Towards Clean Amalgamation of Logic Programs with External Procedures

by

Staffan Bonnier
Jan Maluszynski

RESEARCH REPORT
LOGPRO, January 1988

Postaddress:
Institutionen för datavetenskap
Universitetet i Linköping och
Tekniska Högskolan
581 83 Linköping

Mailing address:
Department of Computer and Information Science
Linköping University
S-581 83 Linköping, Sweden
PhD theses:

* Available at: University of Microfilms Intl., 300 N. Zeeb Road, Ann Arbor, MI 48106, USA.

* No 97 Andrzej Lingas: Advances in Minimum Weight Triangulation, 1983.
* No 109 Peter Fritzson: Towards a Distributed Programming Environment based on Incremental Compilation, 1984.


Licentiate of engineering Theses:


* No 73 Ola Strömfor: A Structure Editor for Documents and Programs. 1986.
* No 108 Rober Bilos: Incremental Scanning and Token-based Editing. 1987
* No 113 Ralph Rönquist: Network and Lattice Based Approaches to the Representation of Knowledge. 1987.
* No 118 Mariam Kamkar, Nahid Shahmehri: Affect-Chaining in Program Flow Analysis Applied to Queries of Programs. 1987
Towards Clean Amalgamation of Logic Programs with External Procedures

by

Staffan Bonnier

Jan Maluszynski

The paper presents a clean approach to the amalgamation of logic programming with external functional procedures. Both the logical semantics and the operational semantics of the amalgamated language are outlined. The operational semantics is based on an incomplete E-unification algorithm which we call S-unification. It is suggested to use the abstract interpretation technique for identifying classes of goals for which the approach is complete. For this purpose a domain of abstract terms is defined, and an abstract unification algorithm used for a compile-time check is developed.
TOWARDS A CLEAN AMALGAMATION OF LOGIC PROGRAMS WITH EXTERNAL PROCEDURES

Staffan Bonnier, Jan Maluszynski
Department of Computer and Information Science, Linköping University
S - 581 83 Linköping, Sweden

ABSTRACT

The paper presents a clean approach to amalgamation of logic programs with external functional procedures. Both the logical semantics and the operational semantics of the amalgamated language are outlined. The operational semantics is based on an incomplete E-unification algorithm which we call S-unification. It is suggested to use the abstract interpretation technique for identifying classes of goals for which the approach is complete. For this purpose a domain of abstract terms is defined, and an abstract unification algorithm used for a compile-time check is developed.

1. INTRODUCTION

This paper presents a systematic approach to the problem of amalgamating logic programs with external procedures. The motivation for this is twofold:

1. To allow for re-using of existing (possibly imperative) software while still preserving the declarative nature of the top-level logic programs.

2. To allow for use of functions in logic programming whenever it is more natural than expressing a function as a relation.

In recent years there have been a number of suggestions concerning combination of functional and logic programming in a single framework (see e.g. [GL] or [BL] for a survey). The approaches can be classified as:

1. Integrating existing programming languages and logic programs (well-known examples of this type are LOGLISP [RS], QLOG [Kom], POPLOG [MH] and APPLOG [Coh]).

2. Construction of new languages which allow one to define functions and relations and to combine functional and relational definitions. (well-known examples are EQLOG [GM 86], LEAF [BBLM] and FUNLOG [SY]).
The main objective within the first approach is often to give access from logic programs to specific features of the underlying programming language, or programming environment. This aspect is usually more important than concern about the declarative semantics of the amalgamation. It may be rather difficult to give such a semantics if low-level features of the underlying system are accessible in the resulting language. On the other hand, within the second approach the objective is often to have a clean logical semantics.

This paper combines the objectives of both approaches by putting some restrictions. The assumption is that the language of functional procedures is given, but we are not specific about its definition, so that our approach could be applied for any language. The task is to integrate the external procedures with logic programs in such a way that the amalgamation has a clean logical semantics and a reasonably efficient operational semantics, as complete as possible.

The basis for construction of an interface between the two different systems is the assumption that terms are their common data structures. A call of a functional procedure is itself a term. Since its execution is assumed to return a term, we can view the underlying programming system as a term rewriting system. This permits the use of the theory of logic programming with equality (see e.g. [JLM]) to give a clean declarative semantics of the amalgamated language, and for application of $E$-unification in its interpreter. However, since we are not specific about the language of the functional procedures, we have no access to the rewrite rules used by the system, and we cannot use them for construction of $E$-unifiers.

Our solution to this problem is an incomplete $E$-unification algorithm, called $S$-unification, outlined first in [LM]. It has the property that whenever it succeeds, the result is a complete set of $E$-unifiers of the arguments (which due to our restrictions is a singleton). It may also fail or report that it is not able to solve the problem of $E$-unification for the given arguments. If the algorithm fails the actual arguments have no $E$-unifier. In [LBM] we give a formal presentation of the $S$-unification algorithm and we prove its properties stated above.

This paper uses the results to develop a sufficient condition for the safe use of amalgamated programs. A compile-time check is presented which for a given amalgamated program $P$ and a class of goals can show that the execution of $P$ with any goal of the class will not abort due to the "do not know" outcome of $S$-unification.

2. THE LANGUAGE
The main feature of the syntax is the partition of functors and predicate symbols into
internal and external symbols. The role of the external symbols is to denote the external procedures to be used in logic programs. The alphabet is, as usual, the union of three disjoint sets of symbols:

① Σ, the set of function symbols (or functors).

② Π, the set of predicate symbols.

③ VAR, the set of variables.

