Abstract

In the thesis we study Tuple Generating Dependencies (TGDs) as a formalism for representing ontologies.

A TGD is a formula of first order logic of the following syntax

$$\forall \bar{x}, \bar{y}\ \Phi(\bar{x}, \bar{y}) \Rightarrow \exists \bar{z}\ \Psi(\bar{z}, \bar{y}),$$

where Φ and Ψ are conjunctive queries. We present four papers where TGD is the core notion.

In the first and the second paper we study ontologies expressed as a set of TGDs. We show Finite Controllability (FC) for some classes of such ontologies. Namely the Sticky Logic and set of TGDs over binary signatures which have BDD property. The FC property is a variant of Finite Model Property (FMP) for heterogeneous formulas. The difference between FC and FMP is important here, because we work with two different languages: The language of TGDs and the language of queries.

In the third paper we study the Chase structure. A TGD rule can be seen as a rewriting rule: whenever in a given structure S there is a substructure isomorphic to $\Phi(\bar{x}, \bar{y})$, add to S a fresh copy of $\Psi(\bar{z}, \bar{y})$. The Chase is a fixpoint of such operation. The chase procedure can be used as a method for data integration. Before we use the chase procedure in practice, one has to have a guarantee that the chase procedure will terminate in a finite number of steps. We show that it is undecidable to test whether for given set of TGDs, chase procedure will terminate for all databases.

In the fourth paper we show undecidability of unrestricted determinacy of conjunctive queries. Surprisingly, this is also about TGDs.

Streszczenie

W rozprawie zajmujemy się zastosowaniami Tuple Generating Dependencies (TGD) w teorii baz danych. Są to zastosowania TGD jako formalizmu do wyrażania ontologii.

TGD jest zdaniem w logice pierwszego rzędu następującej postaci

$$\forall \bar{x}, \bar{y}\ \Phi(\bar{x}, \bar{y}) \Rightarrow \exists \bar{z}\ \Psi(\bar{z}, \bar{y}),$$

gdzie Φ i Ψ są zapytaniami koniunkcyjnymi.

W dwóch pierwszych pracach składających się na rozprawę pokazujemy, że pewne klasy zbiorów TGD mają własność Finite Controllability. Jest to pewien wariant własności modelu skończonego dla heterogenicznych formuł. Różnica między Finite Controllability oraz własnością modelu skończonego jest istotna, gdyż pracujemy z dwoma różnymi językami: język TGD i język zapytań.

Trzecia praca bada strukturę Chase. TGD możemy traktować jako reguły operacyjne: jeśli w strukturze istnieje podstruktura izomorficzna z $\Phi(\bar{x}, \bar{y})$, to dodaj do struktury $\Psi(\bar{z}, \bar{y})$. Chase jest punktem stałym takiej operacji. Procedura Chase może być używana jako sposób intergracji informacji. Aby taka forma migracji danych była użyteczna w praktyce, musimy mieć pewność że ten punkt stały obliczy się w skończonej liczbie kroków. Wynikiem tej pracy jest twierdzenie, że nierozstrzygalnym jest pytanie o skończoność Chase.

Ostatnia praca bada problem determinacji zapytań koniunkcyjnych. Pokazujemy nierozstrzygalność w przypadku gdy również nieskończone struktury traktujemy jako bazy danych. Okazuje się, że ten problem też dotyczy własności TGD.