VUB SCIENCES & BIOENGINEERING SCIENCES

The Research Group

Artificial Intelligence Lab

has the honor to invite you to the public defense of the PhD thesis of

Maxime Jakubowski

to obtain the degree of Doctor of Sciences

Joint PhD with Universiteit Hasselt

Title of the PhD thesis:

Shapes Constraint Language:
Formalization, Expressiveness, and Provenance

Promotors:
Prof. dr. Bart Bogaerts (VUB)
Prof. dr. Jan Van den Bussche (UHasselt)

The defense will take place on

Friday, May 31, 2024 at 4:00 p.m.

UHasselt Campus Diepenbeek, auditorium H1

Members of the jury

Prof. dr. Marc Gyssens (UHasselt, chair)
Prof. dr. Bas Ketsman (VUB, secretary)
Prof. dr. Stijn Vansummeren (UHasselt)
Prof. dr. Anastasia Dimou (KULeuven)
Dr. Shqiponja Ahmetaj (TU Wien, Austria)
Dr. Slawek Staworko (RelationalAI, US)

Curriculum vitae

Maxime Jakubowski obtained his master's degree in Artificial Intelligence from Maastricht University in 2020. After graduating, he began his PhD at UHasselt, within the Data Science Institute under the supervision of Prof. Dr Jan Van den Bussche. It soon became a joint doctorate with the AI Lab at VUB under the supervision of Prof. Dr Bart Bogaerts. This collaboration was established through the joint project of the Flanders AI Research Program, which includes this PhD.
The research focuses on the intersection of database theory and artificial intelligence, using formal logical techniques to better understand existing computer languages.

Abstract of the PhD research

The Shapes Constraint Language (SHACL) is a schema language for expressing structural constraints on RDF graphs, where constraints on nodes are called “shapes”. We study several aspects of this language. First, recent formalizations show a striking resemblance with description logics. We build on top of these formalizations to come to an understanding of SHACL as a logic. Furthermore, because the SHACL specification only defines semantics for non-recursive SHACL, some efforts have been made to allow recursive SHACL schemas. We argue that for defining and studying semantics of recursive SHACL, lessons can be learned from research in non-monotonic reasoning. We look at the proposed semantics from the literature and compare it with techniques from well-established research from non-monotonic reasoning.

Next, SHACL expressions can use three fundamental features that are not so common in similar logics. These features are equality tests; disjointness tests; and closure constraints. It is not clear how the presence of these non-standard features impacts the expressiveness of SHACL. We show that each of the three features is primitive: using the feature, one can express boolean queries that are not expressible without using the feature. We also show that the restriction that SHACL imposes on allowed targets is inessential, as long as closure constraints are not used. In addition, we show that enriching SHACL with “full” versions of equality tests, or disjointness tests, results in a strictly more powerful language.

Lastly, we propose provenance semantics for SHACL. We propose the notion of neighborhood of a node satisfying a given shape in a graph. This neighborhood is a subgraph and provides data provenance for the node and the given shape. We establish a correctness property for the obtained provenance mechanism, by proving that neighborhoods adhere to the Sufficiency requirement articulated for provenance semantics for database queries. We compare neighborhoods with SPARQL queries. We discuss implementation strategies for computing neighborhoods, and present initial experiments demonstrating that our ideas are feasible.