To each function symbol and predicate symbol we assign an arity. Σ partitions into two disjoint sets: one of internal functors (or constructors) and one of external functors. Similarly Π partitions into the three sets of internal predicate symbols, external predicate symbols and { = }, where '=' is a distinguished equality symbol of arity two. The terms are built in the standard way over VAR and Σ. Terms built only from variables and constructors are called structures. The atoms are constructed by filling the argument places of the predicate symbols with terms. In particular, the atoms constructed from '=' are called equations. By a program clause we mean a construction of the form H := B₁, ..., Bₙ where H, called the head of the clause, is an atom built from an internal predicate symbol, and B₁, ..., Bₙ is a possibly empty sequence of atoms, called the body of the clause. By a goal clause we mean a program clause with no head. Finally we define an (amalgamated) program to be a finite set of program clauses.

3. THE LOGICAL SEMANTICS

Our main idea is to allow the use of functional procedures written in any language (be it assembler or be it ML) in logic programs. However, in order to assign a clean formal semantics to the amalgamation and to make it work operationally, some restrictions on the permitted procedures are needed. First of all we restrict ourselves to procedures working on ground (i.e. variable free) terms. It is assumed that whenever a functional procedure is called with all its arguments instantiated to ground terms of the appropriate sort, it terminates and returns a ground term as result. In order for this term to be completely reduced, we also require it to be a structure. In the sequel we distinguish a special class of functional procedures which return truth values (succeed or fail). They will be called tests. Each external functor f in Σ is the name of one non-test procedure, with n arguments, where n is the arity of f. Similarly the external predicate symbols are names of test procedures.

Hence s ( t₁, ..., tₙ ) denotes the application of the procedure named by s to the terms t₁, ..., tₙ, while η ( s ( t₁, ..., tₙ ) ) denotes the result of its evaluation.
The restriction of \( \eta \) to non-test procedures will be denoted by \( \eta_{ nt} \), and its restriction to test procedures will be denoted by \( \eta_{ ts} \). In the rest of the paper we will consider the extension of \( \eta \) obtained by extending \( \eta_{ nt} \) to all ground terms, including structures:

For every \( n \)-ary constructor \( c \) and ground terms \( t_1, \ldots, t_n \), \( \eta_{ nt}(c(t_1, \ldots, t_n)) \) is defined to be \( c(\eta_{ nt}(t_1), \ldots, \eta_{ nt}(t_n)) \).

Two ground terms \( t_1 \) and \( t_2 \) are considered to be equivalent, denoted by \( t_1 \sim t_2 \), iff \( \eta_{ nt}(t_1) = \eta_{ nt}(t_2) \). As a further restriction on the non-test procedures, we require that \( \sim \) is a congruence, that is, for each functor \( f \),

\[
\eta_{ nt}(f(t_1, \ldots, t_n)) = \eta_{ nt}(f(t_1', \ldots, t_n'))
\]

whenever \( \eta_{ nt}(t_i) = \eta_{ nt}(t_i') \) for \( 1 \leq i \leq n \). From a computational point of view, this means that the result obtained by evaluating a term does not change if its subterms are reduced first.

Similarly we require that the relations computed by the test procedures respect \( \sim \), i.e. whenever \( \eta_{ nt}(t_i) = \eta_{ nt}(t_i') \) for \( 1 \leq i \leq n \) and \( P \) is an external predicate symbol,

\[
\eta_{ nt}(P(t_1, \ldots, t_n)) \text{ succeeds iff } \eta_{ nt}(P(t_1', \ldots, t_n')) \text{ succeeds.}
\]

Now the standard construction may be used to obtain the model \( MP \) for each amalgamated program \( P \) (see e.g. [GM87]):

First divide the ground term algebra by \( \sim \) (and get a model for the non-test procedures) and then extend this quotient to the least model for

\[
\{ \forall (B_1 \land \ldots \land B_n \rightarrow H) \mid H : \rightarrow B_1, \ldots, B_n \in P \}
\]

which interprets \( '=' \) as the identity relation and which, for each external predicate symbol \( P \), satisfies \( P(t_1, \ldots, t_n) \) iff \( \eta_{ nt}(P(t_1, \ldots, t_n)) \) succeeds. The model so obtained is called the initial model for \( P \).

4. S-UNIFICATION
This section is primarily concerned with equation solving or \( E \)-unification in the presence of a set \( E \) of equations. For this purpose we develop an incomplete \( E \)-unification algorithm, called \( S \)-unification.

4.1. E-unification
A set \( E \) of equations generates a number of congruence relations on terms, the finest of which is denoted by \( =_E \). By an \( E \)-unifier of two terms \( t_1 \) and \( t_2 \) we mean a substitution \( \sigma \) such that \( t_1 \sigma =_E t_2 \sigma \). A set \( S \) of \( E \)-unifiers of \( t_1 \) and \( t_2 \) is complete iff every \( E \)-unifier \( \sigma \) of \( t_1 \) and \( t_2 \) factors into \( \sigma =_E \theta \gamma \) for some element \( \theta \) of \( S \) and some substitution \( \gamma \).

Here \( =_E \) is defined for substitutions \( \sigma_1 \) and \( \sigma_2 \) by \( \sigma_1 =_E \sigma_2 \) iff \( \chi \sigma_1 =_E \chi \sigma_2 \) for each variable \( \chi \). Traditionally the problem of \( E \)-unification is solved by transforming the (finite set of) equations defining the functions into rewrite rules. By applying a
technique called narrowing [H] or some variant thereof, it is then sometimes possible to construct a complete set of $E$-unifiers for the given terms, or to decide that the terms are not $E$-unifiable. As the central object for our studies, we will use the set $E$ of equations defined by:

$$E = \{ t_1 = t_2 \mid t_1^\gamma = t_2^\gamma \text{ for every ground instantiation } \gamma \text{ of } t_1, t_2 \}$$

This set is closed, in the sense that $t_1 =_{E} t_2$ iff $t_1 = t_2 \in E$. Clearly, in our case we only have access to an infinite number of ground equations which define the functions, and it is therefore impossible to give a reasonably efficient general algorithm, which enumerates a complete set of $E$-unifiers of two terms. Our approach is instead to carefully select some cases when it is possible to construct a complete singleton of $E$-unifiers for the given terms, or to decide that no $E$-unifier of the terms can exist. The incomplete $E$-unification algorithm obtained in this way we call $S$(tructure oriented) unification.

