Transformational query solving

Citation for published version (APA):

Document status and date:
Published: 01/01/1991

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.
Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
Transformational Query Solving

by

Rik van Geldrop
This is a series of notes of the Computing Science Section of the Department of Mathematics and Computing Science Eindhoven University of Technology. Since many of these notes are preliminary versions or may be published elsewhere, they have a limited distribution only and are not for review. Copies of these notes are available from the author.

Copies can be ordered from:
Mrs. F. van Neerven
Eindhoven University of Technology
Department of Mathematics and Computing Science
P.O. Box 513
5600 MB EINDHOVEN
The Netherlands
ISSN 0926-4515

All rights reserved
editors: prof.dr.M.Rem
          prof.dr.K.M.van Hee.
Transformational Query Solving

Rik van Geldrop*

July 12, 1991

Abstract

A transformational programming method is used to derive algorithms for a class of database queries. The purpose is twofold: to illustrate a transformational programming method on the basis of a non trivial example and to give a program derivation for database queries (these are scarce in literature). The results of our efforts are two generic algorithms for files (or databases with only one relation) with minimal restrictions to the system. It is shown how these algorithms may serve as building blocks in solving more complex problems.

*Department of Mathematics and Computing Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands.
1. INTRODUCTION

In the branch of computing science described by information systems, databases play an important role. Several models have been developed for databases, the relational model, [3,4], is one of them. Systems concerned with the implementation of databases are called Database Management Systems (DBMS), the 4th Generation (4G) systems claim to support the relational modelling of databases. In the relational model, based on set theory, specifications for database queries are given in a set theoretical notation. Unfortunately in most 4G DBMS's, the implementation language is a rather restricted set theoretical one and efficiency is mainly a matter of the system. So, if a programmer likes to have control of efficiency, a lower-level system should be used. However, these are restrictive with respect to set operations so that query specifications have to be transformed to the level of files. Since databases may have a complex internal structure, program derivation for database queries becomes very laborious in this (usual) way.

In another branche of computing science, the Bird Meertens formalism (BM), research has been made in the transformation of specifications based on algebraic properties of the underlying data-type [1,2,5]. Here an important role is played by 'structure preserving' functions, so-called morphisms. It is claimed that specifications are programs and executibility depends on the intended machine. Hence one can speak of abstract programs. One of the advantages of such an approach is the separation between the architecture of a solution and the details of its implementation.

Observing that in many database queries the attribute-values to be constructed are resulting from morphism applications, we would like to use the BM tools in our derivations. Program derivation in such a way will be less laborious than a derivation in the usual way.

We don't require that the reader is familiar with the concepts of database theory or BM. The ingredients needed in our derivation will be introduced in section 2. The class of database queries, which we aim to solve, is built up by problems for files. In section 3, the problem class is described and generic algorithms for file problems are derived. How those algorithms apply to database queries is illustrated in section 4. In the algorithms developed, a particular representation is required, and it is claimed that this might be achieved by the system orderings. What can be expected if the representation condition has to be satisfied by the ordering facilities of existing systems is subject of section 5.

2. PRELIMINARIES

In the first part of this section we introduce some notions from database theory and the definition of the (essential part of the) DBMS interface. The second part deals with some basic concepts of BM such as datatypess and transformations (laws). A
link between system interface and BM is subject of the last part of this section.

2.1 Databases and their implementing system.
As mentioned before, we intend to use BM in our derivations. This implies that the algebraic properties of the databasetype (the external behaviour) will be exploited. Therefore, an informal description of some database notions will do.
Databasetypes are used to model the data in some organization. Mostly, the data can be divided into several kinds. In the simplest case, which suffices to introduce the notions needed in section 3, only one kind is involved (in database terms: only one object exists). An object can be described by its relevant characteristics (in database terms: each object is determined by a set of attribute-names). Each attribute name corresponds to an attribute-value set. We will assume that in the organization involved, an occurrence of the object (in database terms: a tuple) can be represented by an element of the (labeled) cartesian product of attribute-value sets, the label set being the attribute names. Usually, the organization deals with several occurrences of an object, i.e. a subset of the cartesian product. But, not every subset may represent a possible state of the object, a so-called table-constraint has to be satisfied. (A frequently occurring table-constraint is the key constraint, e.g. see the definition of SR in section 4.) The subsets of the labeled cartesian product which satisfy the table-constraint constitute the tabletype for the object. Common usage in database theory is the word "relation" for an element of a tabletype. Since we are also interested in (binary) relations on elements of a tabletype, we will prevent confusion by using the word "table" for the database notion "relation". In this simple case, databases are equivalent to tables.
For the moment these notions suffice. In section 4, where we use our schemes to solve database queries, the general case is described and an example databasetype is given.

Implementations for databases are realized by DBMS’s. We prefer to have a hold on efficiency, so 4G systems are left out of consideration. Apart from facilities needed in the solution of our intended problem class, we don’t fix the system involved and specify the implementing interface by:

- databases are table-valued functions over the set of objects
- tables may be considered as
  1. sets with operations:
     - empty, initialization of a table on \( \emptyset \)
     - fetche, retrieval of a tuple, given one of its key-values.
     - store, insertion of a tuple, if the table-constraint is satisfied.
     - delete, deletion of a tuple, if the table-constraint is satisfied.
     (Direct organization of tables.)
  2. files:
     - Let \( \alpha \) be a type, then \( \text{File}(\alpha) \overset{\Delta}{=} [\alpha] \times \mathbb{N} \). (See 2.2 for the definition of \([\alpha]\).)
     - For \( f : \text{File}(\alpha) \) with \( f = (F,p) \), the following operations are available
\[\text{start}(f) \equiv p := 1\]
\[\text{eof}(f) \equiv p = \# F + 1\]
\[\text{fetchn}(f,t) \triangleq \begin{cases} \text{if} & 1 \leq p \leq \# F \rightarrow t := F.p; \ p := p + 1 \\ \text{abort} & \text{if} \ p < 1 \lor p > \# F \end{cases}\]

(Sequential organization of tables.)

There exist ordering facilities for some types. Suppose \(R \subseteq \alpha \times \alpha\), then \(\text{Ordfile}(\alpha, R)\) is the subtype
\[\text{Ordfile}(\alpha, R) \triangleq \{ (F,p) \in \text{File}(\alpha) \mid \forall \ i,j: \ 1 \leq i \leq j \leq \#F : (F.i,F.j) \in R \}\]

In the summary of relations on tables (see appendix) the predicate \((\forall \ i,j: \ 1 \leq i \leq j \leq \#F : (F.i,F.j) \in R)\) is described by ”\(F\) satisfies \(R\).”

Systems equipped with those operations are called File Management Systems (FMS). Some relevant facts about its interfaces are:
- generally they are embedded in imperative languages,
- FMS operations are time consuming, so the objective is to minimize their use.
Our goal will be the derivation of imperative FMS programs.

2.2 BM tools.

The approach of BM applies to a large class of datatypes and exploits only the algebraic properties of those types. The intention is, after fixing an algebraic datatype, to transform the functional expressions, which can be formed in the language of the algebra, in order to exchange expensive operators for cheaper ones. The datatypes we are interested in are lists and sets, but before we introduce the relevant part of BM for these types, we recapitulate some basic facts about algebra.

An algebra is a set together with a family of operators on the set and a (possibly empty) set of equations which have to hold for the operators.

Example. \(M = \langle A, \Sigma, \{ 0_0, 1_1 \} \rangle\) where
\[
\Sigma = \{ + : A \times A \rightarrow A, 0 : \rightarrow A \},
\]
\[
\begin{align*}
0_0 & : \forall x : 0 + x = x + 0 = x \\
1_1 & : \forall x,y,z : (x + y) + z = x + (y + z)
\end{align*}
\]
is an algebra. \(M\) models the monoids. Several instances (models) of a given algebra may exist. E.g. the natural numbers with addition, \(M\), and the booleans with disjunction, \(M'\), are models for the algebra above. Between two models of the same algebra ("two algebras of the same kind") homomorphisms can be defined.

Let \((A, \sigma)\) and \((A', \sigma')\) be algebras of the same kind. A homomorphism \(h\) from \((A, \sigma)\) to \((A', \sigma')\) is a function \(h : A \rightarrow A'\) such that
\[h \circ f = f' \circ h^n\]
for each \(f \in \sigma\) and corresponding \(f' \in \sigma'\). E.g. in our previous example, a homomorphism \(h\) from \(\mathbb{N}\) to \(\mathbb{B}\) is a function \(h : \mathbb{N} \rightarrow \mathbb{B}\) such that
\[ h.0 = \text{false} \]
\[ h.(m + n) = h.m \lor h.n. \]

It is not difficult to prove that (functional) composition preserves homomorphisms. Generally several homomorphisms may exist between two algebras of the same kind, but sometimes there is a model which has exactly one homomorphism to each other model. This (minimal) model is the so-called initial algebra.

The algebra \( M \) of our example has a minimal model and it can be represented by the monoid \( \{1_\oplus\}, \oplus, 1_\oplus \). \( 1_\oplus \) is the (only) generator of the type. A set of generators can be obtained by parametrization with a set \( \alpha \) as follows.

Consider the algebras

\[ C_i = \langle \mathcal{A}(\alpha), \Sigma, I_i \rangle \quad \text{for} \quad i = 0, \ldots, 3 \]

where
\[
\Sigma = \{ + : \mathcal{A}(\alpha) \times \mathcal{A}(\alpha) \to \mathcal{A}(\alpha), \quad 0 : \to \mathcal{A}(\alpha), \quad \tau : \alpha \to \mathcal{A}(\alpha) \}
\]

and
\[
I_i = \{ I_0, \ldots, I_i \}
\]

with
\[
I_0 : \forall x : 0 + x = x + 0 = x
\]
\[
I_1 : \forall x, y, z : (x + y) + z = x + (y + z)
\]
\[
I_2 : \forall x, y : x + y = y + x
\]
\[
I_3 : \forall x : x + x = x
\]

A homomorphism \( h \) from \((A, \{ \oplus, 0, \tau \})\) to \((A', \{ \oplus', 0', \tau' \})\) in \( C_i \) is now defined by

\[ h \circ 0 = 0' \]
\[ h \circ \tau = \tau' \]
\[ h.(x \oplus y) = h.x \oplus' h.y \]

It is clear that a homomorphism in \( C_i \) is a homomorphism in \( C_j \), if \( i \geq j \). Moreover, homomorphisms in \( C_i, i \geq 1 \), are homomorphisms in \( M \). From the theory of algebras it is known, that \( C_i \) has an initial algebra \( IC_i, i = 0, \ldots, 3 \). Together these initial algebras constitute the (Boom) type hierarchy\(^1\). The datatypes that play a role in our application are \( IC_i, i \geq 1 \). We introduce the following representation for them

\[
IC_1 = ([\alpha], [\ ] \cup [\ ] \cup [\ ] \cup [\ ]) \quad \text{with} \quad [\ ].a = [a]
\]
\[
IC_2 = ([\prec a\succ], \psi, \prec \succ, \prec \succ) \quad \text{with} \quad \prec \succ.a = \prec a\succ
\]
\[
IC_3 = ([\alpha], \cup, 0, \{\}) \quad \text{with} \quad \{\}.a = \{a\}
\]

The set of generators of \( IC_1 \) is \( \{ [a] \mid a \in \alpha \} \). \([a]\) consists of elements obtained by finitely many \( \uplus \) applications on elements of this generator set, so \([a]\) denotes the finite lists over \( \alpha \). Similar remarks can be made for the other initial types. Note that a homomorphism on the initial type is completely determined by its behaviour.

