Towards Representing C++ Refactorings as Graph Transformations

László Vidács[1] [2], Martin Gogolla[3], Rudolf Ferenc[2]

[2]*Department of Software Engineering, University of Szeged, Hungary*
{lac,ferenc}@inf.u-szeged.hu

[3]*Department for Mathematics and Computer Science, University of Bremen, Germany*
gogolla@informatik.uni-bremen.de

Abstract

In this paper, we study a metamodel for the programming language C++. We work out refactorings on a C++ metamodel and present the essentials as graph transformations. The refactorings are demonstrated in terms of the C++ source code and the C++ target code as well. Graph transformations allow to capture refactoring details on a conceptual and easy to understand, but also very precise level.

Key words: Metamodel, C++, UML, Graph Transformation, OCL, Refactoring

1 Introduction

The programming language C++ is widely used in industry today. Many applications written in C++ exist which are constantly developed further, for example, to be adapted to modern service-oriented aspects. On the other hand, there is an important current trend in software engineering that focuses on development activities for using models instead of concentrating on code production only.

This contribution tries to narrow the bridge between industrial, code-centric development with C++ and model-centric development employing languages like the Unified Modeling Language (UML). We discuss a C++ metamodel and display first ideas how development and maintenance of C++ artifacts can be based on refactorings performed on instantiations of this C++

[1] László Vidács acknowledges the financial support provided through the European Community's Human Potential Programme under contract HPRN-CT-2002-00275, SegraVis.

metamodel. Our proposal is to express C++ refactorings and development steps as graph transformations. We think it is important to have clearly expressed transformation concepts for an involved domain like C++ software development. Graph transformations possess a sound theoretical basis and allow to express properties on a conceptual level, not only on an implementation level. Surprisingly, graph transformations have not yet been applied for C++ software development.

Graph transformations have been applied for the transformation of metamodels, see for example [8] (among many other works on graph transformation on metamodels). In industry, refactoring techniques [12] are regarded as promising means for software development. Refactoring on C++ code is supported by a variety of tools [13][11][16]. However, refactorings are usually considered from the implementation point of view only, not from a conceptual view. A conceptual view on C++ refactorings on the basis of metamodels and graph transformations allows to express properties like refactoring applicability more precise. A conceptual view also opens the possibility for viewing refactorings on the semantical level, for example, in order to describe semantics preserving refactorings or to test whether they are semantics preserving.

The structure of the rest of this contribution is as follows. Section 2 discusses our C++ language metamodel. Section 3 introduces C++ refactorings on this C++ language metamodel in terms of graph transformations. The contribution ends with a short conclusion.

2 C++ Metamodel

Metamodels, which are also called schemas in the re-engineering community, play a very important role in the process of fact extraction. They define the central repository for the whole process, from where the facts can be reached with the help of different transformations.

There are several researchers working on defining metamodels for C++ programs for reverse engineering purposes (also for program comprehension or to define an exchange format) [3],[1],[6].

The Columbus Schema for C++ [5] satisfies some important requirements of an exchange format. It reflects the low-level structure of the code, as well as higher level semantic information (e.g., semantics of types). Furthermore, the structure of the schema and the used standard notation (UML Class Diagrams) make its implementation straightforward, and what is even more important, an API (Application Programming Interface) is very simple to be realized as well.

Because of the high complexity of the C++ language, the schema is divided into six packages. To introduce all packages is beyond the scope of this paper. We created an excerpt mainly from the *struc* package of the schema, which can be seen in Figure 1. This part represents the scoping structure of a C++ program. Class *Member* is the parent of all kinds of elements which may

