IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 2, February 2016

Application of XML Compressor [EXI] in Medical Image Database (DICOM)

Gaurav Jaiswal[1], Adesh Chandra[1], Niharika Srivastava[1]

Assistant Professor, Department of CS, Shri Ram Murti Smarak College of Engineering & Technology, Unnao, India[1]

Abstract: Recently, many healthcare organizations are planning to adopt Cloud Web services for storing and transmitting of medical information. SOAP (Simple Object Access Protocol) is the basic communication protocol of Cloud Web services, completely based on XML (eXtensible Markup Language) as an encoding language. The XML is a format that is widely used as a tool for data exchange and storage, but XML files are verbose in nature which means that the amount of information has to be stored, transmitted, and queried is often larger than that of other data formats. This huge size introduces bandwidth communication overhead, end to end communication delay, increased processing time and increased storage space. In this paper, we have presented an effective model to compress and to reconstruct digital imaging and communications in medicine (DICOM) images by using EXI compressor. The experimental results proved that the performance of the proposed technique shows better results than any other XML compression standards.

Keywords: Lossless image compression, medical images, SOAP, XML, EXI, DICOM, FI, XML compressor.

I. INTRODUCTION

The health care industry [1], or medical industry, is an aggregation of sectors within the economic system that provides goods and services to treat patients with curative, preventive, rehabilitative, and palliative care. The development of modern equipment's and communication devices has made it possible to take specialty healthcare to the rural and remote population of a country. With the use of latest technologies, "Patient centricity" has become the key trend in healthcare provisioning and is leading to the steady growth in adoption of electronic medical records (EMR), electronic health records (EHR), personal health records (PHR), and technologies related to integrated care, patient safety, point-of-care access to demographic and clinical information, and clinical decision support. In particular, image data acquired from various medical equipment's like X-ray, CT, MRI, etc. is increasingly transmitted through world wide web, mobile, WAN‟s, etc. Availability of data, irrespective of the location of the patient and the clinician, has become the key to both patient satisfaction and improved clinical outcomes. Cloud technologies can significantly facilitate this trend [2]. Methods of transmitting medical images have been studied by many researchers [3] and [4] but all these system face following three common problems, security, speed, and image quality. To access illegal data has become prevalent in wireless and general communication networks and therefore there is a high demand for techniques that protect data being transmitted. The second problem is the time taken to transmit an image from source to destination. This depends largely on network bandwidth and traffic but file size also has a great impact on transmission speed. Most importantly, main challenge faced by the transmission systems is the demand of doctors and healthcare professionals for a lossless image after transmission. It has been noticed in World Wide Web transmissions; users are increasingly using XML formatted files for transmitting data, images and multiple various multimedia documents from one side to another side.

Most medical Web services use SOAP (Simple Object Access Protocol) [5] as communication protocol (Figure1).SOAP is the basic communication protocol of most Cloud Web services. It is completely based on XML as an encoding language for all sent/received messages over the net. Generally SOAP web services often suffer congestions and bottleneck that caused by the required large web requests and responses. Web messages as XML descriptions are bigger than the real payload of the requested services. This is creating high network traffic load result in slowing down the performance of web applications or stopping them completely.

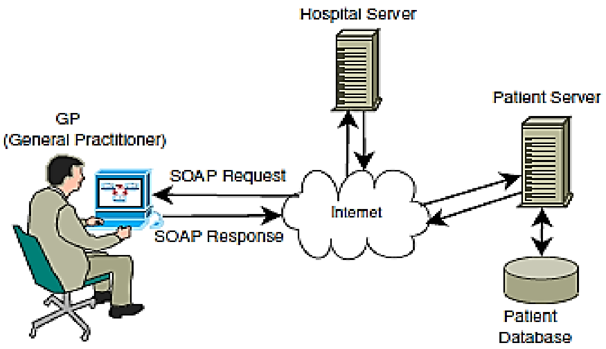


Figure1 Medical SOAP Web Services Scenario

The XML [6] is a markup language for documents containing structured information. Structured information contains both content (words, pictures, etc.) and some indication of what role that contents plays (for example, content in section heading, content in database table, etc.). There is a lot of "redundant" data in XML documents, including element, attributes name and also white spaces.

