

# Developing Agglomerative Hierarchical Clustering System for Planning Port Harcourt Mega City

<sup>1</sup>P. C. Nwosu, <sup>2</sup>H. O. Onungwe and <sup>3</sup>B. O. Eke  
Dept. of Computer Science, Faculty of Science,  
University of Port Harcourt, Nigeria

**Abstract:-** This research was carried out on developing agglomerative hierarchical clustering system for planning Port Harcourt Mega city. The system was implemented using XAMPP technology which incorporates PHP and MySQL as part of the system. Netbeans software was used for the interface design. The designed system provided a means to quickly access, retrieve and update information within the Port Harcourt Mega City database. The system could be accessed from any system or device via the Internet. The system could be beneficial to the government of Rivers State and the administrators of Greater Port Harcourt City Development Authority on whose shoulders lie the responsibility of planning the great city.

**Keywords:-** Clustering, Agglomerative, Hierarchical, Mega City, Dendogram.

## I. INTRODUCTION

Clustering is a task that arranges a set of sample items into groups known as clusters such that these items found in a particular group are most related in any way to one another than to those found in other groups. It is used almost on all field of science associated with experimental statistics. Researchers try to ensure to obtain firsthand information on sample data by discovering sets with “similar behavior” in them [6].

The term clustering is a function of categorizing a collection of elements into groups known as clusters. The hierarchical clustering is referred to as a process of surveying cluster that tries to construct a pecking order of sets of objects in data mining field [10].

Port Harcourt is one of the major cities in Nigerian; it was established in 1912 by Sir Frederick Lugard, the then governor of Nigerian Northern Protectorate and Southern Protectorate. The major reason for its creation was to export base for minerals especially the coal that was discovered in Enugu in very early 90's.

Following the discovery of crude oil in commercial quantity in Oloibiri, a local community in the present Bayelsa State in 1956, Port Harcourt was transformed into a hub of the country's oil business, which is the major backbone of its economy. This led to its modernization and urbanization and by the virtue of its links with the petroleum industry. In addition to its location as the major commercial area as well as a leading industrial town of the then Eastern Region; Port Harcourt is very significant as a backbone of economic, social and cultural life in Rivers State as a whole.

Port Harcourt city is extremely overcrowded because it serves as the only major city in the state, especially for commercial activities. The state House of Assembly under the then Governor Chibuike Rotimi Amaechi passed a law to extend development to the adjoining communities as an effort to decongest the city.

The Port Harcourt Mega City scheme constitutes a major scheme initiated by the former Executive Governor of Rivers State Mr. Chibuike Rotimi Amaechi's administration. It was geared towards expanding and developing Port Harcourt into contemporary city of international standard and control over population in the existing metropolis. This new city's architectural design spreads across more than eight LGA's of Rivers State comprising of Oyigbo, Okirika, Ikwerre, Etche, Eleme Obio/Akpor, Ogu Bolo and the present Garden city.

The proposed area intended represent a contemporary city and a widely overhaul metropolitan city. The proposed municipality will be known as the Greater Port Harcourt City and will be situated near the Port Harcourt International Airport [2].

## II. REVIEW OF RELATED LITERATURES

### A. Clustering Definition

All these definitions share one common basic principle: successively combining similar data objects into groups called clusters [3]. Clustering deals with the assignment of a collection of items into groups known as clusters, in such a way that the items in a particular group are more related to themselves than the items in another groups [8].

It is an unsupervised learning procedure which splits data in a way that related objects are merged collectively in group known as clusters, the procedure is imperative for summarizing as well as ascertaining attributes in the collection of data [9].

