E-ISSN: 2321–9637
Volume 1, Issue 4, November 2013
International Journal of Research in Advent Technology
Available Online at: http://www.ijrat.org

CONTENT DICTIONARIES FOR RELATIONAL ALGEBRA

Kuo-pao Yang[1] , Theresa Beaubouef[2]
[1 2] *Department of Computer Science and Industrial Technology*
[12] *Southeastern Louisiana University*
[1]*kyang@selu.edu*

ABSTARCT: The OpenMath language has been used to create mathematical objects in an XML-based language for establishing content dictionaries for various mathematical concepts. The content dictionaries are an excellent tool for both educational and research purposes. This research investigates the creation of content dictionaries for relational algebra, a theoretical language for describing operations on relational databases.

Keywords: *Content Dictionaries; Relational Algebra; Computer Science Education.*

1. INTRODUCTION

Computer science is a rapidly evolving field with new terminology and techniques being added to the "body of knowledge" almost daily. It is difficult for students and educators to stay abreast of the ever improving technology. Content dictionaries for computer science could help with computer science (CS) education by not only providing a storehouse of the meanings and representations of CS concepts, but also by providing a unified framework which can be used by students and instructors for developing applications for enhancing student learning. Students can look up concepts directly or they can learn to write applications that are generic enough to use the metadata found in the content dictionary. Because the content and information related to the content's representation and visual display are stored together using a markup language, object-oriented concepts are reinforced.

Markup languages are used as a way to store in a text based document "markups" or commands that provide annotation for processing the document. These annotations, called tags, typically define the way document text is to be interpreted. In Hyper Text Markup Language (HTML), a widely used webpage markup language, for example, these annotations describe how data is to be presented in software such as web browsers. It uses predefined structural markers for communicating presentational semantics to the browser software. For example, <head> and </head> are HTML tags placed around text to indicate that it is the heading of a document.

Extensible Markup Language (XML) is another markup language that consists of a set of standardized rules for encoding documents in machine-readable form. Similarly to HTML, XML has been extensively used in representing data structures for structuring, storing, and transporting information. Unlike HTML, however, XML tags do not have predefined semantics. This permits the author of the application to establish unique and specific XML tags and document structure. Commonly applied together, HTML formats and displays data, while XML stores and transports it.

MathML [6] and OpenMath [4] [7] take advantage of the "markup" concept, which they use to represent fundamental mathematics concepts. MathML is based on XML and consists of content markups as well as presentation markups, providing a mechanism for capturing both the structure and content of mathematical expressions, and easily allowing for their display, manipulation, and sharing over the web [1] [8]. The content elements provide the semantics, or meaning, of the mathematical expression and can serve as a learning tool in educational environments, whereas the presentation elements are used for visualization of the expression. In addition to their rendering in web browsers, MathML expressions can be edited in word processors, sent to printers, and evaluated in computer mathematical systems [9].

OpenMath content dictionaries have proven beneficial for educational purposes [3] because they provide a self contained repository of mathematical symbols, operations, and properties stored in a unified framework that allows for easy information lookup, retrieval and visualization. The structure of the content data with tags allows for a variety of custom software to be developed for individual needs while preserving the representation and semantics in a common base of knowledge. Because of this, a field such as computer science, and in particular the area of relational databases, can benefit from the use of content dictionaries. Relational databases [2], because of their strong mathematical foundation and elegance of simplicity have been around for a very long time and almost all information systems are built at least in part on a relational database. They are formally defined based on the mathematical concepts of sets, relations, functions, and formal languages for manipulating the database include the relational algebra and the relational calculus [5].

In this research we are investigating how to apply concepts of content dictionaries originally developed for OpenMath to computer science concepts, and in this paper we discuss the use of a content dictionary for relational algebra. We chose to begin with relational algebra because it has a variety of operations, each with unique symbols and formal properties. The relational model is well defined and based on a strong mathematical foundation, incorporating concepts such as sets and relations. Additionally, knowledge of relational databases are important for computer science students since so many software applications are integrated with them.