\(^1\)The (Boom) type hierarchy, trees - lists - bags - sets, is a hierarchy of binary structures over a given domain. For the definition of binary structures over \( \alpha \), see [7].
on sigletons.

Remark on notation. Here, \( \cdot \) denotes functional application, the notation for functional composition is \( \circ \). In general, the use of \( \cdot \) and \( \circ \) will be omitted.

So far some basic algebraic facts. Now we can be more detailed about BM and its functional transformations. Often, these transformations are based on homomorphisms in \( M \) (morphisms for short). Special forms of morphisms that are of interest for our application are the following:

- **f-map**, denoted by \( f^* \)
  
  Let \( f : \alpha \to \beta \). Then \( f^* : [\alpha] \to [\beta] \) is such that
  
  \[
  f^*([\cdot]) = [\cdot] \\
  f^*([a]) = [f.a] \\
  f^*.(x \oplus y) = f^*.x \oplus f^*.y
  \]

- **\( \oplus \)-reduce**, denoted by \( \oplus / \)
  
  Let \( \oplus : \alpha \times \alpha \to \alpha \) be associative with unit \( 1_{\oplus} \). Then \( \oplus / : [\alpha] \to \alpha \) is such that
  
  \[
  \oplus/.[\cdot] = 1_{\oplus} \\
  \oplus/.[a] = a \\
  \oplus/(x \oplus y) = \oplus/x \oplus \oplus/y
  \]

- **p-filter**, denoted by \( p\triangleleft \)
  
  Let \( p : \alpha \to B \). Then \( p\triangleleft : [\alpha] \to [\alpha] \) is such that
  
  \[
  p\triangleleft.[\cdot] = [\cdot] \\
  p\triangleleft.[a] = \text{if } p.a \to [a] \text{ then } p.a \to [\cdot] \text{ if } \text{ otherwise} \\
  p\triangleleft.(x \oplus y) = p\triangleleft.x \oplus p\triangleleft.y
  \]

\( f^*, \oplus/ \) and \( p\triangleleft \) are (unique) homomorphisms:

\[
\begin{align*}
  f^* &: ([\alpha], \uplus, [\cdot], [\cdot]) \to ([\beta], \uplus, [\cdot], [\cdot]) \\
  \oplus/ &: ([\alpha], \uplus, [\cdot], [\cdot]) \to (\alpha, \oplus, 1_{\oplus}, \text{id}_\alpha) \\
  p\triangleleft &: ([\alpha], \uplus, [\cdot], [\cdot]) \to ([\alpha], \uplus, [\cdot], \text{pf})
\end{align*}
\]

with \( \text{pf}.a = [a] \) if \( p.a, [\cdot] \) otherwise.

For the initial datatypes \( IC_2 \) and \( IC_3 \), f-map, \( \oplus \)-reduce and p-filter are defined in an analogous way. Map, reduce and filter are standard functions in BM. Examples of standard functions which are not necessarily morphisms are the directed reductions. We will use left-reduce.

- **Left-reduce**
  
  Let \( \oplus : \beta \times \alpha \to \beta \) and \( e : \beta \). Then \( (\oplus \to e) : [\alpha] \to \beta \) is such that

  \[
  (\oplus \to e).[\cdot] = e \\
  (\oplus \to e).(x \uplus [a]) = (\oplus \to e).x \oplus a
  \]

  or alternatively

  \[
  (\oplus \to e).[\cdot] = e \\
  (\oplus \to e).([a] \uplus x) = (\oplus \to e (e \oplus a)).x
  \]

For our purpose the relevant transformations are:

L1. the composition of morphisms is a morphism
L2. promotion law. If \( h : (\alpha, \mathcal{O}, 1_\mathcal{O}) \to (\beta, \mathcal{O}, 1_\mathcal{O}) \) is a morphism, then
\[
h \circ/ = \mathcal{O}/ \circ h^*
\]
L3. \( h \) is a morphism iff \( \exists \mathcal{O}, f : h = \circ/ \circ f^* \)
L4. \( f^* g^* = (f g)^* \)
L5. filter-map rule.
\[
\varphi \circ f^* = f^* (\varphi f) \circ
\]
L6. a morphism can be written as a directed reduction.
If \( h = \circ/ f^* \) then \( h = (\mathcal{O} \to 1_\mathcal{O}), \) where
\[
u \circ a = u \circ f.a
\]
L7. formal differentiation. Let \( \mathcal{O} : \beta \times \alpha \to \beta, G : [\beta] \to \gamma, e : [\beta]. \) Suppose
\[
\mathcal{O} : \gamma \times \alpha \to \gamma \text{ is such that } G.(y \mathcal{O} a) = G.y \circ a, \text{ then}
\]
\[
G (\mathcal{O} \to e) = (\mathcal{O} \to G.e)
\]
Not only morphisms can be written as a directed reduction, non-morphisms may have a directed reduction form too. However, finding such a form may cause intricate calculations.

2.3 Linking system interface and BM tools.
We aim at using BM in developing FMS algorithms, so we need some link between the two approaches. As mentioned before, BM is concerned with the algebraic properties of functions, not with the implementation of functions and their evaluations. So, if a programmer wants to compute a function application he has a new problem at hand outside the scope of BM. He has to consider this as a specification in a suitable new programming environment. The use of BM in the development of programs is mainly the calculation of transformations of (possible) equivalent defining forms of the functions involved. In particular, if one is interested in applying standard functions of BM certain standard programs in other environments arise. Since the functions we use are restricted to those that are in fact directed reductions, we will give such a standard program for them only. Our interest is in imperative File algorithms, but due to the few algebraic properties of files, a direct link from BM to File level is rather complicated. An intermediate level may simplify the desired link and Conslists is an appropriate candidate for it:

- morphism applications can be computed by directed reductions
- left reductions are easy to implement as functions over conslists
- operations on files look slightly like conslists ones. \( (2.3.a) \)

In the sequel, we will refer to the following lemma as the standard conversion of a left-reduce.

**Lemma 1**
Let \( \mathcal{O} : \beta \times \alpha \to \beta \) and \( e : \beta. \) Then \( S_1 \) is a correct program fragment for the computation of \( w = (\mathcal{O} \to e)x. \)

\[
S_1: \quad w := e; \ r := x
\]
\begin{verbatim}
; do r ≠ []
   → a := first r; r := tail r
; w := w ⊕ a

od

Invariant: (⊕ ↦ w)r = (⊕ ↦ e)x ∧ r ∈ tails x

Proof. Follows immediately from the definition of left-reduce.
\end{verbatim}

Note that this computation of a left-reduce on conslists leads to an iterative program with time-complexity \( O(N) \), if \# x = N. Since these costs are relative to the costs of \( ⊕ \), an efficiency comparison of the computation of \( (⊕ ↦ f)x \) and \( (⊕ ↦ e)x \) can be made from the definitions of \( ⊕ \) and \( ⊕ \).

3. GENERIC ALGORITHMS FOR FILES

In this section, we solve some problem classes for files. The classes are interrelated: a fundamental problem has an interesting subproblem and can be extended in several ways. The fundamental problem and its subproblem are solved via transformations in 3.2 and 3.1 and with these solutions, we solve the extensions in 3.3. Our transformational solution method consists of two steps: First, a transformation from Sets to Conslists is made by using BM laws. This step is followed by a standard conversion to conslists. Afterwards, the resulting Conslists program is implemented by an FMS algorithm. The correctness of this step is based on state-space transformations.

To avoid repeating assumptions in our solutions, we will use the conventions that \( α \) and \( β \) are types, \( ⊕ \) is an associative, commutative and idempotent operator on \( α \) with unit \( 1_α \), \( h : (\{α}, U, 0) \rightarrow (β, ⊕, 1_β) \) is a morphism, \( E \) is an equivalence relation on \( α \), \( X ∈ \{α} \) and \( q : X → X/E \) defined by

\[ q.t = \{ v ∈ X | v E t \}. \]

is the quotientmap.

3.1 FMS algorithm to compute \( w = hX \).

\begin{verbatim}
Set specification : w = hX

step 1. Let sets of \( α \)-elements be implemented by lists with representation function
\( Rs \)

\[ Rs = U/{–}^* \]

and representation invariant \( Ps(x) \)

\[ Ps(x) : \# Rs x = \# x \]

(Informal: a representation contains no duplicates.)

We have to find a function \( \tilde{h} \) such that \( \tilde{h} = h Rs \).

Graphically
\end{verbatim}
Since $\tilde{h}$ is a composition of morphisms it is a morphism itself. Furthermore, it holds that

$$
\tilde{h} = h \mathcal{R}_s = \{ \text{def } \mathcal{R}_s \} \cup / \{-\}^* = \{ \text{L2} \} \otimes h^* \{-\}^* = \{ \text{L4} \} \otimes/(h \{-\})^*
$$

and using L6, $\tilde{h}$ can be written as a directed reduction:

$$
\tilde{h} = (\otimes \rightarrow 1_\otimes)
$$

(3.1.b)

where $\otimes : \beta \times \alpha \rightarrow \beta$ is defined by $u \otimes a = u \otimes h\{a\}$. Given a representation $x$ of $X$, we may instantiate the standard conversion of a left-reduce (Lemma 1).

Conslist program for (3.1.a)

```plaintext
w := 1_\otimes; r := x
; do r \neq []
  → a := first r; r := tail r
    ; w := w \otimes h\{a\}
  od
```

(3.1.c)

step 2. Let conslists be implemented by files. Let $f : \text{File}(\alpha)$ then $f = (F,p)$. The representation function $\mathcal{R}_c$ is given by

$$
\mathcal{R}_c(F,p) = \begin{cases} 
[\ ] & \text{if } p = \#F + 1 \\
[F,p] \sqcup \mathcal{R}_c(F,p + 1) & \text{otherwise}
\end{cases}
$$

and the representation invariant is

$$
1 \leq p \leq \# F + 1 \land F \text{ is a bijection over } [1..\# F]
$$
Given a representation \( f \) of \( x \), an FMS algorithm for (3.1.c) requires implementations for the

- initialisation \( r := x \),
- guard \( r \neq [ ] \),
- statement sequence \( a := \text{first } r ; r := \text{tail } r \).

Straightforward are \( r := x \rightarrow \text{start}(f) \) and \( r \neq [ ] \rightarrow \neg \text{eof}(f) \). It is not difficult to prove that, from the validity of

\[
\{ Q_{\text{first } r, \text{tail } r}^a \} a := \text{first } r ; r := \text{tail } r \{ Q \}
\]

it follows that

\[
\{ (Q_{\text{first } r, \text{tail } r}^a)^f \text{etchn}(f, a) \} \{ Q_{\mathcal{R}_a(f)}^r \}
\]

holds. So \( \text{etchn}(f, a) \) is a correct implementation of \( a := \text{first } r ; r := \text{tail } r \). Applying those statement transformations to (3.1.c) results in a solution for the subproblem.