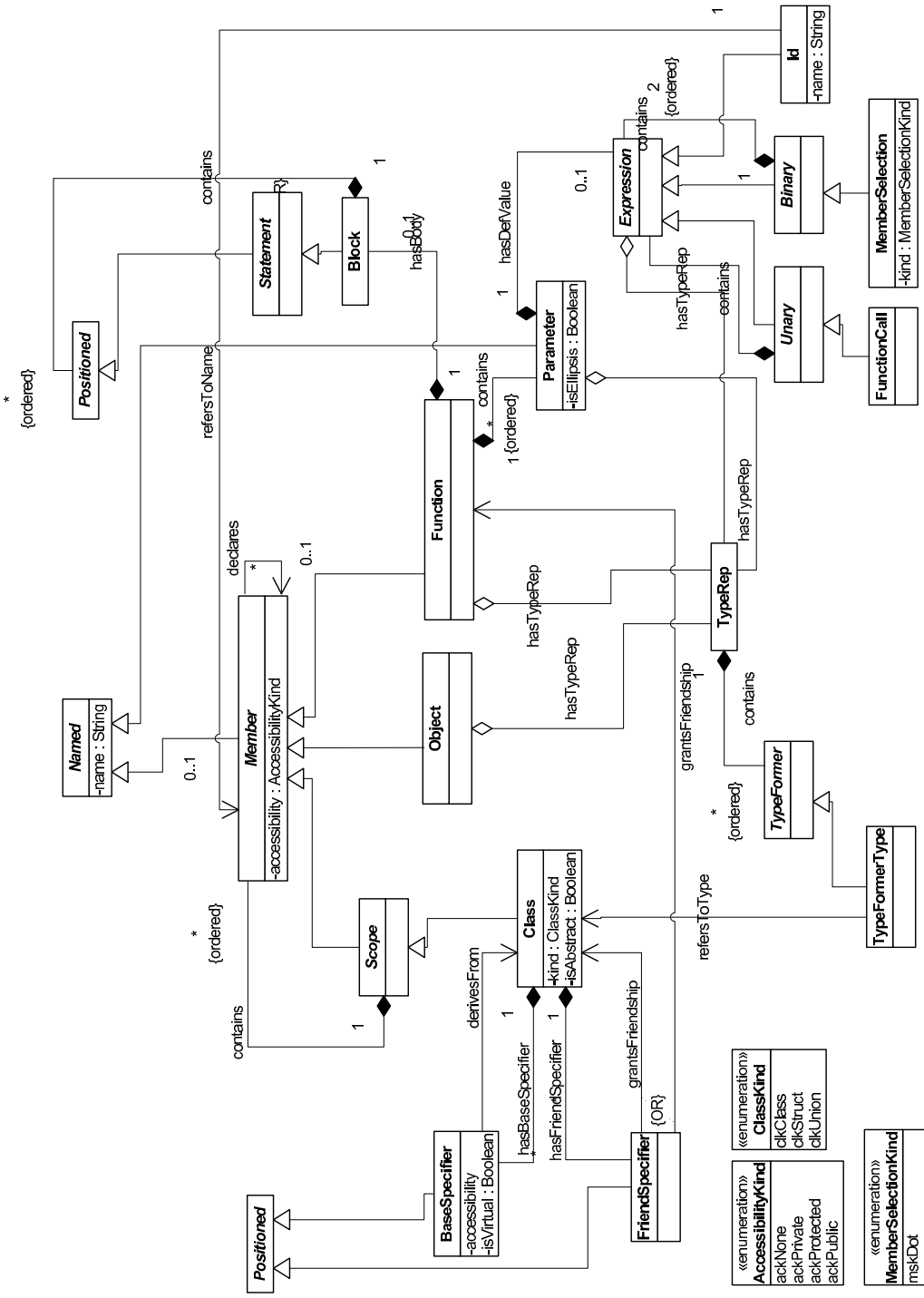

Fig. 1. Excerpt from the Columbus Schema for C++ - UML class diagram

appear in a scope (we use the term "member" in a more general way than usual). In the excerpt there are three important subclasses of *Member*. Class *Class* stands for C++ classes. It may contain further members; it may have base classes and friends. The class *Function* stands for C++ functions. It has body, type(s), parameter(s) and initializers. The class *Object* represents both variables and member fields (Note that it does not represent runtime class instances). These are the main classes used in our examples. For further details please see [7].

As usual in case of complex class diagrams, we cannot express everything using UML class diagram notations easily. For example a function body (block statement) contains ordered *Positioned* nodes. However a *Class* is also *Positioned* but it cannot be a part of a function body. OCL expressions as constraints of the class diagram solve many similar issues. We define the following condition (boolean expression, called invariant) that must be true for all *Block* objects:

```
context Block inv:
not self.contains.oclIsTypeOf(struc_Class)
```

3 Graph Transformation Rules on the C++ Metamodel

In this section we show how refactorings on the C++ metamodel can be described with graph transformation.

We provide examples, which are "classical" refactorings from Fowler's catalog [12]. The basic idea is simple in both cases, however many subtle details arise when realizing these refactorings. We also concentrate on C++-specific issues.

3.1 Graph transformation approach, notation

We use a single pushout approach for graph transformation rules possessing a left and a right hand side. Instead of too complicated NACs (Negative Application Condition) we provide preconditions as OCL expressions.

The definition of directed, attributed graphs are used as usual. Program graph is a directed, labelled, attributed graph where:

- nodes are labelled with class names shown in the UML class diagram
- nodes have attributes which are called as class attributes, the possible values of the attributes are from the corresponding UML types
- edges are labelled with relation names shown in the UML class diagram

Program graphs are introduced using an object diagram-like notation (see 2).

Not all graphs that corresponds the definition above represent C++ programs. For example an undeclared variable may occur according to the schema but the belonging code could not compile. The well formedness is not checked

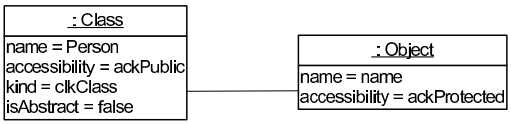

Fig. 2. Object diagram like notation of the graph

in this paper. In re(verse)-engineering context we assume that the starting graph is a well-formed graph and this property is preserved due to the conditions of transformations.

Note that C++ class attributes are modeled with class Object in the metamodel - so the nodes representing them have label *Object*. All *Object* and *Function* nodes have *TypeRep* nodes which show their C++ type. These nodes are omitted from the figures and presented only where it aids comprehension.

