

TITLE OF RESEARCH WORK: Automated XML Schema Representation in Sensor Array Processing System


Research Description:

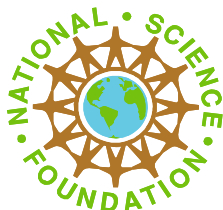
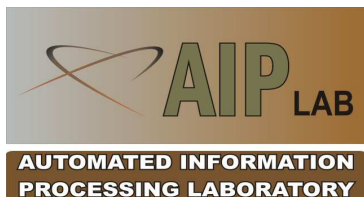
This research aims to provide solution to two very common situations on remote sensing data collection and exchange: the lack of direct physical relationship between data and metadata files and the ambiguity of interpretation between different users with heterogeneous systems (software in which data / metadata will manipulated/processed), and platforms (linux, Windows, Unix,..etc).

Sensed data is data that comes directly from sensors. In addition to the environmental readings from sensors, additional information is crucial for decision making such as sensor identification, reading timestamp, physical location of sensor among others. Metadata is a determinant of how the environment influenced the measurement in case of abnormal findings. Generally, data is taken directly by sensors and some of the metadata as well, but saved in different files. It is very common to have one metadata file for multiple readings. This method is not the most efficient since every reading has its own properties different from the others. Furthermore, in some cases there is a need to alter the metadata file to add more specific information not provided by sensor such as institution or agency conducting the research, principal investigator, contact information, etc. This additional information is helpful for research discovery, back tracking reference in case of data exchange between different users/researches, suggestions input, collaborative efforts and so on.

In addition, different sensors have particular data representation standards that can represent a difficulty for data availability to external users. Understanding the meaning of the content of data and metadata files can be a time consuming task not to mention the dependency of systems and platforms. To deal with the ambiguity of interpretation this work explores XML as a solution. XML (eXtensible Markup Language) is a general purpose markup language capable of describing many different sets of data. XML provides a text-based means to describe and apply a tree-based structure to information. An XML data set can be self sufficient providing the data, the description of the data and the relationships between the data. This flexibility has positioned XML as one of the top choices for data sharing between different systems that might have nothing in common except their XML capabilities.

In conclusion, this research focuses on the design and implementation of a Java-based module to automate the encapsulation of data and metadata files into a single XML file. This new file, due to well known xml standards and protocols, will be suitable for network transmissions and to be share in heterogeneous environments with similar xml capabilities.

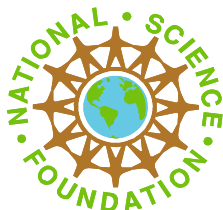
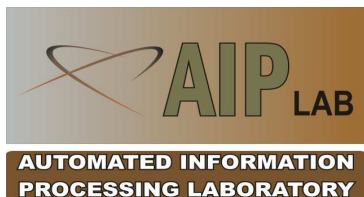
<p>WALSAIP GROUP ASSOCIATION: Automated Information Processing Group</p>	
<p>THESIS TITLE: Automated XML Schema Representation in Sensor Array Processing System</p>	
<p>THESIS ADVISOR: Prof. Domingo Rodriguez</p>	
<p>INSTITUTION: Electrical and Computer Engineering Dept. University of Puerto Rico at Mayaguez</p>	<p>NAME OF RESEACH ASSISTANT: Luz Veronica Acaba-Cuevas</p>
<p>PERSONAL WEBSITE: https://ece.uprm.edu/~s980021/luz_acaba_cuevas/</p>	



RESEARCH PROJECT OUTCOMES:

Publications:

At the moment there are no publications related with this work.



Tools and Applications:

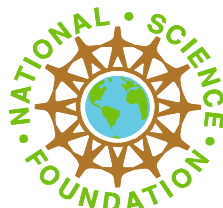
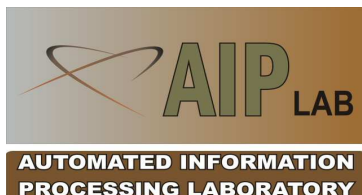
As part of the design and implementation, several tools are being used.