4.2. The Disagreement set

The reduced form of a term $t$ is the term obtained by substituting each ground subterm $u$ of $t$ by $\eta_l(u)$. If for example $c$ and the numerals are constructors, $X$ is a variable and $+$ is an external functor with its usual interpretation, then the reduced form of $c(X, (2+5)+X)$ is $c(X, 7+X)$.

Let $t_1$ and $t_2$ be terms and let $r_1, r_2$ be their respective reduced forms. We define the disagreement set of $t_1$ and $t_2$, denoted by $D(t_1, t_2)$, as the least set which satisfies the following three conditions:

1. If the top symbol of $r_1$ and $r_2$ differs, then $\{r_1, r_2\} \in D(t_1, t_2)$.

2. If $r_1$ and $r_2$ are distinct but have the same external functor as top symbol, then $\{r_1, r_2\} \in D(t_1, t_2)$.

3. If $r_1 = c(\ u_1, \ldots, u_n \ )$ and $r_2 = c(\ u_1', \ldots, u_n' \ )$ where $c$ is a constructor, then $D(u_1, u_1') \subseteq D(t_1, t_2)$ for $1 \leq i \leq n$.

Example:

Let $t_1 = c(X*X, c(2+2, X+2))$ and $t_2 = c(Y*Y, c(7-2, X+2))$ where $c$ and the numerals are constructors, $X$ and $Y$ are variables and $+$, $-$ and $*$ are external functors with their usual interpretations.

The reduced forms of $t_1$ and $t_2$ are $c(X*X, c(4, X+2))$ and $c(Y*Y, c(5, X+2))$ respectively. By two applications of $\odot$ we see that $D(t_1, t_2)$ is the union of
\( D(\{X \times Y, Y \times Y\}, D(4, 5)\) and \(D(\{X + 2, X + 2\})\). From \(\circ\) then follows that \(\{X \times X, Y \times Y\} \in D(\{t_1, t_2\})\), and from \(\circ\) it follows that \(\{4, 5\} \in D(\{t_1, t_2\})\), i.e. \(D(\{t_1, t_2\}) = \{\{X \times X, Y \times Y\}, \{4, 5\}\}\).

\[\square\]

It is not hard to see that the condition \(\forall \{u_1, u_2\} \in D(\{t_1, t_2\} : u_1 u_2 = _E u_2 u_1\), is both necessary and sufficient for \(u\) to be an \(E\)-unifier of \(t_1\) and \(t_2\). Note that this would no longer be true if we choose to skip condition \(\circ\) in the definition of the disagreement set, and instead handled external functors as constructors (this is done in e.g. [SY]). As an example, \(\{X / -Y\}\) is an \(E\)-unifier of \(X \times X\) and \(Y \times Y\) but it is not an \(E\)-unifier of \(X\) and \(Y\), although \(D(\{X \times X, Y \times Y\})\) would be \(\{\{X, Y\}\}\) with the altered definition. If we skipped \(\circ\) we would therefore wind up with an algorithm which sometimes produced an incomplete singleton of \(E\)-unifiers for the given terms, and hence we would not achieve our goal. Our next step is to analyze \(D(\{t_1, t_2\})\), and by means of this analysis formulate three different conditions. The first condition tells us how any \(E\)-unifier of \(t_1\) and \(t_2\) necessarily must bind the variables, and the following two conditions ensure that it is safe to fail:

\(\circ\) If \(\{X, u\} \in D(\{t_1, t_2\})\) where \(X\) is a variable which does not occur in \(u\), then any \(E\)-unifier of \(t_1\) and \(t_2\) must bind \(X\) to a term equivalent to an instance of \(u\). We therefore name the fact that \(X\) is a variable which does not occur in \(u\) by Substcond(\(X, u\)).

\(\circ\) If \(\{u_1, u_2\} \in D(\{t_1, t_2\})\) and \(u_1, u_2\) both have constructors as top symbols, then no \(E\)-unifier of \(t_1\) and \(t_2\) can exist since these constructors must be distinct, and hence the above condition implies that no \(E\)-unifier of \(t_1\) and \(t_2\) exists. We call the existence of such an element in \(D(\{t_1, t_2\})\) by Failcond1(D(\(t_1, t_2\))).

\(\circ\) If \(\{X, u\} \in D(\{t_1, t_2\})\), \(X\) is a variable and \(X\) has at least one occurrence in \(u\) which is outside every subterm of \(u\) having an external functor as top symbol, then no \(E\)-unifier of \(u_1\) and \(u_2\) can exist, and consequently no such exists for \(t_1\) and \(t_2\) either. This case, which we call Failcond2(D(\(t_1, t_2\))), corresponds to the usual occur-check failure.