A simple mathematical definition of clustering as follows: let  $X \in R^{m \times n}$  a collection of data objects denoting a section of  $m$  positions  $X_i$  in  $R^n$ . The aim is to divide  $X$  into  $K$  clusters  $C_k$  in a way that the data objects that are part of same cluster are more similar than data object in other clusters (Fung (2001)). The eventual outcome of the process is a function does not maps distinct objects in its domain with the same object in another domain, thereby preserving distinctness shown as:

$$X \longmapsto C \text{ of data items } X_i \text{ to clusters } C_k$$

Where:

- $X \in R^{m \times n}$  is a group of data objects denoting a set of  $m$  positions  $X_i$  in  $R^n$ .
- $C_k$  represents the  $k$ th cluster.
- $K$  represents the number of clusters.
- $C_k^j$  represents the  $k$ th cluster axis at the iteration  $j$ .

### B. Cluster Validity

Cluster validation is a process, which estimates the outcome of cluster investigation in a measurable and objective mode. A clustering arrangement is said to be authentic or valid if it is uncommon to certain extent. Uncommonness was selected in a statistical framework and requires that probabilities have an objective interpretation to whoever is validating clustering structure [4].

### C. Cluster Cardinality

A complex problem in clustering is to ascertain the number of groups that are form usually referred to as cluster cardinality represented by  $K$ . It is actually the number of basic elements in a group of data samples. Many a times, this  $K$  is usually an excellent assumption based on understanding of the field or occurrences. For  $K$ -means

clustering algorithm, a heuristic approach is introduced for selecting  $K$  then and an effort is put in place to integrate the choice of  $K$  into the unbiased parameter. Occasionally a limit is placed on the range of  $K$  [7].

### D. Similarity/Distance Measures

Hierarchical clustering algorithm just involves determination of the resemblance that exists in the groups of data objects. The objects grouping was carried out using distance metrics [1]. Some distances are such that can be obtained by measuring with a ruler while some are ascertained base on similarity.

### E. Dendrogram

A dendrogram is a unique kind of cluster tree arrangement which gives a well-situated depiction of a hierarchical clustering. A dendrogram consist of series of points, each denoting a cluster [5].

It depicts the way in which the clusters were formed, which could be by merging two separate samples or merging a sample with an existing cluster.

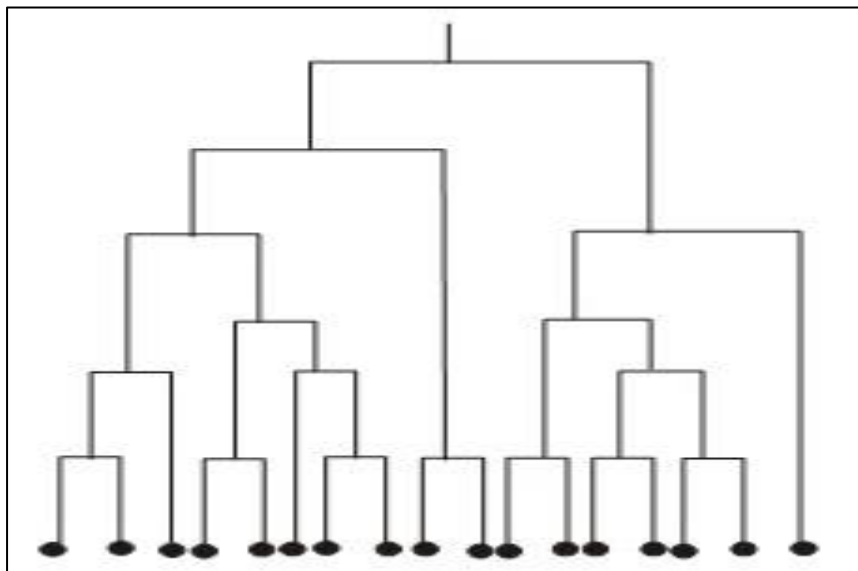


Fig. 1: A Dendrogram

## III. ANALYSIS AND DESIGN

In order to develop the system expected in this project, we had to analyze the system to find out the components parts that are necessary for the system to be properly developed. But before we move into details we may need to specify the concept of analysis and how it relates to the project we carried out in this chapter. In this work, there are two major types of analysis that was carried out:

- System Analysis: This is needed in developing the proposed system.
- Spatial Analysis: This is needed in implementing the Port Harcourt Mega City pattern on a geographical map.

