ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

AN EFFICIENT EFFECTIVE DYNAMIC XML DATA BROADCASTING METHOD IN MOBILE WIRELESS NETWORK USING XPATH QUERIES

Mr. A.Lakshmikandan
Lecturer ,School of CS & IT, DMI- St. John the Baptist University, Malawi

DOI: 10.5281/zenodo.400945

ABSTRACT

The main aim of this paper is to improve energy and latency efficiency of XML dissemination scheme for the mobile computing. It is based on Lineage Encoding using a novel unit structure called G-node for streaming XML data in the wireless environment. It exploits the benefits of the structure indexing and attribute summarization which integrates relevant XML elements into a group. It provides a way for selective access of their attribute values in a dynamic way where broadcasting can be done dynamically supporting Twig Pattern Queries.

KEYWORDS: Twig pattern matching, Wireless Broadcast, XML Streaming

INTRODUCTION

Data mining is the process to extract the data from many resources. Data mining uses to analyze the past data based on particular problem or situation. Data warehouse contains the storage of data in database where the collection of various data called as data warehouse. Data mining is the process of discovery on new knowledge that particular data may come from all parts of business, from the production to the management. Managers also use data mining to decide upon marketing strategies for their product.

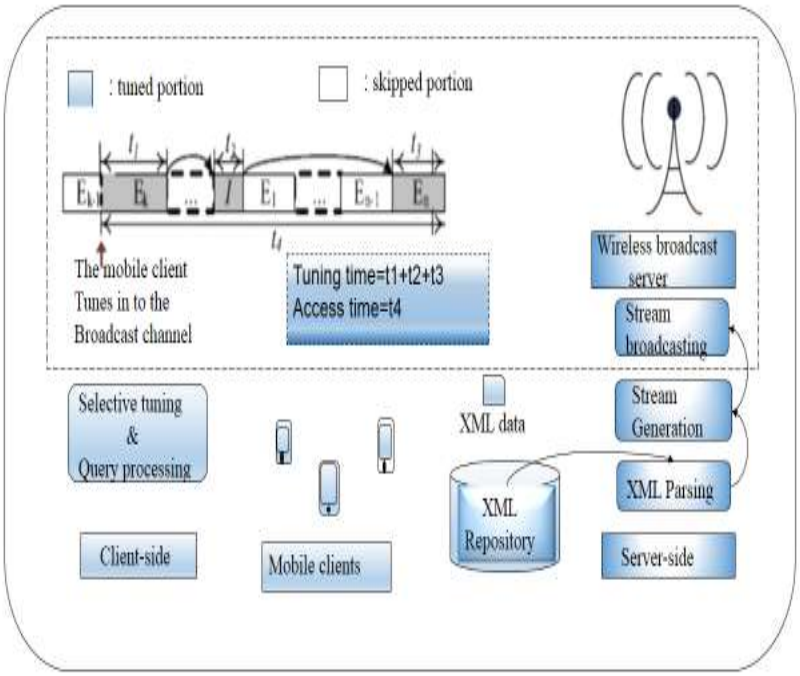
Managers can use data to compare and contrast among competitors. Data mining interprets its data into real time analysis that can be used to increase sales, promote new product, or delete the product that is not value-added to the company.

The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Many people take data mining as a synonym for another popular term, knowledge discovery in database (kdd). Alternatively other people treat data mining as the core process of kdd. Usually there are three processes. One is called preprocessing which is executed before data mining techniques are applied to the right data. The preprocessing also includes data cleaning, integration, selection and transformation. After that comes another process which is called post processing. This evaluates the mining result according to user's requirements and domain knowledge.