Furthermore the notation of multi-nodes is introduced to describe general subgraphs. A multi-node with value `k` represents k piece of nodes of the same type. Usage of multi-nodes is shown in Figure 3.

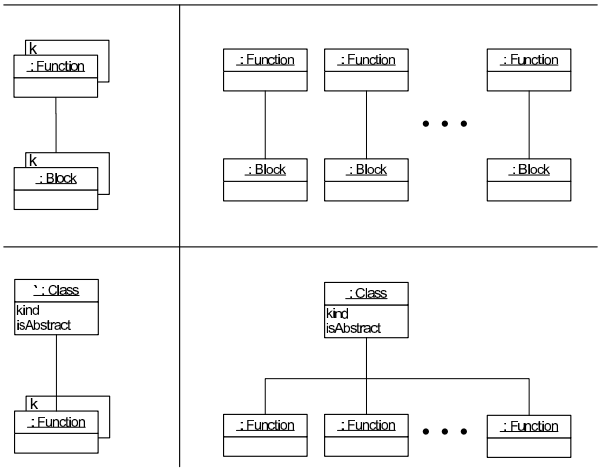

Fig. 3. Usage of multi-nodes

3.2 Extract class

Classes should serve a clear, well-defined aim. During development, classes are growing. In lot of cases, there are new responsibilities added to them. The aim of this refactoring is to extract a separate concept and corresponding data to a new class to improve the quality of the design. The idea is shown in Figure 4 which is taken from [12].

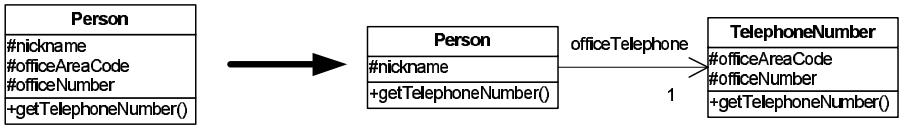

Fig. 4. Extract class refactoring example

There is a long way from this semi-informal description to an applicable

transformation. Based on this figure we make decisions about the context and purpose of the refactoring. After that we formalize it as graph transformation in two steps: the first part concentrates on the structure of the rule and the second part declares conditions using OCL expressions.

The main questions to be considered when realizing this refactoring as a graph transformation is: who can use the new (extracted) class and how can it be used. At first, the old class must somehow access it. If it is the only class that uses the new class then their relationship can be implemented either as the new class is a member in the old class or as a dynamic object creation in the constructor of the old class and as a deletion of the object in the destructor of the old class (in this case the other classes can access the new class through public interface functions of the old class). On the other hand, if the new class is free to be used by other classes then it is more complicated - for example a reference counter can be used. This is related to the question: who can instantiate the new class. Similarly, it has to be determined how the new class can be accessed by other classes: through the old class only or through public functions of the new class as well. Another obstacle is introduced by attributes which are used externally from the old class and are now moved to the new class. We assume that the visibility enables one to access those members through a known interface only (this is not a constraint, using the C++ metamodel, all usages of an attribute can be checked).

In this paper we choose to protect the newly created class. The new class can be instantiated only by the old class. Its properties can be modified by the old class, furthermore the old class provides the public interface to use the new class.

As refactorings can be realized in many different ways so we formalize the extract class first as a rule schema. A rule schema has parameters and multi nodes and can be instantiated in concrete cases. We call rule only these concrete cases when the rule schema has concrete arguments and can be applied directly on a program graph.

The graph transformation rule schema of ExtractClass(... : (Object—Function)) is shown in Figure 5. The attributes and operations to be moved must be determined by analyzing their usage. Therefore these are parameters of the transformation: any number of `Object` and `Function` nodes which are contained by the old class.

On the left hand side there is the old class. Its members are divided into 3 groups: attributes, public (interface) functions and protected (internal) functions. On the right hand side there are both the old class and the new (extracted) class. Attributes and protected (internal) functions are completely moved to the new class.

Public functions are copied from the old class to the new class. The existing implementation of these functions goes to the copy in the new class. The remaining functions in the old class have a new implementation: they only have to call the copied functions in the new class. The new class cannot be