**FMS algorithm for (3.1.a)**

\[
\begin{align*}
\text{w} &:= 1_\emptyset; \text{start}(f) \\
\text{do } &\sim \text{eof}(f) \\
&\rightarrow \text{etchn}(f, a) \\
&\text{; w} := \text{w} \otimes h\{a\} \\
\text{od}
\end{align*}
\]

Some remarks on this derivation

1. **Tupled version for (3.1.a).** Instantiate (3.1.a) with the (product) morphism

\[
h = <h_1, \ldots, h_m> : \{ \alpha \}, \cup, \emptyset \rightarrow (\beta, \otimes, 1_\emptyset),
\]

where \( h_i : \{ \alpha \}, \cup, \emptyset \rightarrow (\beta_i, \otimes_i, 1_{\emptyset_i}) \), \( 1 \leq i \leq m \), is a morphism, \( \beta = \beta_1 \times \ldots \times \beta_m \), \( \otimes = (\otimes_1, \ldots, \otimes_m) \) is defined by

\[
(b_1, \ldots, b_m)(\otimes_1, \ldots, \otimes_m)(c_1, \ldots, c_m) = (b_1 \otimes_1 c_1, \ldots, b_m \otimes_m c_m)
\]

\( 1_\emptyset = (1_{\emptyset_1}, \ldots, 1_{\emptyset_m}) \) and \( h X = <h_1 X, \ldots, h_m X> \). Then the computation of \( h X \) might be done "componentwise".

\[
\begin{align*}
w_1, \ldots, w_m &:= 1_{\emptyset_1}, \ldots, 1_{\emptyset_m}; \text{start}(f) \\
\text{do } &\sim \text{eof}(f) \\
&\rightarrow \text{etchn}(f, a) \\
&\text{; w} := w_1 \otimes_1 h_1\{a\}, \ldots, w_m \otimes_m h_m\{a\} \\
\text{od}
\end{align*}
\]

2. **Conditional version for (3.1.a).** Instantiate (3.1.a) with the morphism \( h \varphi<, \) where \( \varphi : \alpha \rightarrow \mathbb{B} \). I.e. generalize the specification by

\[
w = h \{ x \in X \mid \varphi x \}
\]

(3.1.a')
Again we calculate a directed reduction, now for the morphism \( h \triangleleft R_s \). This yields \( h \triangleleft R_s = (\ominus \rightarrow 1_G) \), where

\[
 u \ominus a = \text{if } \varphi a \rightarrow u \otimes h \{ a \} \rightarrow \varphi a \rightarrow u \varphi,
\]

which results in the following FMS algorithm for (3.1.a')

\[
w := 1_G; \text{start}(f)
\]

; \( \text{do } \neg \text{eoff(f)} \)

\[
\rightarrow \text{fetchn}(f, a)
\]

\[
\text{if } \varphi a \rightarrow \{ w := w \otimes h \{ a \} \\
\neg \varphi a \rightarrow \text{skip} \}
\]

\[
\text{od}
\]

3. **Superfluous conditions on \( h \).** The representation invariant guarantees that \( R_s \) doesn’t appeal to the idempotency of \( \cup \) (the commutativity is used since \( R_s \{ a, b \} = R_s \{ b, a \} \)). This implies that the same derivation can be made for the application of bag homomorphisms. Let \( k : (\langle \alpha \rangle, \cup, \langle \rangle) \rightarrow (\beta, \oplus, 1_G) \) be a morphism with commutative \( \oplus \). Then the following FMS algorithm is correct.

\[
w := 1_G; \text{start}(f)
\]

; \( \text{do } \neg \text{eoff(f)} \)

\[
\rightarrow \text{fetchn}(f, a)
\]

\[
\neg \varphi a \rightarrow \text{skip}
\]

\[
\text{od}
\]

4. For (3.1.a') and (3.1.e) a remark similar to remark 3 can be made. Note that \( k \varphi \triangleleft : (\langle \alpha \rangle, \cup, \langle \rangle) \rightarrow (\beta, \oplus, 1_G) \) is a morphism with commutative \( \oplus \), if \( k \) is. Lemma 2 follows.

**Lemma 2**

Let \( X \in \{\alpha\} \) and \( f \in \text{File}(\alpha) \) such that \( f \) is a representation for \( X \).\(^2\)

Let \( k : (\langle \alpha \rangle, \cup, \langle \rangle) \rightarrow (\beta, \oplus, 1_G) \) be a morphism with commutative \( \oplus \).

Then \( S_2 \) is a correct FMS algorithm to compute \( k X \), where

\[
S_2: \quad [\parallel X : \{\alpha\} \parallel ; \text{f : File(\alpha)} \parallel ; k : (\langle \alpha \rangle, \cup, \langle \rangle) \rightarrow (\beta, \oplus, 1_G) \parallel ; \text{w : } \beta \parallel ; \text{a : } \alpha \parallel ; \text{w := } 1_G \parallel ; \text{start(f)} \parallel ; \text{do } \neg \text{eoff(f)} \parallel ; \text{fetchn(f, a)}]
\]

\(^2\text{We require that the representation invariant will be established by the declaration mechanism of the system.}\)
Conditional and tupled versions of Lemma 2 are straightforward.

3.2 FMS algorithm to compute $W = \{ h.q.t \mid t \in X \}$

Set specification: $W = h^*q^*X$  \hspace{1cm} (3.2.a)

($h^*q^*X$ is the BM notation for $\{ h.q.t \mid t \in X \}$.)

As mentioned before, we use a transformational solution method consisting of two steps. The method is illustrated in 3.1 on the basis of a simple problem. Here we have to tackle a special case of (3.1.a) because $h^*q^*$ is a morphism too. There is no indication to change our former strategy, so we take the same approach as in 3.1.

step 1. Let elements of $\{\alpha\}$ be implemented by lists with representation function $R_\alpha$ and representation invariant $P_\alpha(x)$ as defined in step 1 of 3.1. Assume $\{\beta\}$ is an available type. We have to find a function $fI$ such that $fI = h^*q^*$. Graphically

\[
\begin{array}{c}
\text{abstract level} \\
(\{\alpha\},\cup,\emptyset) \\
\downarrow \text{q*} \\
(\{\{\alpha\}\},\cup,\emptyset) \\
\downarrow \text{h*} \\
(\{\beta\},\cup,\emptyset)
\end{array}
\quad
\begin{array}{c}
\text{concrete level} \\
(\{\alpha\},\text{t},\emptyset) \\
\downarrow \text{fI} \\
(\{\beta\},\cup,\emptyset)
\end{array}
\quad
\begin{array}{c}
\text{concrete level} \\
(\{\alpha\},\text{t},\emptyset) \\
\downarrow \text{H} \\
(\{\beta\},\cup,\emptyset)
\end{array}
\quad
\begin{array}{c}
\text{abstract level} \\
(\{\beta\},\cup,\emptyset) \\
\downarrow \text{id} \\
(\{\beta\},\cup,\emptyset)
\end{array}
\]

$H$ is a morphism and, since $h^*q^*/\{\cdot\}^* = \cup/\{h^*q^*\{\cdot\}\}^*$, it holds that

$\text{H} = (\emptyset \rightarrow \emptyset)$,

where

$\cup \circ a = \cup \cup \{h.q.a\}$  \hspace{1cm} (3.2.b)

At this point in 3.1, we converted the directed reduction to Conslist level. We don’t do this now because we aim at an efficient implementation of $(\emptyset \rightarrow \emptyset)$. To achieve such an implementation we have to minimize the number of $\cup$-operations and find an efficient computation of $h.q.a$ given a representation $x$ of $X$. Although this last task resembles the subproblem of 3.1, it must be noted that the definition of $q.a$ is relative to the implicit universe $X$, i.e. $(E a)\perp x$ is a representation for $q.a$. Without any assumptions about $x$, an $O(N^2)$ algorithm may be obtained for $H x$, if $\# x = N$. Therefore, we shall assume that equivalent elements in $X$ are consecutive in $x$. 12
In the appendix, this (ordering) constraint is described formally and we introduced the notion 'x is an E-segmentation' for it.

**Representation requirement:** Elements of \{\alpha\} are implemented by E-segmentations.

For \(X \in \{\alpha\}\) represented by an E-segmentation, a representation for its partition is easy to construct. Define a function \(\theta: [\alpha] \rightarrow [[\alpha]]\) by

\[
\begin{align*}
\theta(\[]) &= [\[]] \\
\theta(a) &= [[a]] \\
\theta(y \iff [a]) &= \begin{cases} 
& \text{if } a \in \text{first last } \theta y \rightarrow \text{init } \theta y \parallel [\text{last } y \parallel [a]] \\
& \text{if } a \in \text{first last } \theta y \rightarrow \theta y \parallel [[a]] \\
& \text{fi}
\end{cases}
\end{align*}
\]

\(\theta\) has several interesting properties, e.g.

\[
\begin{align*}
q^\ast R_s &= R_s \ast R_s \ast \theta \quad (3.2.c) \\
x = [] \equiv \theta x = [] \\
x \neq [] &\Rightarrow \forall y \in \theta x : y \neq [] \quad (3.2.d)
\end{align*}
\]

Moreover, \(P_s(\theta x)\) holds, i.e. there are no duplicates in \(\theta x\), or equivalently, each equivalence class of \(X\) corresponds to exactly one element of \(\theta x\).

Unfortunately, \(\theta\) is not a morphism from \(([\alpha], \parallel , [\[]\])\) to \(([[[\alpha]]], \parallel , [\[]\])\). (This is easily seen by taking a list of two equivalent elements \([a,b]\). Then \(\theta[a,b] = [[a,b]] \neq [[a],[b]] = \theta[a] \parallel \theta[b]\).) However, from its definition it is clear that there exists a directed reduction form for \(\theta\)

\[
\theta = (\circ \rightarrow \rightarrow [\[]]) \quad (3.2.f)
\]

where

\[
u \circ a = \begin{cases} 
[\alpha] & \text{if } u = [\[]] \\
\text{init } u \parallel [\text{last } u \parallel [a]] & \text{if } a \in \text{first last } u \\
\theta u \parallel [[a]] & \text{if } \neg a \in \text{first last } u \\
\end{cases}
\]

For \(q^\ast R_s\) we know that \(q^\ast R_s = q^\ast \cup \{-\}^* = \cup/(q \{-\})^*, \) so \((L6),q^\ast R_s = (\circ \rightarrow \rightarrow \emptyset)\) where

\[
u \circ a = u \cup \{ q.a \}
\]

Clearly, a computation via \(\theta\) is more efficient, moreover we implicitly satify the remaining task in improving efficiency, the minimization of \(U\)-operations.