Note that the fact that \(X\) occurs in \(u\) is not sufficient for \(X\) and \(u\) not to be \(E\)-unifiable, even if the top symbol of \(u\) is a constructor. Consider for example the case where \(c\) is a constructor, \(f\) is an external functor and \(t\) is a ground term such that \(\eta_f(f(c(t))) = t\), then \(\{X / c(t)\}\) is an \(E\)-unifier of \(X\) and \(c(f(X))\).
4.3. Unification

Next we present the $S$-unification algorithm. The algorithm may succeed and return a substitution as result, it may fail, or it may give an error message:

\[
S\text{-unify}(t_1, t_2) : \\
\text{begin} \\
\sigma := \varepsilon ; \\
D := D( t_1, t_2 ) ; \\
\text{while } D \neq \emptyset \text{ do} \\
\quad \text{if } \text{Failcond}_1(D) \text{ or } \text{Failcond}_2(D) \\
\quad \hspace{1em} \text{then } \text{FAIL} \\
\quad \text{else if } \exists \{ x, u \} \in D : \text{Substcond}( x, u ) \\
\quad \hspace{1em} \text{then } \text{begin} \\
\quad \hspace{2em} \sigma := \sigma \{ x / u \} ; \\
\quad \hspace{2em} D := D( t_1 \sigma, t_2 \sigma ) ; \\
\quad \hspace{1em} \text{end} \\
\quad \text{else } \text{ERROR} \\
\text{else return } \sigma \\
\text{end} .
\]

An error message indicates that the terms are not sufficiently instantiated, and hence that the algorithm can not decide whether they are $E$-unifiable. We do however have the following result [LBM]:

**Theorem:**
For any terms $t_1$ and $t_2$, $S\text{-unify}(t_1, t_2)$ terminates. Moreover:

1. If $S\text{-unify}( t_1, t_2 )$ succeeds and returns $\sigma$, then $\{ \sigma \}$ is a complete set of $E$-unifiers of $t_1$ and $t_2$.

2. If $S\text{-unify}( t_1, t_2 )$ fails, then $t_1$ and $t_2$ have no $E$-unifier.

It should be mentioned that the $S$-unification algorithm reduces to Robinsons unification algorithm [R] whenever it is called with both its arguments instantiated to structures.

**Example:**
Let $t_1 = s( X^X, t )$ and $t_2 = s( X, X )$, where $s$ and the numerals are constructors, $X$ is a variable and $*$ is an external functor with its usual interpretation. $S\text{-unify}( t_1, t_2 )$
will succeed and return \( \{ X / t \} \) iff \( t \) is 0 or 1, it will fail iff \( t \) is any other numeral, and it will give an error message iff \( t \) is nonground.

\( \square \)

5. THE OPERATIONAL SEMANTICS

In this section an interpreter for the amalgamated language is outlined, and some simple examples of amalgamated programs are presented. Throughout this and the following section, internal predicate symbols will be treated as constructors.

5.1. The Interpreter

We start out from a standard Prolog interpreter with its left-to-right computation rule, and exchange its purely syntactic unification for \( S \)-unification. Hence the operational semantics of a program \( P \), which defines how an answer substitution is to be computed from a goal clause, may be given in terms of the least relation \( O_P \) between goal clauses and substitutions, satisfying:

1. \( \square O_P e, \) where \( \square \) is the empty goal.

2. If \( S \)-unify( \( t_1, t_2 \)) succeeds and returns \( \sigma \), then \( :- t_1 = t_2 O_P \sigma \).

3. If \( P \) is an external predicate symbol and \( t_1, \ldots, t_n \) are ground terms such that \( \eta_i(P( t_1, \ldots, t_n )) \) succeeds, then \( :- P( t_1, \ldots, t_n ) O_P e \).

4. If \( A \) is an atom built from an internal predicate symbol, \( H :- B_1, \ldots, B_n \) is a (renamed) clause from \( P \) such that \( S \)-unify( \( A, H \)) succeeds and returns \( \sigma_1 \), and \( :- (B_1, \ldots, B_n)O_1 \sigma_2 \), then \( :- A O_P (\sigma_1 \sigma_2 | V(A)) \), where the last expression denotes the restriction of \( \sigma_1 \sigma_2 \) to the variables occurring in \( A \).

5. If \( :- B_1, \ldots, B_i O_P \sigma_1 \) and \( :- (B_{i+1}, \ldots, B_n)O_1 \sigma_2 \) where \( i \leq n \), then \( :- B_1, \ldots, B_n O_P \sigma_1 \sigma_2 \).

5.2. Admitted Goals of a Program

If, during the execution of a program \( P \) with a goal \( G \), no runtime error caused by \( S \)-unification occurs and no test procedure is called with some nonground argument, then it is not possible at this execution to distinguish our interpreter from SLD-resolution combined with a complete \( E \)-unification algorithm. From the fact that such a combination is a complete deduction system for Horn clause logic with equality [JLM], we may then conclude that we for each program have a class of goals for which our interpreter is complete. If \( P \) is a program and \( G \) is such a goal, we say that \( P \) admits \( G \). More precisely, let \( A \) be a body atom of \( P \) and \( G \), and let \( C_G(A) \) be the set of call
instantiations of $A$, i.e. the set of all possible instantiations of $A$ whenever $A$ becomes the actual subgoal during the processing of $G$. Now $P$ admits $G$ iff, for each bodyatom $A$ of $P$ and $G$, and $\sigma \in C_G(A)$, the following three conditions are satisfied:

1. If $A = t_1 = t_2$, then $S$-unify($t_1\sigma$, $t_2\sigma$) does not give an error message.

2. If $A = \mathcal{P}(t_1, \ldots, t_n)$ where $\mathcal{P}$ is an external predicate symbol, then $t_i\sigma$ is ground for $1 \leq i \leq n$.

3. If $A$ is built from an internal predicate symbol and $H : - B_1, \ldots, B_n$ is any (renamed) clause of $P$, then $S$-unify($A\sigma$, $H$) does not give an error message.