2. BACKGROUND

OpenMath and MathML work together in capturing meaning and expressing visually various mathematical concepts. OpenMath, a general representation XML-based language for communicating mathematical objects, contains semantic definitions and is used to complement MathML, which determines how expressions are elegantly rendered [13]. OpenMath Content Dictionaries (CDs) define a set of related concepts and operations as a set of symbols to interpret the meaning of mathematical formulae [10][11][12]. They are the fundamental mechanism, encoding the semantics of the symbols to build mathematical expressions. Each symbol is defined and placed inside of the CDDefinition element consisting of three parts. The first part is the description written in human language that briefly describes the concept. The remaining two parts deal with mathematical properties associated with the concept: the Commented Mathematical Properties (CMP), and the Formal Mathematical Properties (FMP).

An OpenMath Content Dictionary is an XML file having the extension .ocd. A content dictionary for the factorial operation on integers, for example, is shown in Figure 1.

```
<CD xmlns="http://www.openmath.org/OpenMathCD">
<CDName> integer1 </CDName>
<CDDefinition>
<Name> factorial </Name>
<Role>application</Role>
<Description>
The symbol to represent a unary factorial function on non-negative integers.
</Description>
<CMP> factorial n = product [1..n] </CMP>
<FMP>
  <OMOBJ xmlns="http://www.openmath.org/OpenMath"
version="2.0"cdbase="http://www.openmath.org/cd">
    <OMA>
      <OMS cd="relation1" name="eq"/>
      <OMA>
         <OMS cd="integer1" name="factorial"/> <OMV name="n"/>
      </OMA>
      <OMA>
        <OMS cd="arith1" name="product"/>
          <OMA>
            <OMS cd="interval1" name="integer_interval"/>
            <OMI> 1 </OMI> <OMV name="n"/>
          </OMA>
        <OMBIND>
          <OMS cd="fns1" name="lambda"/>
            <OMBVAR> <OMV name="i"/> </OMBVAR>
            <OMV name="i"/>
        </OMBIND>
```

E-ISSN: 2321–9637
Volume 1, Issue 4, November 2013
International Journal of Research in Advent Technology
Available Online at: http://www.ijrat.org

```
      </OMA>
    </OMA>
  </OMOBJ>
</FMP>
</CDDefinition>
<CDDefinition> … </CDDefinition>
</CD>
```

Figure 1. An OpenMath Content Dictionary (integer1.ocd)

It gives each symbol a human readable description along with examples in either human or machine readable form. For example, the symbol “factorial” is described in prefix form: factorial n = product [1..n]. Examples of OpenMath elements include OMOBJ elements for holding OpenMath Objects, OMA elements for holding OpenMath Applications, OMS elements for holding OpenMath Symbols, OMV elements for holding OpenMath Variables, and OMBIND for encoding OpenMath Binding. For example, the OMS element “cd” refers to the Content Dictionary of “integer1” and the symbol name is “factorial”. The Extensible Stylesheet Language Transformations (XSLT) stylesheets of OpenMath transform the XML-based implementation into presentation markups (integer1.xhtml), which can be displayed in a Firefox browser as shown in Figure 2. The reader is referred to [10][11] for additional detailed discussion of these concepts.