Convinced that our approach can be improved by adding the intermediate level of \(([[[\alpha]]], \parallel , [\[]\])\), we are left with the remaining problem of computing \(h^\ast R_s \ast R_s\) if \(\theta x\) is known. Since \(\theta x\) consists of representations of equivalence classes, we will explore
the use of 3.1, where we showed how morphism application on a set can be computed if a representation of the set is known. Let \( h \) be such as introduced in 3.1, then

\[
R_s h^* = h^* R_s R_s h^*
\]

because both, \( h^* R_s R_s \) and \( R_s h^* \) are morphisms and

\[
\begin{align*}
  h^* R_s R_s [x] &= h^* [x] \\
  R_s [h x] &= R_s h^* [x]
\end{align*}
\]

Summarizing these considerations, we conclude that driven by efficiency the computation of \( \tilde{h} x \) has to be refined by

\[
\begin{align*}
  x_s &= (\ominus \rightarrow [\ ] \rightarrow [\ ]) x \\
  W &= R_s \tilde{h}^* x_s
\end{align*}
\]

Graphically

\[
\begin{array}{cccc}
  \text{abstract level} & \mathcal{R}_s & \text{concrete level} \\
  \{\{\alpha\}\}, U, \emptyset & \overset{q^*}{\Rightarrow} & \{\{\alpha\}\}, U, \emptyset & \overset{\mathcal{R}_s \mathcal{R}_s}{\Rightarrow} & \{\{\alpha\}\}, U, \emptyset & \overset{\emptyset}{\Rightarrow} \\
  \{\{\alpha\}\}, U, \emptyset & \overset{\mathcal{R}_s h^*}{\Rightarrow} & \{\{\beta\}\}, U, \emptyset & \overset{id}{\Rightarrow} & \{\{\beta\}\}, U, \emptyset
\end{array}
\]

In contrast to our previous approach, it is not immediately clear that the successive computations in (3.2.h) can be replaced by the computation of one directed reduction, if a representation \( x \) for \( X \) is given. Therefore, we try to apply formal differentiation, L7.

Since \( \ominus: [\alpha] \times [\alpha] \rightarrow [\alpha] \), \( \mathcal{R}_s h^*: [\alpha] \rightarrow \{\beta\} \) and \( [\ ]: [\alpha] \rightarrow [\ ] \), we look for an operator \( \oplus: \{\beta\} \times [\alpha] \rightarrow \{\beta\} \) such that \( \mathcal{R}_s h^*(\ominus \rightarrow [\ ]) = (\oplus \rightarrow \emptyset) \).

The construction base is correct: \( \mathcal{R}_s \tilde{h}^*(\ominus \rightarrow [\ ]) [\ ] = \emptyset = (\oplus \rightarrow \emptyset) [\ ] \).

Construction hypothesis: \( \mathcal{R}_s \tilde{h}^* \Theta y = (\oplus \rightarrow \emptyset) y \)

Step: \( \mathcal{R}_s \tilde{h}^*(\ominus \rightarrow [\ ]) (y \parallel [a]) \)
case 1 \(=\) \{ case 1: \(y = [ ]\), \(\odot\) \}

\(\mathcal{R}_a \, \tilde{h}^* \, [[a]]\)

\(=\) \{ def * \}

\(\mathcal{R}_a \, [\tilde{h}[a]]\)

\(=\) \{ def \(\tilde{h}\) and \(\mathcal{R}_a\) \}

\(\{h(a)\}\)

case 2 \(\{ \) case 2: \(y \neq [ ] \land a \in E(\text{first last } \theta y)\) \}

\(\mathcal{R}_a \, \tilde{h}^* \, (\text{init } \theta y \models [\text{last } \theta y \models [a]])\)

\(=\) \{ \(\mathcal{R}_a \, \tilde{h}^*\) is a morphism, def * and \(\mathcal{R}_a\) \}

\(\mathcal{R}_a \, \tilde{h}^* \text{init } \theta y \cup \{\tilde{h} \, (\text{last } \theta y \models [a])\}\)

\(=\) \{ \(\tilde{h} = (\oplus \rightarrow 1_\odot),\) (3.1.b) \}

\(\mathcal{R}_a \, \tilde{h}^* \text{init } \theta y \cup \{\tilde{h} \, \text{last } \theta y \otimes h\{a\}\}\)

case 3 \(\{ \) case 3: \(y \neq [ ] \land \neg(a \in E(\text{first last } \theta y))\) \}

\(\mathcal{R}_a \, \tilde{h}^* \, (\theta y \models [[a]])\)

\(=\) \{ \(\mathcal{R}_a \, \tilde{h}^*\) is a morphism, def *, \(\mathcal{R}_a\) and \(\tilde{h}\) \}

\(\mathcal{R}_a \, \tilde{h}^* \theta y \cup \{h\{a\}\}\)

These calculations show that we may succeed in obtaining a directed reduction for (3.2.h) if we tuple the essential components: \(\mathcal{R}_a \, \tilde{h}^* \theta y, \mathcal{R}_a \, \tilde{h}^* \text{init } \theta y, \text{last } \theta y\) and \(\text{first last } \theta y\). Consequently, the type of \(\oplus\) has to be extended and the construction hypothesis will be strengthened in an appropriate way. To adjust the construction base, we define

\(\text{init } [ ] = [ ]\)

\(\text{last } [ ] = [ ]\)

\(\text{first } [ ] = \omega_\alpha\)

where \(\omega_\alpha\) is a fictitious element of \(\alpha\). It follows that

\(\mathcal{R}_a \, \tilde{h}^* \text{init } \theta [ ] = \emptyset\)

\(\tilde{h} \, \text{last } \theta [ ] = 1_\odot\)

\(\text{first last } \theta [ ] = \omega_\alpha\)

We continue our construction for the three additional components.

Strengthened construction hypothesis:

\((\mathcal{R}_a \, \tilde{h}^* \theta, \mathcal{R}_a \, \tilde{h}^* \text{init } \theta, \tilde{h} \, \text{last } \theta, \text{first last } \theta) \, y = (\oplus \rightarrow (\emptyset, \emptyset, 1_\odot, \omega_\alpha)) \, y\)

Step: \(\mathcal{R}_a \, \tilde{h}^* \text{init } \theta (y \models [a])\)

15
case 1,2 \equiv \{ \text{case 1 and 2: } \text{init } y \parallel [a] = \text{init } y \}
\mathcal{R}_s \{ \text{init } y \}
\begin{align*}
\text{case 3} & \quad \{ \text{case 3: } \text{init } y \parallel [a] = y \}
\mathcal{R}_s \{ y \}
\end{align*}

Step:
(\hat{h}, \text{first last } y(\parallel [a]))
\begin{align*}
\text{case 1} & \quad \{ \text{case 1: } \text{last } y(\parallel [a]) = [a], \text{def } \hat{h} \}
(h\{a\}, a)
\end{align*}
\begin{align*}
\text{case 2} & \quad \{ \text{case 2: } \text{last } y(\parallel [a]) = \text{last } y \parallel [a], (3.1.b), (3.2.e) \}
(\hat{h} \text{ last } y \circ h \{a\}, \text{first last } y)
\end{align*}
\begin{align*}
\text{case 3} & \quad \{ \text{case 3: } \text{last } y(\parallel [a]) = [a], (3.1.b) \}
(h\{a\}, a)
\end{align*}

The recognition of case 1 seems to be a final obstacle in the definition of \(\oplus\), but by (3.2.d) it can be shown that
\[ y = [] \equiv \text{init } y = [] \equiv \text{first last } y = \omega_{\alpha} \] (3.2.i)

Summarizing our calculations, we have constructed an operator \(\oplus\) of type
\[ (\{\beta\} \times \{\beta\} \times \beta \times \alpha) \rightarrow (\{\beta\} \times \{\beta\} \times \beta \times \alpha) \] defined by
\[ (V, I, w, pt) \oplus a = \\
\begin{cases}
\{ (h\{a\}, \emptyset, h\{a\}, a) \\
(I \cup \{w \circ h\{a\}, I, w \circ h\{a\}, pt) \\
(V \cup \{h\{a\}\}, V, h\{a\}, a)
\end{cases}
\] if pt = \(\omega_{\alpha}\)
if pt \(\neq\) \(\omega_{\alpha}\) \land pt \(E\) a
if pt \(\neq\) \(\omega_{\alpha}\) \land \neg pt \(E\) a

such that
\[ W = \pi_1 (\oplus \rightarrow (\emptyset, \emptyset, 1_{\emptyset}, \omega_{\alpha})) \times \]

Now we instantiate the standard conversion of a left-reduce
\begin{align*}
W, I, w, pt := \emptyset, \emptyset, 1_{\emptyset}, \omega_{\alpha}; & r := x \\
& \mathbf{do} r \neq []
\quad \rightarrow a := \text{first } r; r := \text{tail } r \\
\quad & \mathbf{if} \quad pt = \omega_{\alpha} \rightarrow W, w, pt := \{h\{a\}\}, h\{a\}, a \\
\quad & \mathbf{if} \quad pt \neq \omega_{\alpha} \rightarrow W, w := I \cup w \circ h\{a\}, w \circ h\{a\} \\
& \mathbf{if} \quad pt E a \rightarrow W, w := W \cup \{w \circ h\{a\}\}, w \circ h\{a\}
\mathbf{fi}
\mathbf{fi}
\mathbf{od}
\end{align*}

This code can be
- smoothened by unfolding the repetition, since only initially pt = \(\omega_{\alpha}\) holds.

16
- improved, because $W = I \cup \{w\}$ is a property of the 4-tuple, if $pt \neq \omega$.

Conslist program for (3.2.a)

\[
W := \emptyset; r := x
\]

; if $r = []$ \rightarrow skip

\[
[ r \neq [] \rightarrow a := \text{first}\ r; r := \text{tail}\ r
\]

; w, pt := h\{a\}, a

; do $r \neq []$

\[
\rightarrow a := \text{first}\ r; r := \text{tail}\ r
\]

(3.2.k)

; if $pt \in a$ \rightarrow w := w \odot h\{a\}

\[
\{} \rightarrow pt \in a \Rightarrow W, w, pt := W \cup \{w\}, h\{a\}, a
\]

\[fi\]

\[
; W := W \cup \{w\}
\]

step 2. Let conslists be implemented by files. To meet the representation requirement, only system orderings can be useful. In the appendix, Lemma p.33, it is proved that an appropriate system ordering $R$ establishes an $R \cap R^-$ segmentation. A representation $f$ of $x$ can be given if an appropriate system ordering exists such that $E = R \cap R^-$.

Assumption: $E = R \cap R^-$ and $c_{\circ R}$ is a system ordering.

Let $f : \text{Ordfile}(\alpha, R)$ and $f = (F, p)$. Let $R_c$ be the representation function as given in 3.1, with the same representation invariant. In the same way as in 3.1, (3.2.k) can be transformed to an FMS algorithm.

FMS algorithm for (3.2.a)

\[
W := \emptyset; \text{start}(f)
\]

; if eof$(f)$ \rightarrow skip

\[
\} \rightarrow \text{eof}(f) \rightarrow \text{fetchn}(f,a)
\]

; w, pt := h\{a\}, a

; do $\} \rightarrow \text{eof}(f)$

\[
\rightarrow \text{fetchn}(f,a)
\]

(3.2.1)

; if $pt \in a$ \rightarrow w := w \odot h\{a\}

\[
\} \rightarrow pt \in a \Rightarrow W, w, pt := W \cup \{w\}, h\{a\}, a
\]

\[fi\]

\[
; W := W \cup \{w\}
\]

\[fi\]

Some remarks on this derivation

1. A tupled version for 3.2.a is obtained by instantiating $h$ with the product morphism defined in 3.1, remark 1. The elaborations are straightforward.
2. Conditional versions for 3.2.a. An instantiation of (3.2.a) with the morphism 
$h \varphi < 1$ yields a conditional computation for $w$ (see 3.1, remark 2), resulting 
in a conditional version of (3.2.1). Independent of this instantiation, we can 
generalize (3.2.a) to the computation of

$$W = h^* q^* \psi < 1 X$$  \hspace{1cm} (3.2.a')

where $\psi \in \alpha \rightarrow \mathbb{B}$ such that

$$u \in t \implies \psi u = \psi t$$  \hspace{1cm} (**)

Note that (**) means that we may define a predicate $\psi'$ on classes such that

$$\psi'< q^* = q^* \psi < 1$$, e.g. $\psi' = \text{some } \psi$, where \text{some } $\psi = \lor \psi^*$. Consequently, we may interchange partitioning and filtering and transform (3.2.a') into

$$W = h^* \psi' < 1 q^* X$$  \hspace{1cm} (3.2.b')

The derivation of an FMS algorithm for (3.2.b') follows the same line as the one for (3.2.a), only some slight modifications have to be made.

