ISSN:0975-9646

Performance Analysis of New Conversion Tool from Relational Datasets to XML Datasets on Selected Websites

[1].B. Hemanth Kumar, [2]. Prof. M. Surendra Prasad Babu.

1. *Associate Professor, Dept., of Information Technology,*
2. *RVR & JC College of Engineering, Guntur, A.P. India*
3. *2.Professor, Dept., of CS & SE, AU college of Engineering, Andhra University, Visakhapatnam, A.P, India.*

Abstract: Over the years internet has become essential part of extensive business applications to exchange or restructure large amounts of business information. Relational database (RDB) is normally used for traditional data processing, where as Extensible Markup Language documents (XML Documents) are widely useful for performing online conversation in B2B applications. In fact XML data formats may be treated as de facto standard of data representation and exchange format for massive amount of enterprise information over the web. Performance of the application depends on both type of the data structures used and the type of the information maintained in the structures. Resource Description Framework (RDF) of semantic web uses XML to describe an Internet resource as a directed graph to provide a mechanism for describing the machine understandable meaning of any document. In this paper performance analysis is conducted for XML conversion tools developed by the authors with some of the existing tools on some selected datasets. In this analysis JDBC is used to extract data from relational datasets and Java XML Parsers to parse and extract the XML datasets. Experiments show that, the tools developed by the authors produce XML documents efficiently with high scalability than the existing tools.

I. INTRODUCTION

The eXtensible Markup Language (XML), a fast maturing technology, has powerful applications in industry, military, government, and scientific application. XML was developed in 1998 and it is accepted by World Wide Web Consortium (W3C)[1] . The anticipated benefits of XML include easier integration of legacy systems, improved application communications, superior knowledge representation, and platform independence.

Application programs request the data sets and the database servers respond to these requests with respect to the type of the data structures available on the server. The extraction of data on the internet depends upon the type of the data structures used to maintain data and type of the application programs used to retrieve data. The application must immediately get the response from the corresponding database to improve the performance of the overall system.

To process a web page intelligently, a computer must understand the data, and it is possible only by using a new generation computing called in Semantic Web [2]. It is an extension of the current World Wide Web which provides both the structure to a Web document and also a way of describing the machine understandable meaning of a document. Semantic Web mainly focuses on the data and information [3].

Hemanth Kumar et al [4] proposed a semantic web system for Information retrieval using J2EE Technologies. A new java based GUI Semantic data conversion tool to convert Relational datasets into Xml datasets has also been developed by them [5]. When creating Xml data sets the tool also generates a DTD for the corresponding Xml datasets. The primary objective of the present paper is to generate the XML datasets using this tool for three typical data sets namely (i). www.indiakisan.net data sets [6]which contains the extensive information about the Indian agriculture, (ii) Gramarogya data sets[7] and (iii) Engineering college data sets. The performance of the proposed tool has been evaluated by comparing the results with the results of generating XML datasets with the existing tools.

2. SELECTED DATASETS

2.1. India Kisan Datasets.

www.indiakisan.net is an Internet application which gives information about the Indian agriculture. This application provides the suitable information for various requests on crop related areas. The India Kisan application server is maintaining a huge amount of data (more than 250 tables) in the form of Relational data sets. It contains information about selection of suitable crops and crop varieties for cropping, soil analysis, fertilizer management, pest and disease control, irrigation management etc., the agrometry advisory system module of the web portal contains information about the meteorological developments in the atmosphere to face a natural eventualities /calamities in an efficient way. The marketing advisory system of the web portal contains information about the prices of their produces in and around them. The web directory service and the discussion forum service provided in the website also maintain some data sets for the purpose. It contains information about the various latest agricultural tools and implements, with their price & availability, so that farmers can buy them and use in their fields for the better productivity. The web portal stores relevant information to provide computational facilities to the agricultural scientists to analyze their laboratory experiments (or on the fields). It stores relevant data and articles for maintain online journals for Indian agricultural scientists to publish their articles. As a whole the web portals stores a rich database in the area of agriculture and help the country to lead the developing world. The home page of the web portal is in the figure 1.

Figure: 1

The available relational tables of the a typical dataset in indiakisan web portal are considered and presented in **table 1.** The number of records are 11937 and size of this dataset is 1098.5 KB. Each relational data set contains a number of tuppels and a number of columns. Initially the conversion tool is applied on these relational datasets which will convert the relational datasets into corresponding XML data sets.

Figure:2

2.2 Gramarogya Datasets:

www.bharathgramaarogya.net is an Internet application which gives health information & knowledge. This application provides the suitable information for various requests on health related areas. The home page is as shown in the figure:2.

