Integrating the ER Approach in an OO Environment*

M. Gogolla, R. Herzig, S. Conrad, G. Denker, and N. Vlachantonis
Technische Universität Braunschweig, Informatik, Abt. Datenbanken
Postfach 3329, D-38023 Braunschweig, Germany
e-mail: gogolla@idb.cs.tu-bs.de

Abstract

We translate Entity-Relationship (ER) schemas into the object-oriented specification language TROLL light. This language describes the Universe of Discourse (UoD) as a system of concurrently existing and interacting objects, i.e., an object community. Thereby two essential aspects, structure and behavior, are integrated in one formalism. By doing the translation from ER to TROLL light we preserve the visual advantages of the former and receive a formalism through the latter which can be mapped to an adequate object-oriented database system. Proceeding this way we hope our proposal for transforming ER schemas into TROLL light specifications provides a valuable link between structural and dynamic modeling.

1 Introduction

Nowadays the Entity-Relationship approach [Che76] has been accepted as a quasi standard [Teo90, BCN92] for the conceptual design of databases and information systems. This approach is undoubtedly a very popular tool for bridging the gap between database designers and users of database systems. One issue in favor of Entity-Relationship design is based on the fact that the graphical representation serves as an intuitively understandable platform for both database designers and users of database systems so that communication between them becomes easier. Another reason for the wide acceptance and application of this approach is due to the fact that an Entity-Relationship schema can be mapped in a systematic way to a relational database schema (see [EWH85, MMR86, TYF86, Hoh89], among others). But most approaches to Entity-Relationship modeling do not deal deeply with dynamic aspects, because the Entity-Relationship approach is used for modeling static structure though it ought not to be separated from the behavioral one. There are temporal Entity-Relationship extensions [KS91b, KS91a, HH91, RS91, Tau91] which add dynamic

*Work reported here has been partially supported by the CEC under Grant No. 6112 (COMPASS) and BMFT under Grant No. 01 IS 203 D (KoSto).
aspects to the Entity-Relationship approach, but most of them are not directed to object-oriented databases.

Recently the advent of object-oriented based technology calls for and demands database design approaches and tools resulting in object-oriented database schemas. The main difficulties in tackling this problem arise from the lack of a common model and widely accepted principles. However, one advantage of object-oriented databases is that they cover both aspects, structure and behavior, and propose in this way an adequate formalism to describe the so-called Universe of Discourse (UoD), i.e., the part of the real world to be modelled. In this framework TROLL light [CGH92, GCH93] was developed as an object-oriented specification language describing the UoD as a system of concurrently existing and interacting objects, i.e., an object community. TROLL light is a dialect of TROLL [JSHS91] and OLOG [SSE87]. Our language follows the decision that it is better to provide a modest number of basic constructs which are clear and easily applicable than to offer a lot of redundant language features.

But because of the textual representation of TROLL light, the language is in some respect less suitable to support the needs of a variety of users, like database designers and system users, in view of intuitive comprehensibility. Nevertheless TROLL light is closer to object-oriented database systems since it provides a lot of language features capable of expressing behavioral aspects of systems. Our aim is to break the traditional way of translating Entity-Relationship schemas to relational databases. We translate Entity-Relationship schemas to TROLL light, preserving the visual advantages of the former and receiving a formalism through the latter which can be mapped to an adequate object-oriented database system like this has been done in [NCB91, Tar92]. After the structural design with the Entity-Relationship approach we suggest an object-oriented mapping to get on the way to an object-oriented schema.

Let us explain the TROLL light approach in more detail. Our specification language comes along with an integrated, open development environment [VHG+93]. The task of this environment is to give support for the creation of correct information systems on the basis of formal specifications. Some important ingredients of the environment are pictured in Figure 1. Rectangles (with upper left corner cut off) represent documents which are generated during the development process; ovals stand for tools which are invoked by the user, and arrows denote the main flow of information. We now look at the single tools item by item.

