Piotr Witkowski

Complexity of Some Logics Extended with Monadic Datalog Programs

PhD Thesis
Supervisor: prof. Witold Charatonik

Institute of Computer Science
University of Wrocław
Wrocław 2014

Piotr Witkowski

Złożoność pewnych logik z monadycznymi programami Datalogowymi

Rozprawa doktorska napisana pod kierunkiem
prof. Witolda Charatonika

Instytut Informatyki
Uniwersytet Wrocławski
Wrocław 2014

Abstract

Since 1930s, when Alonzo Church and Alan Turing proved that the satisfiability problem for first-order logic is undecidable, much effort was put to find decidable subclasses of this logic. The most prominent decidable cases include the guarded fragment GF, the two-variable fragment FO^2 and the universal fragment (also called the Bernays-Schönfinkel class). Still, all these fragments have a limited expressive power and a lot of effort is being put to extend these fragments beyond the first-order logic while preserving decidability. One of numerous ideas was to combine the two-variable fragment of the Bernays-Schönfinkel class ($\forall\forall$ for short) with least fixed points expressible by certain monadic Datalog programs. The obtained logic, called the Bernays-Schönfinkel class with Datalog, was shown decidable in 2NExpTime. It was powerful enough to express some linked data structures (lists, trees etc.) that appear in context of imperative pointer programs verification, to capture semantics of programs without loops. This way it was possible to solve the bounded model checking for discovery of dangling pointers in pointer programs.

In this thesis we generalize the notion of logics with Datalog and prove lower and upper complexity bounds of finite satisfiability problem for some of them. We show that a combination of the two-variable logic with counting quantifiers (C^2) and Datalog is in NExpTime by a polynomial reduction to C^2. Another combination of C^2 with Datalog is reduced to a new logic, C^2 with trees. All combinations of universal two-variable logics with Datalog are shown NExpTime-hard, all combinations of universal three-variable logics — undecidable. The first and the third of these results imply that the (two-variable) Bernays-Schönfinkel class with Datalog is NExpTime-complete.

To motivate our research we argue that the obtained decidable logics improve expressivity of bounded model checking and enable to take the Hoare approach to pointer program verification. In particular we apply the bounded model checking not only do discovery of dangling pointers, but also to variable aliasing, structure intersection and memory leak problems. In the context of Hoare approach we present an algorithm that generates verification conditions that are instances of the entailment problem in logics with Datalog.

A crucial new tool developed in this thesis is a NExpTime satisfiability decision procedure for C^2 with trees, i. e., C^2 on finite structures where two binary predicates are interpreted as forests. A forest is just a disjoint set of trees and an arbitrary number of other binary and unary predicates are allowed. This logic may be of an independent interest since it is a decidable non first order extension of C^2. In particular, it allows to define two successors of finite linear orders, the property that a graph of a bounded degree is connected and even linear cardinality constraints on the interpretations of unary predicates.

Streszczenie

W trzeciej dekadzie XX wieku Alonzo Church i Alan Turing udowodnili, że problem spełnialności formuł logiki pierwszego rzędu jest nierozstrzygalny. Od tego czasu wiele wysiłku włożono w poszukiwania rozstrzygalnych fragmentów tej logiki. Znaleziono wiele takich fragmentów, do najbardziej znaczących należą logika ze strażnikami GF, logika z dwiema zmiennymi FO^2 oraz logika uniwersalna (zwana też fragmentem Bernaysa-Schönfinkla). Logiki te mają jednak niewystarczającą siłę wyrazu, podejmuje się więc próby ich rozszerzania o własności niewyrażalne w logice pierwszego rzędu, przy zachowaniu rozstrzygalności problemu spełnialności. Jedną z takich prób było rozszerzenie uniwersalnego fragmentu logiki z dwiema zmiennymi o monadyczne programy Datalogowe, co umożliwiło wyrażanie własności stałopunktowych, przydatnych np. w weryfikacji imperatywnych programów ze strukturami wskaźnikowymi. Udowodniono, że problem spełnialności otrzymanej logiki należy do klasy 2NExpTime oraz wykazano, że proste struktury wskaźnikowe, takie jak listy i drzewa, są w niej definiowalne. Cechy te pozwoliły na zastosowanie logiki Bernaysa-Schönfinkla z Datalogiem, bo tak nazwano tę logikę, do rozwiązania problemu ograniczonej weryfikacji modelowej (ang. *bounded model checking*) dla programów ze strukturami wskaźnikowymi.

Niniejsza rozprawa rozwija pojęcie logik z Datalogiem oraz pokazuje dolne i górne ograniczenia na złożoność problemu spełnialności dla niektórych z nich, głównie dla logik z dwiema zmiennymi. Udowadniamy, że złożoność kilku rozszerzeń logiki z dwiema zmiennymi i kwantyfikatorami zliczającymi (C^2) o programy Datalogowe jest w klasie NExpTime. Posługujemy się dwiema ściśle ze sobą związanymi wielomianowymi redukcjami. Pierwsza z nich sprowadza problem spełnialności dla logiki z Datalogiem do problemu spełnialności fragmentu C^2, co pokazuje że rozważana logika z Datalogiem jest tylko syntaktycznym rozszerzeniem fragmentu C^2. Druga używa nowego rozstrzygalnego formalizmu, nazwanego logiką C^2 z drzewami. Dowodzimy, że wszystkie rozważane kombinacje logik z dwiema zmiennymi i Datalogiem są trudne w klasie NExpTime, natomiast wszystkie rozszerzenia uniwersalnego fragmentu z trzema zmiennymi o programy Datalogowe są nierozstrzygalne. W szczególności, powyższe wyniki implikują NExpTime zupełność logiki Bernaysa-Schönfinkla z Datalogiem.

Zainteresowanie logikami z Datalogiem uzasadniamy zastosowaniami w weryfikacji programów ze wskaźnikami. Pokazujemy na przykładach, że badane w tej pracy rozstrzygalne logiki umożliwiają weryfikację modelową większej klasy problemów niż fragment Bernaysa-Schönfinkla z Datalogiem. Pokazujemy również szkielet automatycznego systemu weryfikacji programów ze wskaźnikami metodą niezmienników Hoare'a. System ten w istotny sposób korzysta z rozstrzygalności problemu spełnialności dla pewnej logiki z Datalogiem.

Wzmiankowana powyżej logika C^2 z drzewami to logika C^2 interpretowana w klasie skończonych struktur, w których dwa wyróżnione predykaty binarne reprezentują lasy rozłącznych drzew o zadanym z góry stopniu; nie ograniczamy przy tym wystąpień innych unarnych i binarnych predykatów ani stałych. W logice tej można wyrazić takie własności wyższego rzędu jak spójność grafów o ograniczonym stopniu, następniki dwóch skończonych porządków liniowych, a nawet pewne liniowe ograniczenia na rozmiar struktur. Z tego względu logika C^2 z drzewami wydaje się interesująca sama w sobie, a nie tylko jako narzędzie służące rozstrzyganiu problemu spełnialności logik z Datalogiem. Pokazujemy NExpTime zupełność tej logiki.