DOI 10.17148/IJARCCE.2016.5202

Most XML documents are stored in file systems, so there is a need of efficient way to stored file-based XML. Its self-describing format brings flexibility, but comprises efficiency.

A large number of XML compression techniques have been proposed in the literature of recent years. But among all, Efficient XML Interchange is best XML compressor [7]. In this paper, we proposed an architecture, where we convert the all DICOM format medical images to XML files and compress this files with the help of EXI and tried to remove the medical image transmission problem and evaluate the performance of EXI compressor on the basis of some parameters such as memory utilization, CPU usage time, compression ratio, decompression ratio, and compressed, decompressed file size.

The remainder of this paper is organized as follows. Section 2 briefly introduces the review of literature, Section 3 provides the analysis regarding XML compression techniques, and Section 4 presents the proposed solutions, Section 5 provides experimental results and conclusion is given in Section 6.

II. BACKGROUND

The Digital Imaging and Communications in Medicine (DICOM) [8] standard was created by the National Electrical Manufacturers Association (NEMA) [9] to aid the distribution and viewing of medical images, such as CT scans, MRIs, and ultrasound. DICOM isn‟t just a digital image-coding format. It is a comprehensive set of standards for handling, storing and transmitting information in medical imaging. Image compression is the application of Data compression on digital images. Process of reducing the amount of data needed for storage or transmission of a given piece of information (text, graphics, video, sound, etc.), typically by use of encoding techniques. Data compression stands for compressing data or files containing data so that they can be stored in much less memory space that they had been stored in their original form.

Extensible Markup Language (XML) [6] is a simple, very flexible text format derived from SGML (ISO 8879).Originally designed to meet the challenges of large-scale electronic publishing, XML is also playing an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere. XML is a text-based markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable. XML is a flexible way to create common information formats and share both the format and the data on the World Wide Web, intranets, and elsewhere. For example, computer makers might agree on a standard or common way to describe the information about a computer product (processor speed, memory size, and so forth) and then describe the product information format with XML. Such a standard way of describing data would enable a user to send an intelligent agent (a program) to each computer maker's Web site, gather data, and then make a valid comparison. XML can be used by any individual or group of individuals or companies that wants to share information in a consistent way.

Unlike other data storage techniques that use a binary format for data representation, XML uses a text format, which enables users with nothing more than a rudimentary text-editor, and not necessarily the same editor that created the XML, to view and modify XML. Generally, humans should not directly read XML, though the capability is present. Due to XML‟s text format and tagging method, XML documents are usually larger than comparable binary formats. However, the XML designers did this consciously.

XML is "extensible" because, unlike HTML, the markup symbols are unlimited and self-defining. XML is actually a simpler and easier-to-use subset of the Standard Generalized Markup Language (SGML), the standard for how to create a document structure. Both HTML and XML use tags to delimitate data boundaries (words bracketed by '<' and '>'). XML files are portable. They can be transferred from one platform to another without any changes or modification. To make XML easy for people to read, computers end up carrying a lot of unnecessary human baggage. This burden can be reduced on our computer by replacing human-readable elements with concise binary encoding technology such as Efficient XML Interchange (EXI).

III. XML COMPRESSOR

Compression is the application of techniques to a file that reduces the size of the file. Information compression comes in two broad categories of lossy and lossless, where the former does not decompress back to the identical original file and latter does. The choice of lossy or lossless is specific to the domain application. In many cases, a lossy compression is perfectly acceptable, such as JPG imagery where image quality remains acceptable, and in others cases lossy it is not acceptable. For this paper, the concept of compression enables more data transferring under a predefined bandwidth condition. Compression also enables greater file storage capacity when archiving XML information.

XML compressors can be classified with respect to two main characteristics. [10]

1. The first classification is based on their awareness of the structure of the XML documents. According to this classification, compressors are divided into two main groups:

- ➢ General Text Compressors
- ➢ XML Conscious Compressors

2. The second classification based on their ability of supporting queries, compressors are :

- ➢ Non-queriable XML compressors
  - Schema dependent
  - Schema independent
- ➢ Queriable XML compressors
  - Homomorphic
  - Non Homomorphic

DOI 10.17148/IJARCCE.2016.5202

A. Efficient XML Interchange (EXI)

Efficient XML Interchange [11] standard is a W3C recommendation for enhancing the communication of XML data between computer systems. According to the definition of the W3C, "EXI is a very compact, high performance XML representation that was designed to work well for a broad range of applications. It simultaneously improves performance and significantly reduces bandwidth requirements without compromising efficient use of other resources such as battery life, code size, processing power, and memory [12]".

