

## Modeling Agro Data repository using XML

N.K.Gupta<sup>1</sup>, R.K.Isaac<sup>2</sup>

<sup>1</sup>Sam Higginbottom Institute of Agriculture, Technology and Sciences, Naini, Rewa Road, India

[narendra.nkg2010@gmail.com](mailto:narendra.nkg2010@gmail.com)

<sup>2</sup> Sam Higginbottom Institute of Agriculture, Technology and Sciences, Naini, Rewa Road, India

[isaac\\_rk@hotmail.com](mailto:isaac_rk@hotmail.com)

### ARTICLE INFO

Received: 21 April. 2016

Accepted 02 May 2016

#### Corresponding Author:

**N.K.Gupta**

Sam Higginbottom Institute of  
Agriculture, Technology and Sciences,  
Naini, Rewa Road, India

**Email:** [narendra.nkg2010@gmail.com](mailto:narendra.nkg2010@gmail.com)

**Key words:** Agro, W3C, NXD,  
XPath, XUpdate

### ABSTRACT

Today everywhere there is data, if we want to use that data we have to maintain data repository for that, in such a way so that it can easily and efficiently handled by the user as well as by researcher. But it is a challenging task to design data repository of such type because there are different types of data available as structured, semi-structured and unstructured data. After long analysis it is found that Agro data belongs to the semi-structured type because of its varying nature of attributes in it for same concept. Agro data means data that belongs to the Agricultural domain. As we know XML as a flexible and self-describing technique for modeling semi structured data. In this research article XML is chosen for agro data modeling since the structure of XML may vary from a flat regular data-centric structure to a deep irregular document centric structure.

©2016, IJICSE, All Right Reserved

### INTRODUCTION

This time nearly all fields are using database system to make data more efficient and more useful because its saves data in a proper suitable structure. But data present in the world all are not same type. There are mainly three types of data available as structured data, semi-structured data and unstructured data. Structured data refers to any data that resides in a fixed field with in a record or file. Structured data is managed using relational database, SQL (structured query language)[17]. It uses schema to manage data with database. Unstructured data is all those things that can't be readily classified and fit into a neat box. If the object to be stored carries no tags and has no established schema, ontology, glossary or consistent organized it is unstructured. Unstructured data include free text such as document produced in your company, image and videos, audio files, and some type of social media.

Semi-structured data is in between structured and unstructured but it is close to unstructured data. If the data has any organizational structure (schema) or carries a tag then it is somewhat easier to organize and analyze and because it is more accessible for analysis may make it more valuable. Some type of data that appears to be unstructured but are actually semi-structured are text (XML, EDI), sensor data, web

server logs and search pattern. Semi-structured data model is a tree like hierarchical model, but it is a graph. There is a root element called root node. There is a corresponding XML document for a hierarchical model. Other internal nodes are either object or attributes. Leaf node contains data.

Agro data is a semi-structured type data. Agro data can be considered as information technology applied to the management and analysis of agricultural data. It is the science to develop computer databases and algorithms for purpose of growth and enhance agriculture. Agro data stands for agricultural data, this type of data is not stored in proper design, there is no schema. Various types of agricultural data present are different in their value and attributes. Agricultural data means data about the soil, land, water, crop, metrology and many more. The aim of modeling Agro database is to enhance agriculture system and applies new technology to the agriculture by keeping all data in a proper structure and uses it gently. In agricultural data we have consider the data about the crops (quality, type), and data about the area and the location of the land, it also consider data about the water supply and water resources. Data about soil related attribute and metrology.

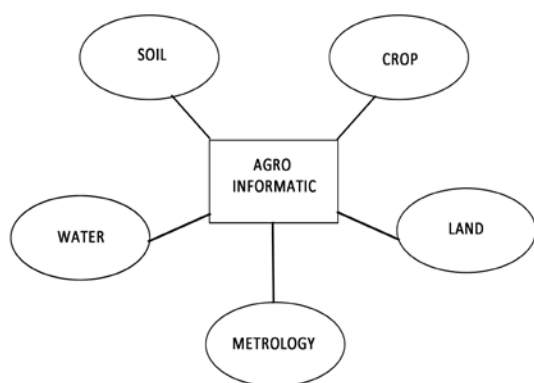


Fig. 1 Agro data model

Figure 1:

All these data are related to each other logically and make sense to maintain and growth of agriculture. This database also helps to maintain agriculture economy. To design Agro database we are using XML. There are three technique to design XML database namely XML-enabled database, Native XML database, and Hybrid XML database. XML-enabled database stores XML document is a relational or object-oriented database that has been extended to hold XML data. The XML-enabled database may only store part of the XML document. Native XML document [7] is a database that stores an XML document as a unit of logical storage. A Native XML document stores all components of a XML model intact-tree hierarchy is maintained and document go in and come out. Basically an NXD falls into two main categories (1) document based storage (2) Node based storage. Document based storage will store the entire XML document in text form and provide some database transaction support such as indexing, materialized view and so on in accesing the document. A simply strategy for this might store the the document as a Binary Large Object(BLOB) or Character Large Object (CLOB) in a relational table or as a file in a file system and provide XML-aware indexes over the document. In contrast, a node based storage model the XML document as the internal model such as Document Object model or simple API for XML (SAX) from the document and stores this model. How the model is stored depends on the physical underlying database [8]. A hybrid XML database provides XML data management in both native and XML-enabled fashion. Most modern RDBMSs today are hybrid and allow, depending on the application requirements, composition of the XML document to and from relational tables as well as native storage.