Technologies

- Java – the entire solution will be coded in Java. Java is widely accepted by open source community. In addition, Java is portable (meaning it does not depend on any architecture or operating system). Java is object oriented. Also, easily integrated by the use of well documented APIs with well known technologies.
- FTP (File Transfer Protocol) - FTP allows files of any type to be transferred between a server and a client across a TCP/IP network. If any client or information source wants to receive data transmitted, it will be necessary to have FTP services running in combination with a FTP server.
- XML (eXtensible Markup Language) – is a general purpose markup language capable of describing many different sets of data. XML provides a text-based means to describe and apply a tree-based structure to information. XML will be used to add the *description labels* to every piece of information in the merged data/metadata file. This will ensure ambiguity of interpretation and data availability to heterogeneous users.

Software

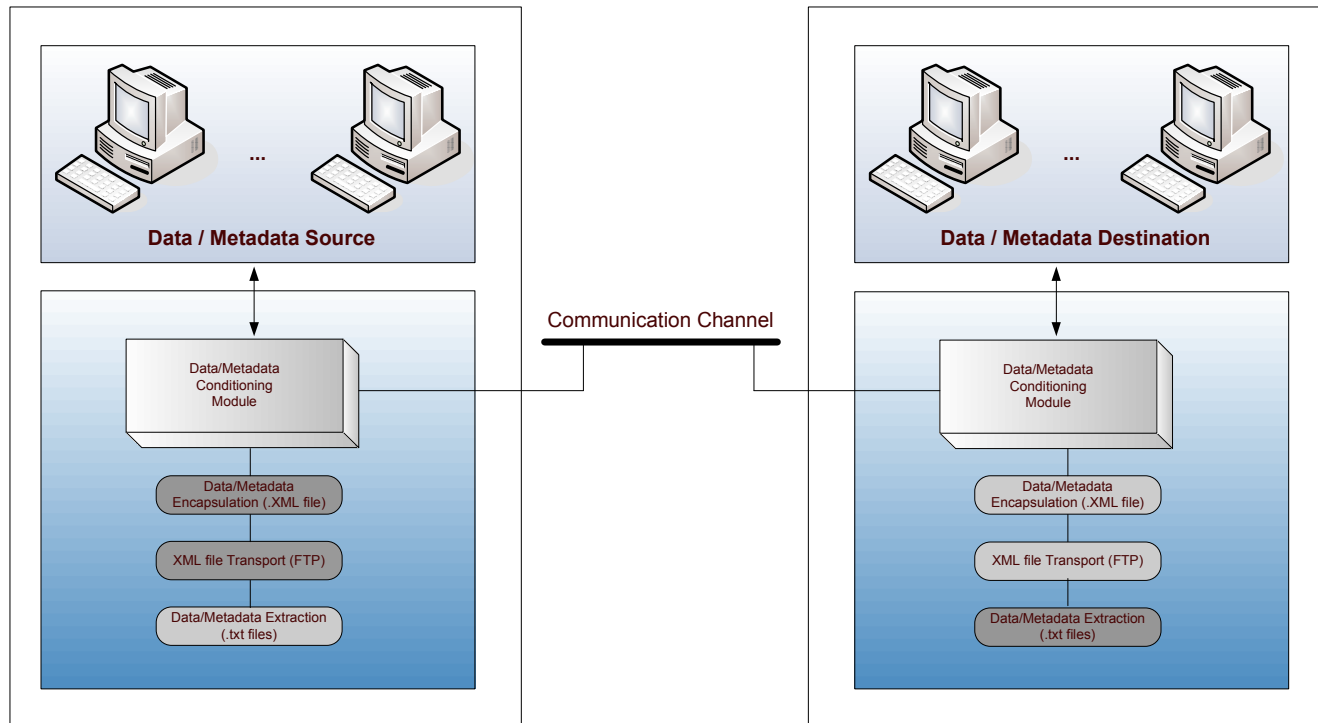
- IBM Rational Software Architect version 6.0 – IBM Rational Software Architect is an integrated design and development tool that leverages model-driven development with the UML for creating well-architected applications and services.
- JFTP 4.0 – JFTP from jMethods is a graphical FTP client application for transferring files between local workstations and an FTP site. JFTP can connect to any system that has a valid Internet address and an FTP server program, allowing transferring files between a wide variety of systems, including Windows, Unix and iSeries (AS/400) systems. JFTP integrates the mostly widely used and widely accepted security mechanism, SSL (Secured Sockets Layer), to securely transfer critical and sensible data.