Mobile computing fetches an efficient XML dissemination scheme for the energy and latency concepts in the electronic devices such as Laptop,ipod, Tab, Android mobile and in all the devices. Inconsequence of receiving the data in sequence of streaming information, this paper aims to develop the conventional XML query processing into a binary format. Also it feasibly shrinks out the structural tree of indexing data into shortened tree format. By utilizing the XPath query in conventional XML query processing, users can get the data frequently, especially in downloading information from website either in Laptop or in mobile devices. Thereby twig pattern queries are used under the title of LineageEncoding that encompasses the lightweight and effective encoding scheme.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

Figure:



Multi-cloud Architecture

The users define a novel unit naming as G-node for sequencing XML data in the wireless environment. Hence this project paves a way for the twig pattern queries which are namely Data mining and XML mining. Therefore, it is a boon feeding technology for the users to facilitate the required information of delivering information.

Withthe rapid development of wireless network technologies, wireless mobile computing has become popular. Users communicate in the wireless mobile environment so as to reduce the latency while downloading the pages from internet. Therefore, developers define a novel unit structure called G-node for streaming XML data in the wireless environment. It exploits the benefits of the structure indexing and attributes summarization that can integrate relevant XML elements into a group. It provides a way for selective access of their attribute values and text content.

Users also propose a lightweight and effective encoding scheme, called Lineage Encoding, to support evaluation of predicates and twig pattern queries over the stream. Users need to consider energy conservation of mobile clients when disseminating data in the wireless mobile environment, because they use mobile devices with limited battery-power (i.e., energy-efficiency). The overall query processing time must also be minimized to provide fast response to the users. The goals of conventional query processing on streamed XML data are to minimize computation costs and filtering time.

LITERATURE REVIEW

I. WIRELESS XML STREAM GENERATION ALGORITHM

Input: A well formed XML document D
Output:Wireless XML Stream XS
Steps:

1. **Insert the attributes for current element into the content handler**
2. **Generate the stream generator**
3. **Constructs the element as *Start Element* and *End Element* for reducing the time task for downloading pages in the internet.**

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

4. **Apply the comparison of elements in lineage coding**
5. **Construct Lineage codes based on (V, H) the document parsing.**
6. **Use the client to predict the downloaded page with necessary information in downloading.**

Decision tree classification

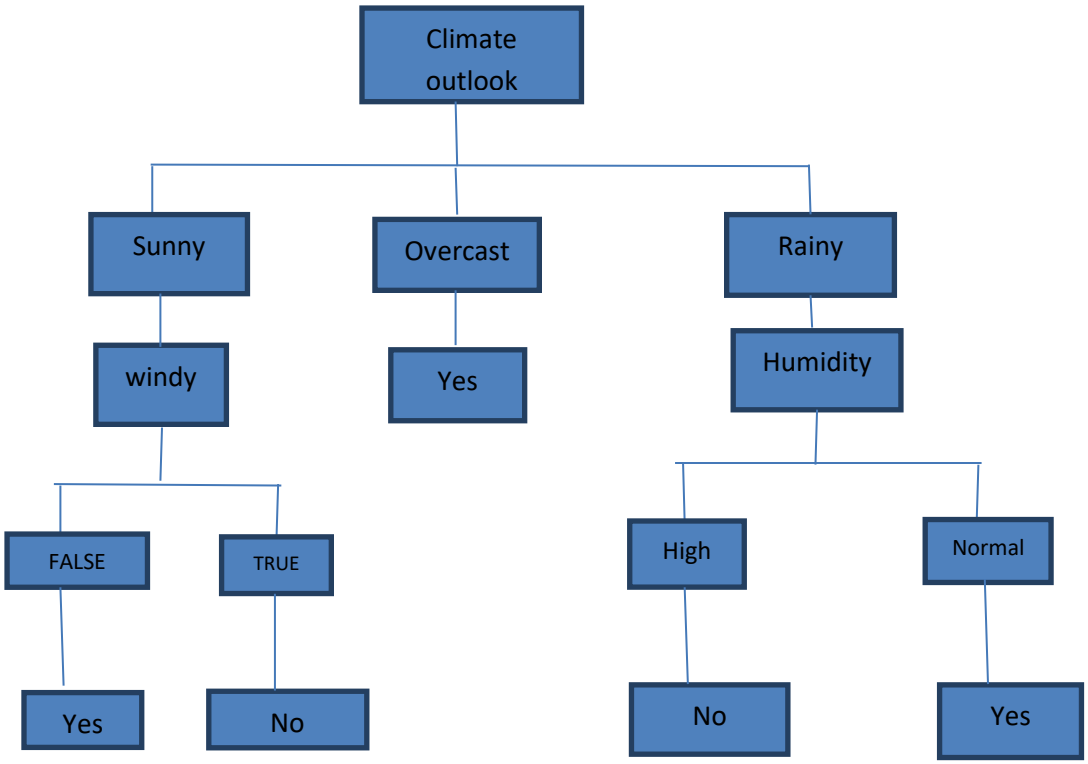
Decision tree builds the tree structure under the classification of regression models. It splits the dataset into subset through an associated decision tree in its incremental development. Decision nodes and leaf nodes are the final result in the decision tree.