5.3. Examples
Next we present two simple examples of amalgamated programs:

Example 1:
Suppose $+$ and $-$ are external functors, and let $\leq$ be an external predicate symbol, all with their standard interpretations. The predicate $\text{less}$ has the same formal semantics as $\leq$ on nonnegative integers, but may be called with its first argument uninstantiated:

$$\text{less}(0, X).$$
$$\text{less}(Y+1, X) : - 1 \leq X, \text{less}(Y, X-1).$$

Hence $\mathcal{P}$ would yield the enumeration $X = 0, X = 1$ and $X = 2$. However, the program does not admit goals of the form $\mathcal{P}$ where $n$ and $m$ are integers, since $S$-unification of $\text{less}(n, m)$ and $\text{less}(Y+1, X)$ would give ERROR as a result. Neither does it admit goals where the second argument is uninstantiated, due to the symbol $\leq$ occurring in the second clause.

Example 2:
By a partition of a nonnegative integer $n$, we mean a sequence of positive integers which sums up to $n$. The following program, which uses Prolog's standard notation for lists, enumerates all partitions of a given integer:

$$\text{part}([], 0).$$
$$\text{part}([H|T], X) : - \text{less}(H, X), H \neq 0, \text{part}(T, X-H).$$

The program uses $\text{less}$ from the previous example and the external predicate symbol
with its obvious interpretation. As an example, the goal: \texttt{part}(X, 2), would yield the following enumeration: \( X = [1, 1], X = [2] \).

6. STATIC ANALYSIS OF AMALGAMATED PROGRAMS

This section outlines a technique for static analysis of amalgamated programs. The problem is whether a given program admits a given class of goals. This problem is undecidable. This can be proved by reducing it to the halting problem of Turing machines, in a similar way as it was done in [DM] for the occur-check problem. Therefore a sufficient condition is suggested, which results in a relatively simple check procedure. The approach is similar to that of abstract interpretation, see e.g. [M], [JS], [B], [N]. For the infinite domain of terms, a finite family of subdomains is defined. Its elements, called kinds are the following:

- The empty set, denoted by \( e \).

- The set of variables, denoted by \( v \).

- The set of all ground terms, denoted by \( g \).

- The set of all terms having a structure as reduced form, denoted by \( s \).

- The set of all terms, denoted by \( t \).

The concretization function \( \Delta \) on the set \( K = \{ e, g, v, s, t \} \) of kinds, associates with each kind the corresponding set of terms. By the assumption that the external procedures always terminate and return a structure, \( K \) forms a lattice which is illustrated in the following figure:

![Lattice Diagram](image)

An abstract substitution \( \pi \), characterizes a set of substitutions \( S(\pi) \) by specifying:

- \( D_\pi \): a finite set of variables, the domain of \( \pi \).
\( \kappa_n : D_n \rightarrow K \) : the kind assignment which specifies the kind of terms which are allowed to be assigned to the variables, i.e. for every \( x \in D_n \) and \( \sigma \in S(n) \), \( x_\sigma \in \Delta(\kappa_n(x)) \).

\( \alpha_n \subseteq D_n \times D_n \) : the aliasing relation ; \( x \alpha_n y \) iff there is a substitution \( \sigma \) in \( S(n) \) such that \( x_\sigma \) and \( y_\sigma \) have a common variable.

It is assumed that the classes of goals checked for admission are characterized by abstract substitutions. Thus, the starting point for static analysis is a program \( P \), a goal \( G = \vdash B_1, \ldots, B_n \) and an abstract substitution \( n_0 \) defined on the variables of \( B_1 \). \( n_0 \) specifies all bindings of the variables of \( B_1 \) which are allowed at the invocation time, i.e. when the execution of \( B_1 \) starts. Assume now that an abstract substitution is given not only for \( B_1 \) but also for \( B_2, \ldots, B_n \) and for every body literal of \( P \). The abstract substitution \( n_L \) assigned to a literal \( L \) is said to be a correct annotation for \( P, G \) and \( n_0 \) iff at every invocation of \( L \) during the execution of \( \vdash B_1, \ldots, B_n \) where \( \sigma \in S(n_0) \), the binding of the variables of \( L \) are in \( S(n_L) \). Correct annotations will be generated by abstract interpretation, but this is outside the scope of this paper. Here we confine ourselves with the definition of an abstract unification algorithm which will be used for checking admissibility of goals. The algorithm works on abstract terms. By an abstract term we mean a pair \( \langle t, n \rangle \), where \( t \) is a term and \( n \) is an abstract substitution on a domain including all variables of \( t \). \( \langle t, n \rangle \) denotes all terms \( t_\sigma \) such that \( \sigma \in S(n) \).

For each abstract term \( \langle t, n \rangle \), we define the kind \( \kappa_n(t) \) as follows:

1. If \( t \) is a constant, then \( \kappa_n(t) = g. \)

2. If \( t \) is a variable \( x \), then \( \kappa_n(t) = \kappa_n(x). \)

3. If \( t = c(t_1, \ldots, t_n) \) where \( c \) is a constructor, then

   \[ \kappa_n(t) = \bigvee \{ g, \kappa_n(t_1), \ldots, \kappa_n(t_n) \}. \]

4. If \( t = f(t_1, \ldots, t_n) \) where \( f \) is an external functor,

   then \( \kappa_n(t) = g \) if \( \kappa_n(t_1) = \ldots = \kappa_n(t_n) = g \), and \( \kappa_n(t) = t \) otherwise.