RELATION OF RESEARCH WORK TO WALSAIP PROJECT:

WALSAIP Project is developing a new conceptual framework for the automated processing of information arriving from sensors in a generalized wide-area, large-scale distributed network infrastructure. This research will provide the mechanism to exchange data and metadata acquired from sensor ensuring physical relationship between data and metadata at the same time making contents fully comprehensible for external users. This module will be installed on every data and metadata transmitter/receiver server part of the WALSAIP cyber-infrastructure and will be the standard application for data/metadata encapsulation and transport to other WALSAIP servers and/or external users.

IMAGE REPRESENTATIVE OF RESEARCH WORK:



The diagram above shows, the interaction between Data/Metadata Conditioning Module and Source/Destination of data. Every component of the WALSAIP infrastructure will use this module in order to exchange data. The main flow of data in WALSAIP cyber-infrastructure is between Raw Data Server (RDS) and Computed Data Server (CDS). Is primarily designed to be used within WALSAIP project but it can be used for external users to incorporate their particular data into WALSAIP servers. There are three main services provided by the Data/Metadata Conditioning

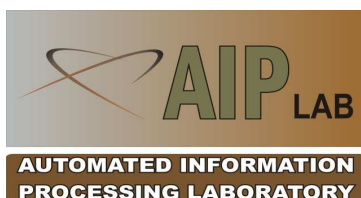
Module: Data/Metadata Encapsulation into XML file, XML file transport, and Data/Metadata Extraction. All files that are sent through the communication channel needs to be processed first by the Data/Metadata Encapsulation service. This will merge both data and metadata files into a single XML document ready to be transmitted trough the communication channel. Not all files that are encapsulated for transmission purposes. There is the option to encapsulate two files and keep a copy on the local workstation. XML file Transport service will enable the user to send the XML file to another workstation. The workstation that received the file has the option to keep the data and metadata encapsulated in XML. If the software package to process data and metadata does not support XML file for processing, the user has the option to revert the encapsulation to put data and metadata in two different text files suitable for input.

RESEARCH DEMONSTRATION:

A functional prototype will be available on mid fall in order to validate the concepts of data / metadata encapsulation and extraction explained earlier. The prototype will automate the encapsulation of data and metadata files in a single XML file suitable for transmission purposes avoiding ambiguity of interpretation.

Listing of Functional Requirements:

- Encapsulation: This action will take as an input two files, data and metadata and generate an XML file as the output. This function will allow the user to specify the names or labels of each piece of information to be transferred. FTP (File Transfer Protocol) technology will be used as the protocol to transfer files. Encapsulation can be performed even if the user doesn't desire to transfer data. The user will have the option of encapsulate data/metadata and save it locally on the workstation for later transmission with this tool or with other mechanism such as email.
- Extraction: Take as the input one XML file and split it in two files: data and metadata. There is no warranty that the systems installed on any particular server will support XML for data processing purposes. For that reason, the tool provide the feature of separating data and metadata into two separated files in order to have data in the most common way used without forcing any particular data format to the user.
- Transport: This feature will provide a list of the servers that are available for data destination. Then the XML file that will be transmitted is selected trough a browse feature. Only XML files can be transmitted.
- GUI: The user will have access to all functions trough a graphical user interface (GUI). This will enable the user to interact with the functions through direct manipulation of graphical



images and widgets in addition to text. Usability guidelines will be used for the design and implementation of the GUI in order to reduce training time and maximize ease of use.

Quality requirements or other requirements:

- Usability – The use of usability guidelines for this solution seek to shortening the time to accomplish task by reducing the number of mistakes made and reducing learning. This will lead to get peoples satisfaction with the system.
- Reliability – the main three activities to ensure software reliability that will be applied to the design and implementation are:
 - Error Prevention – the application will provide mechanism to avoid users to commit errors. For example, instead of entering text, the user can choose values from lists (where applicable).
 - Fault detection and removal – the solution will provide error correction/detection code to validate entries from users.
- Supportability – XML and Java will provide the platform to make this solution available to heterogeneous users with different data architectures, systems, and platforms.