ID3 Algorithm

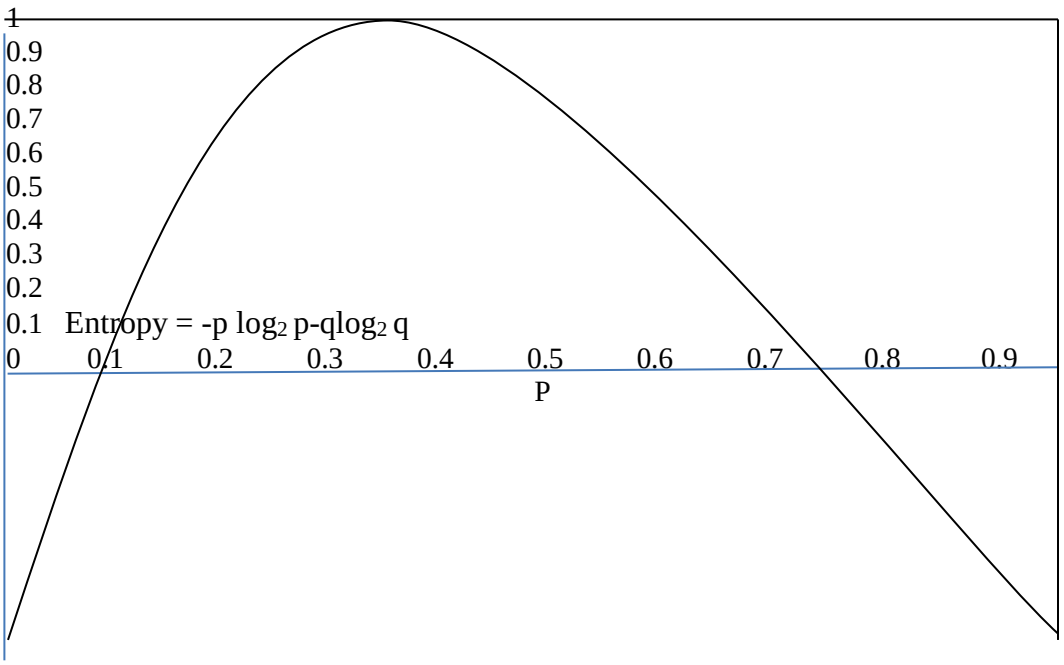
In decision tree learning, **ID3** means **Iterative Dichotomiser 3. ID3** is an algorithm invented by Ross Quinlan · It used to generate a decision treefrom a dataset. ID3 is the precursor to the C4.5 algorithm which is typically used in the machine learning and natural language processingdomains.

ID3 Algorithm applies the top down nodes in greedy search through possible branches using no backtracking. ID3 uses Entropy and Information Gain to construct a decision tree.

ID3 Algorithm entropy calculates the sample homogeneity. Therefore if sample is entirely homogeneous, then the entropy is zero. And if sample is an equally divided, the entropy value is one.



ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7



The developers have calculated two types of entropy using frequency tables as follows

a) Entropy uses the frequency table of one attribute

$$E(S) = \sum_{i=1}^{c} - \text{Pi} \log_2 Pi$$

Play cricket	
Yes	No
9	5

Entropy(PlayCricket)=entropy(5,9)

=Entropy(0.36,0.64)

= $-(0.36 \log_2 0.36) - (0.64 \log_2 0.64)$

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

b) Entropy uses the frequency table of two attributes

$$E(T,X) = \sum_{c \in X} P(c)E(c)$$

		Play Cricket		
		yes	No	
Outlook	Sunny	3	2	5
	Overcast	4	0	4
	Rainy	2	3	5
				14

E(Play Cricket ,Outlook) = P(Sunny)*E(3,2)+P(Overcast)*E(4,0)+P(Rainy)*E(2,3)

= (5/14)*0.971+(4/14)*0.0+(5/14)*0.971

= 0.693

Information Gain

Once the dataset is segregated on an attribute, information gain is based on after the decreasing of entropy.Decision tree is constructed to find all attribute which returns the highest information gain. It is the most homogeneous branches.

Step1: Calculate the targeted entropy

Entropy (Play Cricket) = Entropy (5, 9)

= Entropy (0.36, 0.64)

= - (0.36$\log_2$ 0.36)-(0.64 $\log_2$ 0.64)

= 0.94

Step2: the dataset is divided on different attributes

- The entropy is calculated for each branch. So to obtain the total entropy on split, it is added proportionally.
- As a result, the information gain receives the decreased entropy with subtracted results.**

		Play Cricket	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3
Gain = 0.247			

		Play Cricket	
		Yes	No
Humidity	High	3	4
	Normal	6	1
Gain =0.152			

		Play Cricket	
		Yes	No
Temp.	Hot	2	2
	Mild	4	2
	Cool	3	1
Gain = 0.029			

	Play Cricket

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

		Yes	No
Windy	False	3	4
	True	6	1
Gain = 0.048			

Gain (T, X)= Entropy (T)-Entropy(T,X)

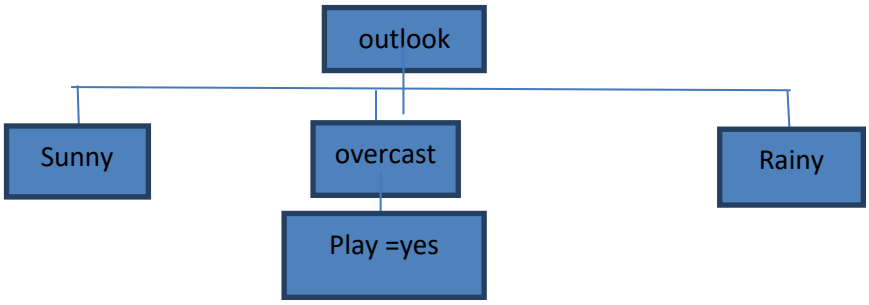
G(Play Cricket, Outlook)= E(playCricket,outlook)
= 0.940-0.693=0.247

Step 3: decision node has the largest information gain as the chosen attribute

		Play Cricket	
		Yes	No
outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3
Gain = 0.247			

Step 4a: A branch with entropy of 0 is a leaf node.