EXI [13] represents the contents of an XML document as an EXI stream. Streams are the basic structure of EXI documents and consist of an EXI Header followed by an EXI Body. The EXI Header contains document metadata information such as the format version and options that were used during encoding. The EXI Body contains a sequence of "events" which comprise the contents of the body. Figure 2 shows the EXI structure.

EXI is coded directly with formal grammars instead of using a higher-level language API. This is a key point that is important to realize because EXI itself is not an API, but an algorithm definition, and therefore the grammar process must be directly implemented. Figure 2 shows the EXI structure.

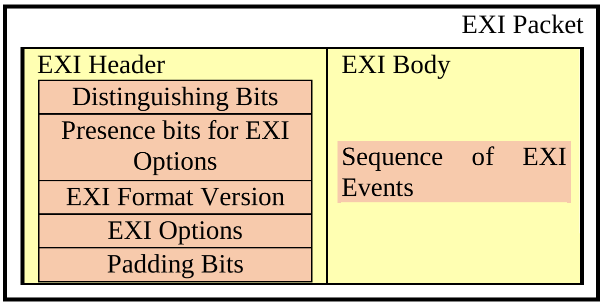


Figure2 EXI Structure

The list of publicly available implementation of the EXI 1.0 specification, in alphabetical order is given in Table I. All these specifications are open-source i.e., freely available.

Table I. Available EXI 1.0 specification

Name	Organisation	Features
Efficient XML [15]	Agile Delta, Inc	Works with fully structured, semi-structured, and unstructured XML data.
EXIficient [16]	A G Siemens (an open source project)	Command-line program that encodes text XML files into binary EXI files.
EXI Processor [17]	EISLAB, Lulea University of Technology	Efficient, typed and low-level API.
Open EXI [18]	Fujitsu (Open source project)	open source Java implementations of EXI

IV. PROPOSED SOLUTION

To remove the current problem related to healthcare industry to transmit the heavy size medical images in lossless format, we proposed a system where by adopting the some steps we can be able to send and receive large size images in less time and to get the medical images in lossless format while transmitting over the WEB.

In this proposed scheme we convert the all medical related images with DICOM format to XML file by using the YAKAMI DICOM TOOLS. This XML file size is smaller than the original DICOM file but larger for transmission process. Thus to overcome this problem we can compressed this XML file by using the EXI Processor in schema-informed mode for getting higher compression ratio. At the receiver side the EXIP decompression process is performed to get the original XML file from this compressed version of the XML file and again by using Yakami Tool, we get our original medical image (in DICOM format). Figure 3 illustrated the main steps involved in the proposed methodology.