The data type interpreter supports the development of data type specifications. It allows to test the data type specification at hand and to validate that the specification meets the application requirements. The TROLL light editor is understood to be a syntax-directed tool which also already checks context-sensitive conditions a TROLL light specification has to fulfill. The TROLL light animator is designed to simulate the specified object community. It provides object windows which enables the user for instance to observe attributes, to initiate state transitions by clicking events or to formulate queries against object states. By this the informal view of the real-world fragment to be modeled is validated against the current specification. The TROLL light prover is given verification tasks by the user. It checks whether the desired properties are fulfilled by the specified TROLL light templates. The TROLL light transformer takes the specification of an object community and interactively generates executable code in an object-oriented programming language
supported by an object-oriented database system. The ER schema editor allows to design database schemas interactively. It fully supports all graphical features of the ER model and thus utilizes all advantages of the ER approach. Schemas obtained in this way can automatically be translated into TROLL light specifications by the ER schema translator. The advantage of this intermediate step lies in the fact that the obtained TROLL light descriptions can be processed with the other tools in the development environment like the TROLL light animator or the TROLL light prover.

Currently we are mainly working on the animator and the prover. But our environment is designed as an open environment. This means that new tools can be added. Nevertheless our environment is integrated in the sense that all tools have a common user interface and that they communicate with each other.

The main purpose of the present paper is to work out the basic ideas behind the ER schema translator. The remainder of the paper is organized as follows. In Section 2 we will give a short introduction to the main concepts of TROLL light. Afterwards we will reflect on various constructs of the Entity-Relationship approach and its extensions and show how they can be expressed in TROLL light. Our translation reveals that ER modelling and object-orientation can go hand in hand and support each other. Finally we give our estimation of the applicability of this proposal and its result on methodical decisions.
2 Concepts of TROLL light

2.1 Overview

TROLL light is a language for describing structural and dynamic properties of objects. With respect to the description of structural properties TROLL light offers concepts of semantic data models. Attributes describe observable properties of objects. Simple attributes are data-valued or object-valued. Over and above that complex attributes are obtainable by the application of predefined sort constructors like SET or TUPLE. TROLL light objects are organized into object hierarchies induced by sub-object relationships. Thereby a sub-object relationship must be understood as an exclusive “part-of” relationship. Normally attribute values are directly affected by event occurrences. Besides that derived attributes can be specified the state of which depends on other stored or derived information. In order to state derivation rules TROLL light incorporates an SQL-like query calculus. The query calculus of TROLL light also allows to specify static integrity constraints.

The description of dynamic properties is based on the specification of events. Events are abstractions of state-modifying operations on objects. Events on objects are described by a finite set of event generators. Each event generator may be accompanied by a list of parameter sorts. An event generator together with actual parameter values denotes an event. The effect of events on attributes is described by valuation rules. Events in different objects can be synchronized by interaction rules. Possible event sequences can be restricted to admissible ones by the means of CSP-like process descriptions.

All in all a template has the following structure:

```
TEMPLATE
  DATA TYPES          declaration of data types used in current template
  TEMPLATES           declaration of templates used in current template
  SUBOBJECTS          sub-object relationships
  ATTRIBUTES          observable properties
  EVENTS              event generators
  CONSTRAINTS         static integrity conditions
  VALUATION           effect of event occurrences on attributes
  DERIVATION          derivation rules for derived attributes
  INTERACTION         synchronization of events in different objects
  BEHAVIOR            description of admissible life cycles
END TEMPLATE
```

In the next section we show by a small example how objects in an information system may be specified by writing templates.

2.2 Examples

Let us assume that an information system is to be described which shall contain data about authors. For every author the name, the date of birth, and the number of books sold by year have to be stored. An author may change the name only once in the life.
An appropriate TROLL light specification would be:

```
TEMPLATE Author
  DATA TYPES String, Date, Nat;
  ATTRIBUTES Name:string; DateOfBirth:date;
               SoldBooks(Year:nat):nat;
  EVENTS   BIRTH create(Name:string, DateOfBirth:date);
           changeName(NewName:string);
           storeSoldBooks(Year:nat, Number:nat);
  DEATH    destroy;
  VALUATION [create(N,D)] Name=N, DateOfBirth=D;
             [changeName(N)] Name=N;
             [storeSoldBooks(Y, NR)] SoldBooks(Y)=NR;
  BEHAVIOR PROCESS authorlife1 =
           ( storeSoldBooks -> authorlife1 |
             changeName -> authorlife2 |
             destroy -> POSTMORTEM );
  PROCESS authorlife2 =
           ( storeSoldBooks -> authorlife2 |
             destroy -> POSTMORTEM );
           ( create -> authorlife1 );
END TEMPLATE;
```

Process descriptions as used in the BEHAVIOR section can be transformed to event-driven sequential machines (see Figure 2). These machines will be called o-machines (o for object) in the sequel. The states of the o-machine denote behavior states of objects.

