MOBILE CONSTRAINTS FOR SEMANTIC WEB APPLICATIONS

Peter Gray, Kit Hui, and Alun Preece
University of Aberdeen, Computing Science Department, Aberdeen, Scotland
pgray,khui,apreece@csd.abdn.ac.uk

Abstract We present a framework for semantic web applications based on constraint interchange and processing. At the core of the framework is a well-established semantic data model (P/FDM) with an associated expressive constraint language (Colan). To allow data instances to be transported across a network, we map our data model to the RDF Schema specification. To allow constraints to be transported, we define a Constraint Interchange Format (CIF) in the form of an RDF Schema for Colan, allowing each constraint to be defined as a resource in its own right. We show that, because Colan is essentially a syntactically-sugared form of first-order logic, and P/FDM is based on the widely-used extended ER model, our CIF is actually very widely applicable and reusable. Finally, we outline a set of services for constraint fusion and solving, which are particularly applicable to business-to-business e-commerce applications. These services are accessible using the CIF.

1. Introduction and Motivation

The *semantic web* vision is to enable rich machine-processing of web information sources [2]. Most of the work done to date in realising this vision has focussed upon developing "web friendly" representations for structured data. Building on the XML standard, a number of proposals for expressing data schemas have appeared, including RDF, RDFS, and XML Schema [4]. These approaches support the representation and communication of entity-relational information in web applications, for example, allowing a set of instances of some entity type to be gathered from several structured web pages, and transported for storage in a web-connected database.

The next significant stage in the realisation of the semantic web lies in extending the basic data schema representations with information on how the data can and should be used. This information can take various

forms, including logical axioms, rules, constraints, or even functional and procedural representations. This work will create the *reasoning services* for the semantic web. Berners-Lee's vision for these reasoning services is that they are built on top of the data schema representations. Early work in this direction included inference engines for large-scale ontologies [4, 8], mechanisms for representing and reasoning with business rules [11], and mobile constraint languages and constraint-solving frameworks [13, 7].

In this paper, we present a Constraint Interchange Format (CIF) aimed specifically at business-to-business e-commerce applications on the semantic web. The constraint language is well-established, having been used successfully in diverse and challenging application domains such as representing and reasoning about protein structures in biomedical applications [6], and configuring telecommunications network services [5]. Earlier versions of the CIF were based on Prolog term structures; this paper presents a new XML encoding of the CIF, designed to be more open and less platform-dependent than its predecessor. The new XML-CIF has been modelled using the RDF Schema specification, so that constraints become web resources about which statements can be made (for example, statements about the authorship and context of a constraint). In addition, the CIF has been designed to refer to RDF Schema definitions so that the terms referred to in a constraint can have corresponding RDF Schema descriptions. The motivation here is to allow applications of RDF Schema to use our CIF and constraint-solving services, in keeping with Berners-Lee's vision of reasoning services atop the data representation layer.

The paper is organised as follows: the next section introduces our constraint language, Colan, and the semantic data model upon which it operates; the following section describes the design of the XML-CIF encoding of Colan expressions; the penultimate section gives an overview of the constraint-solving services already available for semantic web applications using the CIF; the final section concludes.

2. The Colan Constraint Language

An entity-relationship (ER) diagram (or UML class diagram) is commonly thought of as being useful chiefly as a diagrammatic guide or visualisation of the types of entities and relationships in a database (or an object-oriented application); it is less commonly seen as an operational basis for forming queries or specifications. In our previous work [1, 6] we showed how quantified constraints can be expressed in our Colan language, in a very readable form of first order logic, including evalu-

```
constrain each t in tutor such that astatus(t)="research"
  so that each s in advisee(t) has grade(s) > 60;
```

First-order logic version:

$(\forall t,s,g)\ tutor(t) \wedge astatus(t,\text{'research'}) \wedge advisee(t,s) \wedge grade(s,g) \Rightarrow g>60$

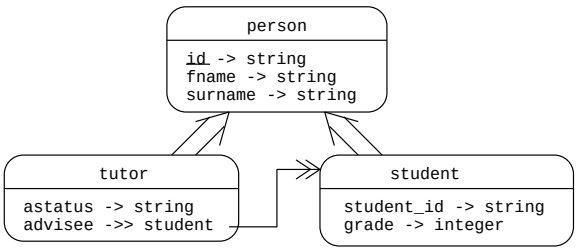

Figure 1. Example Colan constraint and ER diagram for corresponding schema.