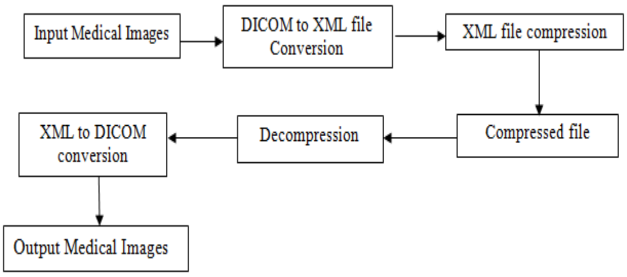


Figure 3 Proposed Medical Image Transmission Architecture over WEB

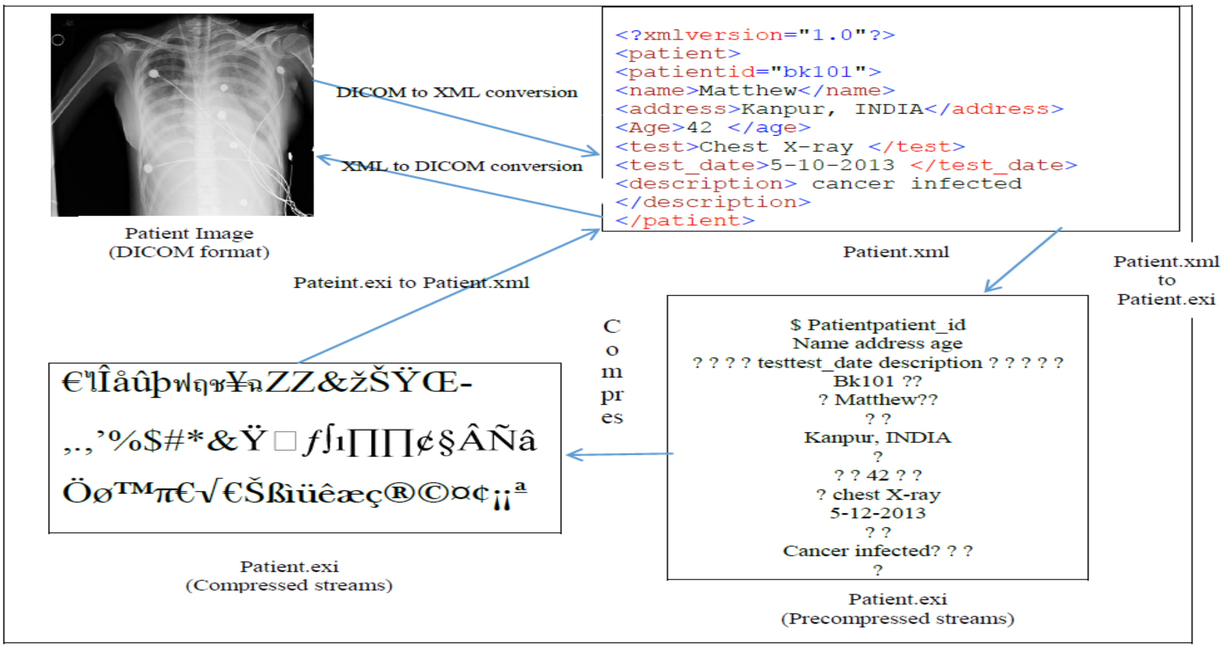
The aforementioned binary XML compression mechanisms were analyzed in terms of their performance. To ensure the consistency of the performance behavior of the evaluated XML compressor, experiments have been performed on the high computing resource capability machine environment for getting the high compression ratio and better results. Table II lists the setup details of high and low environment.

Table II. Setup details of powerful resources environment

Operating System	Windows 10
CPU	Intel Core i5-3210M, CPU @ 2.51 GHz
Hard Disk	500GB, Western Digital
RAM	4 GB

We analysed the compression ratio for different sizes of XML files, the memory used by the compression tools, CPU time for performing their operation and the compressed and decompressed file sizes. In this way, our analysis work is divided into three parts measuring – The file size after compression and decompression, the memory utilization for each operation, and total time taken to perform operations.

Figure 4 describes the complete step by process performed by the proposed architecture.



```
<?xmlversion="1.0"?>
<patient>
<patientid="bk101">
<name>Matthew</name>
<address>Kanpur, INDIA</address>
<Age>42 </age>
<test>Chest X-ray </test>
<test_date>5-10-2013 </test_date>
<description> cancer infected
</description>
</patient>
```

Figure4 Proposed Architecture Process

The work is performed on EXI Processor [13] Table III depicts the experimental values EXI in schema less mode, EXI schema mode. Before using the EXIP on files, we convert the available patient file on DICOM format to XML file format with the help of YAKAMI converter.

A. Performance metrics

We measure and compare the performance of the XML compression tool using the following metrics:

Compression ratio: represents the ratio between the sizes of compressed and uncompressed XML documents as given by:

Compression Ratio = (Compressed Size) / (Uncompressed Size).

Compression Time: represents the elapsed time during the compression process, i.e. the period of time between the start of program execution on a document until all the data are written to disk.

Decompression Time: represents the elapsed time during the decompression process i.e. the period of time between the start of program execution on reading the decompressed format of the XML document until delivering the original document.