A typical dataset in Gramarogya portal is considered and presented in the table 2. The number of records are 4311 and size of this dataset is 1074.9 KB.

Table: 1

Table Name	No. of columns	No. of Rows	Size in KB	Table Name	No. of columns	No. of Rows	Size in KB
advisory_crop	3	8	2.1	rice_adv_fert	2	3	1.1
advisory_crops	3	16	2.4	rice_adv_fertsub	4	14	5.5
advisory_cropsuggestion	6	33	2.8	rice_adv_water	3	8	4.4
advisory_districts	3	3	2.1	rice_rain_disease	4	44	16.0
biodiesel_tb_fertilizer	3	4	1.4	rice_soiltypes	3	2	3.1
biodiesel_tb_functional_uses	2	4	2.5	rice_transplantation	4	3	8.5
biodiesel_tb_general_inf	2	39	5.5	rice_zones	2	7	2.2
biodiesel_tb_harvesting	2	3	1.2	Rules	2	7	3.4
biodiesel_tb_irrigation	2	3	1.2	Sdepth	6	18	1.9
biodiesel_tb_keyfactors	1	7	1.1	Season	2	5	1.9
biodiesel_tb_land	2	2	1.4	seed_rate	3	2	1.1
biodiesel_tb_nursery_det	2	46	6.1	Sgadvice	5	95	7.0
biodiesel_tb_nursery_mgnt_id	3	14	1.5	sgadvice_sys	2	68	6.6
biodiesel_tb_spacing	3	12	1.3	Sgirritems	3	10	1.6
biodiesel_tb_weed_mgnt	2	3	1.3	Sgrainitems	3	5	1.3
border_measer	5	12	1.4	sgrain_sub1	7	83	22.3
coleus_tb_about	2	15	3.0	Soil	2	4	2.1
coleus_tb_climate	2	4	1.6	Soiltype	1	3	1.1
coleus_tb_economics	2	2	2.7	space_moisture	3	3	1.1
coleus_tb_general	2	29	4.6	sprink_rate	4	40	2.7
coleus_tb_irrigation	2	5	1.7	States	3	1	2.0
crop_and_oil	5	5	1.3	surf_spec	4	7	1.6
j_alt_irr	10	16	2.3	Wdepth	6	18	2.1
j_cotton_irr	9	2	1.1	Webdirectory	15	14	2.0
j_toba_rain	15	16	2.9	wheat_adv_adopt	3	6	4.2
kg_sit	2	3	2.1	wheat_adv_weed	4	9	6.2
nc_sit	2	3	2.1	wheat_crop	3	4	1.2
nutrient_mgt	3	5	1.2	wheat_disease	2	8	1.2
Rainfed	4	2	1.1	wheat_seed	2	6	1.1
Rate	6	30	2.4	wheat_water	2	2	1.1
Researcharticals	3	2	2.4	wheat_weed	2	2	1.0
Result	5	26	2.2	Zone	2	2	2.1
rf_type	2	2	2.1	z_crop	2	3	2.1

Table 2.

Table Name	No. of Rows	No. of columns	Size in KB	Table Name	No. of Rows	No. of columns	Size in KB
Brinjal_cure	6	3	16	mango_disease	7	4	16.0
brinjal_symptoms	6	11	16	mango_symptoms	13	2	16.0
Cardio	5	16	2.1	nephro_db	354	2	48
cashew_copt	10	2	16	Neurodatasmell	7	8	3.7
cashew_rt	113	3	16	Neurodynamic	4	8	2.6
cashew_solt	12	2	16	neurogradual1	7	6	3.1
chilli_deccon	12	2	16	Neurohemiplezia	15	8	6.9
Chilli_dlist	12	2	16.0	neurosurgery_brain_head	4	3	16.0
Chilli_otable	15	2	16.0	neurosurgery_brain_tumor	4	3	16.0
Chilli_rlist	106	3	16.0	neurosurgery_brain_vasc	3	3	16.0
Cotton_disease	16	3	16.0	neurosurgery_spine_comp	3	3	16.0
Decfrcyc	7	10	7.8	neurosurgery_spine_infection	3	3	16.0
Dermdata	4	19	1.6	Nose	3	16	3.4
Ear	3	14	3.1	obst_db	56	2	16
horticulture_contact	2	6	16.0	Psycht	3	12	2.8
Incfrecyc	4	7	4.3	Stable	29	8	4.2
infections_table	29	2	16	Symptoms	2	28	2.4
ladyfinger_disease	5	4	16.0	Throat	3	16	3.4
lillydisease_cure	3	2	16.0	tomato_cure	6	3	16.0
lillydisease_symptom	10	2	16.0	tomato_sym	6	10	16.0
lillyvirus_cure	3	3	16.0	turmeric_farm	36	14	16.0
lillyvirus_symptom	33	10	16.0	turm_prevention	35	3	16.0
lily_prevention	33	3	16.0	Urodata	5	19	1.7
				Users	32	7	5.0