Spatial analysis involves the techniques for the analyzing the positions, features as well as interactions among geographical data samples using exploratory

methods so as to solve a problem or deduce meaningful facts from the data. It deduces or produces additional facts in geographical objects samples that can be used in building clusters and pattern that can be applied in various situations such as the planning of Port Harcourt Mega city. These types of survey have become general and widespread because Geographic Information System (GIS) provides desirable features such as low cost and easy to use platforms.

Geographical Information System (GIS) allows geographical conception of object samples like building agglomeration as well as clusters within an area or location by employing graphical representation of the objects. In order to accomplished this goal, a proper database and a geographical information source such as city map was used.

A. Input/Output Design

**Directory Information : Data Entry**

**Name:**

**Address:**

**Zone:**  Select Item ▼ **Category:**  Select Item ▼

**Telephone No.**

**Keyword:**

**Save**

Fig. 2: Data Entry Form

Enter Search Keyword  **Search**

Residency Information					
	Name	Address	Zone	Mobile	Category
1					
2					
3					
4					
5					
6					

Fig. 3: Output Display Form

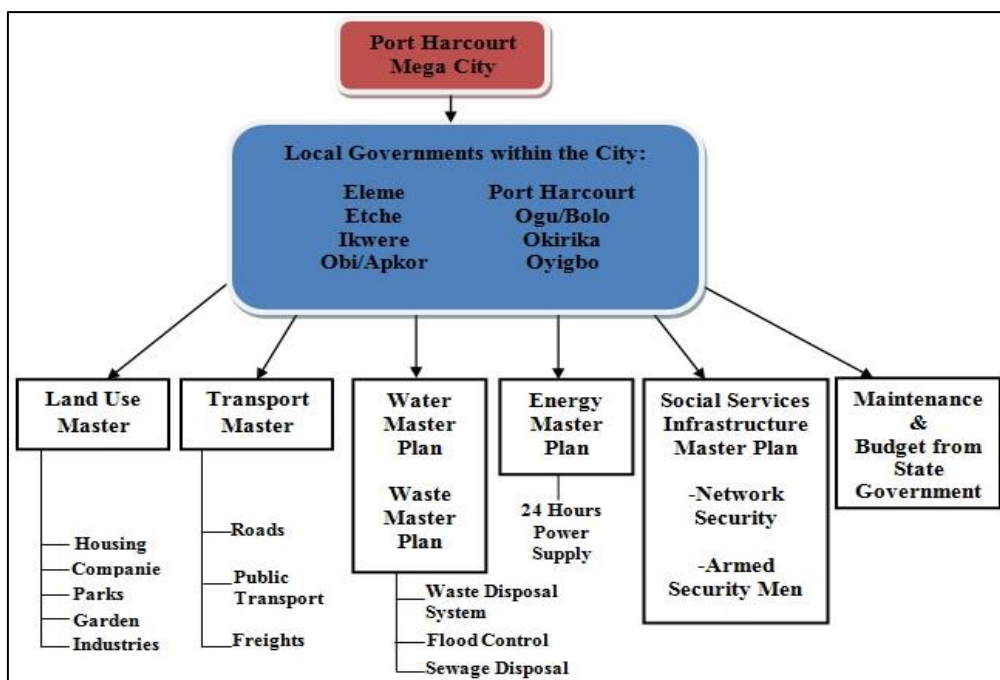


Fig. 4: Hierarchical Model of the New Port Harcourt Mega City

#### IV. SYSTEM IMPLEMENTATION

- XAMPP
- PHP Hypertext Preprocessor
- My SQL
- Web Browser such as Mozilla Firefox or Internet Explorer
- Netbean

PHP Hypertext Preprocessor programming language was essentially applied in the creation of the web-based application. Netbean was used to design the system interface.