The abstract unification algorithm, which we call \( AT \)-unification, applies to two abstract terms \( \langle t_1, n \rangle \) and \( \langle t_2, n \rangle \). Our design objective for \( AT \)-unify is that it should be able to recognize whenever there exists a substitution \( \sigma \) in \( S(n) \) for which \( S \)-unify( \( t_1, t_2 \sigma \) ) gives an error message. Due to the observation that the term component of an abstract term describes the top level structure of all its instances, we start by constructing the disagreement set \( D( t_1, t_2 ) \). In this way we do exactly the
same job as $S$-unification would have done for any terms in the denotations of $(t_1, n)$ and $(t_2, n)$. First we note that if $\text{Failcond}_1(u_1, u_2)$ or $\text{Failcond}_2(u_1, u_2)$ applies for some $\{u_1, u_2\} \in D(t_1, t_2)$, then no instance of $t_1$ and $t_2$ will $S$-unify, and hence $AT$-unify may safely fail. However, now $D(t_1, t_2)$ contains pairs of sets of terms rather than pairs of terms. Therefore the analysis of $D(t_1, t_2)$ has to be done in a quite different way than for $S$-unification. Our intention is to pick elements $\{u_1, u_2\}$ from $D(t_1, t_2)$ which are safe in the sense that for no $\sigma$ in $S(n)$, $S$-unify$(u_1\sigma, u_2\sigma)$ gives an error message. It is not hard to see that each of the following two conditions, which we call $\text{Substcond}_1(u_1, u_2)$ and $\text{Substcond}_2(u_1, u_2)$ respectively, are sufficient for this to be true:

1. One of $u_1$ and $u_2$ is a variable $X$, i.e. $\{u_1, u_2\} = \{X, u\}$, and either $\kappa_n(X), \kappa_n(u) \leq s$ and $X$ does not occur in $u$, or $\kappa_n(X) = v$ and for no variable $Y$ that occurs in $u$, $X \alpha_n Y$.

2. One of $u_1$ and $u_2$ is a nonvariable term to which $\kappa_n$ assigns the kind $s$, and the other is a term, with an external functor as top symbol, to which $\kappa_n$ assigns the kind $g$.

Now we assume that $u_1$ and $u_2$ are instantiated by a $\sigma \in S(n)$ in such a way that $S$-unify$(u_1\sigma, u_2\sigma)$ succeeds with a substitution $\theta$ as result. Of course this assumption may be wrong. We do know however that $S$-unify$(u_1\sigma, u_2\sigma)$ will not give an error message. Our next step is then to simulate the composition of $t_1\sigma$ and $t_2\sigma$ with $\theta$. This is done by the construction of two new abstract terms, denoted $(t_1, n) + \{u_1, u_2\}$ and $(t_2, n) + \{u_1, u_2\}$ and defined as follows:

3. If one of $u_1$ and $u_2$ is a variable $X$, so that $\{u_1, u_2\} = \{X, u\}$, then $t_i + \{u_1, u_2\} = t_i\{X / u\}$ for $i = 1, 2$, and $D_\pi + \{u_1, u_2\} = D_\pi - \{X\}$.

4. If both $u_1$ and $u_2$ are nonvariable terms, then $t_i + \{u_1, u_2\} = t_i$ for $i = 1, 2$, and $D_\pi + \{u_1, u_2\} = D_\pi$.

In order to construct $\kappa_n + \{u_1, u_2\}$ and $\alpha_n + \{u_1, u_2\}$, four different conditions are defined. A variable $X$ in $D_\pi$ may or may not satisfy some of them:

$\text{Condition 1}$: For no variable $Y$ that occurs in $u_1$ or $u_2$, $X \alpha_n Y$ holds.

$\text{Condition 2}$: $\kappa_n(u_1) = v$ and $X$ occurs in $u_2$.

$\text{Condition 3}$: $X$ occurs in $u_1$ or $u_2$ and at least one of $\kappa_n(u_1)$ and $\kappa_n(u_2)$ is $g$.

$\text{Condition 4}$: None of condition 1 to 3 holds for $X$. 

Now \( \kappa_n + \{u_1, u_2\}(X) \) is defined for each \( X \in D_n + \{u_1, u_2\} \) by:

1. If Condition 1 or Condition 2 holds for \( X \), then \( \kappa_n + \{u_1, u_2\}(X) = \kappa_n(X) \).
2. If Condition 3 holds for \( X \), then \( \kappa_n + \{u_1, u_2\}(X) = g \).
3. If Condition 4 holds for \( X \), then \( \kappa_n + \{u_1, u_2\}(X) = \kappa_n(X) \lor \kappa_n(u_1) \lor \kappa_n(u_2) \).

Finally, we let \( S \) be the subset of \( D_n + \{u_1, u_2\} \), consisting of all variables \( X \) such that \( \kappa_n + \{u_1, u_2\}(X) \) is \( v, s \) or \( t \). \( a_n + \{u_1, u_2\} \) is defined to be the least equivalence relation on \( S \) which satisfies:

1. If \( X a_n Y \) then \( X a_n + \{u_1, u_2\} Y \).
2. If \( u_1 = Y \) and \( X \) satisfies Condition 2, then \( X a_n + \{u_1, u_2\} Y \).
3. If Condition 4 holds for \( X \), and \( Y \) occurs in \( u_1 \) or \( u_2 \), then \( X a_n + \{u_1, u_2\} Y \).