Table 3:

Table Name	No. of Rows	No. of columns	Size in KB	Table Name	No. of Rows	No. of columns	Size in KB
ITATT2010	4439	49	512	CEMARKS2010	835	99	384
ITATT2011	4376	49	512	CEMARKS2011	935	99	384
ITATT2012	3724	49	512	CEMARKS2012	1056	99	384
ITMARKS2012	985	99	384	MEATTD2012	4969	49	640
CSMARKS2011	927	99	384	CHEATTD2011	2434	49	256
CSMARKS2012	1180	99	512	CHEATTD2012	2321	49	256
ECATT2010	5048	49	640	CHEMARKS2010	420	99	256
EEMARKS2010	1010	99	384	MCAATTD2010	2304	49	384
EEMARKS2012	1071	99	384	MCAATTD2012	2358	49	384
CEATTD2010	4578	49	512	MCAMARKS2010	552	75	256
CEATTD2011	5229	49	640	MCAMARKS2011	447	75	128
CEATTD2012	4779	49	640	MCAMARKS2012	492	75	256

2.3 Engineering College Datasets:

Atypical data set consisting of the information about the student marks and the attendance of an Engineering college for three consecutive years are collected from the Department of Information Technology, RVR & JC, Guntur and presented in table 3. The number of records is 125124 and the size of the dataset is 13824 KB

3. DATA CONVERSION TOOLS:

Bob Lyons [8] developed a XML conversion tool, namely, Unidex XML converter2.2, which provides B2B integration solutions to its clients in the financial, insurance and telecommunications industries. This converter performs the conversions: (a) flat file to XML document (b) XML document to flat file and (c) flat file to flat file. It runs on any platform for which a Java Virtual Machine is available. Microsoft foundations developed a Windows based graphical transformation engine, namely, RustemSoft XML Converter Standard Edition Converter [9] for creating XML documents from any database. It is designed for 32-bit operating systems and runs only on Microsoft.NET2.0 Framework.

Altova an XML development software company developed Altova MapForce [10] another any-to-any

graphical data mapping, conversion, and integration tool that maps data between any combination of XML, database, flat file, EDI(Electronic Data Interchange), Excel, XBRL(eXtensible Business Reporting Language**)**, and/or Web service and transforms data instantly or auto generates royalty-free data integration code for the execution of recurrent conversions.

The XML Conversion Utility [11] allows you to convert large XML files to delimited text files. The XML Conversion Utility converts an XML file into a pipe (|) delimited output file. This tool is extremely helpful for users who want to import large XML files into a tool that reads delimited text, such as Excel, and to extract reports from XML files.

Boston based eXcelon Corporation proposed a tool named CSV-to-XML Adapter in Stylus Studio package[12], which converts a Coma Separated Value Data file(CSV File) into XML through one-time data conversion operation. Softanic[13] developed Easy XML Converter, which Convert XML to Excel, Text, HTML, Access, XML, support batch mode**.** The software has an information window that displays related data from the different levels of the XML file. Perfect Data Solutions developed a conversion tool 'Access to XML Converter'[14], which converts access database to XML format. Effective XLS to XML converter software scan whole access database and successfully convert to XML file. Ecrion Software Inc developed a tool 'Ecrion Data Architect'[15], which provides a complete solution for integrating data from multiple sources and generate a unified XML representation. It maps data from multiple sources including databases (Oracle, SQL Server, MySQL, etc.), web services, XML files, spreadsheets etc.

The authors proposed new conversion tool [5] that can dynamically generate the XML datasets for the given Relational datasets. It can convert any relational datasets into XML datasets and generates corresponding DTD to validate the XML datasets. It can be used on any operating system which supports Java Virtual Machine (JVM).

3.1. Features of the Proposed XML Conversion Tool:

The proposed tool is a graphical interface that can dynamically generate the xml datasets for the given Relational datasets. This has the following features.

- It can convert dynamically any relational datasets into XML datasets.
- Can use this tool on any operating system which supports JVM.
- It generates the XML file and corresponding DTD for the generated XML file.
- It can validate the generated xml file.
- It minimize the of the size of the generated XML data.
- It can display the RDB datasets in a tabular format and XML datasets in a tree structure format.
- It can sort the data on a particular field.

3.2. Features of the RustemSoft XML Converter Standard Edition:

Using this software, enterprises can share data with customers, suppliers or partners as well as distribute data across business units. This tool has the following features.