- A generalization of (3.2.g) is needed

$$h^* R_s \varphi < 1 R_s = R_s h^* \varphi < 1$$  \hspace{1cm} (3.2.g')

Its proof is completely analogous to that of (3.2.g).

- (3.2.h) has to be modified into

$$x = (\otimes \rightarrow [ ] ) x$$  \hspace{1cm} (3.2.h')

$$W = R_s h^* \psi^* < 1$$

where $\psi''$ is used as an abbreviation for $\psi' R_s$. The correctness of this refinement follows immediately from the filter-map rule and (3.2.g')

$$h^* \psi' < 1 R_s^* R_s = \{ L5 \}$$

$$h^* R_s^* (\psi' R_s) < 1 R_s = \{ (3.2.g') \}$$

$$R_s h^* (\psi' R_s) < 1$$

- The construction for $\oplus$ must be adapted. The reader may easily check the need for tupling $R_s h^* \psi' < 1 \theta y$, $R_s h^* \psi^* < 1 \text{init } \theta y$, $h \text{last } \theta y$, first last $\theta y$ and the modified definition of $\oplus$

$$(V, I, w, pt) \oplus a =$$  \hspace{1cm} (3.2.j')

$$\{ (h\{a\}, \emptyset, h\{a\}, a) \} \quad \text{if } \psi''[a] \land pt = \omega_o$$

$$\{ (\emptyset, \emptyset, h\{a\}, a) \} \quad \text{if } -\psi''[a] \land pt = \omega_o$$

$$\{ (I \cup \{ w \otimes h\{a\}\}, I, w \otimes h\{a\}, pt) \} \quad \text{if } \psi''(\text{last } \theta y \mid [a]) \land pt \neq \omega_o \land pt \text{Ea}$$

$$\{ (V, I, w \otimes h\{a\}, pt) \} \quad \text{if } -\psi''(\text{last } \theta y \mid [a]) \land pt \neq \omega_o \land pt \text{Ea}$$

$$\{ (V, I, h\{a\}, a) \} \quad \text{if } -\psi''[a] \land pt \neq \omega_o \land pt \text{Ea}$$

$$\{ (V, V, h\{a\}, a) \} \quad \text{if } -\psi''[a] \land pt \neq \omega_o \land pt \text{Ea}$$

Remains the computation of $\psi''$. With the definition of $\psi'$ as suggested above,
we find $\psi'' = \text{some } \psi$ and a computation rule for our specific applications might be

$$t \in z \Rightarrow \psi'' z = \psi t$$

In particular, if $\theta y \neq [\ ]$, then

$$\psi'' \text{ last } \theta y = \psi \text{ first last } \theta y$$

The standard conversion for leftreduce can be instantiated and smoothened as before. Again an improvement can be made, since

$$(\psi pt \Rightarrow V = I \cup \{w\}) \land (\neg \psi pt \Rightarrow V = I)$$

is a property of the 4-tuple, if $pt \neq \omega$. The following FMS algorithm results

$${W := \emptyset; \text{start}(f) \quad \text{if } \text{eof}(f) \quad \rightarrow \text{skip} \quad \text{if } \neg \text{eof}(f) \rightarrow \text{fetchn}(f, a) \quad \text{do } \text{end} \text{ if } \text{eof}(f) \rightarrow \text{fetchn}(f, a) \quad \text{if } \neg \text{if } pt \in a \quad \rightarrow w := w \otimes h\{a\} \quad \text{if } \neg \psi pt \rightarrow W := W \cup \{w\} \quad \text{if } \neg \psi pt \rightarrow \text{skip} \quad \text{fi} \quad \text{w, pt := h}\{a\}, a \quad \text{fi} \quad \text{do} \quad \text{end} \text{ if } \text{pt \in a} \quad \rightarrow W := W \cup \{w\} \quad \text{if } \neg \psi pt \rightarrow \text{skip} \quad \text{fi}$$

3. **Superfluous conditions on h.** As explained in 3.1, remark 3, the scheme of (3.2.1) solves a generalization of (3.2.a)

$$W = k^*q^*X$$

where $k : (\alpha \times \alpha, \psi, \langle \rangle) \rightarrow (\beta, \oplus, 1_{\oplus})$ is a morphism with commutative $\oplus$. Remember that $k \varphi : (\langle \alpha \rangle, \psi, \langle \rangle) \rightarrow (\beta, \oplus, 1_{\oplus})$ is a morphism with commutative $\oplus$, if $k$ is.

4. For tupled and conditional versions a remark similar to remark 3 can be made and Lemma 3 follows.

**Lemma 3**

Let $X \in \{\alpha\}$ and E an equivalence relation on $\alpha$ with quotientmap q. Let $R \subseteq \alpha \times \alpha$ be a cgo satisfying $E = R \cap R^{-1}$ and $f \in \text{Ordfile}(\alpha, R)$ such that $f$ is a representation for $X$. Let $k : (\langle \alpha \rangle, \psi, \langle \rangle) \rightarrow (\beta, \oplus, 1_{\oplus})$ be a morphism with commutative $\oplus$,

---

3 The system ordering $R$ in this lemma is a means to meet the representation requirement that elements of Ordfile($\alpha$, R) are E-segmentations.
and $\psi$ a predicate on $\alpha$ satisfying

$$u \in E \Rightarrow \psi u = \psi t.$$  

Then $S_3$ is a correct FMS algorithm to compute $k^*q^*\psi q X$, where

$$S_3: \left[ \begin{array}{|l|} 1 \vert X : \{\alpha\} ; E, R : \alpha \times \alpha ; f : \text{Ordfile}(\alpha, R) ; \\
\vert k : (\langle \alpha \rangle, \cup, \langle \rangle) \to (\beta, \emptyset, 1_{\emptyset}) ; \psi : \alpha \to B ; \\
\vert W : \{\beta\} \\
\vert\vert w : \beta ; pt, a : \alpha ; \\
\vert W := \emptyset ; \text{start}(f) \\
\vert\vert\text{if } \text{eof}(f) \to \text{skip} \\
\vert\vert\text{if } \neg\text{eof}(f) \to \text{fetchn}(f, a) \\
\vert\vert\quad ; w, pt := k\{a\}, a \\
\vert\vert\quad ; \text{do } \neg\text{eof}(f) \\
\vert\vert\quad \to \text{fetchn}(f, a) \\
\vert\vert\quad \text{if } \neg\text{pt} \in a \quad \to w := w \cup k\{a\} \\
\vert\vert\quad \text{if } \neg\psi pt \to W := W \cup \{w\} \\
\vert\vert\quad \text{fi} \\
\vert\vert \text{fi} \\
\vert\text{od} \\
\vert\text{if } \neg\psi pt \to W := W \cup \{w\} \quad \text{fi} \\
\vert\text{fi} \\
\end{array} \right]$$

Conditional and tupled versions of Lemma 3 are straightforward.

### 3.3 Extensions.

There are several problems which contain the former ones as a subproblem. Sometimes the solution of the subproblem may serve as a building block for the overall solution. We will treat two of them.

#### 3.3.1 FMS algorithm to compute $W = \{h.q.t, g.q.t \mid t \in X\}$

where $g = (\emptyset \not\rightarrow e)$.

**Solution.** Instantiate Lemma 3 with $\psi \equiv \text{true}$. Add the invariant for the computation of $(\emptyset \not\rightarrow e) Q$ where $Q$ is the representation of $q.pt$.

Then the correctness of the following FMS algorithm is easily checked.

$$W := \emptyset ; \text{start}(f) \\
\text{if } \text{eof}(f) \to \text{skip} \\
\text{if } \neg\text{eof}(f) \to \text{fetchn}(f, a) \\
\quad ; w, pt, g0 := h\{a\}, a, e \cup a \\
\text{do } \neg\text{eof}(f)$$

20
3.3.2 FMS algorithm to compute

\[ W_1 = \{ h_1.q_1.t \mid t \in X \} \land W_2 = \{ h_2.q_2.t \mid t \in X \} \]

where \( h_1 : (\{\alpha\}, \cup, \emptyset) \rightarrow (\beta_1, \otimes, \mathfrak{I}) \) and \( h_2 : (\{\alpha\}, \cup, \emptyset) \rightarrow (\beta_2, \otimes, \mathfrak{I}) \) are morphisms, \( E_1 \) and \( E_2 \) equivalence relations on \( \alpha \) such that \( E_2 \subseteq E_1 \) and \( q_1 \) and \( q_2 \) the quotients map of \( E_1 \) and \( E_2 \).

Solution. Assume \( X \) is an \( E_1 \) and \( E_2 \) segmentation. Instantiate Lemma 3 with \( \psi = \text{true} \) for \( E_1 \) for \( E_2 \) simultaneously. Using the fact that \( \text{pt}_2 E_1 \text{pt}_1 \land \neg \text{pt}_1 E_1 t \Rightarrow \neg \text{pt}_2 E_2 t \), we obtain

\[
W_1, W_2 := \emptyset, \emptyset; \text{start}(f) \\
\text{if eof}(f) \rightarrow \text{skip} \\
\[]{\neg \text{eof}(f) \rightarrow \text{fetchn}(f,a)} \\
\quad ; w, \text{pt}_1, w_2, \text{pt}_2 := h_1\{a\}, a, h_2\{a\}, a \\
\quad ; \text{do} {\neg \text{eof}(f)} \\
\quad \rightarrow \text{fetchn}(f,a) \\
\quad ; \text{if pt}_1 E_1 a \rightarrow \\
\quad \quad \text{if pt}_2 E_2 a \rightarrow w := w_2 \otimes h_2\{a\} \\
\quad \quad \{\neg \text{pt}_2 E_2 a \rightarrow W_2, w, \text{pt}_2 := W_2 \cup \{w_2\}, h_2\{a\}, a \} \text{ fi} \\
\quad \quad ; w := w_1 \otimes h_1\{a\} \\
\quad \quad \{\neg \text{pt}_1 E_1 a \rightarrow \\
\quad \quad W_1, w_1, \text{pt}_1 := W_1 \cup \{w_1\}, h_1\{a\}, a \\
\quad \quad ; W_2, w_2, \text{pt}_2 := W_2 \cup \{w_2\}, h_2\{a\}, a \} \text{ fi} \\
\quad \text{od} \\
\quad ; W_1, W_2 := W_1 \cup \{w_1\}, W_2 \cup \{w_2\} \text{ fi}
\]

Note that it is not difficult to adjust (3.3.2.b) to the computation of

\[ W_1 = \{ [h_1.q_1.t, \{ h_2.q_2.u \mid u \in q_1.t \}] \mid t \in X \} \]

4. GENERIC FILE ALGORITHMS APPLIED TO DATABASES

Now, we shall illustrate the use of the algorithms for database queries. In section
2.1, we described a simple databasetype and introduced some database notions. Generally, databasetypes are more complex and additional notions are involved. We will continue our informal database description of section 2.1 for the general case and illustrate the several notions on the basis of our example databasetype SR defined below.