We may assume that the information system will be populated by lots of authors so that these authors must be managed by a higher-level object. To this end usually classes are introduced in many object-oriented approaches. There is some confusion concerning the notion of class in object-oriented programming and object-oriented databases. A class is often used to cover two aspects, namely an object factory and an object container [Ban88]. In TROLL light object factories appear as templates while object containers are merely understood as composite objects, and composite objects have to be described by templates again. The following template Authorclass for such composite objects provides sub-object names for local identification of author objects.

```
TEMPLATE Authorclass
```
DATA TYPES    String, Date, Nat;
TEMPLATES    Author;
SUBOBJECTS     Authors(No:nat):author;
ATTRIBUTES-derived NumerofAuthors:nat;
EVENTS     BIRTH create;
            addAuthor(No:nat, Name:string,
                    DateOfBirth:date);
            removeAuthor(No:nat);
DERIVATION NumerofAuthors= CNT(Authors);
INTERACTION addAuthor(N,S,D) >> Authors(N).create(S,D);
                removeAuthor(N) >> Authors(N).destroy;

END TEMPLATE;

Every template implies a corresponding object sort. We use names starting with an upper case letter to denote templates whereas the same name starting with the corresponding lower case letter is used to denote the corresponding object sort.

An object of sort authorclass will hold many author objects as private sub-objects (or components). In this example a natural number serves as a key for the identification of sub-objects. In contrast to many other data models, in TROLL light such keys need not be related to attributes of objects they identify. Adding an object to an author class means to create a new sub-object, and removing an object from an author class means to destroy the sub-object.

An object of sort authorclass may be sub-object of a higher object again, and so on. For example, a template Libraryworld may be specified having a sub-object of sort authorclass as sub-object beside other sub-objects like a book index and several libraries. Each of these libraries may have a collection of documents as sub-objects which refer to the books in the common book index (for more details see [CGH92, GCH93]).

As already mentioned it is possible to specify explicit integrity constraints within TROLL light templates. For this TROLL light contains a query calculus which is an offspring of a calculus for extended Entity-relationship schemas [HG88, GH91]. Besides static integrity constraints preconditions for valuations and for event occurrences can also be given by means of this calculus.

Let us consider an example for an integrity constraint. If we want to restrict the authors belonging to the same authorclass object to have together not more than 50,000 books sold in 1993, this can be expressed as follows:

\[
50,000 \leq \sum \left( \text{SELECT SoldBooks(PRJ2(A),1993)} \right) \\
\text{FROM} \quad \text{A IN Authors} \\
\text{WHERE} \quad \text{DEF(SoldBooks(PRJ2(A),1993))} \\
\]

This constraint must be inserted into the CONSTRAINTS section of the Authorclass template. For understanding this formula some remarks are necessary. For parameterized sub-object and attribute names we assume the sub-object resp. attribute symbol without any parameters to stand for a set of tuples consisting of parameter values and a sub-object resp. an attribute value. Only those tuples are members in
that set for which the sub-object exists resp. for which the attribute value is currently defined. For instance, Authors stands for a term of sort \( \text{SET}(\text{TUPLE(nat, author}) \). In consequence, the variable \( A \) which becomes bound to Authors in the FROM clause is of sort \( \text{TUPLE(nat, author}) \). Therefore, we use \( \text{PRJ2}(A) \) to project onto the second component, i.e., the author component, out of such a tuple. If we want to refer to attributes of other objects we have to add a parameter denoting the object the attribute of which we are interested in. In the given constraint we want to know the value of the attribute SoldBook(1993) of some sub-object. Because \( \text{PRJ2}(A) \) is evaluated to some author object we may use this term as additional parameter in SoldBooks(P RJ2(A), 1993).

Finally we note that our query calculus is safe because we require variables to be bound to finite domains, i.e., to stored information. Thereby, universal and existential quantifiers may be used as well. In [CGH92, GCH93] TROLL light is presented in more detail.