- It can produce static XML documents.
- It is designed for 32-bit operating systems and runs only on Microsoft.NET2.0 Framework.
- Because it is developed on Microsoft.NET it runs on Microsoft Windows family.
- It is not portable. i.e. you cannot run this application on any platform.

4. IMPLEMENTATION OF DATA CONVERSION TOOLS TO SOME TYPICAL DATASETS:

Three experiments are conducted to implement data Conversion tools mentioned in Section 3 to generate the corresponding XML datasets. The first experiment is conducted for some of the datasets available on India Kisan application database server. The second experiment is conducted for some of the datasets available on Gramarogya application database server. The third experiment is conducted for some of the datasets available on Engineering College database server. The results are presented in table 3, table 4 and table 5 respectively. In these tables the root element represents the table name and each row represents one sub-element of the root element. This sub-element name is treated as record, i.e., for each **row** of the table there exists one element of type record in the XML data set. The columns of the tables are treated as child elements of the element record. The first column represents the Name of the dataset, second column represents the corresponding access time in relational dataset format, third column represents the access time of the corresponding XML dataset format and fourth column gives the differences between the access time in milliseconds. The value of the child element is the corresponding value of the relational data set. The schema of the relational data set represents the Document Type Definition of the XML data set.

4.1. Conversion of Indiakisan RDB Datasets into XML Datasets:

In the first experiment the proposed conversion tool is applied to generate the XML datasets corresponding to the relational datasets for the indiakisan application. The results are presented in the table 4. The access time in milliseconds for each dataset is shown separately in the table4.

Table :4

Table Name	Relational Datasets Access time (milliseconds)	Xml datasets Access time (milliseconds)	Difference in access time (milli seconds)	Table Name	Relational Datasets Access time (milli seconds)	Xml datasets Access time (milli seconds)	Difference in access time (milli seconds)
advisory_crop	339225516	254655550	84569966	rice_adv_fert	306482527	161645125	144837402
advisory_crops	324364410	168854980	155509430	rice_adv_fertsub	358210052	202537752	155672300
advisory_cropsuggestion	351613962	189325459	162288503	rice_adv_water	343921923	200193320	200193320
advisory_districts	302522527	159078598	143443929	rice_rain_disease	495157066	346851625	148305441
advisory_irrigation	318144065	159145646	158998419	rice_rain_pest	357460236	183562156	173898080
advisory_months	333299623	166357456	166942167	rice_ricecharacteristics	575699705	451866368	123833337
Bdensity	314884154	178321547	136562607	rice_seasons	366340974	221980167	144360807
biodiesel_tb_economics	311353818	164665341	146688477	rice_seed	348005403	187033827	160971576
biodiesel_tb_fertilizer	324138962	160732998	163405964	rice_soiltypes	338862341	173985521	164876820
biodiesel_tb_functional_uses	318103279	167316517	150786762	rice_transplantation	362905620	219498847	143406773
crop_and_oil	341690354	164413075	177277279	surf_spec	306530579	170694040	135836539
crop_fert	840907993	735603826	105304167	swtools_brix_values	429067382	374106485	54960897
crop_soil	341690354	164413075	177277279	swtools_pol_values	716869882	699999301	16870581
Dist	307054109	163490611	143563498	tobacco_adv_adop	304165753	177085076	127080677
drip_state	307176192	161290053	14588613	tobacco_adv_variety	336502545	187342246	149160299
Items	307341855	267666980	39674875	water_eff	294979110	165873875	129105235
j_alt_irr	335506887	188069434	147437453	Wdepth	303230718	18684134	116389373
nc_sit	320727076	163483627	157243449	wheat_crop	313611367	164243780	149367587
nutrient_mgt	303047454	169363983	133683471	wheat_disease	314930529	163550395	151380134
Onlineuser	345338025	211754846	133583179	wheat_fert	318830466	160762890	158067576
Organic	321202275	184446068	136756207	wheat_irr_crop	352719689	177372264	175347425
org_goals	299932812	164004922	135927890	wheat_irr_fert	331915090	187163732	144751358
org_result	337857186	188157992	149699194	wheat_irr_land	339978963	167353113	172625850
org_sub1	332452867	184079821	148373046	wheat_irr_water	330028258	171914587	158113671
org_sub2	301125702	173009699	128116003	wheat_land	292428228	25487541	37552818
org_sub3	315521386	160198014	155323372	wheat_pest	301852889	162385723	139467166
pest_mgt	311164407	160303055	150861352	wheat_raindisease	403714159	219463368	184250791
pest_tab_crops	292477675	162650561	129827114	wheat_rain_pest	421215546	259353353	161862193
pest_tab_crop_defects	495838438	495838438	115370809	wheat_rain_seed	351408070	184581839	166826231
Rainfed	299673282	163250358	136422924	wheat_seed	283266170	162058586	121207584
Rate	338537719	205367162	133170557	wheat_water	288915770	158584680	130331090
Researcharticals	324341502	165215691	159125811	wheat_weed	300470590	159707448	140763142
Result	332952652	178473522	154479130	Zone	318909526	161115729	157793797
rf_type	307151049	162748897	144402152	z_crop	305157220	155906686	149250534