SR models the registration of students and their examinations. Two objects, ST (student) and RES (result), are distinguished. The attribute-names for ST and RES are inventoried and fixed in the so-called database-skeleton $g$, [3]. STR gives the relation between attribute-name and corresponding attribute-value set for ST. For RES this is done in RESR. The tabletype for ST consists of subsets $T$ of the labeled cartesian product $\prod STR$, which satisfy the table-constraint $\text{key}(\{RNR\}, T)$. For RES the tabletype RESF is given.

To arrive at a databasetype, we have to consider the (labeled) cartesian product of tabletypes, the label set being the set of objects (in our example: $\prod DK$). Often, not every combination represents a possible state, a so-called database-constraint has to be satisfied (in our example: $\forall t \in v(RES) \exists s \in v(ST) : t(RNR) = s(RNR)$). Finally, the databasetype consists of all combinations which satisfy the database-constraint (in our case: SR).

Definition of SR

database-skeleton $g$

$$g = \{ \text{ST} ; \}
\begin{align*}
&\{ \text{RNR} \quad \text{registration number} \\
&\text{Y1} \quad \text{first year of registration} \\
&\text{ENQ} \quad \text{entrance qualification} \\
&\text{DEPTC} \quad \text{department code} \\
&\text{SEV} \quad \text{additional information} \\
\} \\
&\text{RES} ; \\
&\{ \text{RNR} \\
&\text{Ci} \quad \text{course index} \\
&\text{EDATE} \quad \text{date of examination} \\
&\text{MARK} \}
\}$$

tabletype for ST is STF

$$\text{STF} = \{ T \in \{\prod \text{STR}\} \mid \text{key}(\{ \text{RNR}\}, T) \}$$

tabletype for RES is RESF

$$\text{RESF} = \{ T \in \{\prod \text{RESR}\} \mid \text{key}(\{ \text{RNR}, \text{Ci}, \text{EDATE}\}, T) \}$$
where

\[ \text{STR} = \{ (\text{RNR} ; [1..1000000]) \]
\[ (\text{Y1} ; [00..99]) \]
\[ (\text{ENQ} ; \text{CHAR}(80)) \]
\[ (\text{DEPTC} ; [1..100]) \]
\[ (\text{SEV} ; \text{CHAR}(80)) \} \]
\[ \text{RESR} = \{ (\text{RNR} ; [1..1000000]) \]
\[ (\text{CI} ; \text{CHAR}(5)) \]
\[ (\text{EDATE} ; \text{integer}) \]
\[ (\text{MARK} ; \text{integer}) \} \]

and \( \text{key}(S,T) \equiv (\forall t_1, t_2 \in T : t_1 | S = t_2 | S \Rightarrow t_1 = t_2) \)

\text{databasetype SR}

\[ \text{SR} = \{ v \in \prod \text{DK} | \forall t \in v(\text{RES}) \exists s \in v(\text{ST}) : t(\text{RNR}) = s(\text{RNR}) \} \]

with \( \text{DK} = \{ (\text{ST} ; \text{STF}), (\text{RES} ; \text{RESF}) \} \).

Furthermore a tabletype \( \text{AVF} \) exists modelling averages of results per course index:

\[ \text{AVF} = \{ T \in \prod \text{AVR} | \text{key} \{\text{CI}, T\} \} \]

and \( \text{AVR} = \{ \) (CI ; CHAR(5))
\[ , (\text{NB} ; [1..1000000]) \text{ number of students} \]
\[ , (\text{AV} ; \text{real}) \text{ average result} \} \]

Given a database \( v \in \text{SR} \). Construct a table \( W : \text{AVF} \) containing the following information:

For each course with one or more results for A-students
- course index
- number of A-students with one or more results in this course
- average number of results for this course per A-student

Here an A-student is a student with entrance qualification 'A'.

\textbf{Solution}

\textbf{Specification in set notation}

\[ W = \{ ((\text{CI}, w_1), (\text{NB}, w_2), (\text{AV}, w_3)) | t \in v(\text{RES}) \land w_0 > 0 \} \]

where

\[ w_1 = t(\text{CI}) \] (4.a)
\[ w_2 = \# \{ u(\text{RNR}) | u \in v(\text{RES}) \land u(\text{CI}) = t(\text{CI}) \land \varphi(u) \} \]
\[ w_3 = w_0 / w_2 \]

23
with \[ \varphi(u) = \exists s \in v(ST) : s(RNR) = u(RNR) \land s(ENQ) = 'A' \]

Analyzing this specification, we may conclude that part of the attribute-values to be computed arise from morphism applications on equivalence classes, so we direct our solving strategy to the application of some version of Lemma 3.

Instantiate \( \alpha \) with \( \prod RESR \).

Since \( RESF \subseteq \{ \prod RESR \} \) it holds that \( v(RES) \in \{ \prod RESR \} \). Define
\[
\begin{align*}
\imath_{E_1} & \triangleq u(Cl) = t(Cl)
\end{align*}
\]

Clearly, \( E_1 \) is an equivalence relation on \( \prod RESR \). Let \( q_1 \) be the quotientmap of \( E_1 \) on \( v(RES) \). Let \( R \) be a linear ordering on \( \text{CHAR}(5) \), then we define \( c_0 : R_1 \) by
\[
\begin{align*}
u_1, u_2 & \equiv u_1(Cl) R u_2(Cl)
\end{align*}
\]

It holds that \( E_1 = R_1 \cap R_1^- \), see appendix. Assume that \( R_1 \) is a system ordering.

Let \( f \in \text{Ordfile}(\prod RESR, R_1) \) such that \( f \) is a representation for \( v(RES) \).

Given an element \( t \in v(RES) \), \( w_0 \) and \( w_1 \) can be defined as morphism applications on \( q_1.t \), as follows.
- Consider the monoid \( M_0 = (\mathbb{N},+,0) \). \( + \) is associative and commutative.

Define \( k_0 : (\{ \prod RESR \}, \cup, \emptyset) \rightarrow M_0 \) by
\[
\begin{align*}
k_0\{u\} & = 1
\end{align*}
\]

Then \( w_0 = k_0 \{ u \in q_1.t \mid \varphi(u) \} \)

- Let \( b \in \text{CHAR}(5) \). Consider the two-element set \( \{ 1_0, b \} \) with operation \( \circ \) defined by
\[
\begin{array}{c|cc}
\circ & 1_0 & b \\
1_0 & 1_0 & b \\
b & b & b
\end{array}
\]

then \( \circ \) is associative, commutative and idempotent and \( M_1 = (\{ 1_0, b \}, \circ, 1_0) \) is a monoid. Define \( k_1 : (\{ \prod RESR \}, \cup, \emptyset) \rightarrow M_1 \) by
\[
\begin{align*}
k_1\{u\} & = u(Cl)
\end{align*}
\]

Then \( w_1 = k_1 q_1.t \)

The predicate \( w_0 > 0 \) contains a free variable \( t \) and trivially satisfies the required condition for \( \psi \), since \( u_1 E_1 u_2 \equiv q_1 u_1 = q_1 u_2 \).

A first approximation for a partial solution for (4.a) is obtained by instantiating a tupled version for Lemma 3, with a conditional version for one of its components.

first approximation
\[
\begin{align*}
| f : \text{Ordfile}(\prod RESR, R_1) \& W : \text{AVF}
\end{align*}
\]

24
Two tasks remain before a complete solution for (4.a) is obtained: the computations of \( \varphi \) and of \( W_2 \). But firstly we mention that (4.b) can be improved, due to the definition of \( <> \). The computation of \( W_1 \) can even be omitted, since \( W_1 = pt_1(CI) \).

- Task 1: computation of \( \varphi \)

From the definition of SR it follows that there exists an \( s \in v(ST) \) such that \( s(RNR) = a(RNR) \). Since \( \text{key}(\{RNR\}, v(ST)) \) holds we can retrieve \( s \) via the fetchn operation. So a correct implementation for \( \varphi \) is

\[
\text{fetchc}(v(ST), a\{\text{RNR}, s\})
\]

\( \varphi = s(\text{ENQ}) = 'A' \)

There are two reasons why we do not plug in this computation in (4.b)

1. The code would be unnecessarily obscure. We propose the introduction of a procedure \( \phi \) with input parameter \( r : \Pi \text{RESR} \), denoted by \( | r | \), and output parameter \( b : \text{IB} \), denoted by \( | b | \), to define by

\[
\text{proc } \phi = (| a : \Pi \text{RESR}, | b : \text{IB}, | s : \Pi \text{STR};
\]

\]
2. It may lead to redundant table-operations. Once the entrance qualification of
a student is known, this suffices for its other results (of this course). So we
decide for a clustering per student within an equivalence class. Formally, this
results in the introduction of a new equivalence relation

Define equivalence relation $E_2$ by

$$ u_1 E_2 u_2 \equiv u_1(CI) = u_2(CI) \land u_1(RNR) = u_2(RNR) $$

with quotient map $q_2$. It holds that

$$ u_1 E_2 u_2 \Rightarrow \varphi u_1 = \varphi u_2 $$

and we may define a predicate $\varphi'$ such that

$$ u \in q_2.t \Rightarrow \varphi' q_2.t = \varphi u $$

Therefore we choose for an extension like 3.3.2, however now the elements of the
refined relation are of temporary interest and we do not explicitly construct the
variable $W_2$. Let $\leq$ the usual ordering on $\mathbb{N}$. Define $c q_0 R_2$ by

$$ u_1 R_2 u_2 \equiv u_1 R_1 u_2 \land (u_1(CI) \neq u_2(CI) \Rightarrow u_1(RNR) \leq u_2(RNR)). $$

Then $E_2 = R_2 \cap R_2^{-}$. Assume $R_2$ is a system ordering and $f \in \text{Ordfile}(\prod RESR, R_2)$ such that $f$ is a representation for $v(RES)$. A refinement of (4.b) is then given by

second approximation

$$ f : \text{Ordfile}(\prod RESR, R_2) $$

$$ W : \text{AVF} $$

$$ w_0 : \mathbb{N} ; pt_1, pt_2, a : \prod RESR; s : \prod STR; b : \mathbb{B} ; $$

$$ \text{empty}(W); \text{start}(f) $$

\[ \text{if } \text{eof}(f) \Rightarrow \text{skip} \]

\[ \neg \text{eof}(f) \Rightarrow \text{fetchn}(f,a) \]

\[ \text{phi}(a,b) \]

\[ \text{if } b \rightarrow w_0 := 1 \Rightarrow \neg b \rightarrow w_0 := 0 \Rightarrow \text{fi} \]

\[ \text{pt}_1, \text{pt}_2 := a, a \]

\[ \text{do } \neg \text{eof}(f) \]

\[ \rightarrow \text{fetchn}(f,a) \]

\[ \text{if } \text{pt}_1 E_1 a \Rightarrow \text{if } \text{pt}_2 E_2 a \Rightarrow \text{skip} \]

\[ \neg \text{pt}_2 E_2 a \Rightarrow \text{phi}(a,b); \text{pt}_2 := a \Rightarrow \text{fi} \]
- Task 2: computation of \( W^2 \).

