Journal of Symbolic Computation 67 (2015) 42–67

Contents lists available at ScienceDirect

Journal of Symbolic Computation

www.elsevier.com/locate/jsc

Regular expression order-sorted unification and matching

Temur Kutsia [a], Mircea Marin [b]

[a] RISC, Johannes Kepler University Linz, Austria
[b] West University of Timişoara, Romania

ARTICLE INFO

Article history:
Received 13 April 2014
Accepted 12 June 2014
Available online 7 August 2014

Keywords:
Unification
Matching
Ordered-sorts
Regular expressions

ABSTRACT

We extend order-sorted unification by permitting regular expression sorts for variables and in the domains of function symbols. The obtained signature corresponds to a finite bottom-up unranked tree automaton. We prove that regular expression order-sorted (REOS) unification is of type infinitary and decidable. The unification problem presented by us generalizes some known problems, such as, e.g., order-sorted unification for ranked terms, sequence unification, and word unification with regular constraints. Decidability of REOS unification implies that sequence unification with regular hedge language constraints is decidable, generalizing the decidability result of word unification with regular constraints to terms. A sort weakening algorithm helps to construct a minimal complete set of REOS unifiers from the solutions of sequence unification problems. Moreover, we design a complete algorithm for REOS matching, and show that this problem is NP-complete and the corresponding counting problem is #P-complete.

1. Introduction

Order-sorted algebra has been introduced in Goguen (1978), motivated by searching a better way to treat errors in abstract data types and to speed up certain theorem proving methods. In order-sorted algebras, variables and arguments of function symbols range over certain subsets of the universe of terms, specified by the sorts. Walther (1988) gave a syntactic unification algorithm for order-sorted terms, and characterized the relationship between sort hierarchies and the cardinality of minimal complete sets of unifiers.

Schmidt-Schauß (1989) extended Walther's work, permitting term declarations in sorted signatures. He studied syntactic unification algorithms and their complexities in various kinds of signatures.

http://dx.doi.org/10.1016/j.jsc.2014.08.002

Table 1
Comparison with the order-sorted language from Walther (1988).

The language in Walther (1988)	The language in this paper
The set of basic sorts $\mathcal{B}$, partially ordered with $\preceq$.	The finite set of basic sorts $\mathcal{B}$, partially ordered with $\preceq$.
Sets of variables $\mathcal{V}_{\mathsf{s}}$ for each $\mathsf{s} \in \mathcal{B}$.	Sets of variables $\mathcal{V}_{\mathsf{R}}$ for each $\mathsf{R} \in \mathcal{R}_{\mathcal{B}}$.
Sets of function symbols $\mathcal{F}_{\mathsf{w} \to \mathsf{s}}$ for $\mathsf{w} \in \mathcal{B}^*$, $\mathsf{s} \in \mathcal{B}$.	Sets of function symbols $\mathcal{F}_{\mathsf{R} \to \mathsf{s}}$ for $\mathsf{R} \in \mathcal{R}_{\mathcal{B}}$, $\mathsf{s} \in \mathcal{B}$.
The sets of function symbols and variables are pairwise disjoint.	The sets of function symbols are not required to be disjoint.

He also gave a complete procedure for sorted equational unification. Frisch and Cohn (1992) gave an abstract version of the sorted unification algorithm, independent of the sorted language being used, and reformulated Schmidt-Schauß's results in this setting. Uribe (1992) proved decidability of sorted unification in the so called semi-linear sort theories: A problem which Schmidt-Schauß left open. Weidenbach (1996) further generalized Schmidt-Schauß's and Uribe's results for syntactic sorted unification to more complex sort theories.

Since the original work by Goguen, several variants of the order-sorted algebra have been proposed, see Goguen and Diaconescu (1994) for a survey. Some of these variants permit overloaded function symbols. A desirable property of overloaded order-sorted algebras is the existence of a least sort for terms. Goguen and Meseguer (1992) gave conditions on the signature to guarantee the existence of such a sort. Equational unification algorithms for overloaded order-sorted algebras have been proposed in Kirchner (1988), Meseguer et al. (1989), Schmidt-Schauß (1989), Boudet (1992), Hendrix and Meseguer (2012).

All the above mentioned work was done for order-sorted algebras over *ranked signatures,* where function symbols have a fixed arity. Comon (1989) observed an interesting relation between such signatures and tree automata: A finite ranked order-sorted signature is a finite bottom-up ranked tree automaton. Based on this observation, Comon and Delor (1994) used some strong properties of regular languages (decidability of emptiness and finiteness, stability under intersection, union and complement) to bring together the order-sorted framework and simplification of first-order equational formulas.

In this paper, we move from ranked to unranked signatures. Unranked terms/trees are commonly used as an abstract model of XML documents, program schemata, multithreaded recursive program configurations with an unbounded number of parallel processes, variadic functions in programming languages, etc. Rewriting, programming, model checking, knowledge representation techniques over unranked expressions have also been explored. Solving equations in one form or another is a fundamental problem in these applications. This is the problem we address in this paper.

More precisely, we generalize unification from ranked order-sorted terms without overloading to unranked order-sorted terms with overloading. Our sorts for variables and for function domains are described by regular expressions over basic sorts. Table 1 shows the detailed comparison of our language with the one in Walther (1988). The basic sorts in both papers are partially ordered. We consider the set $\mathcal{R}_{\mathcal{B}}$ of regular expressions over a poset $(\mathcal{B}, \preceq)$ of basic sorts, extend the partial order $\preceq$ to $\mathcal{R}_{\mathcal{B}}$, and, like Walther, restrict ourselves to syntactic unification.

We abbreviate the regular expression order sorts used in the current paper as REOS. To guarantee the existence of a least sort, we extend the condition of *preregularity* defined for ranked order-sorted signatures in Goguen and Meseguer (1992) to REOS signatures. The *finite overloading property* of the REOS signature (the same function symbol can belong only to finitely many different sets of function symbols) guarantees that a least sort is effectively computable.

Table 1 reveals that our variables have regular expression sorts, thus they may be instantiated with term sequences by sort-preserving substitutions. The problem of unification in an unsorted language where variables stand for term sequences (sequence unification, SEQU) has been studied earlier, see, e.g. Kutsia (2007) and the discussion on related work thereof. Our work can be seen as