4.2. Conversion of Gramarogya RDB Datasets in to XML Datasets:

In the second experiment the proposed conversion tool is applied to generate the XML datasets corresponding to the relational datasets for the Gramarogya application. The results are presented in the table 5. The access time for each dataset is shown separately in the table5.

Table : 5

Table Name	Relational Datasets Access time (milli seconds)	Xml Datasets Access time (milli seconds)	Difference in access time (milli seconds)	Table Name	Relational Datasets Access time (milli seconds)	Xml Datasets Access time (milli seconds)	Difference in access time (milli seconds)
Brinjal_cure	307071150	164125608	142945542	mango_disease	325867952	160373455	165494497
brinjal_symptoms	308643976	168598964	140045012	mango_symptoms	309958109	161735919	148222190
Cardio	302666680	177420874	125245806	nephro_db	332427724	161236973	171190751
cashew_copt	320006593	163627500	156379093	Neurodatasmell	326092841	163646777	162446064
Ear	316694719	156506204	160188515	obst_db	379633724	204058616	175575108
Ees	324138962	160732998	163405964	Onco	323553691	170302091	153251600
exotic_disease	320434022	160436592	159997430	ongole_disease	419755583	289534564	130221019
garlic_disease	357022191	171542473	185479718	Oph	307365601	158337442	149028159
homeo_db	358613735	175262498	183351237	Pharma	359996312	177312200	182684112
horticulture_contact	328989019	165402865	163586154	Psycht	310561817	159822268	150739549
horticulture_userinfo	318103279	167316517	150786762	rose_prevention	328116562	164534040	163582522
horticulture_views	328905489	183184176	145721313	rose_symptoms	326548486	162703361	163845125
Incfrecyc	312830262	179368887	133461375	Stable	341690354	164413075	177277279
infections_table	307054109	163490611	143563498	Symptoms	303274020	161837608	141436412
ladyfinger_disease	354118750	197477333	156641417	Throat	302720877	161889850	140831027
lillydisease_cure	335585110	181952455	153632655	tomato_cure	326894061	326894061	167658739
lillydisease_symptom	307176192	161290053	14588613	tomato_sym	303729105	164367538	139361567
lillyvirus_cure	318860079	176358727	142501352	turmeric_farm	341626380	158937798	182688582

Table : 6

Table Name	Relational Datasets Access time (milli seconds)	Xml datasets Access time (milli seconds)	Difference in access time (milli seconds)	Table Name	Relational Datasets Access time (milli seconds)	Xml datasets Access time (milli seconds)	Difference in access time (milli seconds)
ITATT2010	2452194902	1714616522	737578380	CEMARKS2010	1168794358	1110285246	58509112
ITATT2011	2422078783	1743347828	678730955	CEMARKS2011	1243049326	1149675727	93373599
ITATT2012	2151860920	1414937652	736923268	CEMARKS2012	1338689135	1279908759	58780376
ITMARKS2010	1225736791	1121311508	104425283	MEATTD2010	2561580033	1733826226	827753807
CSMARKS2012	1541887587	1405455721	136431866	CHEATTD2012	1467031425	1021245437	445785988
ECMARKS2012	1534416525	1490793104	43623421	BTATTD2012	956037150	400915200	555121950
EEATT2010	2821280917	1903610375	917670542	BTMARKS2010	562958421	365666307	197292114
EEATT2011	2561663842	1791865726	769798116	BTMARKS2011	489922881	336279611	153643270
EEATT2012	2221131736	1607491176	613640560	BTMARKS2012	386609751	208588243	178021508
CEATTD2010	2451137226	1756217341	694919885	MCAMARKS2010	677372403	483777966	193594437
CEATTD2011	2781396239	1936626303	844769936	MCAMARKS2011	667623678	404897829	262725849
CEATTD2012	2534185769	1821643812	712541957	MCAMARKS2012	627538696	436802164	190736532

4.3. Conversion of Engineering College RDB Datasets in to XML Datasets:

In the third experiment, the proposed conversion tool is applied to generate the XML datasets corresponding to the relational datasets for the college database. The results are presented in the table 6. The access time for each dataset is shown separately in the table6.