Given that \( f \) is an \( E_1 \) and an \( E_2 \) segmentation, \( W^2 \) can be defined by

\[
W^2 = (\Theta \rightarrow 0) q_1.t
\]

where

\[
u \Theta a = \begin{cases} u & \text{if } (\varphi a \land pt_2 E_2 a) \lor \neg \varphi a \\ u + 1 & \text{if } (\varphi a \land \neg pt_2 E_2 a)\end{cases}
\]

Using an extension like 3.3.1 and adding some details, (4.a) is solved.

Solution of (4.a)

\[
\begin{array}{l}
[[ f : \text{Ordfile}(\prod \text{RESR}, R_2) \\
W : \text{AVF} ]
\end{array}
\]

\[
\begin{array}{l}
[[ w_0, w_2 : \mathbb{N} ; pt_1, pt_2, a : \prod \text{RESR} ; s : \prod \text{STR} ; b : \mathbb{B} ]
\end{array}
\]

\[
\begin{array}{l}
\text{proc phi = (} a : \prod \text{RESR} , s : \prod \text{STR} | \\
\text{fetchc}(v(ST), \text{a[RNR]}, s) \\
\text{if } s(\text{ENQ}) = 'A' \rightarrow b := \text{true} \\
\text{if } s(\text{ENQ}) \neq 'A' \rightarrow b := \text{false} \\
\text{fi }]
\end{array}
\]

\[
\begin{array}{l}
\text{empty}(W); \text{start}(f) \\
\text{if } \text{eof}(f) \rightarrow \text{skip}
\end{array}
\]

\[
\text{do } \neg \text{eof}(f) \rightarrow \text{fetchn}(f,a)
\]

\[
\begin{array}{l}
\text{phi}(a,b) \\
\text{if } b \rightarrow w_0 := w_0 + 1 \neg b \rightarrow \text{skip}
\end{array}
\]

\[
\begin{array}{l}
\text{store}(W,((\text{CI}, pt_1(\text{CI})))),
\text{if } w_0 > 0 \rightarrow \text{store}(W,((\text{CI}, pt_1(\text{CI})))),
\text{if } w_0 \leq 0 \rightarrow \text{skip}
\end{array}
\]

\[
\begin{array}{l}
\text{phi}(a,b) \\
\text{if } b \rightarrow w_0 := 1 \neg b \rightarrow w_0 := 0 \\
\text{if } pt_1, pt_2 := a, a
\end{array}
\]

\[
\begin{array}{l}
\text{do } \neg \text{eof}(f) \\
\text{fetchn}(f,a) \\
\text{if } pt_1 E_1 a \rightarrow
\end{array}
\]

27
if \( pt_2 \triangleright a \rightarrow \) skip
\[ [\triangleright \neg pt_2 \triangleright a \rightarrow \]
\[ \phi(a,b) \]
\[ ; \text{if } b \rightarrow w_2 := w_2 + 1 [\triangleright b \rightarrow \text{skip } fi \]
\[ ; \text{pt}_2 := a \]
\[ fi \]
\[ ; \text{if } b \rightarrow w_0 := w_0 + 1 [\triangleright b \rightarrow \text{skip } fi \]
\[ [\triangleright \neg pt_1 \triangleright a \rightarrow \]
\[ \text{if } w_0 > 0 \rightarrow \]
\[ \text{store}(W,\{(CI,pt_1(CI)),(NB,w_2),(AV,w_0/w_2)\}) \]
\[ [\triangleright w_0 \leq 0 \rightarrow \text{skip} fi \]
\[ ; \phi(a,b) \]
\[ ; \text{if } b \rightarrow w_0, w_2 := 1, 1 [\triangleright b \rightarrow w_0, w_2 := 0, 0 fi \]
\[ ; \text{pt}_1, \text{pt}_2 := a,a \]
\[ fi \]
\[ od \]
\[ ; \text{if } w_0 > 0 \rightarrow \text{store}(W,\{(CI,pt_1(CI)),(NB,w_2),(AV,w_0/w_2)\}) \]
\[ [\triangleright w_0 \leq 0 \rightarrow \text{skip} fi \]
\[ fi \]
\[ ]] ] ]

5. SYSTEM-REQUIREMENTS

In our problem solution we supposed the existence of a ordering \( R \) such that \( E = R \cap R^\rightarrow \). Here, we shall discuss some pragmatic aspects of ordering facilities in the available systems. Ordering facilities are given by syntactical constructs whose semantics require some primary facts about tabletypes:

- a tabletype is defined over a heading, i.e. a set of attribute names or table indices.
- each attribute(name) uniquely determines a domain, the attributevalue-set.

So let \( TT \) be a tabletype, then we shall denote its heading by \( H(TT) \) while \( F(A) \) is used for the domain of an attribute \( A \in H(TT) \). Let \( T : TT \).

In the syntactical constructs to achieve relations on \( T \), the user is only allowed to give an enumeration of \( D \subseteq H(TT) \). This so-called ordering list is part of the basis for some lexicographical relation \( S \) on \( T \) \( \triangleright D \) which in its turn defines (the semantics of) the relation \( R \) on \( T \):

\[
\text{u R t \equiv u \triangleright D S t \triangleright D} \quad (\ast)
\]

We continue with \( D = \{A_1, A_2\} \) and orderingslist \( ed = [A_1, A_2] \) in our illustrations. A complete basis for a lexicographical relation consists of a unique (labeled) cartesian product together with a relation for each of its factors. Such a unique cartesian product is implicitly derived from the ordering list

\[
\Pi (A_1 : F(A_1), A_2 : F(A_2))
\]

28
The relations on the factors are taken to be standard orderings available by the system. This implies the existence of at most one such ordering per domain. Generally those standard orderings are linear ones and a user must give its preference with respect to ascending or descending traversal by adding an adjective to each element of the ordering list, e.g. $ed = [asc A_1, asc A_2]$.

Given that standard orderings are linear, it follows that $S$ is linear and consequently, relations $R$ on $T$ defined by (*) have interesting properties. However, this puts its restrictions on the possibilities for $E$, even so that the only relation satisfying our requirement is the discrete relation on $D$. For readers familiar with the relational approach, the following proof will do:

Let $R_1$ and $R_2$ be linear orderings on $F(A_1)$ and $F(A_2)$ respectively, then $S = R_1 \parallel R_2$. By the definition of backward image relation, it holds that

$$R = \ll^{-} \circ S \circ \ll$$

which equals, fact 10,

$$R = S \parallel \pi$$

Now, we calculate

$$E \equiv R \cap R^{-} \equiv (S \parallel \pi) \cap (S \parallel \pi)^- \equiv \{ \cap \text{ distributes over } \parallel \} \equiv (S \parallel \pi) \cap (S^- \parallel \pi) \equiv \{ S \text{ is linear, fact 5 } \} \equiv I \parallel \pi$$

**6. CONCLUSIONS**

- We constructed a program for a database query by instantiating appropriate schemes for files. Since the problems which were solved in these schemes, relate to the application of structure preserving functions, a transformational programming method could be used in the derivation of those schemes.

- The advantage of schemes is trivial: no correctness proofs have to be given, instantiation of the schemes suffices. Especially in the development of imperative programs for database queries one may profit by schemes, because deriving imperative programs for databases is almost unfeasible in the usual way.

- Since imperative programs were our goal we had to put up with extensive code for our schemes. If abstract programs would suffice as solution for our problems,
then we could have terminated the derivations at the point where the definition of the directed reduction is known. In both cases, the transformation of the directed reduction to file level is the same and can be considered as a standard one.

- $S_2$, the generic scheme in Lemma 2, can be viewed as a (4G system) conversion to file level, if tables are implemented by sets. Each function $k$ satisfying the conditions of Lemma 2 can be made available as a standard function. However, if tables are implemented by bags, a function such as $k_0$ in section 4, is not correctly implemented by $S_2$.

- An optimal use of $S_3$, the generic scheme in Lemma 3, can only be made if the programmer is acquainted with the theory of relations, in particular relations on cartesian products. E.g. during our derivation the need for an extra representation requirement (E segmentation) arises. Having only system orderings to satisfy such a condition on the data we had to prove that these suffice to meet the requirement. See appendix, Lemma, p.33.

Acknowledgments
I would like to thank Kees Hemerik and Jaap van der Woude for their useful comments on earlier drafts of this paper. Their suggestions relating to both accuracy and clarity considerably improved the presentation.

---

4We could not find an appropriate reference for this subject. Therefore we collected the relevant notions and facts ourselves and added this as an appendix to this note. For ease of manipulation we chose for a pointless notation.
APPENDIX.

For reader's reference we give a short overview of our approach to relations on
tables and some definitions and facts from the theory of relations, [6], as used in our
calculations. Although several notions have a wider scope of applicability, our main
interest is in homogeneous relations.

For each tabletype TT, there exists a labeled cartesian product LCT such that
LCT is the smallest product such that ∪ TT ⊆ LCT.
By fixing the ordering of the labelset, LCT is isomorphic to a cartesian product CT.
Relations defined on CT induce relations on each element of TT.

Relations are mathematical objects on which, apart from the set operations ∪,
∩ and ∩, the following operations are defined:

- unary: reversion denoted by
- binary: composition denoted by

Since the definitions of those operators is common knowledge, it suffices to give
priority rules with respect to their use
- unary operators bind stronger than binary ones,
- composition has a higher priority than ∪ or ∩

Some concrete relations are frequently used. Let X be a set, then

Ix = \{(x, x) | x ∈ X\}
\Pi X = X × X

are relations. We omit subscripts and infer type information from the context.

Several laws hold for the relational structure, e.g.:

R ∩ I = I ∩ R = R
R ⊆ S ⇒ (R ∩ T ⊆ S ∩ T ∧ T ∩ R ⊆ T ∩ S)
¬ (R− ) = (¬ R)−
○ distributes forward and backward over ∪

Relations may have several properties. Let R be a relation, then

**Definition [functional properties]**

R ∩ R− ⊆ I

R− ○ R ⊆ I

R− ○ R ⊇ I

R ○ R− ⊇ I

R ○ R− ⊆ I

R is functional

R is injective

R is total

R is surjective

Well-known combination: R is bijective ∆ R ∩ R− = R− ○ R = I

**Definition [ordering properties]**
I $\subseteq R$ $\triangleq$ R is reflexive
$R \circ R \subseteq R$ $\triangleq$ R is transitive
$R \subseteq R^\sim$ $\triangleq$ R is symmetric
$R \cap R^\sim \subseteq I$ $\triangleq$ R is antisymmetric
$\neg R^\sim \subseteq I \cup R$ $\triangleq$ R is connective

Apart from the well-known combinations, linear ordering (reflexive, transitive, antisymmetric and connective) and equivalence relation (reflexive, transitive and symmetric), we use:

- R is a quasi ordering (qo) $\triangleq$ R is reflexive and transitive
- R is a cqo $\triangleq$ R is a connective qo

A special instance of a bijection is found in

**Definition [enumeration]**
Let $V$ be a set and $\#V = N$.

F is an enumeration of $V \triangleq F : \{1..N\} \rightarrow V$ is a bijection.