Table III. Experimental result value (Using EXI Processor)

	EXI without schema				EXI with schema			
Input Patient XML file name and original size (in MB)	**Compressed file size (in MB)**	**Compression ratio (in %)**	**Memory usage (in K)**	**CPU usage time (in sec)**	**Compressed file size (in MB)**	**Compression ratio (in %)**	**Memory usage (in K)**	**CPU usage time (in sec)**
Patient 1 (98.4)	3.35	29.4	1320	0.34	3.22	30.5	1348	0.32
Patient 2 (52.4)	2.65	19.7	1315	0.28	2.48	21.1	1344	0.25
Patient 3 (30.84)	2.51	12.28	1312	0.24	2.20	14.01	1334	0.20
Patient 4 (12.08)	2.17	5.5	1310	0.20	1.44	8.3	1330	0.14
Patient 5 (2.52)	0.36	7.0	1332	0.14	0.34	7.3	1310	0.13

V. RESULTS AND DISCUSSION

In this section we reported and discussed the results obtained by running our exhaustive set of experiments.

The Figures 5 to 8 represent an important part of the results of our work. Several remarks and guidelines may be observed. Some remarks are given as follows –

- The result of figure 5 represent that EXIP compressor in schema-mode gives more compressed file.
- Figure 6 depict that the EXIP compressor in schema-mode achieves the best compression ratio over the different XML documents.
- Figure 7 show that EXI has the best performance in terms of compression time and decompression time metrics.
- Figure 8 illustrate that the limited memory resources is sufficient for EXI.

DOI 10.17148/IJARCCE.2016.5202

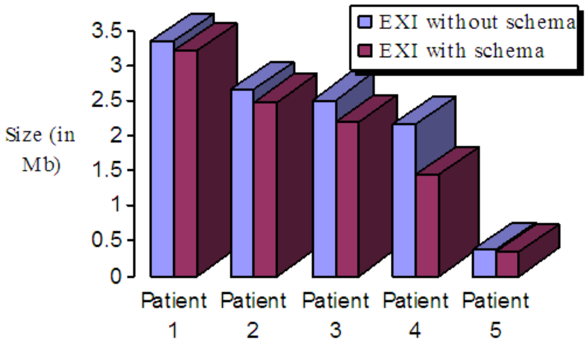


Figure 5 Compressed File Size (in MB)

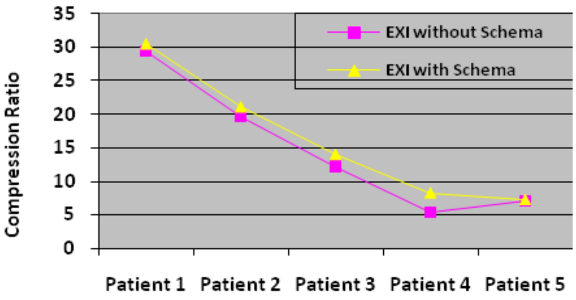


Figure 6 Compression Ratio

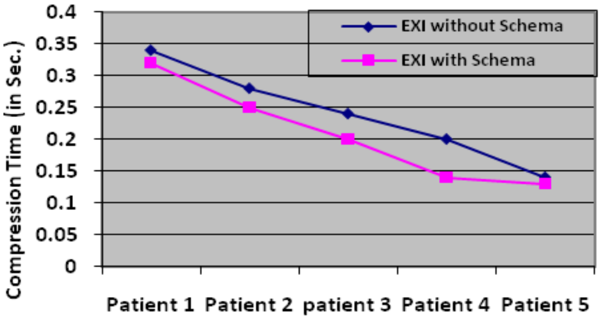


Figure 7 Compressions Time/ Decompression Time (in Sec)

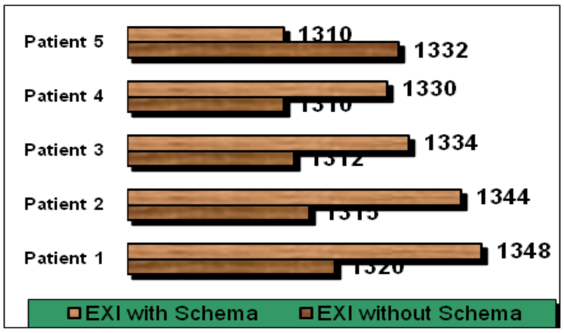


Figure 8 Memory Resources (in K)

VI. CONCLUSION

This paper has shown that XML is an extremely powerful and widely embraced data-representation format that delivers system-to-system interoperability, but does so at the cost of being verbose and often complex to process. These costs of XML can prevent network edge devices from performing real-time network operations due to limited bandwidth and processing capabilities constrained by small CPUs and limited memory capacity. However, through the EXI techniques alternative XML format, the verboseness and processing cost of XML are shown to be significantly reduced, enabling real-time network operations by edge devices.