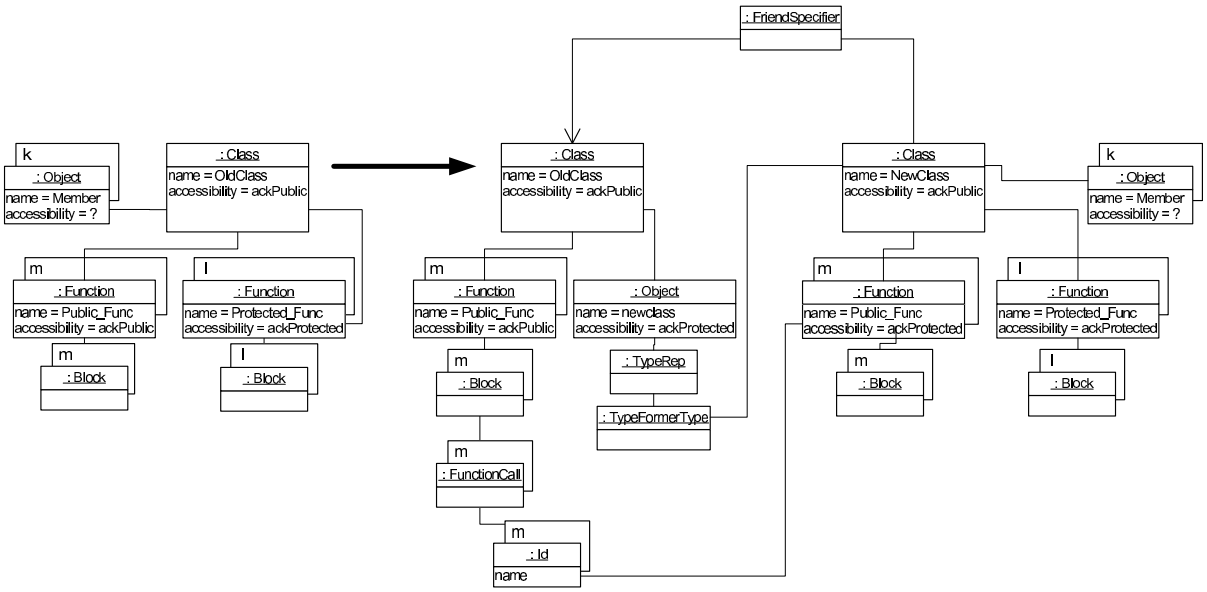

Fig. 5. Extract class refactoring as graph transformation

accessed from outside, so the copied public functions became protected.

To ensure the connection between the two classes we have a new member (*Object*) in the old class, its type is the new class. This is represented by new *TypeRep* and *TypeFormerType* nodes. (The two layers for types are required to represent complex types in C++ [5].) Now we have to let the old class access the new one, which has protected members/functions. This is done by giving friendship grant to the old class. This is represented by the *FriendSpecifier* node.

Concrete rule and OCL conditions

The left hand side and right hand side of a concrete rule is given in the figures below. The C++ code before the transformation with the left hand side is shown in Figure 6. The graph is in fact the object model of the code, so it contains a node (nickname) which does not take part in the transformation. The resulting source code and the right hand side is shown in Figure 7.

```
class Person {
protected:
  string nickname;
  string officeAreaCode;
  string officeNumber;
public:
  string getTelephoneNumber();
};
```

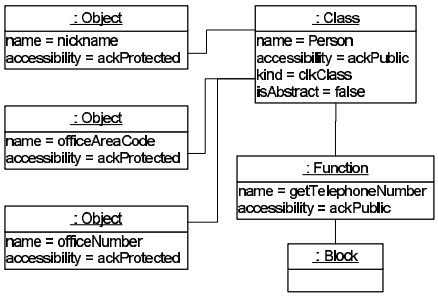

Fig. 6. Extract class - C++ code and model instance before the transformation

Moving function from one class to another requires careful examinations of the body. Class members or functions referenced from the body may become

```
class Person {
protected:
  string nickname;
  TelephoneNumber _TelephoneNumber;
public:
  string getTelephoneNumber();
};
class TelephoneNumber {
protected:
  TelephoneNumber();
  string officeAreaCode;
  string officeNumber;
  string getTelephoneNumber();
  friend class Person;
};
```

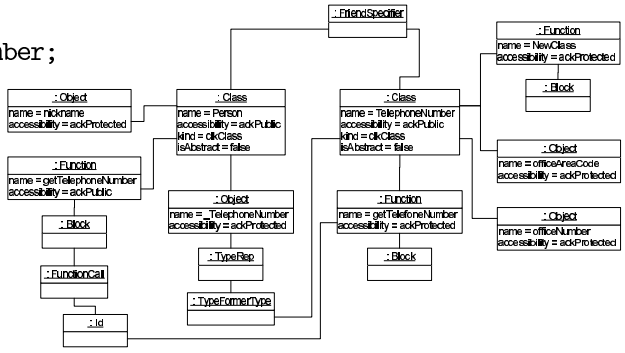

Fig. 7. Extract class - transformed C++ code and model instance

unaccessible because of changing the class (they are "foreign" in the new class). (Note that the friendship relation is not symmetric, only the old class can reach the new class.) To prevent accessing unreachable class members we have to check the subgraph of the function body. References to class members/functions are classified based on the relation of the old class and the referenced class as follows: inside the class, class hierarchy (base classes upwards) and outer classes. Bad references that prevent applying the rules are the follows:

- reference to protected/private member of the old class which is not among the parameters of the rule (if the referenced member is public it means that the extract class refactoring is not so reasonable here)
- reference to protected/private member of one of the base classes (in public case the above note applies)
- reference to protected/private member of an outer class which is a friend of the old class

According to the schema a reference is an Id node which has a refersToName relation to a Member, especially to a Function or Object. The referenced Member is contained by a class which is the class we are looking for. To distinguish the 3 different cases mentioned above it is not enough to search Ids in the subgraph in special container nodes like MemberSelection or FunctionCall. It may happen that a member selection contains this pointer and the function call without a memberselection may reference a function in a base class. So we have to scan all Ids and find the referenced class members/functions and their container classes. Note that in the case of MemberSelection expressions the container class can be found through the left hand side child Id as well.