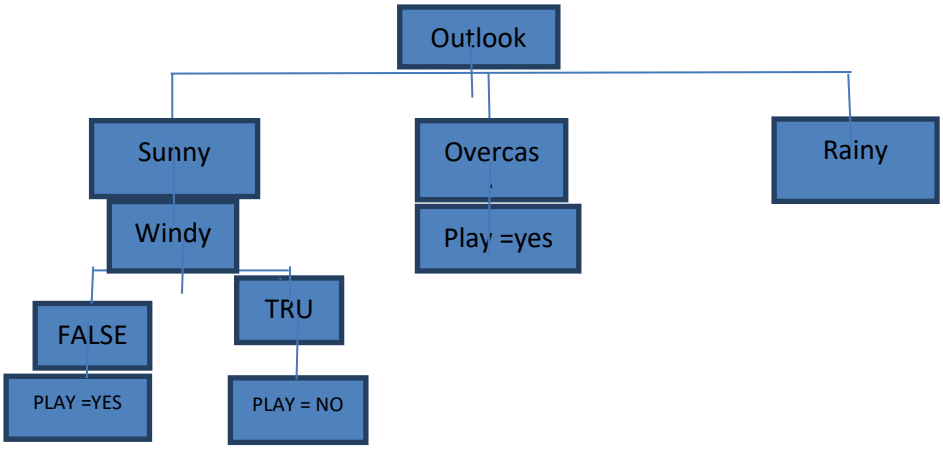
Temperature	Humidity	Windy	Play Cricket
Hot	High	FALSE	Yes
Cool	Normal	TRUE	Yes
Mild	High	TRUE	Yes
Hot	Normal	FALSE	Yes
Hot	High	FALSE	Yes



Step 4b: A branch with entropy more than 0 needs further splitting.

Temp	Humidity	Windy	Play Cricket
Mild	High	FALSE	Yes
Cold	Normal	FALSE	Yes
Mild	Normal	FALSE	Yes
Cool	Normal	TRUE	No
Mild	High	TRUE	No

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

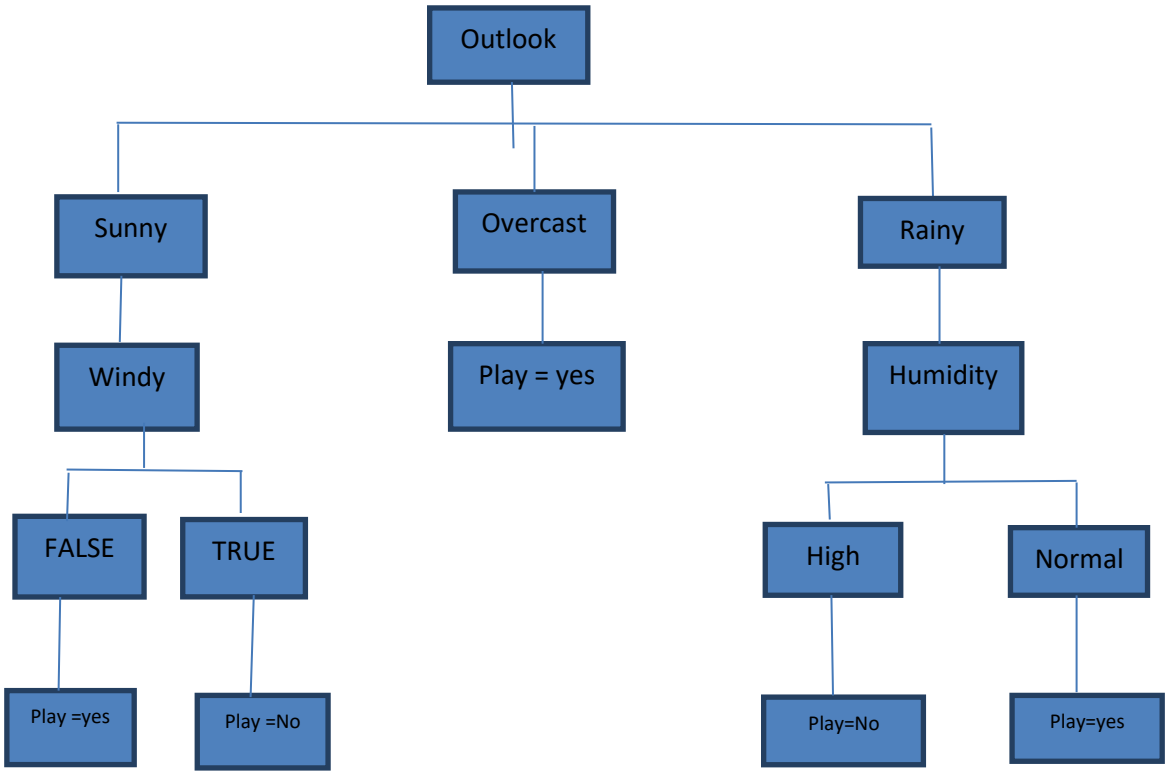


Step 5: the ID3 algorithm is running recursively on the non-leaf branches, until all data is classified

Decision Tree to Decision Rules

A decision tree easily transforms the set of rules in mapping from the root node to leaf node one by one.