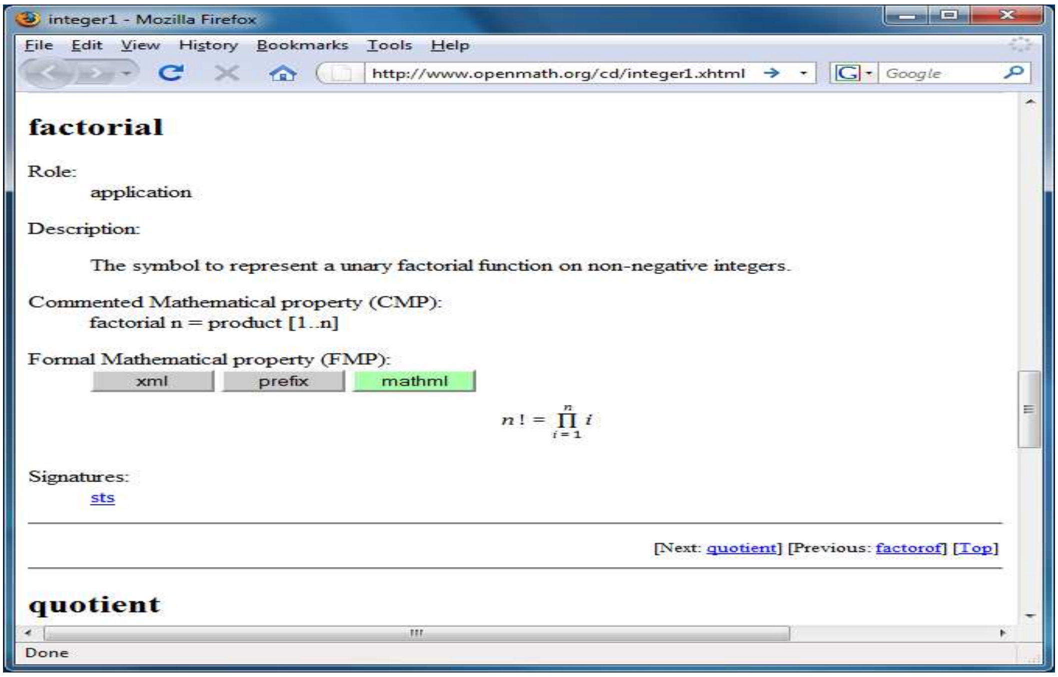

Figure 2. An OpenMath Content Dictionary (integer1.xhtml) in Mozilla Firefox

3. IMPLEMENTATION

We now discuss the implementation of relational algebra operations using the OpenMath approach. We have implemented a comprehensive set of basic axioms in terms of the fundamental relational algebra in this OpenMath format. The developed XML-based Content Dictionary (relational_algebra1.ocd) for Relational Algebra (RA) is shown in the Figure 3:

```
<CD xmlns="http://www.openmath.org/OpenMathCD">
<CDName> relational_algebra1</CDName>
<CDDefinition>
```

E-ISSN: 2321–9637
Volume 1, Issue 4, November 2013
International Journal of Research in Advent Technology
Available Online at: http://www.ijrat.org

```
<Name>Union</Name>
<CMP>Let R and S be two union compatible rough relations...</CMP>
<FMP> …
   <OMS cd="logic1" name="implies"/>
   <OMA>
      <OMS cd="relational_algebra1" name="union"/>
      <OMV name="R"/> <OMV name="S"/>
   </OMA>
   <OMA>
      <OMS cd="set1" name="suchthat"/> <OMV name="T"/>
      <OMBIND>
         <OMS cd="fns1" name="lambda"/>
         <OMBVAR> <OMV name="t"/> </OMBVAR>
         <OMA>
            <OMS cd="logic1" name="or"/>
            <OMA> <OMS cd="set1" name="in"/>
               <OMV name="t"/> <OMV name="R"/>
            </OMA>
            <OMA>
               <OMS cd="set1" name="in"/> <OMV name="t"/> <OMV name="S"/>
            </OMA>
         </OMA>
      </OMBIND>
   </OMA>
</FMP>
<Example> … </Example>
</CDDefinition>
<CDDefinition>Difference … </CDDefinition>
<CDDefinition>Intersection … </CDDefinition>
<CDDefinition>Join … </CDDefinition>
<CDDefinition>Selection … </CDDefinition>
<CDDefinition>Projection … </CDDefinition>
</CD>
```

Figure 3. Content Dictionary for Relational Algebra (relational_algebra1.ocd)

Knowledge of Relational Algebra (RA)is placed inside the CDDefinition element of OpenMath. The symbol elements such as union, difference, intersection, join, selection, and projection are added to introduce relational algebra concepts. The CMP tag defines logical laws of the RA theory and describes the mathematical properties in text format. The FMP tag implements XML-based code in OpenMath format, which is then transformed into MathML and can be displayed in a web browser.

The XSLT style sheet (RA_product1.xsl) for RA products describes how presentation markups display in web browsers as follows:

```
<xsl:template match="om:OMS[@cd= relational_algebra1' and @name=' union']">
        <xsl:call-template name="infix">
                <xsl:with-param name="mo"><mo>&#8746;</mo> </xsl:with-param>
                …
        </xsl:call-template>
</xsl:template>
```

To display infix form, the OpenMath symbol "union" in the Content Dictionary (relational_algebra1.ocd) calls the infix stylesheet template. The MathML presentation element tag, <mo>, helps to display ∪, or 8746.

The Content Dictionary for Relational Algebra (relational_algebra1.ocd), using its stylesheet (relational_algebra1.xsl) and OpenMath stylesheets, is transformed into XHTML format (relational_algebra1.xhtml) by Apache Ant, a Java-based build tool. This Content Dictionary for RA displayed in a Firefox browser is shown in Figure 4.