### WHY XML?

XML stands for Extensible markup language is the World Wide Web Consortium's (W3C). It is easy to use and it is human readable because of strictly

distinguishing content and markups that we call tags. There are two methods of processing of XML file: DOM [2] (Document object model) and SAX (Simple API for XML). The Document Object Model (DOM) is a cross-platform and language-independent convention for representing and interacting with objects in HTML, XHTML and XML documents. The nodes of every document are organized in a tree structure, called the DOM tree. SAX is an event driven online algorithm for parsing XML document. While XML serves as a good hierarchical abstraction over relational data, XML shows its real strength in managing semi-structured data where there is insufficient schema or structure to describe data so that it can be fitted into tabular form [10]. XML database can deal with any type of data, but it is most suited for semi-structured type data because it is less rigid than the traditional relational format. Agro data is also semi-structured type data so we are using XML. XML is both schema ful and schema less, in this all the object and attributes are represented as a tag in a document, there is only tag and document in XML. We are using X query and X update [14] to connect two documents and extract data from document. X path, X link, X pointer also use to do the operation with or within document. There are three different ways to design an XML database namely XML enabled Database, Native XML Database and Hybrid XML Database. There are two types of document format Data-centric [1] and Document-centric [1]. Data-centric data those are regular in structure or no mixed content. These characteristic can be found in some documents and sit very well with the regular and strict model for relational databases. Document-centric Data that are less regular structure, mixed content, larger grained data.

To use or extract values form stored data in document. X path and X update is two XML database language. X path is a (W3C-standard) language used to address a section of an XML document. A single string of X path can address any section of an XML document. X path uses a compact, non-XML syntax to facilities use of X path within URIs. X update is an XML-syntax that is used to specify modifications of XML document. X update uses X path expressions to select nodes for processing [13]. All X updates [15] command have a select attributes which uses X path to specify an XML node to act upon. X pointer [15] is system for addressing component of XML based internet media it is used to point data rather than document. X link is used to create hyperlink in XML documents. Any element in an XML document can behave as a link. With X link, the links can be defined outside the linked files. X query is the language for querying XML data not just XML files but anything that can appear as XML including database, it is like SQL

for database. X query is built on X path expression.

**AGRO DATABASE MODEL**

Agro data model is a semi-structured data model. To represent agro data there is a tree like hierarchical structure. It is a tree like structure but it is a graph. There is one root node that is object and various internal nodes that represent either object or attributes, most last nodes called leaf node that contain data. Each links between nodes is either an attributes link or relationship link. It is tree like structure but it is a graph. Leaf node has atomic value and internal nodes have arc or more arcs. Name of attributes type and relationship are labeled as arcs.[8]

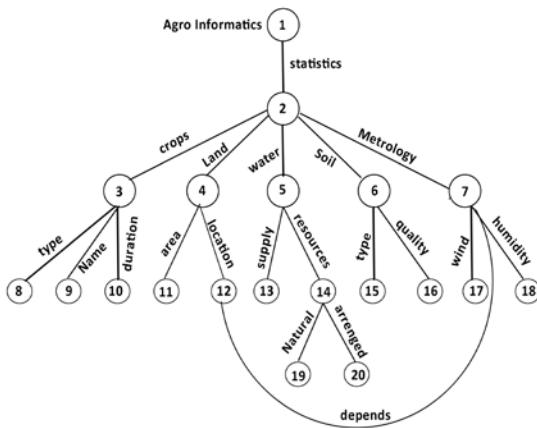


Fig. 2 edge-labeled Agro data tree

Figure 2:

XML document has a single root element which is written in first tag of document. Other objects and attributes are written in further tags. Nesting of tags are allowed with same start and end tag. Tags are text surrounded by brackets, tags come in matching pair. XML tags describe content/data. Well-formed XML document if it has own matching tags and it is schema less. Valid XML involves a document type definition (DTD)[16] that tells the labels and gives a grammar for how they may be nested. The purpose of a DTD is to define the legal building blocks of an XML document. It defines the document structure with a list of legal elements. A DTD can be declared inline in your XML document, or as an external reference. We can model the database with a document node and its associated element node. Node in a hierarchical tree has hierarchical relationship to each other. It can be traversed without having the knowledge of the exact structure of the tree and without having the knowledge of type of data contained within.

XML representation of Agro data model [8]

```
<Agro informatics>
  <statistics>
    <crop>
```

```
      <type></type>
      <name></name>
      <duration></duration>
    </crop>
    <land>
      <area></area>
      <location></location>
    </land>
    <water>
      <supply></supply>
      <resources>
        <natural></natural>
        <arranged></arranged>
      </resouces>
    </water>
    <soil>
      <type></type>
      <quality></quality>
    </soil>
    <metrology>
      <wind></wind>
      <humidity></humidity>
    </metrology>
  </statistics>
</Agro informatics>
```

Agro informatics is root tag. Crop, water, land, soil, metrology are objects and all other are data field. There is a path based query language to extract or update data, travel tag by tag to extract data from atomic tags.