The scan can be implemented as an OCL expression.

```
let Old : struc_Class = B1.Rl48.Rl18.oclAsType(struc_Class)
in
let Bases : Bag(struc_Class)  = Old.Rl65.Rl64
```

```
in
let M : Bag (struc_Member)  = B1.Rl177->select(i | i.oclIsTypeOf(expr_Id)).
        oclAsType(expr_Id).Rl149.select(oclIsTypeOf(struc_Object) or
        oclIsTypeOf(struc_Function))
in
M.iterate( m : struc_Member; res : Boolean = true |
let Cont : struc_Class = m.Rl18.oclAsType(struc_Class)
in
--condition A : referenced members are in the old class but they are not
--              among the parameters of the transformation rule
if ((m.accessibility='protected') or (m.accessibility='private') and
    (Cont=Old) and not (Bag{O2,O3,F1}.exists(i | i=m)) )
then
res and false
else
--condition B : referenced members are in the base classes
  if ((m.accessibility='protected') or (m.accessibility='private') and
      (Bases.exists(bc | bc=Old)))
  then
  res and false
  else
    --condition C : referenced members are outer friends
    let Friends : Bag(struc_Class) = Old.Rl67.Rl66
    in
    if ((m.accessibility='protected') or (m.accessibility='private') and
        (Friends.exists(fc | fc=Cont)))
    then
    res and false
    else
    res
    endif
  endif
endif
)
```

The expression checks for wrong references in a function body B1 (argument of the expression). In case of any occurrence of a wrong reference the expression returns false and prevents the rule to be applied. This example show the expressiveness of the OCL. OCL expressions may contain searches through collections which cannot be easily formulated with simple NACs. A general subgraph notation needed for instance to check whether a class is one of the base classes of the old class.

3.3 Other refactorings

There are several refactorings which can be implemented on our metamodel as graph transformation in similar way. In this paper we give only the (detailed) case study of extract class not only because of space. It contains many structural changes and also complex preconditions. In [15] there is a list of 15 formalized refactorings in 29 pages. Papers usually demonstrate two refactorings like in [9] pull up method and encapsulate variable, in [2] extract code

and move method refactorings are studied. We may say that none of them can say more then extract class regarding formalization of either the structural changes or the preconditions, which was the aim of the current contribution.

4 Implementation

The extract class refactoring graph transformation was implemented using the USE (UML-based Specification Environment) software [14] instead of an existing graph transformation engine. USE is a system for the specification of information systems. It is based on a subset of the Unified Modeling Language (UML). A USE specification contains a textual description of a model using features found in UML class diagrams (classes, associations, etc.). Expressions written in the Object Constraint Language (OCL) are used to specify additional integrity constraints on the model. A model can be animated to validate the specification against non-formal requirements.

The USE specification of C++ schema contains enumerations, classes and relations corresponding to the UML class diagrams of the schema. Every C++ program can be instantiated as an object diagram (a graph based on the specification). The USE can fairly handle and display our metamodel and model instances before and after the transformation. But the transformation itself can not be handled directly in the environment. We wrote a textual description of both sides of the rule instead. The description is processed by a script[2] which creates a sequence of basic graph operations from it (create/delete nodes, insert/delete edges). OCL expressions are used as postconditions to modify attributes of nodes. The script (based on the description) also generates a function that can list the possible nodes on which the rule can be applied and a function which applies the transformation (see Appendix A). There is a generated class (called RuleCollection) in the USE model which contains information and functions regarding to transformation rules. To apply the rule one have to pass the appropriate nodes as parameters.

The result of the extract class refactoring in USE can be seen in Figure 8

The execution of the script was quick because the parameters determined the place of the transformation so the modifications were made locally (below 1 sec - without measuring). Future work is to try this implementation on real life software systems. Running time of the rule-creator script depends on the size of the rule. In general the most time consuming part is identification of the places where the transformation is applicable. In case of refactorings in most cases the programmer has to consider and choose a place to apply the refactoring. The decision is made based on criteria which are not easy to formalize. For instance extract class refactoring may be applicable on almost every classes (which have members or functions) but only in few cases it is useful to apply. Thus in a real life software the identification of the big