**Fact 1.** If $R$ is a qo then $R \cap R^\sim$ is an equivalence relation.

**Fact 2.** An equivalence relation $E$ on $X$ defines a quotient map $q_E : X \rightarrow X/E$ by

$q_E(x) = \{ y \in X \mid x E y \}$

$q_E$ is total, surjective and functional. (If no confusion can occur, subscripts are omitted.)

**Fact 3.** If $R$ is a cqo, $R \subseteq S$ and $S$ is transitive, then $S$ is a cqo.

**Fact 4.** If $f$ is a function, then

$f^\sim \circ (R \cap S) = f^\sim \circ R \cap f^\sim \circ S$

$(R \cap S) \circ f = R \circ f \cap S \circ f$

**Fact 5.** If $R$ is linear, then $R^\sim$ is and $R \cap R^\sim = 1$

The following common way of defining relations is of particular interest for sets on which no ordering structure is imposed.

**Definition [backward image relation]**
Let $f : X \rightarrow Y$ and $T \subseteq Y \times Y$. The backward image relation $S$ of $T$ under $f$ is defined by

$S = f^\sim \circ T \circ f$

**Definition [forward image relation]**
Let $f : X \rightarrow Y$ and $R \subseteq X \times X$. The forward image relation $S$ of $R$ under $f$ is defined by

$S = f \circ R \circ f^\sim$

Depending on the $^\sim$ (functional) properties of $f$, the (ordering) properties of $R$ or $T$
are transferred to $S$. For our purpose we mention

**Fact 6.** The image of a linear ordering under $f$ is linear, if $f$ is a bijection.

E.g. If $F$ is an enumeration of $V$ then $F \circ \leq \circ F^\ast$ is linear.

**Fact 7.** If $f$ is a total, surjective function and $T$ is linear, then $S$ is a c.qo.

**Fact 8.** If $R$ is a c.qo and $q$ is the quotient map of $R \cap R^\ast$, then $S$ is linear.

On cartesian products, relations are often defined via relations on the factors. Using projection functions, the well-known product and lexicographical relation can be introduced via the backward image relation. Let $X$, $Y$ and $Z$ be sets and $R \subseteq X \times X$ and $T \subseteq Y \times Y$. Then

**Definition [projection functions $\ll$ and $\gg$]**

$\ll : X \times Y \to X$ and $\gg : X \times Y \to Y$ are defined by

$\ll (x, y) = x \quad \gg (x, y) = y$

$\ll$ and $\gg$ are total and surjective functions.

**Definition [product relation and lexicographical relation, $\parallel$ and $\|\|$]**

On $X \times Y$, the product relation $R \parallel T$ and the lexicographical relation $R \| T$ are defined by

\[
R \parallel T = (\ll \circ R \circ \ll) \cap (\gg \circ T \circ \gg)
\]

\[
R \| T = (R \parallel T) \cap ((- I \parallel T) \cup (T \parallel T))
\]

$X \times (Y \times Z)$ differs from $(X \times Y) \times Z$, so there is no associativity for $\parallel$ and $\|$ in the usual way. However, the cartesian products are isomorphic and therefore we model associativity of those operators by isomorphism.

Still using implicit typing we mention some relevant laws for the product and lexicographical relations

\[
\ll^{-1} \circ \Pi \circ \ll = \Pi = \gg^{-1} \circ \Pi \circ \gg
\]

\[
\Pi \parallel \Pi = \Pi
\]

$\parallel$ and $\|$ distribute forward and backward over $\cap$ and $\cup$

$\ll^{-1}$ distributes over $\parallel$ and $\|$

**Fact 9.** If $R$ and $T$ are linear, then $R \| T$ is.

**Fact 10.** $R \parallel \Pi = \ll^{-1} \circ R \circ \ll = R \| \Pi$

**Definition [monotonicity]**

Let $f : X \to Y$, $R \subseteq X \times X$ and $T \subseteq Y \times Y$.

$f$ is monotonic w.r.t. $R$ and $T \triangleq R \subseteq f^{-1} \circ T \circ f$

**Fact 11.** If $f$ is monotonic w.r.t. $R$ and $T$ and $g$ is monotonic w.r.t. $T$ and $U$, then
g ◦ f is monotonic w.r.t. R and U.

**Fact 12.** If f is total, then f is monotonic w.r.t. R and its forward image.
E.g. If F is an enumeration of V, then F is monotonic w.r.t. ≤ and F ◦ F−.

**Fact 13.** || and ∥ are monotonic w.r.t. ⊆, in both arguments.

**Definition [F satisfies R]**

Let F be an enumeration of V and R ⊆ V × V, then

\[
F \text{ satisfies } R \triangleq \leq \supseteq F \circ R \circ F
\]

**Fact 14.** F satisfies R ⊃ F is monotonic w.r.t. ≤ and R

**Fact 15.** If F satisfies Rand R ~ S then F satisfies S.

**Definition [E-segmentation]**

Let F be an enumeration of V and E ⊆ V × V an equivalence relation with quotientmap q. Let ⊆ be a linear ordering of q*V, then

\[
F \text{ is an E-segmentation of } V \triangleq q \circ F \text{ is monotonic w.r.t. } \leq \text{ and } \subseteq.
\]

**Lemma** If F satisfies R, then

\[
R \text{ is a c.q.o } \Rightarrow F \text{ is an } R \cap R^- \text{ segmentation}
\]

**Proof**

\[
F \text{ is an } R \cap R^- \text{ segmentation} \\
\iff q \circ F \text{ is monotonic w.r.t. } \leq \text{ and } q \circ R \circ q^- \text{ is linear} \\
\iff q \text{ is monotonic w.r.t. } R \text{ and } q \circ R \circ q^- \text{ and } q \circ R \circ q^- \text{ is linear} \\
\iff R \text{ is a c.q.o}
\]

**Corollary** If R ⊆ S and S is transitive, then

\[
F \text{ satisfies c.q.o } R \Rightarrow F \text{ is a } S \cap S^- \text{ segmentation.}
\]

**Proof.** Immediately from facts 3, 15.
9. BIBLIOGRAPHY


In this series appeared:

89/1 E.Zs.Lepoeter-Molnar
Reconstruction of a 3-D surface from its normal vectors.

89/2 R.H. Mak
A systolic design for dynamic programming.

89/2 P.Struik

89/3 H.M.M. Ten Eikelder
Some category theoretical properties related to
C. Hemerik a model for a polymorphic lambda-calculus.

89/4 J.Zwiers
Compositionality and modularity in process
W.P. de Roever specification and design: A trace-state based
approach.

89/5 Wei Chen
Networks of Communicating Processes and their
T.Verhoeff (De-)Composition.

89/6 T.Verhoeff
Characterizations of Delay-Insensitive
J.T.Udding Communication Protocols.

89/7 P.Struik
A systematic design of a parallel program for

89/8 E.H.L.Aarts
Dirichlet convolution.
A.E.Eiben
K.M. van Hee

89/9 K.M. van Hee
Discrete event systems: Dynamic versus static
P.M.P. Rambags topology.

89/10 S.Ramesh
A new efficient implementation of CSP with output
K.M. van Hee guards.

89/11 S.Ramesh
Algebraic specification and implementation of infinite
K.M. van Hee processes.

89/12 A.T.M.Aerts
A concise formal framework for data modeling.
K.M. van Hee

89/13 A.T.M.Aerts
A program generator for simulated annealing
K.M. van Hee problems.
M.W.H. Hesen

89/14 H.C.Haesen
ELDA, data manipulatie taal.

89/15 J.S.C.P. van
Optimal segmentations.
der Woude

89/16 A.T.M.Aerts
Towards a framework for comparing data models.
K.M. van Hee

89/17 M.J. van Diepen
A formal semantics for Z and the link between
K.M. van Hee Z and the relational algebra.
Formal methods and tools for the development of distributed and real time systems, p. 17.

Dynamic process creation in high-level Petri nets, pp. 19.


Decomposition of delay-insensitive circuits, p. 25.

On the delay-sensitivity of gate networks, p. 23.


A logic for one-pass, one-attributed grammars, p. 14.

Receptive Process Theory, p. 16.

Combining the functional and the relational model, p. 15.

A formal semantics for Z and the link between Z and the relational algebra, p. 30. (Revised version of CSNotes 89/17).

A proof system for process creation, p. 84.

A proof theory for a sequential version of POOL, p. 110.

Proving termination of Parallel Programs, p. 7.

A proof system for the language POOL, p. 70.

Compositionality in the temporal logic of concurrent systems, p. 17.

A fully abstract model for concurrent logic languages, p. p. 23.

On the asynchronous nature of communication in logic languages: a fully abstract model based on sequences, p. 29.
<table>
<thead>
<tr>
<th>Year</th>
<th>Authors</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>91/02</td>
<td>R.P. Nederpel, H.C.M. de Swart</td>
<td>Implication. A survey of the different logical analyses &quot;if...then...&quot;, p. 26.</td>
</tr>
<tr>
<td>91/03</td>
<td>J.P. Katoen, L.A.M. Schoenmakers</td>
<td>Parallel Programs for the Recognition of P-invariant Segments, p. 16.</td>
</tr>
<tr>
<td>91/05</td>
<td>D. de Reus</td>
<td>An Implementation Model for GOOD, p. 18.</td>
</tr>
<tr>
<td>91/06</td>
<td>K.M. van Hee</td>
<td>SPECIFICATIEMETHODEN, een overzicht, p. 20.</td>
</tr>
<tr>
<td>91/07</td>
<td>E. Poll</td>
<td>CPO-models for second order lambda calculus with recursive types and subtyping, p.</td>
</tr>
<tr>
<td>91/12</td>
<td>E. van der Sluis</td>
<td>A parallel local search algorithm for the travelling salesman problem, p. 12.</td>
</tr>
<tr>
<td>91/14</td>
<td>P. Lemmens</td>
<td>The PDB Hypermedia Package. Why and how it was built, p. 63.</td>
</tr>
<tr>
<td>Page</td>
<td>Authors</td>
<td>Title</td>
</tr>
<tr>
<td>------</td>
<td>---------------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>91/16</td>
<td>A.J.J.M. Marcelis</td>
<td>An example of proving attribute grammars correct: the representation of arithmetical expressions by DAGs, p. 25.</td>
</tr>
<tr>
<td>91/18</td>
<td>Rik van Geldrop</td>
<td>Transformational Query Solving, p. 35.</td>
</tr>
<tr>
<td>91/19</td>
<td>Erik Poll</td>
<td>Somewhere categorical properties for a model for second order lambda calculus with subtyping, p. 21.</td>
</tr>
<tr>
<td>91/23</td>
<td>K.M. van Hee, L.J. Somers, M. Voorhoeve</td>
<td>Z and high level Petri nets, p. 16.</td>
</tr>
<tr>
<td>91/25</td>
<td>P. Zhou, J. Hoorman, R. Kuiper</td>
<td>A compositional proof system for real-time systems based on explicit clock temporal logic: soundness and completeness, p. 52.</td>
</tr>
<tr>
<td>91/27</td>
<td>F. de Boer, C. Palamidessi</td>
<td>Embedding as a tool for language comparison: On the CSP hierarchy, p. 17.</td>
</tr>
<tr>
<td>91/28</td>
<td>F. de Boer</td>
<td>A compositional proof system for dynamic process creation, p. 24.</td>
</tr>
</tbody>
</table>