xWCPS: Bridging the Gap Between Array and Semi-structured Data

Panagiotis Liakos[1(✉)], Panagiota Koltsida[1], George Kakaletris[1], and Peter Baumann[2]

[1] Athena Research and Innovation Center, Artemidos 6 and Epidavrou, 15125 Maroussi, Greece
{p.liakos,gkakas}@di.uoa.gr, p.koltsida@imis.athena-innovation.gr
[2] Jacobs University Bremen, Campus Ring 1, 28759 Bremen, Germany
p.baumann@jacobs-university.de

Abstract. The ever growing amount of information collected by scientific instruments and the presence of descriptive metadata accompanying them calls for a unified way of querying over array and semi-structured data. We present `xWCPS`, a novel query language that bridges the path between these two different worlds, enhancing the expressiveness and user-friendliness of previous approaches.

Keywords: Array databases · Query language · Scientific data

1 Introduction

Earth scientists are increasingly overwhelmed by the large volume of data produced during their observations and experiments, e.g., due to the extensive use of sensor arrays. While scientific data conveys a significant diversity of data structures, it is usually multidimensional and complex in nature, it is accompanied with metadata regarding its provenance, and its updates are mostly append-only [1]. Implementing interoperable infrastructures to provide unified access to this kind of data is a challenging problem requiring advances in various domains.

The inefficiency of traditional database management systems (DBMSs) to handle arrays, has led to the development of array DBMSs like `rasdaman` [2], which close this gap by extending the set of supported data structures of relational databases with multidimensional arrays of unlimited size. This enables the efficient storage of n-D spatiotemporal scientific data generated by satellites, telescopes and sensors. Search, retrieval and processing operations on top of array DBMSs has been enhanced through standards specified by the Open Geospatial Consortium (OGC)[1], which defines directives to implement simple

This research has been partially supported by *EC FP7 Work programme (2007-2013)* under grant agreements n. 283610 *"European Scalable Earth Science Service Environment (EarthServer)"* and n. 262693 *"International Network for Terrestrial Research and Monitoring in the Arctic (INTERACT)"*.

[1] www.opengeospatial.org

P. Lambrix et al. (Eds.): EKAW 2014 Satellite Events, LNAI 8982, pp. 120–123, 2015.
DOI: 10.1007/978-3-319-17966-7_14

and easy-to-use web-based interfaces. As far as the metadata is concerned, the World Wide Web Consortium (W3C)[2] provides standard definitions, such as `XML` and `XQuery` [4], that enable the encoding and querying of semi-structured data.

Our approach aims at interconnecting the aforementioned technological advances to offer a unified way of querying array data and their respective metadata. We propose `xWCPS`, an innovative query language that combines ideas from previous approaches and evaluate its syntactic power with the use of a running example. In particular, we consider an array database giving access to *coverages*[3] and their accompanying `XML` metadata, we process as `PNG` images those that satisfy a certain condition and project their *boundingbox* element.

In Section 2 we discuss the buildings blocks of `xWCPS` and in Section 3 we present it formally. Section 4 concludes this study.

2 Background and Motivation

The fundamental idea of this work is the merging of two extensively used standards into a unified language, which provides a more expressive way to query array databases. In this section, we discuss the details of the components used as the building blocks of our query language, the `XQuery` and `WCPS` languages, and explore their use and limitations.

XQuery. The use of `XML` data has exploded in recent years and now a colossal amount of information is stored in `XML` databases or documents in a filesystem. `XQuery` [4] is a query language designed by the `W3C` to enable the efficient utilization of this information. It allows the user to select the `XML` data elements of interest, reorganize and possibly transform them, and return the results in a structure of her choosing. The basic structure of its queries is the `FLWOR` (pronounced *'flower'*) expression. The acronym derives from the keywords permitted in `XQuery` expressions, namely `for`, `let`, `where`, `order by` and `return`. The first query in Fig. 1 provides a sample `XQuery`. Suppose we wish to fetch the *boundingbox* element of all the *coverages* that have *'Clay'* as the title for their *lithology* element. This information is part of our descriptive metadata, so we can simply iterate over the names using `for`, apply our condition with `where`, and retrieve our results with `return` after they are ordered using `order by`. However, `XQuery` does not provide any means to process the *coverages* as images.

WCPS. The Web Coverage Processing Service (`WCPS`) Interface Standard is defined by OGC[4] and specifies a language for retrieval and processing of multidimensional geospatial data [3], with `rasdaman` [2] as the reference implementation. The primary structure of `WCPS` comprises the `for`, `where`, and `return`

[2] http://www.w3.org/

[3] *Coverages* are digital geospatial information representing space/time-varying phenomena.