able functions which can be computed over data values represented in the ER diagram. Hence, the underlying data model is called the Functional Data Model (FDM). The semantics of the objects referred to in Colan constraints are described in terms of this extended ER data model, which is of the kind in widespread use in UML and in database schemas. Our FDM, P/FDM (Prolog/Functional Data Model) is a Semantic Data Model based on Shipman's original data model [12]. (For full details of P/FDM, Colan [1], and related work, see: www.csd.abdn.ac.uk/∼pfdm)

An example Colan constraint is shown in Figure 1, which expresses a constraint on a university database containing student records. A variable `t` ranges over an entity type `tutor` which is populated with stored object instances. Each of these instances may be related to instances of student entities through the relationship `advisee`, which delivers a set of related entities as in an object-oriented language. Thus functions may also represent a derived relationship, or method. The entity classes can form part of a subtype hierarchy, in which case all properties and methods on the superclass are inherited by each subclass. Method definitions may be overridden, but not constraints.

This is significant for semantic web applications, since it means that information represented in this way is not restricted to human inspection — it can be proof-checked mechanically, transformed by symbol manipulation, or sent to a remote constraint solver. Moreover, given a standardised interchange format, data and attached constraints from multiple sources can be gathered together, checked for compatibility, and used to derive new information. Because the P/FDM data model

is an extended ER model, it maps very easily onto the RDF Schema specification, as we shall show in the following section.

Colan is as expressive as the subset of first-order logic that is useful for expressing integrity constraints: namely, *range-restricted* constraints [1, 6]. This class of constraints includes those first-order logic expressions in which each variable is constrained to be a member of some finite set of values. Figure 1 shows the Colan constraint in the equivalent first-order logic representation.

2.1. Constraints in Semantic Data Models

One of our original motivations for the introduction of constraints into P/FDM was to enrich the capability of the data model in capturing semantics. Data models are more than type systems (which they resemble syntactically) because they also represent constraints on the data. Thus a powerful constraint language has enabled us to capture the semantic information that others have endeavoured to express by extending the diagrammatic notation of the ER model. Thus, although the original FDM was somewhat lacking in semantics, we have been able to more than make up for that by introducing constraints.

In doing this, we have capitalised on a cardinal virtue of the FDM, that it enables one to make well-formed mathematically precise computations over data stored as instances of entities related by an ER diagram. Thus we based the Colan formalism for constraints on just these well-formed computations, revising and extending it to use the full expression syntax of P/FDM's Daplex query language. Crucially, because the expressions are fully quantified and referentially transparent, it is straightforward to move them into other contexts and transform them in ways which preserve their semantics. This would not be so if our expression of data semantics had been confined to a diagrammatic notation.

In summary, we have introduced our constraint language, Colan, and the semantic data model upon which it operates, and we have argued for the applicability of this language and data model to semantic web applications. The following section describes our new XML encoding of Colan in the form of a Constraint Interchange Format (CIF) based on the RDF Schema specification.

3. XML Constraint Interchange Format

In defining our Constraint Interchange Format, we were guided by the following design principles:

- the CIF would need to be serialisable into XML, to make it maximally portable and open;
- constraints should be represented as resources in RDF, to exploit RDF's widely-used data model and ease of parsing, and so that RDF statements can be made about the constraints themselves;
- there must be no modification to the existing RDF and RDF Schema specifications, so that the CIF would be layered cleanly on top of RDF;
- it must be possible for constraints to refer to terms defined in any RDF Schema, with such references made explicit.

As we showed in the previous section, the entity-relational basis of both our P/FDM data model and RDF made it relatively straightforward to map from the former to the latter. In building the RDF Schema for our CIF we were guided by the existing grammar for Colan [6] which relates constraints to entities, attributes and relationships present in the ER model. This grammar serves as a *metaschema* for the Colan constraints (such metaschemas are very common in relational and object database systems). A number of issues arose in developing the RDF Schema for CIF, discussed in the following subsections.

