Available online at www.sciencedirect.com

ScienceDirect

Electronic Notes in Theoretical Computer Science 328 (2016) 3–16

Electronic Notes in Theoretical Computer Science

www.elsevier.com/locate/entcs

On Regular Paths with Counting and Data Tests

Everardo Bárcenas[1]

CONACYT
Universidad Veracruzana
Xalapa, Veracruz, Mexico

Edgard Benítez-Guerrero[2]

Facultad de Estadística e Informática
Universidad Veracruzana
Xalapa, Veracruz, Mexico

Jesús Lavalle[3]

Benemérita Universidad Autónoma de Puebla
Instituto Nacional de Astrofísica, Óptica y Electrónica
Puebla, Puebla, Mexico

Abstract

Regular path expressions represent the navigation core of the XPath query language for semi-structured data (XML), and it has been characterized as the First Order Logic with Two Variables (FO^2). Data tests refers to (dis)equality comparisons on data tree models, which are unranked trees with two kinds of labels, propositions from a finite alphabet, and data values from a possibly infinite alphabet. Node occurrences on tree models can be constrained by counting/arithmetic constructors. In this paper, we identify an EXPTIME extension of regular paths with data tests and counting operators. This extension is characterized in terms of a closed under negation Presburger tree logic. As a consequence, the EXPTIME bound also applies for standard query reasoning (emptiness, containment and equivalence).

Keywords: Modal Logics, XPath, Automated Reasoning, Data Trees, Counting

1 Introduction

XPath is a W3C standard query language for semi-structured data (XML), and it also takes an important role in many XML technologies, such as, XProc, XSLT,

[1] iebarcenaspa@conacyt.mx
[2] edbenitez@uv.mx
[3] jlavalle@cs.buap.mx

http://dx.doi.org/10.1016/j.entcs.2016.11.002

and XQuery [19,2]. The navigation core of XPath, also known as regular path queries, has been recently characterized by the First Order Logic of Two Variables (FO^2) [19]. Models for this logic are unranked trees, where nodes are labeled by propositions from a finite alphabet. Data tests, also known as data joins in databases community, on XPath queries are expressions of the forms $\rho_1 \equiv \rho_2$ and $\rho_1 \not\equiv \rho_2$. These expressions hold whenever data values (propositions from an infinite alphabet) contained in path ρ_1 are (dis)equal to data values contained in path ρ_2, respectively. Another important constructors on XPath queries concerns counting: $\rho_1 \# \rho_2$, where $\# \in \{\leq, >, =, \neq\}$. These expressions hold whenever the number of nodes denoted by ρ_1 and ρ_2 satisfies constraint $\#$. There are several recent works studying regular path extensions with either data tests or counting [14,4,13,12,2]. However, as far as we know, the current work represent the first study on regular path extensions concerning both constructors, data tests and counting. More precisely, we give a characterization of a regular paths with data test with respect to constants $\rho \equiv k$ ($\rho \not\equiv k$) and with counting operators on children paths. For this characterization we use a modal tree logic equipped with a fixed point operator, converse modalities and Presburger arithimetic constraints [3]. Due to this characterization, the EXPTIME bound from the logic is imported to standard query reasoning (emptiness, containment, and equivalence) with counting and data tests.

1.1 Related works

There are several extensions of FO^2 with data tests [16,8,6,7]. In [16], $FO^2(<, +1, \equiv)$ for data trees is introduced: $<$ stands for descendants and following sibling relations, $+1$ refers to child and next sibling relations, and $\equiv$ is a binary predicate for data tests. Decidability, without any complexity analysis, for $FO^2(<, +1, \equiv)$ in data trees is first shown by a reduction to the reachability problem of a counter tree automata model. Previously in [6], the same result was obtained for data words (one branched tree), more precisely, $FO^2(<, +1, \equiv)$ for data words was shown decidable by a reduction to reachability problem of Petri nets. A 3NEXPTIME upper bound for $FO^2(<, +1, \equiv)$ is implied in [6]. Even earlier in [7], $FO^2(+1, \equiv)$ for trees was introduced and shown decidable also in 3NEXPTIME. Regarding model theoretic results of extensions of FO with data models, in [8] several cases are studied: data words, data trees and data graphs. In another direction, regarding regular paths (XPath navigation core), it is well know data test on full navigation regular paths is undecidable [13]. Several fragments (downward, forward, transitive) of regular path expressions with data tests are studied [11,15,12,13,14]. With their corresponding complexity ranging from EXPTIME to non elementary. Contrastingly, in this paper, instead of restricting navigation on queries, we study the full navigation (children, parents, following and previous sibling, descendants and ancestros) regular path expressions, but we restrict data tests to constants only, that is, expressions of the form $\rho \equiv k$ ($\rho \not\equiv k$).

Regarding regular paths with counting, there are several recent studies [19,2,4]. In [19], it was show the extension of regular paths with counting is in general undecidable. EXPTIME fragments (counting with respect to constants) were later