5. PERFORMANCE ANALYSIS:

In this paper an application program is also developed in Java to evaluate the access time for the given datasets. The inputs for this application program are relational data set and XML data set. The same type of data and the same number of fields are maintained in relational data set and XML data set.

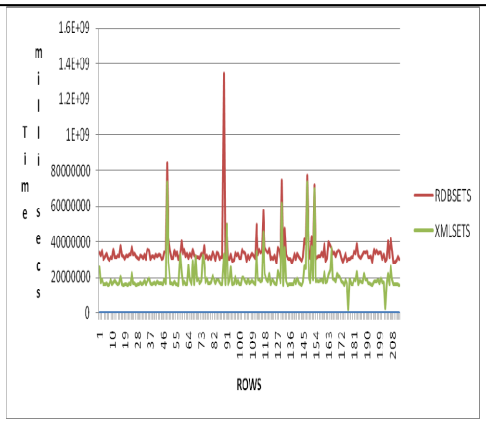
This application program will read the data from the relational dataset as well as from corresponding XML dataset which is already generated by the conversion tool. The output of this application program is the access time to process the data from each dataset. The access time for each dataset is presented in table4, table5 and table6 respectively.

5.1 Performance Analysis on Indiakisan Datasets:

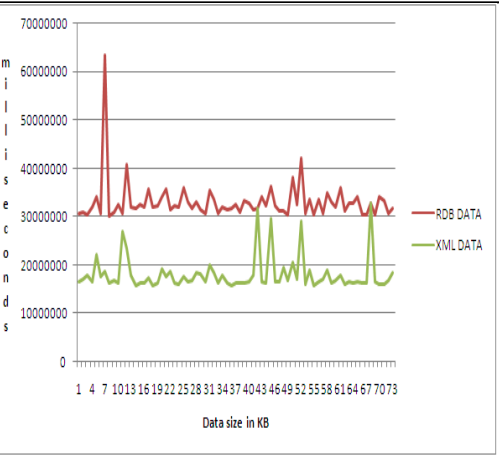
Access times of Indiakissan Datasets in RDB format and XML format are given in the table 4. The performance of the proposed tool with respect to access time is as shown in the graph 1. In this graph '~'line indicates access time on RDB datasets and '~' line indicates access time on XML datasets. From the graph, it can be noted that, the time line to access the XML datasets is always below to the time line to access the relational datasets and hence it can be concluded that it takes less time to access the XML datasets than Relational Datasets.

5.2 Performance Analysis on Gramarogya Datasets.

Access time of Gramarogya Datasets in RDB format and XML format are given in the table 5. The performance of the proposed tool with respect to access time is as shown in the graph 2. In graph '~' line indicates access time on RDB datasets and '~' line indicates access time on XML datasets. From graph1, it can be noted that, the time line to access the XML datasets is always below to the time line to access the relational datasets and hence it can be concluded that it takes less time to access the XML datasets than Relational Datasets.



Graph :1 Performance analysis on Indiakisan Datasets

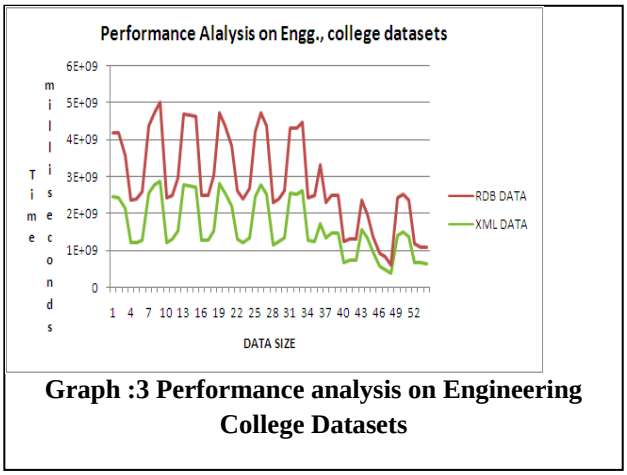


Graph :2 Performance analysis on Gramarogya datasets.

5.3 Performance Analysis Engineering College Datasets:

Access time of Engineering College Datasets in RDB format and XML format are given in the table 6. The performance of the proposed tool with respect to access

time is as shown in the graph 3. In graph ~ line indicates access time on RDB datasets and ~ line indicates access time on XML datasets. From graph3, it can be noted that, the time line to access the XML datasets is always below to the time line to access the relational datasets and hence it can be concluded that it takes less time to access the XML datasets than Relational Datasets.