3.1. Metaclasses for Entities and Relations

Our implementation of the P/FDM semantic data model makes use of an `entmet` class that holds information on all *entity* classes, and a `propmet` class that holds information on relationships (functions), both stored and derived. The metaschema is fully queryable, and for this purpose the property values of members of these metaclasses are all held as strings (as is common in data dictionaries), so that the answer to a query on them returns the name of an entity or relation and not the contents of the entity or relation. The P/FDM Daplex definitions of the `entmet` and `propmet` classes are shown in Figure 2, together with their superclass, `objmet`.

The property `rdfname` on the `entmet` and `propmet` classes holds the unique URI for an RDF resource, and thus provides an explicit link to the RDF Schema definition for the corresponding RDF Schema class or property. Thus, constraints carry explicit relationships to the domain ontology (as represented by an RDF Schema) for the terminology to which they refer. We chose to employ our own `entmet` and `propmet` classes rather than use the RDF `Class` and `Property` classes directly, in order to maintain close compatibility with our original Colan metaschema,

```
% objmet - superclass of entity and property metaclasses
declare objmet ->> entity
declare oname(objmet) -> string

% entmet - the metaclass of all entity classes
declare entmet ->> objmet
declare super(entmet) -> entmet
declare rdfname(entmet) -> string % link to RDF Schema

% propmet - the metaclass of all properties (functions)
declare propmet ->> objmet
declare fname(propmet) -> string
declare firstargtype(propmet) -> entmet
declare resulttype(propmet) -> entmet
declare has_inv(propmet) -> boolean
declare rdfname(propmet) -> string % link to RDF Schema
```

Figure 2. P/FDM Daplex definitions for entity and property metaclasses.

and support extensibility in the CIF (for example, Figure 2 shows that `propmet` has additional properties not required by RDF `Property`).

In the RDF Schema we introduce the corresponding metaclasses, `entmet` and `propmet`, which will record the graph of classes and properties. Thus there will be one instance of the `entmet` class for each actual class representing a real-world entity in the instance level RDF Schema for a given application domain. These metaclasses then provide the natural result types for properties used in constraints. Thus, for example, we can use them to say that an atomic boolean value in a predicate in a constraint can be derived by comparing the property value of a variable which holds an entity identifier with another value given by an expression. This entity and property must be known. We could even write a metalevel constraint to require their consistency, as checked by a type checker.

Figure 3 shows the RDF Schema definitions corresponding to the Daplex definitions of the `objmet` and `entmet` classes from Figure 2. It is worth noting that, because properties in RDF are global, some of the original local P/FDM property names must be renamed (for example, `entmet_rdfname` in Figure 3, renamed from `rdfname` in Figure 2).

The basic rules we used when mapping the P/FDM declarations to RDF Schema are as follows:

- a P/FDM class c defined as an `entity` (declared as c `->> entity`) maps to an RDF resource of type `rdfs:Class` (where `rdfs` is the namespace prefix for the RDF Schema descriptions);

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999
   /02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/
   2000/01/rdf-schema#">

<rdfs:Class rdf:ID="objmet">
  <rdfs:subClassOf rdf:resource=
   "http://www.w3.org/2000/01/
    rdf-schema#Resource"/>
</rdfs:Class>

<rdf:Property rdf:ID="oname">
  <rdfs:domain rdf:resource=
   "#objmet"/>
  <rdfs:range rdf:resource=
   "http://www.w3.org/2000/01/
    rdf-schema#Literal"/>
</rdf:Property>
```

```
<rdfs:Class rdf:ID="entmet">
  <rdfs:subClassOf
   rdf:resource="#objmet"/>
</rdfs:Class>

<rdf:Property rdf:ID="super">
  <rdfs:domain
   rdf:resource="#entmet"/>
  <rdfs:range
   rdf:resource="#entmet"/>
</rdf:Property>