3 Translating ER Schemas into TROLL light Templates

3.1 Translation of Basic ER schemas

Basic ER schemas according to Chen [Che76] are given by a set of entity types \( E = \{E_1, \ldots, E_p\} \) and a set of relationship types \( R = \{R_1(E_{1_1}, \ldots, E_{1_{n_1}}), \ldots, R_q(E_{q_1}, \ldots, E_{q_{n_q}})\} \) such that \( E_{i,j} \in E \). Both entity types and relationship types can be associated with data-valued attributes.

We propose the following general transformation scheme.

1. An entity type \( E \) (or a relationship type \( R \)) with attributes \( A_1:d_1, \ldots, A_n:d_n \) is mapped to the following template.

   \[
   \text{TEMPLATE } E/R \\
   \text{DATA TYPES } \ldots; \\
   \text{ATTRIBUTES } A_1:d_1, \ldots, A_n:d_n; \\
   \text{EVENTS } \text{BIRTH} \text{ create;} \\
   \hspace{1cm} \text{update}A_i(d_i); \\
   \text{DEATH} \text{ destroy;} \\
   \text{VALUATION} \text{ [update}A_i(D)] A_i=D; \\
   \text{END TEMPLATE;}
   \]

   Here we assume that immediately after the creation of an object all attributes take undefined values. Of course this can be avoided by providing the birth event \text{create} with appropriate parameters.

2. ER schemas are viewed as objects in TROLL light for which the following object description can be given.
TEMPLATE ERSchema

DATA TYPES   Nat, ...

TEMPLATES $E_1, ..., E_p, R_1, ..., R_q$

SUBOBJECTS $E_1.s(nat) : e_1, ..., E_p.s(nat) : e_p$

$R_1.s(e_{1, i_1}, ..., e_{1, m_1}) : r_1, ..., R_q.s(e_{q, i_1}, ..., e_{q, m_q}) : r_q$

EVENTS

BIRTH create;

add$E_j(nat)$;   add$R_k(e_{k, 1}, ..., e_{k, n_k})$;
update$E_j(A_{j, i}(nat, d_{j, i})$; update$R_k(A_{k, i}(e_{k, 1}, ..., e_{k, n_k}, d_{k, i})$;
remove$E_j(nat)$;  remove$R_k(e_{k, 1}, ..., e_{k, n_k})$;

DEATH destroy;

INTERACTION add$E_j(N)$   $>> E_j.s(N).create$;
update$E_j(A_{j, i}(N, D)$   $>> E_j.s(N).updateA_{j, i}(D)$;
remove$E_j(N)$   $>> E_j.s(N).destroy$;
add$R_k(E_1, ..., E_{n_k})$   $>> R_k.s(E_1, ..., E_{n_k}).create$;
update$R_k(A_{k, i}(E_1, ..., E_{n_k}, D)$   $>> R_k.s(E_1, ..., E_{n_k}).updateA_{k, i}(D)$;
remove$R_k(E_1, ..., E_{n_k})$   $>> R_k.s(E_1, ..., E_{n_k}).destroy$;

END TEMPLATE;

Note that there is a special birth event for an ER schema which can be interpreted as the installation of an ER object base, and a death event denoting the shutdown of this base.

The transformation scheme has been given under the following premises. (1) On schema level entities are identified by a natural number whereas relationships are identified by the participating entities. (2) Since the behavior of objects is not subject to ER modeling in the narrow sense, object dynamics can only be reflected by canonical updates concerning the insertion and deletion of entities or relationships and attributes updates. (3) Since we assume all users to be confronted with the schema object all operations concerning the insertion and deletion of entities or relationships and attribute updates can be triggered from the schema level.

Of course, other choices could be taken as well. (1) Other keys are possible for the identification of entities, for example by referring to attributes of the objects in mind (see key concept below). (2) Proposals for the modeling of object behavior can be reflected by non-generic events as shown in Section 2. (3) Event triggering can be restricted to object level.

Another point of choice concerns the particular representation of relationships. Instead of representing relationships by objects, sometimes it seems to be more convenient to map relationship types to attributes. This is discussed in the next subsection.