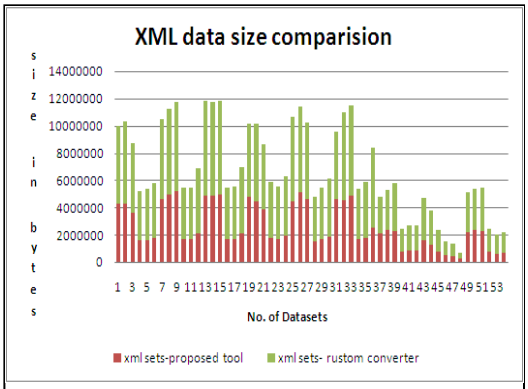
Graph :3 Performance analysis on Engineering College Datasets

6. Proposed Tool Comparison with RustomSoft XML Converter:

The Rustom Soft XML converter is applied on Engineering College datasets to convert relational datasets to corresponding XML datasets. It is also applied to generate the DTD structure for the generated XML datasets. The size of the generated XML datasets and DTD structures using proposed tool and Rustom Soft XML converter are presented in table7. First column represents serial number, the second column represents the Name of the dataset, third column represents the number of datasets, fourth column represents size of the XML dataset by using proposed tool, fifth column represents size of the XML dataset by using Rustom Soft conversion tool, sixth column represents size of the DTD by using proposed tool, and seventh column represents size of the DTD by using Rustom conversion tool.

Table 7: Comparison of XML File Size and DTD File Size After Conversion by Proposed Tool and Rustomsoft XML Converter on Engineering Students Datasets.

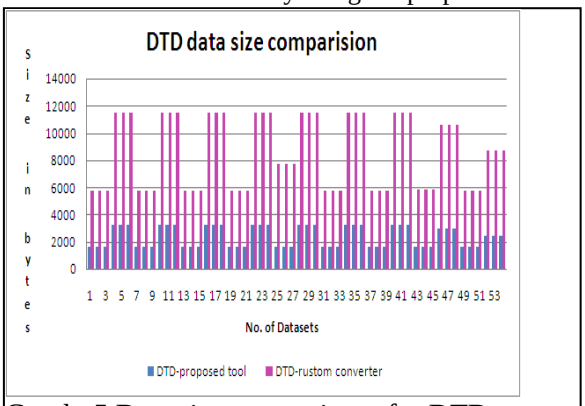
S.No	Table Name	NO. of data sets	Converted XML file size in bytes		Converted DTD file size in bytes	
			Using proposed conversion tool	Using Rustom converion tool	Using proposed conversion tool	Using Rustom converion tool
1.	ITATT2010	4439	4310971	5684318	1633	5762
2.	ITATT2011	4376	4252081	6017516	1633	5762
3.	ITATT2012	3724	3617161	5096031	1633	5762
4.	ITMARKS2010	902	1606126	3612451	3193	11498
5.	CSATT2010	4769	4630533	5884025	1633	5762
6.	CSATT2011	5121	4972016	6238713	1633	5762
7.	CSATT2012	5376	5219422	6486003	1633	5762
8.	CSMARKS2010	945	1682925	3787620	3193	11498
9.	CSMARKS2011	927	1642839	3770904	3193	11498
10.	CSMARKS2012	1180	2103193	4793284	3193	11498
11.	EEMARKS2010	1010	1797870	4060556	3193	11498
12.	EEMARKS2011	952	1696396	3832094	3193	11498
13.	CHEATTD2012	2321	2253606	3536239	1633	5761
14.	CHEMARKS2010	420	746980	1686683	3193	11497
15.	CHEMARKS2011	462	816114	1881882	3193	11497
16.	CHEMARKS2012	459	815089	1866859	3193	11497
17.	BTATTD2010	1765	1592375	3101990	1633	5811
18.	BTATTD2011	1405	1271921	2479524	1633	5811
19.	BTATTD2012	870	785643	1530657	1633	5811
20.	BTMARKS2010	282	468526	1063312	2953	10585
21.	BTMARKS2011	244	403545	928677	2953	10585
22.	BTMARKS2012	127	210192	482880	2953	10585
23.	MCAMARKS2011	447	614591	1358346	2441	8746
24.	MCAMARKS2012	492	678236	1483658	2441	8746



Graph: 4 Data size comparisons for XML

The data size comparisons for XML datasets and corresponding DTD structures are as shown in the Graph:4 and Graph:5 respectively. In graph4 ▬ line indicates XML data size using proposed tool and ▬ line indicates XML data size using Rustom conversion tool. In graph 5 ▬ line indicates DTD size using proposed tool and ▬ line indicates DTD size using Rustom conversion tool.

From graph4 and graph 5, it can be concluded that the size of data, converted by using Rustom Soft XML converter is more than the data converted by using the proposed tool.