To define the role of each element in a formal model, we have to design Document type definition (DTD) of related document. A DTD defines the structure of the content of an XML document, thereby allowing you to store data in a consistent format[9]. corresponding DTD for Agro data document [6]

```
<!DOCTYPE Agro_informatics [
  <!ELEMENT Agro_informatics (statistics *)
  <!ELEMENT statistics (crop, land, water, soil,
metrology)
  <!ELEMENT crop (type, name, duration)
  <!ELEMENT type (#PCDATA)
  <!ELEMENT name (#PCDATA)
  <!ELEMENT duration(#PCDATA)
  <!ELEMENT land (area, location)
  <!ELEMENT area (#PCDATA)
  <!ELEMENT location (#PCDATA)
  <!ELEMENT water (supply, resources)
  <!ELEMENT supply (#PCDATA)
  <!ELEMENT resource (natural, arranged)
  <!ELEMENT natural (#PCDATA)
  <!ELEMENT arranged (#PCDATA)
  <!ELEMENT soil (type, quality)
  <!ELEMENT type (#PCDATA)
  <!ELEMENT quality (#PCDATA)
```

```
<!ELEMENT metrology (wind, humidity)
<!ELEMENT wind (#PCDATA)
<!ELEMENT humidity (#PCDATA)
```



Agro\_informatics is root tag. #PCDATA after an element name means it is text value. Its component are the tags that appear nested within , in the order specified. Inside Agro\_informatics (statistics\*) means there may be number of statistics exist.

## EXPERIMENT

The node represented here for agro data is a portion of complete node. The complete node populated with data is stored in sql server 2000 and various querying mechanisms like (X query, X path, DOM and LINQ) were applied on this data repository. The GUI is developed by using .Net platform and language for programming used as C#. Execution time is recorded for different types of query based on level of complexity. In addition a separate independent relational model is developed for the same set of data. Necessary measures were adopted during database creation viz.

Normalization for this relational model. Various types of queries based on complexity were executed in relational model also and the response time is recorded for each query 5 times.

## CONCLUSION AND FUTURE WORK

Developed model is used for various types of queries, execution time is recorded and analyzed for different approaches applied for data retrieval, It is observed that X path works well among others approaches. So it is advisable to use Xpath approach for data manipulation in XML data repository for agro data model. In future the same data repository can be stored in various native XML databases available and the performance analysis can be achieved for the same data, and the optimized database can be selected.

Agro data is very useful to maintain agricultural issues and helps to develop the new concepts in agricultural research.

## REFERENCES

1. XML databases, George Papamarkos, Lucas Zamboulis, Alexandra Poulouvassilis, Brikbeck college, University of London, 2007.
2. Moving from XML documents to XML database, fengdong Du, Dalian University of Technology, China. The University of British Columbia, march 2004.
3. EPCC, introduction to XML Database Technologies, Matthew Egbert.
4. From Semi-structured data to XML, Roy Goldman, Jason McHugh, Jennifer Widom, Stanford University.
5. Semi-Structured Data (XML), CS561-SPRING 2012, WPI, MOHAMED ELTABAKH.
6. Semi-Structure Data and XML, Naci Akkok, Pal Halvorsen, Arthur M. keller, Vera Goebel.
7. A STUDY OF NATIVE XML DATABASES, M. Mercedes Mart'inez-Gonz'alez, Miguel A. Mart'inez-Prieto and Mar'ia Mu~noz-Nieto. WEBIST 2009.
8. H. Su-Cheng, L.Chien-Sing, Data storage practices and query processing in XML databases: A survey, Knowledge based system 24(2011) 1317-1340.
9. eXtensible markup language, student guide, ETB/PF/XML/0380, NIIT 2012 .
10. L.Zhen Hua, M.ravi, A Decade of XML data management: An Industrial Experience Report from Oracle, USA, 2009 IEEE.
11. R. Bourret. XML databases. <http://www.rpbourret.com/xml/XMLanddatabases.html>, 2015.
12. W3C, XML Path Language (Xpath) version 1.0, <http://www.w3.org/TR/xpath/>, November 2009.
13. W3C, XML Path Language (Xpath) version 2.0, <http://www.w3.org/TR/xpath20/>, January 2007.
14. W3C, X query 1.0: an XML Query language, <http://www.w3.org/TR/xquery/>, January 2007.
15. Xupdate (<http://en.wikipedia.org/wiki/Xupdate>, 2015
16. S.Haw, Bridging XML and Relational database: mapping choices and performance evaluation, IETE Technical Review, VOL 27, ISSUE 4, JUL-AUG 2010.
17. Eric Pardede, SQL/XML Hierarchical Query Performance analysis in an XML-Enabled database system, Journal of universal computer science, VOL. 15, no 10 (2009).