Ultimately, if EXI is widely implemented, the Healthcare organization will be able to deploy XML based network traffic further, specifically to the Cloud Web Service or other small mobile and wireless network edge devices. Through the use of Cloud Web Service in Hospital organization it will be beneficial for all patients and also for Doctors who can access patient information like patient medical report, previous test images and other information for instant medical treatment. This paper proposes a novel technique/model using EXI compressor for compressing XML files in a lossless fashion. EXI passes the “More-Faster-Better” litmus test for modern technology development and adoption.

After analyzing the experimental result we concluded that EXI schema informed mode compression delivers superior results for compressing the DICOM format medical images after converting them into XML files.

REFERENCES

[1] R. Kooper, A. Shirk, S.-C. Lee, A. Lin, R. Folberg, and P. Bajcsy, “3d medical volume reconstruction using web services,” Web Services, 2005. ICWS 2005. Proceedings. 2005 IEEE International Conference on, p. 716, july 2005.

[2] C. B. Christian Werner and S. Fischer, “Compressing soap messages by using differential encoding,” In Proceedings of the IEEE International Conference on Web Services. San Diego, CA, USA, pp. 540–547, 2004.

[3] Lee, S., Lee, T., Jin, G. and Hong, J., 2008, ―An Implementation of Wireless Medical Image Transmission System on Mobile Devices‖, Journal of Medical Systems, Vol. 32, No.6, pp. 471-480.

[4] Lee, Y., Kwak, J., Cho, S., Kim, J. and Kang, H., 2007, ―A Study on the Medical Image Transmission Service Based on IEEE 802.15.4a‖, Advances in Hybrid Information Technology, Vol. 4413, pp. 159-167.

[5] M. Nakagawa, K. Nozaki, and S. Shimojo, “Web-based distributed simulation and data management services for medical applications,” Computer-Based Medical Systems, 2006. CBMS 2006. 19th IEEE International Symposium on, pp. 125 –130, 0-0 2006.

[6] Veldman, I., Keijzer, A. and Keulen, M., 2009, ―Compression of Probabilistic XML Documents‖, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, Vol. 5785, pp. 255-267.

[7] Jaiswal G., Mishra M., 2013, “Why Use Efficient XML Interchange Instead of Fast Infoset”, 3rd IEEE International Advance Computing Conference (IACC), pp. 925-930.

[8] DICOM (Digital Imaging and Communications in Medicine), PS 3.5 Data Structures and Encoding, NEMA, Rossyln VA, 1999. "http://medical.nema.org/dicom/1999/draft/99_05dr.pdf"

[9] DICOM (Digital Imaging and Communications in Medicine), PS 3.11 Media Application Profiles, NEMA, Rossyln VA, 1999. "http://medical.nema.org/dicom/1999/draft/99_11dr.pdf"

[10] Sakr S., 2009, “XML Compression Techniques: A survey and comparison” Elsevier :Journal of Computer and System Science, pp. 303-322.

[11] W3C: 2011 [Online]. Available: http://www.w3.org/TR/2011/REC-exi-20110310/#principles.

[12] W3C: 2009 [Online]. Available: http://www.w3.org/XML/EXI/.

[13] W3C: 2009 Efficient XML Interchange Measurement Note. Available: http://www.w3.org/TR/2009/WD-exi primer-20091208.

[14] W3C: 2009 [Online]. Available: http://www.w3.org/TR/2009/CR-exi-20091208/.

[15] EfficientXML, “Welcome to AgileDelta Homepage” retrieved 2013, from http://www.agiledelta.com/product_efx.html.

[16] EXIficient, “Welcome to the SourceForge Homepage”, retrieved 2013, from, http://exificient.sourceforge.net/.

[17] EXIProcessor, “Welcome to the SourceForge Homepage”, retrieved 2013, from, http://sourceforge.net/projects/exiprocessor/.

[18] OpenEXI, “Welcome to the SourceForge Homepage”, retrieved 2013, from, http://openexi.sourceforge.net/.

DOI 10.17148/IJARCCE.2016.5202