[2] Thanks to Fabian Büttner

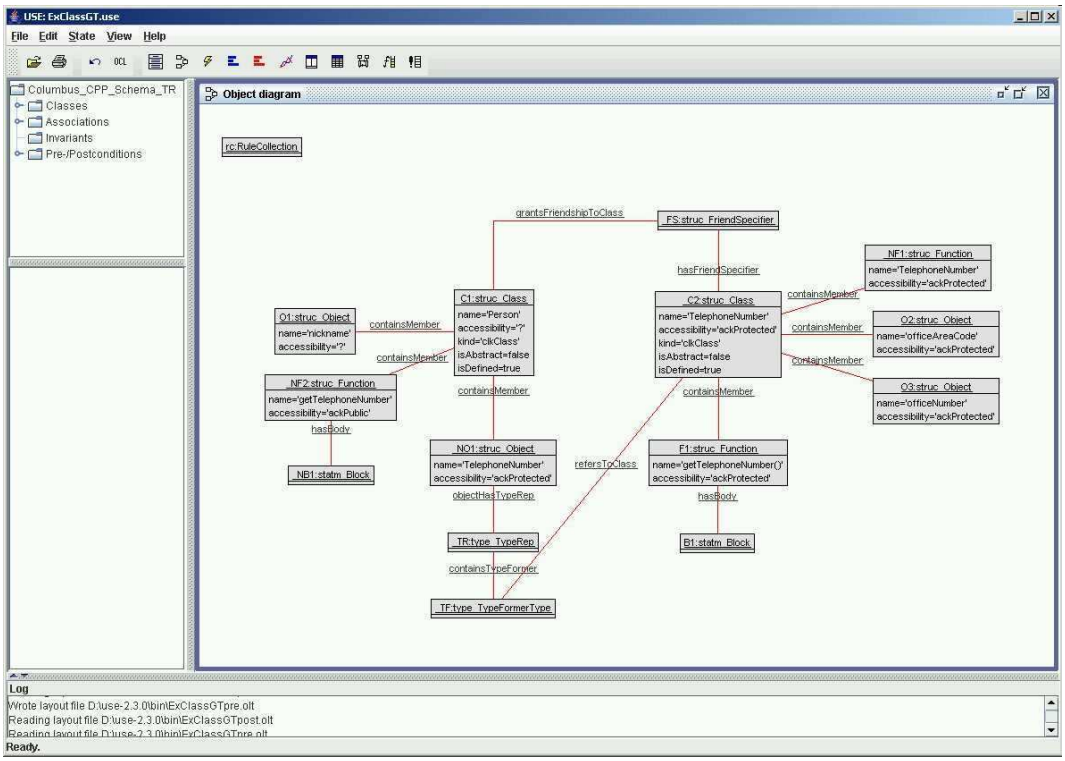

Fig. 8. Object diagram in USE after the refactoring

class can be done using other visualization or analyzer tools. After we have identified the place (parameters of the rule), the actual refactoring can be applied quickly - like in our example.

5 Related work

Bottoni [2], Mens [9], Fanta [4], Eetvelde [15], Opdyke [10]

6 Conclusion

This contribution discusses metamodels for the C++ language and the C++ preprocessor. It uses graph transformations to express C++ refactorings in a clear way and on a conceptual level. Although graph transformations have been used for metamodel transformation of various languages, they have not been extensively for C++. For a successful use of graph transformations, it is necessary to work on cumbersome subjects like C++ refactorings and the accompanying nasty details. We are aware of the fact that many of our concepts are already known from other successful application areas of graph transformation.

Future work will elaborate further C++ language and C++ preprocessor refactorings. We will also implement these refactorings in a graph rewriting machine resp. modeling tool. We think that such an implementation will give

insight into the application conditions and properties of C++ refactorings on a conceptual level. Such an implementation will thus enable a more concise understanding of C++ refactorings.

References

[1] Bell Canada Inc., Montréal, Canada, "DATRIX – Abstract semantic graph reference manual," Version 1.2 edition (2000).

[2] Bottoni, P., F. Parisi-Presicce and G. Taentzer, *Specifying integrated refactoring with distributed graph transformations.*, in: *AGTIVE*, 2003, pp. 220–235.

[3] Ebert, J., B. Kullbach, V. Riediger and A. Winter, *GUPRO - Generic Understanding of Programs*, Electronic Notes in Theoretical Computer Science **72** (2002).

[4] Fanta, R. and V. Rajlich, *Reengineering object-oriented code*, in: *ICSM '98: Proceedings of the International Conference on Software Maintenance* (1998), p. 238.

[5] Ferenc, R., Á. Beszédes, M. Tarkiainen and T. Gyimóthy, *Columbus - reverse engineering tool and schema for C++*, in: *ICSM 2002* (2002), pp. 172–181.

[6] Ferenc, R., S. E. Sim, R. C. Holt, R. Koschke and T. Gyimóthy, *Towards a Standard Schema for C/C++*, in: *WCRE 2001* (2001), pp. 49–58.

[7] *Homepage of FrontEndART Ltd*, `http://www.frontendart.com` (2005).

[8] Gogolla, M., *Graph Transformations on the UML Metamodel*, in: *GVMT'2000* (2000), pp. 359–371.

[9] Mens, T., N. Van Eetvelde, S. Demeyer and D. Janssens, *Formalizing refactorings with graph transformations*, Journal on Software Maintenance and Evolution: Research and Practice (2005).

[10] Opdyke, W. F., "Refactoring Object-Oriented Frameworks," Ph.D. thesis, Urbana-Champaign, IL, USA (1992).

[11] *Homepage of Ref++*, `http://www.refpp.com` (2005).