<rdf:Property
 rdf:ID="entmet_rdfname">
  <rdfs:domain
   rdf:resource="#entmet"/>
  <rdfs:range rdf:resource=
   "http://www.w3.org/2000/
    01/rdf-schema#Literal"/>
</rdf:Property>
...
</rdf:RDF>
```

Figure 3. RDF Schema definitions for the `objmet` and `entmet` classes.

- a P/FDM class c declared to be a subtype of another class s (declared as c `->>` s) maps to an RDF resource of type `rdfs:Class`, with an `rdfs:subClassOf` property the value of which is the class named s;
- a P/FDM function f declared on entities of class c, with result type r (declared as $f(c)$ `->` r) maps to an RDF resource of type `rdf:Property` with an `rdfs:domain` of c and an `rdfs:range` of r.

3.2. Representation of Variables in Constraints

A fundamental notion in Colan is that a variable is always introduced in conjunction with a set that it ranges over. Thus terms such as `(p in pc)` and `(e in employee)` are common, as in the example expressions:

```
(p in pc) such that name (p) = "xxx"
(e in employee) such that salary (e) > 5000 and age (e) < 50
```

This is represented in the syntax by the `setmem` metaclass, while variables themselves are described by the `variable` class, both defined as shown in Figure 4. An instance of the `variable` class is a legal instance of an `expr` class (representing an expression) by virtue of a series of subclass relationships. There is also a semantic constraint that such an instance must already have been introduced with a quantifier; this is not

```
<rdfs:Class rdf:ID="variable">
  <rdfs:subClassOf
   rdf:resource="#operand"/>
</rdfs:Class>

<rdf:Property rdf:ID="varname">
  <rdfs:domain
   rdf:resource="#variable"/>
  <rdfs:range rdf:resource=
   "http://www.w3.org/2000/01/
    rdf-schema#Literal"/>
</rdf:Property>

<rdfs:Class rdf:ID="setmem">
  <rdfs:subClassOf
   rdf:resource="#boolprim"/>
</rdfs:Class>
```

```
<rdf:Property rdf:ID="var">
  <rdfs:domain
   rdf:resource="#setmem"/>
  <rdfs:range
   rdf:resource="#variable"/>
</rdf:Property>

<rdf:Property rdf:ID="set">
  <rdfs:domain
   rdf:resource="#setmem"/>
  <rdfs:range
   rdf:resource="#setexpr"/>
</rdf:Property>
```

Figure 4. RDF Schema definitions relating to the `setmem` metaclass.

currently captured in the RDF Schema but could possibly be represented using the RDF Schema `ConstraintResource` extension mechanism. An example XML-CIF fragment corresponding to the Colan fragment `(p in pc)` is shown in Figure 5.

Space prohibits us from including the full CIF RDF Schema here, but both it and the full P/FDM schema are available at: www.csd.abdn.ac.uk /research/akt/cif/ together with example constraints. In summary, our CIF RDF Schema serves the purpose of describing what are valid constraints, themselves expressed at an instance level in RDF. It combines the information in a grammar, which is normally used by a syntax checker or a parser, with information normally held in a database schema and used to validate database queries. The interesting thing is that the P/FDM data description language is expressive enough to capture this, especially cardinality constraints. RDF is not so expressive, although it does provide for schema constraints such as cardinality through the `ConstraintResource` extension mechanism.

It should be noted that the metaschema makes a clean separation between the description of *constraints* (both universal and existential) and *expressions*. Constraints and their boolean components are a representation of first-order logic, with the usual connectives. Any knowledge source that uses FOL should be able to understand this. Expressions refer to facts about entities, their subtypes, attributes and relationships, and is based on the widely-used concepts of an ER model. The corresponding data could be held in XML or a variety of storage schemas.

The following section describes our existing framework for using CIF constraints, as used in KRAFT [10] and for which communication and