3.2 Alternative Representations of Relationship Types

The representation of a relationship as an object is especially suited when the relationship shows distinct properties of a higher-level object, for example by having attributes which can be updated. Especially in cases where a relationship type has no attributes, the representation as an object may produce unnecessary overhead. In this case the representation of relationships as values may be indicated. Here value representation means that object relationships are expressed by object-valued attributes.
These attributes can either be stated at schema level or entity level. First we give an example for modeling relationships at schema level.

Assume that the attribute Salary of CE is dropped. Then the following TROLL <i>light</i> specification of the example ER schema has the same ”information capacity” as the TROLL <i>light</i> specification given before.

**TEMPLATE** ERSchema

```
...  
SUBOBJECTS  Companies(nat):company;
            Employees(nat):employee;
ATTRIBUTES  CEs:SET(TUPLE(company, employee));
...  
VALUATION  [create]  CEs=EMPTY;
            [addCE(C,E)]  CEs=ADD(CEs, (C,E));
            [removeCE(C,E)]  CEs=REMOVE(CEs, (C,E));
...  
END TEMPLATE;
```

Since CE is now modeled as an attribute the corresponding interaction rules have been replaced by valuation rules. Note that the sort expression SET(TUPLE(company, employee)) corresponds to a relation in the relational data model. Indeed, the representation of relationships given by this example resembles the way relationships are represented in a relational database.

Instead of interrelating objects at schema level object-valued attributes can only be used at entity level. This representation is particularly recommended for functional relationships.

Assume that an employee may only work for one company at a time. Then this relationship can adequately be described by an object-valued attribute of Employee.

**TEMPLATE** Employee

```
...  
TEMPLATES  Company;
ATTRIBUTES  ...; worksFor:company;
EVENTS     ...; updateCompany(company);
VALUATION  ...; [updateCompany(D)] worksFor=D;
END TEMPLATE;
```

Object-valued attributes at entity level may also be used for representing N:M relationships. Then the single-valued attribute worksFor has to be replaced by a multi-valued attribute. The inverse relationship employs : Company → set(Employees) can be expressed by an inverse attribute. Interestingly object-oriented database management systems like ObjectStore [LLOW91] offer support for attributes and their inverses to implement N:M relationships.

### 3.3 Extended ER Schemas

There have been many proposals in extending ER schemas into the direction of general semantic data models [dSNF80, ABLV83, EWH85, EKTW86, Tha90, WT91,
EGH+92, PS92]. In contrast to the original ER approach these models mostly support object-valued attributes, complex attributes, derived attributes, constraints, keys, component relations, and ISA relations.

Object-valued attributes: In the original ER model attributes are restricted to data domains. It has been argued that this restriction is important from an ER design methodology point of view since it justifies the conceptual difference between entity and relationship types. On the other hand from a conceptual modeling point of view functional relationships between objects can be modeled more naturally by object-valued attributes. Hence many semantic data models do already support the modeling of object-valued attributes. Object-valued attributes are supported in TROLL light by referring to object sorts in the specification of attributes.

Complex attributes: Many non-standard applications, for example the modeling of multi-media objects, require non-standard data domains. In principle, there are two possible ways for the integration of such domains into an object model: first by referring to arbitrary user-defined abstract data types, second by providing sort constructors which directly allow the specification of multi-valued or composite attributes. Both ways are supported by TROLL light.

Derived attributes: Derived attributes are attributes the values of which depend on other stored or derived information. The specification of derived attributes is directly supported in TROLL light by the DERIVATION section of a template.

Constraints: The query calculus of TROLL light can also be used to describe a wide range of static integrity constraints on object states (e.g. cardinality constraints) in the CONSTRAINTS section of a template.

Key concept: In principle, the particular identity of an entity is abstract, i.e., it should be invisible to the user [KC86]. But suppose that an attribute of a certain object is to be updated. Then without a specific notion of object identity at hand a user can only describe the object in mind associatively through some related objects or values. However there will never be a guarantee that this description can be unique.

In order to facilitate object access, many models offer a key concept providing abstract objects with an additional logical identity. In its simplest form some attributes of an object are marked as keys. Here it is important to note that in the framework of TROLL light the specification of keys does not belong to object descriptions in the narrow sense since keys are irrelevant from the viewpoint of a single object. Keys get only important when many objects of a given sort are to be distinguished, for example by a higher-level object such as a schema. So in many cases needs for the simple identification of similar objects led to new attributes. Take for example a person’s name which turns out to be more artificial than a person’s color of eyes.