[12] *Refactoring catalog*, `http://www.refactoring.com/catalog/` (2005).

[13] *Homepage of Slickedit*, `http://www.slickedit.com/` (2005).

[14] *Homepage of USE*, `http://www.db.informatik.uni-bremen.de/projects/USE/` (2005).

[15] Van Eetvelde, N. and D. Janssens, *Refactorings as graph transformations*, Technical report, University of Antwerp (2005), uA WIS/INF 2005/04.

[16] *Homepage of Xrefactory*, `http://xref-tech.com` (2005).

Appendix

A USE description

```
------------------------------------------------------------------------
-- Extract Class refactoring
------------------------------------------------------------------------
rule ExClass
left
  C1:struc_Class
  O1:struc_Object
  O2:struc_Object
  F1:struc_Function
  (C1,O1): containsMember
  (C1,O2): containsMember
  (C1,F1): containsMember
right
  C1:struc_Class
  C2:struc_Class
  O1:struc_Object
  O2:struc_Object
  NO1:struc_Object
  F1:struc_Function
  NF1:struc_Function
  NF2:struc_Function
  NB1:statm_Block
  FS:struc_FriendSpecifier
  TR:type_TypeRep
  TF:type_TypeFormerType
  (C2,O1): containsMember
  (C2,O2): containsMember
  (C2,F1): containsMember
  (C2,NF1): containsMember
  (C1,NF2): containsMember
  (NF2,NB1): hasBody
  (C1,NO1): containsMember
  (NO1,TR): objectHasTypeRep
  (TR,TF): containsTypeFormer
  (TF,C2): refersToClass
  (C2,FS): hasFriendSpecifier
  (FS,C1): grantsFriendshipToClass
  -- postconditions
  [C2.name = 'TelephoneNumber']
  [C2.accessibility = 'ackProtected']
  [C2.kind = 'clkClass']
  [C2.isAbstract = false]
  [C2.isDefined = true]
  [NF1.name = 'TelephoneNumber']
  [NF1.accessibility = 'ackProtected']
  [F1.accessibility = 'ackProtected']
  [NO1.name = 'TelephoneNumber']
  [NO1.accessibility = 'ackProtected']
  [NF2.name = 'getTelephoneNumber']
```

```
  [NF2.accessibility = 'ackPublic']
end

-----------------------------------------------------------------------
-- Extract Class refactoring
-- ExClassGT_ExClass.cmd
-- ExClass(_C1,_O1,_O2,_F1)
-- the 'match' parameter can be bound with '!let match = ...'
-----------------------------------------------------------------------
!let _C1 = match->at(1)
!let _O1 = match->at(2)
!let _O2 = match->at(3)
!let _F1 = match->at(4)
!openter rc ExClass(_C1,_O1,_O2,_F1)

!create _C2 : struc_Class
!create _NO1 : struc_Object
!create _NF1 : struc_Function
!create _NF2 : struc_Function
!create _NB1 : statm_Block
!create _FS : struc_FriendSpecifier
!create _TR : type_TypeRep
!create _TF : type_TypeFormerType
!insert(_C2,_O1) into containsMember
!insert(_C2,_O2) into containsMember
!insert(_C2,_F1) into containsMember
!insert(_C2,_NF1) into containsMember
!insert(_C1,_NF2) into containsMember
!insert(_NF2,_NB1) into hasBody
!insert(_C1,_NO1) into containsMember
!insert(_NO1,_TR) into objectHasTypeRep
!insert(_TR,_TF) into containsTypeFormer
!insert(_TF,_C2) into refersToClass
!insert(_C2,_FS) into hasFriendSpecifier
!insert(_FS,_C1) into grantsFriendshipToClass
!set _C2.name := 'TelephoneNumber'
!set _C2.accessibility := 'ackProtected'
!set _C2.kind := 'clkClass'
!set _C2.isAbstract := false
!set _C2.isDefined := true
!set _NF1.name := 'TelephoneNumber'
!set _NF1.accessibility := 'ackProtected'
!set _F1.accessibility := 'ackProtected'
!set _NO1.name := 'TelephoneNumber'
!set _NO1.accessibility := 'ackProtected'
!set _NF2.name := 'getTelephoneNumber'
!set _NF2.accessibility := 'ackPublic'
!delete(_C1,_O1) from containsMember
!delete(_C1,_O2) from containsMember
!delete(_C1,_F1) from containsMember

!opexit
```