```
<cif:setmem>                                    <cif:entmet>
  <cif:var>                                       <cif:entmet_rdfname>
    <cif:variable ID="#p">                          http://www.aktors.org/
      <cif:varname>p</cif:varname>                   domain/pc_config#pc
    </cif:variable>                               </cif:entmet_rdfname>
  </cif:var>                                    </cif:entmet>
  <cif:set>                                   </cif:entclass>
    <cif:entset>                            </cif:entset>
      <cif:entclass>                      </cif:set>
                                        </cif:setmem>
```

Figure 5. XML-CIF fragment corresponding to the Colan fragment `(p in pc)`.

coordination are achieved by the exchange and fusion of constraint knowledge.

4. Fusion of Constraints from Different Sources

Having a semantic data model (P/FDM) extended with constraints (Colan) and mapped into an open interchange format (RDF Schema and XML-CIF) supports a range of applications in which information needs to be moved across a network with rich metalevel information describing how the information can be used. An example application common in *business-to-business e-commerce* involves the composition of a package product from components selected from multiple vendors' catalogues (for example, consumer electronic equipment, package holidays, fitted kitchens, or financial products). In each case, there are various kinds of constraints which must be aggregated and solved over the available component instances: constraints representing customer requirements ("I want a PC with a colour printer"), constraints representing rules for what constitutes an acceptable package ("any printer must have a driver that is compatible with the PC OS"), and constraints representing restrictions on the use of particular components ("this printer has drivers only for Windows OSes").

In this last case, the ability to store constraints together with data in P/FDM allows instructions to be attached to the class descriptor for data objects in a product catalogue database. When a data object is retrieved, these attached instructions must also be extracted to ensure that the data is properly used. Thus the attached constraint becomes mobile knowledge which is transported, transformed and processed in a distributed environment. An advantage of this approach is that we can use constraint logic solving techniques to see if a complex set of constraints, whose form is not known until runtime, does have a solution.

4.1. A Constraint Fusion Example

To demonstrate constraint fusion from different sources, consider a configuration problem where a PC is built by combining components from vendors. All of the constraints below are expressed in Colan and so would be exchanged using the XML-CIF representation described in the previous section.

The user specifies his requirement in the form of constraint through the user-agent. In this example, he specifies that the PC must use a "pentium2" processor but not the "win95" OS:

```
constrain each p in pc
  to have cpu(p)="pentium2" and name(has_os(p)) <> "win95"
```

For the components to fit together, they must satisfy certain constraints imposed by the solution database. For example, the size of the OS must be smaller or equal to the hard disk space for a proper installation:

```
constrain each p in pc
    to have size(has_os(p)) =< size(has_disk(p))
```

Now the candidate components from different vendors may have instructions attached to them as constraints. In the vendor database of operating systems, "winNT" requires a memory of at least 32 megabytes:

```
constrain each p in pc such that
    manufacturer(p) = "HAL" and name(has_os(p))="winNT"
        to have memory(p) >= 32
```

When we fuse all constraints together so that they apply to the solution database, we get the description of an equivalent constraint satisfaction problem (note the *conditional constraint* in the last line):