In the generic TROLL light description of an arbitrary ER schema we simply used numbers for the identification of entities, but this is not obligatory. Specific attributes of the objects which are to be identified can be used as keys as well.

Component relations: It has been recognized that component (or part) relations constitute an extremely useful modeling primitive [HGP92]. Component relations can be classified either to be exclusive or shared. Second there can be a dependency from the part to the whole or vice-versa.
Following the ER paradigm component relations are represented by normal relationship types, but to reduce semantic overloading separate modeling primitives may be advantageous. In TROLL light exclusive part relations, where the part is furthermore dependent on the whole, can be easily described by vertical structuring (hierarchical composition of object communities).

**TEMPLATE** Car

**TEMPLATES** Engine;  
**SUBOBJECTS** Motor:engine;  
**EVENTS** ...; getMotor; dropMotor; ...;  
**INTERACTION** getMotor >> Motor.create;  
dropMotor >> Motor.destroy; ...;

**END TEMPLATE;**

Some semantic data models like GSM [HK87] or IFO [AH87] support the modeling of higher-order object types like aggregation types (a motor-boat is composed of a motor and a hull) or grouping types (a convoy is given by a set of vehicles). In many ways higher-order object types resemble component relations. Indeed, they can be viewed as special component relations where the whole depends on the parts. In this way relationship types can be viewed as special higher-order object types too. Consequently the representation of higher-order object types in TROLL light follows the representation of relationship types.

**TEMPLATE** ERschema

**TEMPLATES** MotorBoat, Motor, Boat ...;  
**SUBOBJECTS** MotorBoats(motor, boat):motorBoat; ...;  
**EVENTS** ...;  
addMotorBoat(motor,boat);  
removeMotorBoat(motor,boat);  
...;  
**INTERACTION** addMotorBoat(M,B) >> MotorBoats(M,B).create;  
removeMotorBoat(M,B) >> MotorBoats(M,B).destroy; ...;

**END TEMPLATE;**

**ISA relations:** The specialization or ISA relation has been proved to be a fundamental concept of semantic data models [HK87]. This relation allows to model aspects of a given object, for example the role student of person. In the original ER model ISA relations would be represented by relationship types, but again to reduce semantic overloading additional modeling primitives show to be helpful.

With regard to a TROLL light representation it is important to observe that aspects show a similar dependency on their origin as the dependency found in part relations. Hence vertical structuring of TROLL light might be also used to represent aspects.

**TEMPLATE** Person

**TEMPLATES** Student;  
**SUBOBJECTS** RoleStudent:student;  
**EVENTS** ...; becomeStudent; ceaseToBeStudent; ...;  
**INTERACTION** becomeStudent >> RoleStudent.create;  
ceaseToBeStudent >> RoleStudent.destroy; ...;
END TEMPLATE;

Remark: Both component and ISA relations may be accompanied by the propagation of attributes and events. This is called inheritance. However, it is observable that the particular direction in which attributes or events should be propagated often depends on the concrete application. In TROLL light the propagation of attributes must be made explicit by the specification of derived attributes, whereas the propagation of events is specified by interaction rules. The advantage of this additional effort can be seen in the increased flexibility of this approach.

4 Conclusions

The specification of large information systems with TROLL light requires in the first development phase a quite complete and detailed overview of the UoD. Experience shows that information system developers do not have such an overview at the beginning of the development. For overcoming this problem we have selected the ER model as a means for creating a first structural description of the UoD.

Our proposal for transforming ER schemas into TROLL light specifications provides an important link between structural and dynamic modeling. It is quite amazing to see how nice the concepts of the ER model can be formulated in an object-oriented specification language. The translation shows that ER modeling and object-orientation can assist each other.

Nevertheless it will be necessary to integrate the transformation of ER schemas into TROLL light also methodically into the whole software development process. In our opinion this can be achieved because the proposed transformation ideas give enough freedom for also including functional and even non-functional requirements into the transformation process which results in the possibility of generating alternative solutions.

References