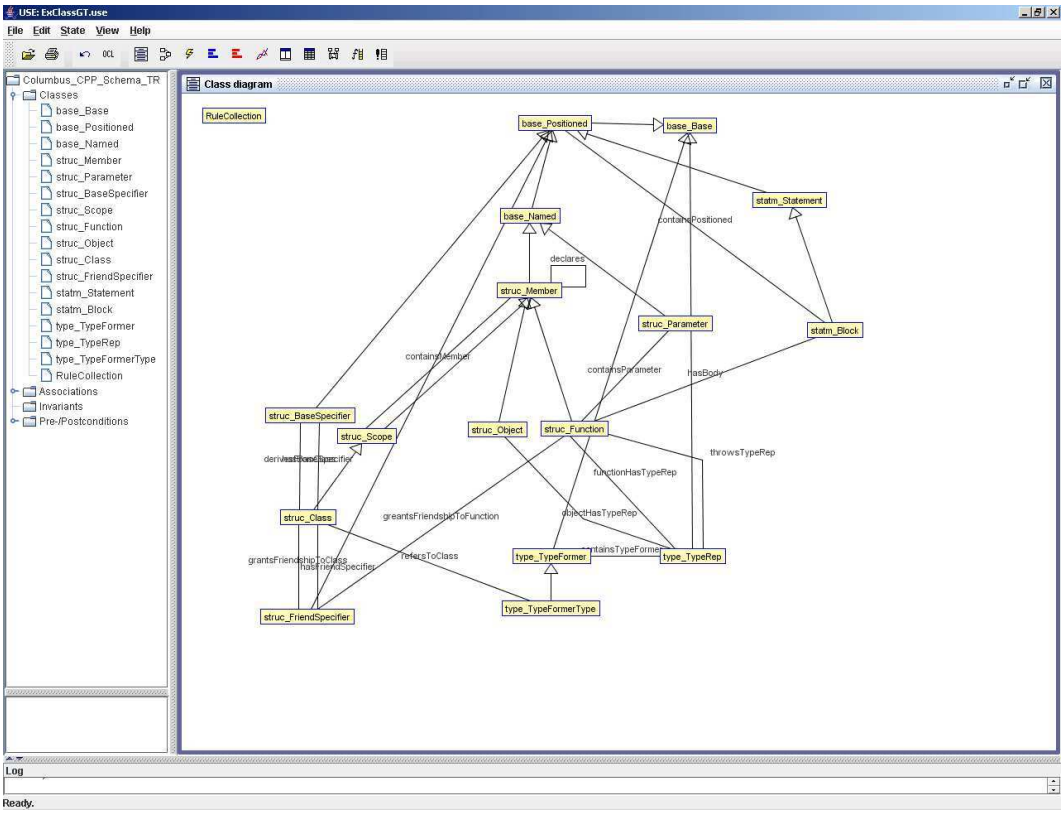

Fig. A.1. Class diagram of the C++ schema excerpt in USE

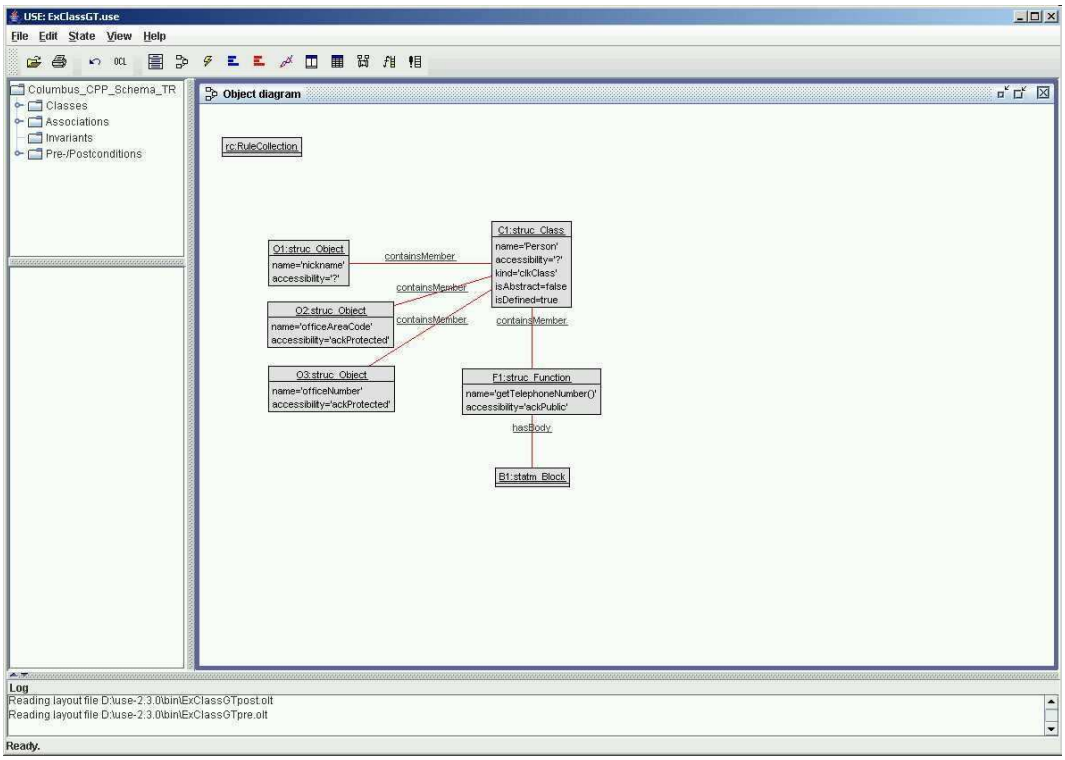

Fig. A.2. Object diagram in USE before the refactoring