Graph: 5 Data size comparisons for DTD

7. CONCLUSIONS:

Performance analysis of the three data sets presented in the above experiments indicates that the efficiency of the system with respect to data access time is less when the data maintained in XML format than RDB. All the Applications which will process only text data can directly interact with XML datasets. As per W3C recommendations, in Semantic web applications the data will be represented in RDF[16][17] format which follows the XML syntax. So the XML datasets can be used in the Semantic web systems more effectively.

The disadvantage of maintaining the text information in RDB is that the application program has to identify the database server and again the database within that server for each query. It means that the application needs two operations to extract information from the relational database which is a time consuming process. But when data is represented in an XML format, the XML data and the application program can be maintained within the same root. Therefore to extract the information from the XML data the application needs only one operation i.e. identifying the XML data. Consequently in web applications it is better to use XML datasets than RDB datasets. The performance of all the dynamic web services depends up on the access time. By minimizing the access time we can increase the performance of a system.

REFERENCES:

[1]. Dr. Peter Chen, Dr. Mladen Vouk , "XML and the Semantic Web: What is the future?", Proceedings of the IEEE 36th Hawaii International Conference on System Sciences (HICSS'03), 6-9 Jan. 2003, pages 123-124.

[2]. Zhaoyan Jin, Quanyuan Wu, Hua Fan, "Restructuring XML Document --from the Perspective of Sorting" IEEE 3rd International Conference on Communication Software and Networks (ICCSN), Xi'an , 27-29 May 2011, Pags 529-533..

[3]. B. Hemanth Kumar, M. Surendra Prasad Babu, "An Implementation of Semantic Web System for Information retrieval using J2EE Technologies" (IJCSE) International Journal on Computer Science and Engineering, Vol. 3 (6), 2011,2498-2505.

[4]. B. Hemanth Kumar, M. Surendra Prasad Babu, "Design and Development of a New JAVA Based Conversion Tool from Relational Database toXML Datasets" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (6), 2012,5439-5446.

[5]. www.Indiakisan.net

[6]. www.bharathgramaarogya.net .

[7]. B2B Integration Solutions from Unidex: http://www.unidex.com/xflat.htm.

[8]. Rustem Soft XML Converter Standard Edition: http://www.rustemsoft.com/

[9]. http://www.ecrion.com/landingpage/databasetoxml.aspx.

[10]. Stefan Decker, Prasenjit Matura, Sergey Melnik, "Framework for the Semantic Web: An RDF Tutorial " IEEE INTERNET COMPUTING, IEEE Computer Society, November-December 2000, pages 68-73, http://computer.org/internet

[11]. Ora Lassila, Ralph R. Swick, "Resource Description Framework (RDF) Model and Syntax Specification" W3C recommendation, Feb. 1999, http://www.w3.org/TR/1999/REC-Rdf-Syntax-19990222

[12]. Chen Zhifeng, "Research on the Application of XML in Web Services", 2009 IEEE International Conference on Multimedia Information Networking and Security, Wuhan, China, Vol:1, pages 680-683.

[13]. Ranjit Bose, "Applying Semantic Web Services to Web-based Information Systems and Applications", Proc. of Fifth IEEE International Conference on Information Technology: New Generations, 2008, Las Vegas, Nevada, USA. Pages 1150-1155.

[14]. T. Bray, J. Paoli, C.M.Sperberg, McQueen, E. Maler, "XML 1.0 Extensible Markup Language (XML) 1.0 (Second Edition)", W3C Recommendation, 6 October 2000,

[15]. Martin Hep, "Semantic Web and Semantic Web Services Father and Son or Indivisible Twins?" IEEE Internet Computing, Vol: 10, No: 2, March-April 2006 pages 85-88.

[16]. Tim Berners-Lee, James Hendler, Ora Lassila, "The Semantic Web", Scientific American Degital Magazine , Vol:284, No:5, May 2001, pages 29-37.

[17]. Wei Zhang and Robert A. van Engelen "An Adaptive XML Parser for Developing High-Performance Web Services" , IEEE Fourth International Conference on eScience, 7-12 Dec. 2008, Indianapolis, IN,pages 672 - 679.

[18]. Silberschatz, Henry F. Korth, "Database System Concepts ", McGraw-Hill.

[19]. Brett McLaughlin, "Enterprise Applications in Java and J2EE", O'Reilly Media, 07-01-2000.

[20]. http://www.oracle.com/technetwork/java/javase/jdbc/index.html.

[21]. XML Schema Part 0: Primer, W3C Working Draft,

[22]. Zhaoyan Jin, Quanyuan Wu, Hua Fan " Restructuring XML Document -from the Perspective of Sorting" 978-1-61284-486-2/111$26.00 ©2011 IEEE pages 529-5.