E-ISSN: 2321–9637
Volume 1, Issue 4, November 2013
International Journal of Research in Advent Technology
Available Online at: http://www.ijrat.org

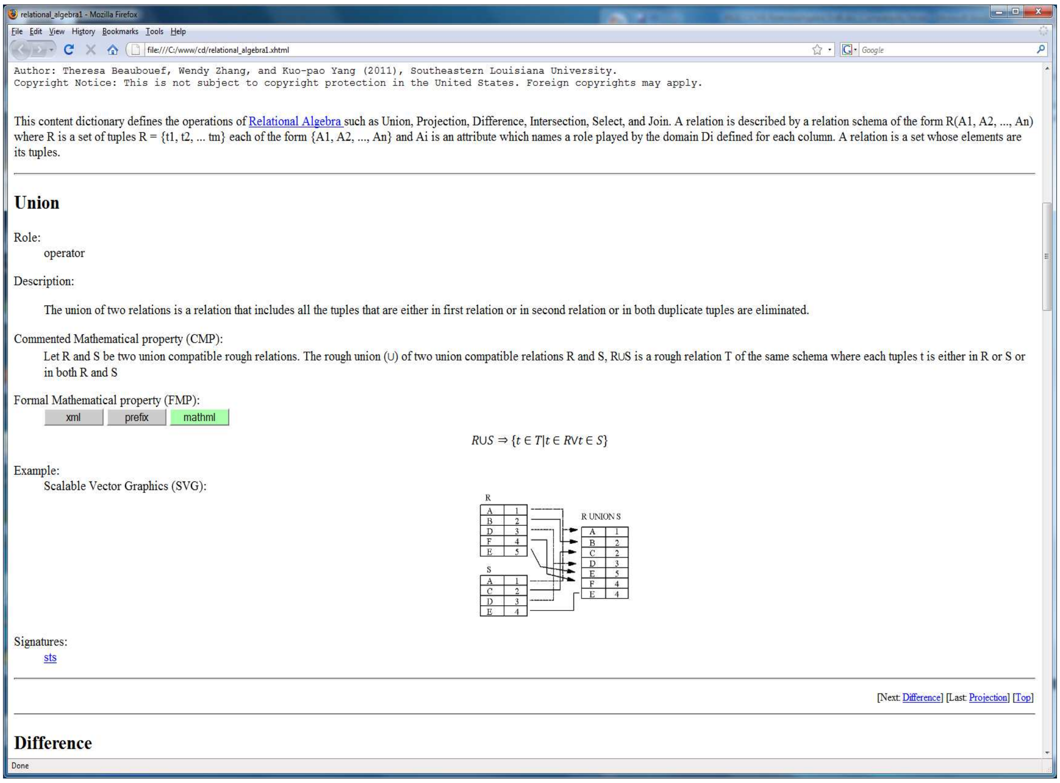

Figure 4. Content Dictionary for Relational Algebra (relational_algebra1.xhtml)

For the Union operator, the Description tag illustrates that the union of two relations is a relation that includes all the tuples that are either in the first relation or in the second relation or in both, and duplicate tuples are eliminated. The Commented Mathematical Property (CMP) describes the mathematical attributes of the Union operation in text format. The union (∪) of two union compatible relations R and S, R∪S is a relation T of the same schema where each tuple t is either in R or S or in both R and S. The Formal Mathematical Property (FMP) has three toggle switch buttons to display XML code, prefix form, and MathML presentation. The layout of MathML presentation in a web browser appears as: $R \cup S$ □ $\{t \in T \mid t \in R \vee t \in S\}$. Each content dictionary is linked to its signature file.

The new stylesheets of RA enable one to write external references and vector graphics in OpenMath Content Dictionary (OCD) files after modifying the original stylesheets of OpenMath. To make an external reference in an XHML file, it is required to write <a> tags in OCD file, for example:

```
<a xmlns="http://www.w3.org/1999/xhtml" href="http://en.wikipedia.org/wiki/relational_algebra">
        Relational Algebra
</a>
```

This external link, Relational Algebra, is enabled in the web browser after transferring this OpenMath Content Dictionary (OCD) file into an XHTML file.