R_1: IF (Outlook=Sunny) AND (Windy = FALSE) THEN Play=Yes
R_2 :IF (Outlook =Sunny)AND (Windy = TRUE) THEN Play=No
R_3 : IF(Outlook=Overcast)THEN Play=Yes
R_4 : IF(Outlook =Rainy) AND (Humidity =High) THEN Play =No
R_5 : IF(Outlook = Rain)AND (Humidity = Normal)THEN Play= Yes



Decision trees- Issues

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

- They are working with continuous attributes(binging)
- They are avoiding over fitting
- Super attributes have an attributes with many values.
- It is working with missing values.

METHODOLOGY

1. XML DATA & MANIPULATION

An XML document can be represented as a rooted, ordered, and labeled tree. Nodes represent elements, attributes, texts, and the parent-child relationships are represented by edges in the XML tree. It shows a simple XML document that will be used as a running example in the paper.

A server retrieves an XML document to be disseminated from the XML repository and it generates wireless XML stream by using SAX (Simple API for XML), which is an event-driven API. SAX invokes content handlers during the parsing of an XML document.

Structured Indexing approach integrate multiple elements of the same path into one node, thus, the size of data stream can be reduced by eliminating redundant tag names thereby enabling Twig Pattern Query Processing.

2. LINEAGE ENCODING

Lineage Encoding used to support queries which involve predicates and twig pattern matching. In the proposed scheme, two kinds of lineage codes, i.e., vertical code denoted by Lineage Code (V) and horizontal code denoted by Lineage Code (H), are used to represent parent-child relationships among XML elements in two G-nodes.

We also define relevant operators and functions that exploit bit-wise operations on the lineage codes. To the best of our knowledge, our scheme is the first wireless XML streaming approach that completely supports twig pattern query processing in the wireless broadcast environment.

3. ATTRIBUTE SUMMARIZATION

The Attribute Value List (AVL) generated in Attribute Summarization with lineage-encoded data is the key to process the Twig Pattern Queries in Selective tuning approach in the mobile end. In XML, an element may have multiple attributes, each of which consists of a name and value pair; there is structural characteristic that elements with the same tag name and location path often contain the attributes of the same name. Attribute summarization eliminates repetitive attribute names in a set of elements when generating a stream of G-nodes.

4. G-NODE & XML DISSEMINATION

We define a streaming unit of a wireless XML stream, called G-node. The G-node structure eliminates structural overheads of XML documents, and enables mobile clients to skip downloading of irrelevant data during query processing.

The group descriptor is a collection of indices for selective access of a wireless XML stream. Node name is the tag name of integrated elements, and Location path is an XPath expression of integrated elements from the root node to the element node in the document tree. Child Index (CI) is a set of addresses that point to the starting positions of child G-nodes in the wireless XML stream. Attribute Index (AI) contains the pairs of attribute name and address to the starting position of thevalues of the attribute that are stored continuously in Attribute Value List.

The components of the group descriptor are used to process XML queries in the mobile client efficiently. Specifically, Node name and Location path are used to identify G-nodes. Indices relating to time information such as CI, AI, and TI are used to selectively download the next G-nodes, attribute values, and text. Finally, Lineage Code(V, H) is used to handle axis and predicate conditions in the user's query & Attribute Value List (AVL store attribute values of the elements represented by the G-node, respectively .All the G-Node data's are Broadcasted with the help of a Wifi device which can be received by any android devices in its coverage.