```
constrain each p in pc
    to have cpu(p)="pentium2"
    and name(has_os(p)) <> "win95"
    and size(has_os(p)) =< size(has_disk(p))
    and if manufacturer(p) = "HAL" and name(has_os(p))="winNT"
        then memory(p)) >= 32 else true
```

The process of solving the application problem, therefore, is to retrieve data from other databases and populate the solution database while satisfying (i) all the integrity constraints attached to the solution database, (ii) constraints on data objects and (iii) user requirement constraints. This process corresponds to a generate-and-test approach where invalid candidates are rejected by database integrity constraints. A more efficient prune-and-search approach can be achieved by exporting constraint

fragments to a constraint fusing mediator which composes the overall description as a constraint satisfaction problem (CSP) for a configuration task so that it may plan the solution. The CSP is then analysed and decomposed into database queries and constraint logic programs which are fed across to distributed databases and constraint solvers, under the control of a mediator [9].

5. Discussion and Conclusion

In this paper, we have presented a framework for semantic web applications based on constraint interchange and processing. At the core of the framework is a well-established semantic data model (P/FDM) with an associated expressive constraint language (Colan). To allow data instances to be transported across a network, we have mapped our data model to the less expressive (but adequate) RDF Schema. To allow constraints to be transported, we have provided a Constraint Interchange Format (CIF) in the form of an RDF Schema for Colan, allowing each constraint to be defined as a resource in its own right.

Because Colan is essentially a syntactically-sugared form of first-order logic, and P/FDM is based on the widely-used extended ER model, our CIF is actually very widely applicable and reusable. Through the KRAFT project, we have available a set of services for constraint fusion and solving, which are particularly applicable to business-to-business e-commerce applications. All of these services can be accessed using the CIF.

We targeted our approach for use with RDF Schema, and in particular the XML encoding of RDF Schema, in an effort to maximise the applicability of our work. RDF Schema is the simplest and most universal of the semantic web data representations, while still being adequately expressive for our purposes. In linking Colan to RDF Schema, we also allow its usage with more expressive data modelling languages built on top of RDF Schema, including DAML+OIL [4], and the anticipated recommendations of the W3C WebOnt initative (www.w3c.org). However, a basic requirement of our approach in defining the RDF Schema for Colan expressions was that it should in no way require modification to the underlying RDF definitions (this is in contrast to the OIL approach, which requires modification at the RDF layer in order to capture certain kinds of expression [3]).

Our constraint interchange and solving services are being incorporated into the AKT infrastructure, as one of the basic knowledge reuse mechanisms in the AKT service layer. Further information on this work can be found at: www.aktors.org

Acknowledgements. This work is funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under the Advanced Knowledge Technologies (AKT) Interdisciplinary Research Collaboration (IRC). The AKT IRC comprises the Universities of Aberdeen, Edinburgh, Sheffield, Southampton and the Open University. The constraint fusion services were developed in the context of the KRAFT project, funded by the EPSRC and British Telecom. An earlier version of this paper was presented at the IJCAI-01 workshop *E-Business and the Intelligent Web*.

References

[1] N. Bassiliades and P.M.D Gray. CoLan: a Functional Constraint Language and Its Implementation. *Data and Knowledge Engineering*, 14:203–249, 1994.

[2] Tim Berners-Lee, editor. *Weaving the Web*. Orion, 1999.

[3] Stefan Decker, Sergey Melnik, Frank van Harmelen, Dieter Fensel, Michel Klein, Jeen Broekstra, Michael Erdmann, and Ian Horrocks. The semantic web: The roles of XML and RDF. *IEEE Internet Computing*, Sept–Oct:63–74, 2000.

[4] Dieter Fensel, Ora Lassila, Frank van Harmelen, Ian Horrocks, James Hendler, and Deborah L. McGuinness. The semantic web and its languages. *IEEE Intelligent Systems*, November/December 2000.

[5] N. J. Fiddian, P. Marti, J-C. Pazzaglia, K. Hui, A. Preece, D. M. Jones, and Z. Cui. A knowledge processing system for data service network design. *BT Technical Journal*, 17(4):117–130, October 1999.

[6] P. M. D. Gray, S. M. Embury, K. Hui, and G. J. L. Kemp. The evolving role of constraints in the functional data model. *Journal of Intelligent Information Systems*, 12:113–137, 1999.

[7] P. M. D. Gray, K. Hui, and A. D. Preece. Finding and moving constraints in cyberspace. In *Intelligent Agents in Cyberspace*, pages 121–127. AAAI Press, 1999. Papers from the 1999 AAAI Pring Symposium Technical Report SS-99-03.

[8] Jeff Heflin and James Hendler. Dynamic ontologies on the web. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI-2000)*, pages 443–449, Menlo Park, CA, 2000. AAAI Press.

[9] Kit-ying Hui and Peter M. D. Gray. Developing finite domain constraints – a data model approach. In *Proceedings of the 1st Internatinal Conference on Computational Logic (CL2000)*, pages 448–462. Springer-Verlag, 2000.

[10] A. Preece, K. Hui, A. Gray, P. Marti, T. Bench-Capon, Z. Cui, and D. Jones. KRAFT: An agent architecture for knowledge fusion. *Inernational Journal of Cooperative Information Systems*, 10(1 & 2):171–195, 2001.

[11] D. Reeves, B. Grosof, M. Wellman, and H. Chan. Toward a declarative language for negotiating executable contracts. In *Artificial Intelligence for Electronic Commerce: Papers from the AAAI-99 Workshop*, Menlo Park, CA, 1999. AAAI Press.

[12] D.W. Shipman. The Functional Data Model and the Data Language DAPLEX. *ACM Transactions on Database Systems*, 6(1):140–173, March 1981.

[13] M. Torrens and B. Faltings. Smart clients: constraint satisfaction as a paradigm for scaleable intelligent information systems. In *Artificial Intelligence for Electronic Commerce: Papers from the AAAI-99 Workshop*, Menlo Park, CA, 1999. AAAI Press.