Developers of Content Dictionaries can write in Scalable Vector Graphics (SVG) format, an open XML-based standard that has been under the development by the World Wide Web Consortium (W3C) since 1999. This requires the incorporation of <svg> tags in the OCD file. For example, the image in Figure 4 (union.png) in Portable Network Graphic (PNG) format is enclosed in the following SVG code and can be displayed in a web browser after transforming it into XHML format.

```
<svg xmlns="http://www.w3.org/2000/svg" xmlns:xlink="http://www.w3.org/1999/xlink">
        <image width="100%" height="100%" xlink:href="img/union.png"/>
</svg>
```

4. CONCLUSION

Computer science (CS) encompasses a broad spectrum of topics covering such things as hardware, architecture, design, software engineering, programming, databases, testing, applications, and user issues. Because of the vast amounts of background knowledge and the ever evolving nature of the field of computer science, it is difficult for computer professionals to remain up-to-date on the all latest terminology and practices. There also exist situations where the integration of systems takes place. In these cases, it is important that concepts from the field of computer science are well defined and well-understood so that consistency across applications can be maintained.

A Content Dictionary for computer science can serve as a repository for the formal description of concepts and their relationships to other concepts. It can also be useful as a learning tool where CS students can find information about computer science concepts. Students can also learn more about markup languages and how to develop applications that can make use of the metadata stored along with the various content information. Moreover, the content dictionary aims to provide some structure and standardization to CS concepts that are ever evolving, so that there can be a common ground for their description and use, thus aiding developers, students, and educators.

Because the relational database is such an important concept, not just for computer science students and instructors, but also professionals, we have developed a content dictionary for the relational algebra, a formal language for expressing legal operations on a relational database. We have shown that a markup language similar to OpenMath can be developed to create this content dictionary, and we have discussed its ability to capture meaning and express operations suitable for the relational algebra. This is only the first step in a much more ambitious project—we plan to investigate other aspects of computer science for their suitability to a much larger Content Dictionary for all of computer science, which we will call OpenCS.

References

[1] Carlisle, D. (2000): OpenMath, MathML and XSL, ACM Special Interest Group on Symbolic and Algebraic Manipulation (SIGSAM), 34 (2): pp. 6-11.

[2] Codd, E.F. (1970): A Relational Model of Data for Large Shared Data Banks, Communications of the ACM, 13 (6): pp. 377–387.

[3] Davenport, J. (2000): On Writing OpenMath Content Dictionaries, ACM Special Interest Group on Symbolic and Algebraic Manipulation (SIGSAM), 34 (2): pp. 12-15.

[4] Dewar, M. (2000): OpenMath: An Overview, ACM Special Interest Group on Symbolic and Algebraic Manipulation (SIGSAM), 34 (2): pp. 2-5.

[5] Elmasri, R.; Navathe, S. (2007): Fundamentals of Database Systems, 5th ed., Pearson / Addison Wesley.

[6] MathML. (2013): Mathematical Markup Language, http://www.w3.org.

[7] OpenMath.(2013): OpenMath Content Dictionaries, http://www.openmath.org.

[8] Ross, M.; Rugh, W. (2003): Interactive Mathematics on the Web: MathML for Signals and Systems Demonstrations, Proceedings of American Control Conference, pp. 706-709.

[9] Wright, F. (2000): Interactive Mathematics via the Web Using MathML, ACM Special Interest Group on Symbolic and Algebraic Manipulation (SIGSAM), 34 (2): pp. 49-57.

[10] Yang, K.P.; Zhang, W.; Petry, F. (2009): Building Content Dictionaries for Geometric Algebra in OMDoc Format, Proceedings of the 47th ACM Southeast Conference, Clemson, South Carolina, March, Article No. 45.

[11] Yang, K.P.; Petry, F.; Collins, J. (2010): Building Content Dictionaries for Geometric Algebra in OpenMath Format, Journal of Computing Sciences in Colleges, 25 (4): pp. 22-29.

[12] Yang, K.P.; Zhang, W.; Petry, F. (2012): Physics Markup Approaches Based on Geometric Algebra Representations, Journal of Informatica, ISSN 0350-5596, 36 (1), pp. 91-102.

[13] Zhang, W.; Yang, K.P.; Petry, F., Zambo, S.; Shrestha, R. (2010): Using OpenMath to Present Semantics of Geometric Algebra, Proceedings of the 42nd IEEE Southeastern Symposium on System Theory (SSST), pp. 362-366.