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

5. TREE FORMATION & SELECTIVE TUNING

In this section, we describe how a mobile client can retrieve the data of its interests. Assuming that there is no descendant axis in the user query, query processing algorithms for a simple path query and a twig pattern query are presented.

I. Simple Path Query Processing, Algorithm shows the simple path query processing over the wireless XML stream. Given a query, the mobile client constructs a query tree. Then, it starts to find relevant G-nodes over the wireless XML stream. The mobile client downloads a group descriptor of the G-node which corresponds to the query node. If the current node is the leaf node, the mobile client downloads

1. Twig Pattern Query Processing, query over the proposed wireless XML stream. In the Tree traversal phase, the mobile client first constructs a query tree. Then, traversing the query tree in a depth-first manner, it selectively downloads group descriptors ofthe relevant G-nodes into the nodes in the query tree.
2. Our Selective tuning approach is dynamic and it eases the client to minimize the tuning time and thereby reducing access time. It dynamically chooses between the Twig Pattern Query and Normal Query and process to render the data.
3. Tuning is optimized with the help of the XPath Query pattern that holds the predicates.

6 .DYNAMIC DATA PROCESSING

Our XML Automation tool, used for customized XML creation , enables the server to tobroadcast the customized data as and when needed without relying on the third party for XML files. Our Implementation support to dynamic customized XML is a major advantage of the wireless streaming in mobile environment.

In this paper, we use XPath as a query language. A location path selects the results of an XPath query. A location path consists of location steps. Processing each location step selects a set of nodes in the document tree that satisfy axis, node test and predicates described.

A GNode the novel attribute of our system can be added dynamically to the broadcast channel without interrupting the streaming of XML data. This feature enables to dynamically add events in the existing channel.
Dynamic addition of GNode ensures the credibility of the Broadcast system efficiently proposed by our approach.
AVL tree and Structured Indexing process will be handled that will probably affect the XML document in temporary buffer .

Dynamic modification of Attribute value enables to change any data on the broadcast stream whenever needed and is achieved by the Attribute summarization mechanisms and the Structured Indexing of XML data handled in our system.

IV.FLOW DIAGRAM

The flow diagram represents the stages involved in reducing the task time of downloading pages into lineage encoding.

Reducing latency is the main goal of this project. The system can remove the unwanted pages in the URL while downloading the pages from internet. Therefore, the energy consumption is done through android mobile for saving energy through solar energy concept. Also it encompasses the twig pattern queries which reserve the discovering new knowledge into data mining. Lineage encoding is used to support queries involving predicates and twig pattern matching

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

Figure:

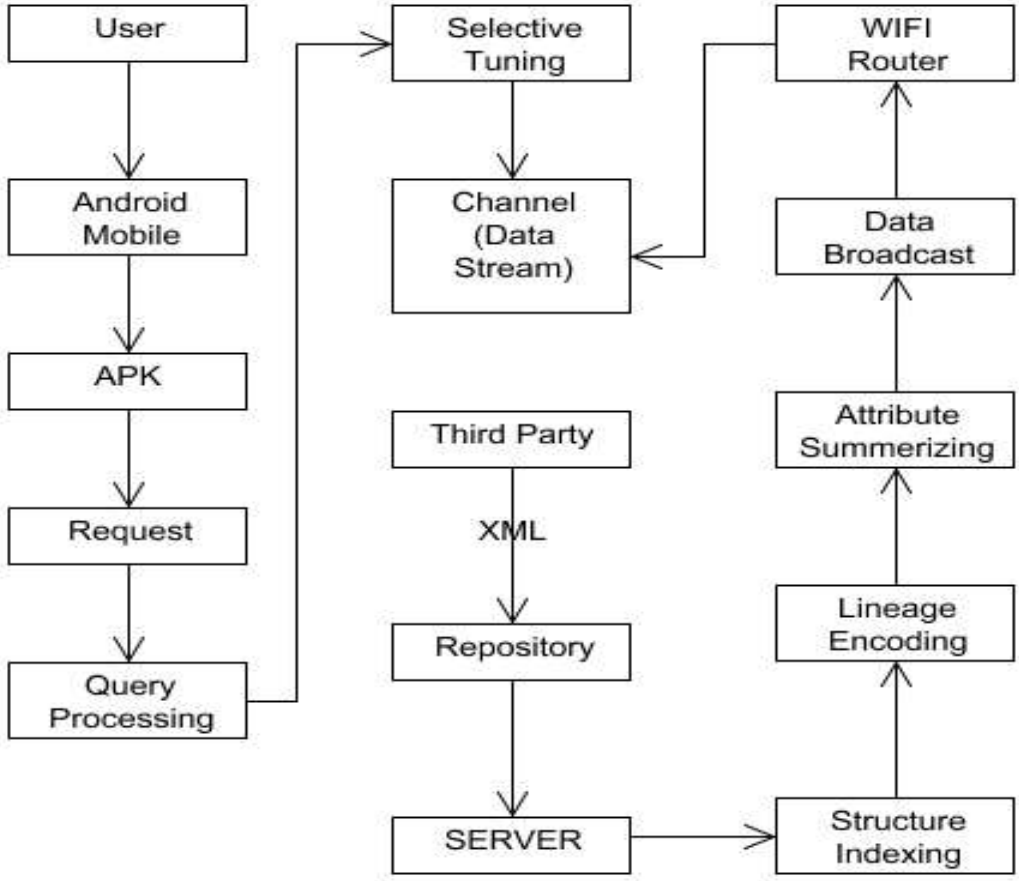


Fig 4. Flow diagram for lineage Encoding

CONCLUSION

Lineage encoding is used to support queries involving predicates and twig pattern matching. Hence, the wireless generation algorithm supplies the sufficient accessing of information in broadcasting for playing cricket scores. The mobile client can retrieve the required data satisfying the given twig pattern by performing bit wise operations on lineage encodes in the relevant G-nodes, by providing both energy and latency efficiencies. Thereby ID3algorithm accessed to attain the highest information gain in cricket match.

REFERENCES

1. M. Altinel and M. Franklin, "Efficient Filtering of XML Documents for Selective Dissemination of Information," Proc. Int'l Conf. Very Large Data Bases (VLDB),
2. C. Zhang, J.F. Naughton, D.J. DeWitt, Q. Luo, and G.M. Lohman, "On Supporting Containment Queries in Relational Database Management Systems," Proc. ACM SIGMOD Int'l Conf. Management of Data Conf., 2001.
3. S. Al-Khalifa, H.V. Jagadish, N. Koudas, J.M. Patel, D. Srivastava, and Y. Wu, "Structural Joins: A Primitive for Efficient XML Query Pattern Matching," Proc. Int'l Conf. Data Eng. (ICDE), pp. 141-152, Feb. 2002.
4. N. Bruno, D. Srivastava, and N. Koudas, "Holistic Twig Joins: Optimal XML Pattern Matching," Proc. ACM SIGMOD Int'l Conf. Management of Data Conf
5. A. Gupta and D. Suciu, "Stream Processing of XPath Queries with Predicates," Proc. ACM SIGMOD Int'l Management of Data Conf.,

ISSN: 2277-9655
Impact Factor: 4.116
CODEN: IJESS7

6. .P. Park, C.-S. Park, and Y.D. Chung, "Attribute Summarization: A Technique for Wireless XML Streaming," Proc. Int'l Conf. Interaction Sciences, pp. 492-496, Dec. 2009
7. https://en.wikipedia.org/wiki/**ID3_algorithm**
8. https://www.cise.ufl.edu/~ddd/cap6635/Fall-97/Short-papers/2.htm
9. https://en.wikipedia.org/wiki/ID3_algorithm
10. J. Ziv, and A. Lempel, "A universal algorithm for sequential data compression", IEEE Transaction on Information Theory, Volume 23, Number 3, May 1997, pages 337-343