[4] www.opengeospatial.org

clauses. To illustrate, we continue with our example and process two *coverages* as PNG images with the second query in Fig. 1. WCPS allows the selection of *coverages* by specifying their names, which is very limiting when one wants to query for all the available *coverages*. Moreover, the `where` clause supports only expressions evaluating criteria regarding the array data. Therefore, the descriptive information is inaccessible through WCPS.

(a)

```
for $c in /server/coverages/coverage
where $c//lithology/@title = 'Clay'
order by $c
return $c//boundingbox
```

(b)

```
for $c in (coverageA,coverageB)
return encode($c, "png")
```

Fig. 1. Sample XQuery (a) and WCPS (b) queries

Motivation. Having briefly discussed XQuery and WCPS, it is now evident that the simple task of our example has no practical solution with existing approaches. One needs to use XQuery to filter the *coverages* and retrieve the metadata part of the results, and then compose a WCPS query with the correct names to fetch the images. The results should be then merged manually. Limitations such as the aforementioned have motivated us to propose xWCPS, a query language able to resolve them effortlessly.

3 xWCPS: A Novel Query Language

The xWCPS query language aims towards merging the path two widely adopted standards, namely XQuery and WCPS, have paved, into a new construct, which enables search on both XML-encoded metadata and OGC *coverages*. Additionally, by using xWCPS as the subsuming language we avoid alienating our targeted users, i.e., those already familiar with WCPS. Our language follows XQuery's FLWOR expression syntax and adds two clauses to WCPS, which along with the increased syntactic power instilled into the ones the two languages share, enables the composition of more expressive queries.

The enriched set of clauses is outlined below. Clauses `for` and `let` can appear any number of times in any order. Clauses `where` and `order by` are optional, but must appear in the predefined order should they be used.

- `for`: can create iterator variables holding *coverages* through the WCPS syntax or metadata, which can be *coverages* as well, through the XQuery syntax.
- `let`: can initialize variables following either of the two syntaxes in the assignment expression. The use of the `let` clause can greatly reduce repetitiveness, making xWCPS extremely less verbose than WCPS.
- `where`: can specify criteria on both data and metadata, thus enriching significantly the selecting and filtering capabilities.
- `order by`: can order the results depending on the expression provided.

– `return`: supports both `XQuery` and `WCPS` expressions and enables the creation of *mixed results*, through a newly defined function, viz. `mixed()`. The current available *mixed* formats are `XML`, `HTML` and `ZIP`.

The query in Fig. 2 illustrates how easy it is to retrieve the desired results of our example with `xWCPS`. We iterate over our desired *coverages* with `for`, apply our condition with `where`, order the results with `order by`, and retrieve both processed images and selected metadata. Cases where conditions on array data are needed as well, or results should be pure `WCPS` or `XQuery`, are also handled by `xWCPS`, as noted above.

```
for $c in /server/coverages/coverage
where $c//lithology/@title = 'Clay'
order by $c
return mixed(encode($c, "png"), xquery($c//boundingbox), "xml")
```

Fig. 2. A sample `xWCPS` query

In the context of the EarthServer[5] project, we have developed an `xWCPS` engine, able to parse queries and execute them over cooperating array and `XML` databases. Our partners, forming a very experienced evaluation team in the earth data domain, have verified the effectiveness of our innovative language.

4 Conclusion

We presented a novel query language which allows seamless integration of multidimensional *coverage* data with metadata for retrieval and processing applications. The merits of `xWCPS`, which are visible through the use case provided, indicate that merging these two different worlds is a very promising direction.

References

1. Ailamaki, A., Kantere, V., Dash, D.: Managing scientific data. Commun. ACM **53**(6), 68–78 (2010)
2. Baumann, P., Dehmel, A., Furtado, P., Ritsch, R., Widmann, N.: The multidimensional database system rasdaman. In: Proc. of the 1998 ACM SIGMOD Int. Conf. on Management of Data, Seattle, Washington, USA, pp. 575–577, Jun 1998
3. Baumann, P.: The OGC web coverage processing service (WCPS) standard. Geoinformatica **14**(4), 447–479 (2010)
4. Chamberlin, D.: XQuery: A query language for XML. In: Proc. of the 2003 ACM SIGMOD Int. Conf. on Management of Dat, San Diego, California pp. 682–682, Jun 2003

[5] http://www.earthserver.eu/