2020. The effects of database complexity on SQL query formulation. *Journal of Systems and Software* 165 (2020), 110576.

[10] Toni Taipalus, Mikko Siponen, and Tero Vartiainen. 2018. Errors and Complications in SQL Query Formulation. *ACM Transactions on Computing Education* 18, 3 (2018).

[11] Cynthia Taylor and Saheel Sakharkar. 2019. '); DROP TABLE textbooks;– An Argument for SQL Injection Coverage in Database Textbooks. In *Proceedings of the 50th ACM technical symposium on computer science education*. 191–197.

[12] Vivian Van Der Werf, Min Yi Zhang, Efthimia Aivaloglou, Felienne Hermans, and Marcus Specht. 2023. Variables in Practice. An Observation of Teaching Variables in Introductory Programming MOOCs. In *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1* (Turku, Finland) *(ITiCSE 2023)*. Association for Computing Machinery, New York, NY, USA, 208–214. https://doi.org/10.1145/3587102.3588857