Now the \( AT \)-unification algorithm may be defined. At each iteration it picks one element \( \{u_1, u_2\} \) from \( D(t_1', t_2') \) for which one of the substitution conditions applies, and constructs \( \langle t_i', n' \rangle + \{u_1, u_2\} \) for \( i = 1, 2 \). If Substcond\(_1\)\((u_1, u_2)\) holds, then \( \{u_1, u_2\} \) will not occur again in \( D(t_1' + \{u_1, u_2\}, t_2' + \{u_1, u_2\}) \). If however Substcond\(_2\)\((u_1, u_2)\) applies, then \( \{u_1, u_2\} \) will be in \( D(t_1' + \{u_1, u_2\}, t_2' + \{u_1, u_2\}) \) with the kind \( g \) assigned to all the variables occurring in \( u_1 \) and \( u_2 \). Therefore we can not have \( D(t_1', t_2') = \emptyset \) as termination condition. If however \( \kappa_n'(u_1) = \kappa_n'(u_2) = g \) holds for each \( \{u_1, u_2\} \in D(t_1', t_2') \), then we may safely stop without giving an error message. We call this condition Stopcond\((D)\).
\[ \text{AT-unify}(\langle t_1, n \rangle, \langle t_2, n \rangle) : \]
\begin{align*}
\text{begin} \\
\langle t_1', n' \rangle := \langle t_1, n \rangle \\
\langle t_2', n' \rangle := \langle t_2, n \rangle \\
D := D(t_1', t_2'); \\
\text{while not } \text{Stopcond}(D) \text{ do} \\
\quad \text{if } \text{Failcond}_1(D) \text{ or } \text{Failcond}_2(D) \\
\quad \quad \text{then } \text{FAIL} \\
\quad \text{else } \text{if } \exists \{ u_1, u_2 \} \in D : \text{Substcond}_1(u_1, u_2) \text{ or } \text{Substcond}_2(u_1, u_2) \\
\quad \quad \text{then } \text{begin} \\
\quad \quad \quad \langle t_1', n' \rangle := \langle t_1', n' \rangle + \{ u_1, u_2 \} \\
\quad \quad \quad \langle t_2', n' \rangle := \langle t_2', n' \rangle + \{ u_1, u_2 \} \\
\quad \quad \quad D := D(t_1', t_2'); \\
\quad \quad \end{align*}
\text{end} \\
\text{else } \text{ERROR} \\
\text{end}.

For the AT-unification algorithm we have the following theorem:

**Theorem:**
For any term patterns \( \langle t_1, n \rangle \) and \( \langle t_2, n \rangle \), \( \text{AT-unify}(\langle t_1, n \rangle, \langle t_2, n \rangle) \) terminates. Moreover:

1. If \( \text{AT-unify}(\langle t_1, n \rangle, \langle t_2, n \rangle) \) fails, then \( \text{S-unify}(t_1\sigma, t_2\sigma) \) fails for all \( \sigma \in S(n) \).

2. If, for some \( \sigma \in S(n) \), \( \text{S-unify}(t_1\sigma, t_2\sigma) = \text{ERROR} \), then \( \text{AT-unify}(\langle t_1, n \rangle, \langle t_2, n \rangle) = \text{ERROR} \).

\[ \square \]

**Example:**
Let \( t_1 = c(x, 3+y) \), \( t_2 = c(z+z, 3) \) and let \( D_n = \{ x, y, z \} \), \( \kappa_n(x) = \kappa_n(z) = v \) and \( \kappa_n(y) = g \). If \( x \alpha_n z \) then \( \text{AT-unify}(\langle t_1, n \rangle, \langle t_2, n \rangle) \) will terminate with \( \text{ERROR} \), otherwise the algorithm will stop without error message.

\[ \square \]

Let \( P \) be a program, let \( G \) be a goal and let \( n_0 \) be the initial abstract substitution. If each body atom \( A \) of \( P \) and \( G \) is correctly annotated for \( P, G \) and \( n_0 \) then it suffices to check the following for each body atom \( A \) of \( P \) and \( G \), in order to be sure that \( P \) admits each goal defined by \( G \) and \( n_0 \):
① If \( A = t_1 = t_2 \), then \( AT\)-unify(\( \langle t_1, \pi_A \rangle, \langle t_2, \pi_A \rangle \)) does not give an error message.

② If \( A = P(t_1, \ldots, t_n) \) where \( P \) is an external predicate symbol, then \( \pi_A \) assigns the kind \( g \) to each variable in \( A \).

③ If \( A \) is built from an internal predicate symbol and \( H := B_1, \ldots, B_n \) is any (renamed) clause of \( P \), then \( AT\)-unify(\( \langle A, \pi_A', \rangle, \langle H, \pi_A' \rangle \)) does not give an error message.

Here \( \pi_A' \) is an extention of \( \pi_A \) which includes all variables in \( H \), so that no two distinct variables of \( H \) are aliases and all of the variables in \( H \) are assigned the kind \( v \).

7. CONCLUSION

The paper presents a clean approach to amalgamation of logic programs with external functional procedures. The logical semantics of the amalgamated language is outlined as well as the operational semantics, based on an incomplete \( E \)-unification algorithm which we call \( S \)-unification. It is suggested to use the abstract interpretation technique for identifying classes of goals for which the approach is complete. For this purpose a domain of abstract terms is defined, and an algorithm of abstract unification, which is used for a compile-time check is developed.

An idea close to \( S \)-unification appeared independently in Le Fun [ALN] but in a purely operational setting. The idea of residuations used in Le Fun, and the idea of delays used in CLP [JM] makes it possible to avoid some errors of \( S \)-unification. Implicitly it is present also in our description, but only in the limits of one unification : notice that our abstract presentation of the algorithm concerns selection of certain pairs in the disagreement sets. A more realistic version should probably use left-to-right selection with residuation. Extending this technique beyond the scope of one unification is also possible, but a more elegant version would be some kind of static analysis of the program leading to some program transformation.

The logical semantics of the language is based on well-known results presented in many papers, e.g. [GM87], [JLM]. However, our objectives are different. EQLOG attempts to exploit the narrowing technique, but we cannot do that since our functional procedures are black-boxes. Therefore our approach is inherently incomplete.

Since the language of functional procedures is left unspecified, we deal as a matter of fact with a family of logic programming languages. In that sense there is a similarity to
the CLP family of languages. However, in contrast to the CLP project we do not plan to use additional information about the domains during the execution. Thus our approach is "less complete" but applies to any domain. The future work includes the following topics:

- Efficient algorithms for S-unification.
- Methods for static analysis of amalgamated programs, including automatic generation of annotations, and program transformation for improving admissibility of goals.
- Relaxation of the restrictions on external procedures, e.g. considering lazy evaluation.
- Application of the idea of S-unification to higher-order procedures.
- Amalgamated programs with negation.

8. REFERENCES


[LBM] J.Leszczylowski, S.Bonnier, J.Maluszynski, Logic programming with


Towards Clean Amalgamation of Logic Programs with External Procedures

Staffan Bonnier, Jan Maluszynski

The paper presents a clean approach to the amalgamation of logic programming with external functional procedures. Both the logical semantics and the operational semantics of the amalgamated language are outlined. The operational semantics is based on an incomplete E-unification algorithm which we call S-unification. It is suggested to use the abstract interpretation technique for identifying classes of goals for which the approach is complete. For this purpose a domain of abstract terms is defined, and an abstract unification algorithm used for a compile-time check is developed.
A Selection of Previous Research Reports.

LiTH-IDA-R-87-20  Krzysztof Kuchciński, Zebo Peng: Parallelism Extraction from Sequential Programs for VLSI Applications. This paper is to appear in *Microprocessing and Microprogramming, the Euromicro Journal*, 1988.


LiTH-IDA-R-87-18  Henrik Nordin: Reuse and Maintenance Techniques in Knowledge-Based Systems.

LiTH-IDA-R-87-17  Tony Larsson: Specification and Verification of VLSI Systems Actional Behaviour. This is a close version of a paper presented at *the 8th international conference on Computer Hardware Description Languages, CHDL, 87*.


LiTH-IDA-R-87-15  Nils Dahlbäck: Kommunikation med datorer i naturligt språk - vad är det och vem behöver det?


LiTH-IDA-R-87-10  Andrzej Lingas: On Parallel Complexity of the Subgraph Isomorphism Problem.

LiTH-IDA-R-87-09  Andrzej Lingas, Marek Karpinski: Subtree Isomorphism and Bipartite Perfect Matching are Mutually NC Reducible.


LiTH-IDA-R-87-07  Peter Hanecloau: A Formal Approach to Reason-maintenance Based on Abstract Domains.

LiTH-IDA-R-87-06  Johan Hultman: COPPS - A Software System for Defining and Controlling Actions in a Mechanical System.

LiTH-IDA-R-87-05  Christer Bäckström: Logical Modelling of Simplified Geometrical Objects and Mechanical Assembly Processes.


LiTH-IDA-R-87-02  Rober Bilos: A Token-Based Syntax Sensitive Editor. Also presented at the *Workshop on Programming Environments - Programming Paradigms*, Roskilde, Denmark, October 22-24, 1986.


organizes undergraduate and graduate studies in Computer Science, Telecommunication and Computer Systems, and Administrative Data Processing. Research activities have an emphasis on advanced software technology and computer systems design and are organized in a number of research laboratories:

- **ACTLAB - Laboratory for Complexity of Algorithms**, which is concerned with the design and analysis of efficient sequential and parallel algorithms, and complexity theory, especially in the areas of computational geometry, data structures on bounded domains and graph algorithms.

- **AIELAB - Artificial Intelligence Environments Laboratory**, which conducts research on representation and manipulation of knowledge, problem solving techniques and expert systems with mechanical engineering applications.

- **ASLAB - Application Systems Laboratory**, which studies design of advanced support systems for interactive use of computers, including tools for automated construction of applications software.

- **CADLAB - Laboratory for Computer-Aided Design of Electronics**, which concentrates its research activities around tools for integrated development of hardware and software, graphics-based modelling and simulation techniques.

- **LIBLAB - Laboratory for Library and Information Science**, which studies methods for access to documents and the information contained in the documents, concentrating on catalogs and bibliographic representations, and on the human factors of library use.

- **LOGPRO - Laboratory for Logic Programming**, which concentrates its research activities around foundations of logic programming, relations to other programming paradigms and methodology.

- **NLPLAB - Natural Language Processing Laboratory**, which conducts research related to the development and use of natural language interfaces to computer software.

- **PELAB - Programming Environments Laboratory**, which works with design of tools for software development, specific functions in such tools and theoretical aspects of programs under construction.

- **RKLLAB - Laboratory for Representation of Knowledge in Logic**, which covers issues and techniques such as non-monotonic logic, probabilistic logic, modal logic and truth maintenance algorithms and systems.

Research Reports 1987 and 1988

<table>
<thead>
<tr>
<th>Report</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>LiTH-IDA-R-88-03</td>
<td>Staffan Bonnier, Jan Maluszynski: Towards a Clean Amalgamation of Logic Programs with External Procedures.</td>
<td></td>
</tr>
<tr>
<td>LiTH-IDA-R-88-02</td>
<td>Ulf Nilsson: Inferring Restricted AND-Parallelism in Logic Programs using Abstract Interpretation.</td>
<td></td>
</tr>
<tr>
<td>LiTH-IDA-R-87-26</td>
<td>Jonas Löwgren: Applying a Rapid Prototyping System to Control Panel Dialogues.</td>
<td></td>
</tr>
<tr>
<td>LiTH-IDA-R-87-24</td>
<td>Sven Moen: Drawing Dynamic Trees.</td>
<td></td>
</tr>
</tbody>
</table>