The steps involved in carrying out the system implementation as well as setting up of the programming language used for the implementation were properly examined. Data was collected and entered into the system; the result of the implementation was also properly highlighted.

#### V. DISCUSSION AND RESULT

The system could be accessible from any web-browser, preferably internet Explorer 9 and above or Mozilla Firefox 35 and above. All that is needed is to use the software to store records. These records could be updated or deleted by the system administrators and users can then access these information at any time from any location with a computer system or any device connected to the internet using appropriate versions of common web browsers.

Access to data serves as ultimate goal of the new system because it is a useful purpose as the current system exists as a map located somewhere in the Rivers State government house. Access to this map may take days, weeks, months or years as the case maybe to be accessed when needed. In some cases, access may not be guaranteed depending on the individual that needs to access the map. This project provided an online system that can make Port Harcourt Mega city data available to people just with few mouse clicks. The system could be accessible through any device via the Internet platform.

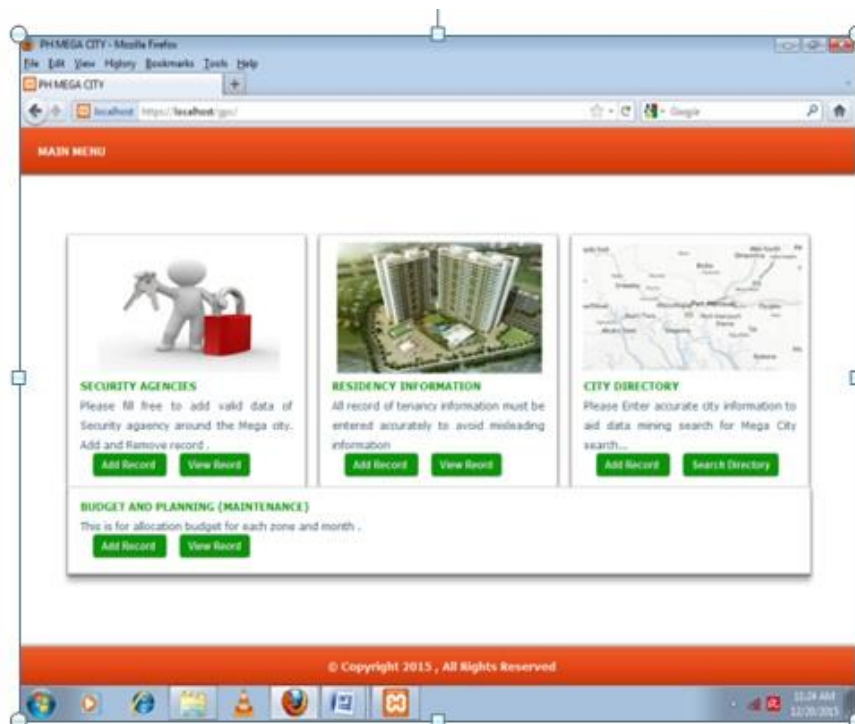


Fig. 5: User Interface of the System

#### VI. RECOMMENDATIONS

The work carried out in this research is recommended to Rivers State Government, through the department of Land and Housing in the state whose goal is to transform the city to a world-class city and make it a tourism center to generate more internal revenue for the state.

It is also recommended to construction companies that are looking for a methodology to apply in the development of larger and advanced city in any part of the world. They will be able to use this work as a stepping stone to achieve their goals and put better standardized structures in place.

#### REFERENCES

- [1]. Blei, D. M. (2008): Hierarchical Clustering. COS424 Princeton University. Extracted on from [http://protocols.netlab.uky.edu/~liuj/teaching/CS685\\_s18/06-Clustering-hierarchical-demo.pdf](http://protocols.netlab.uky.edu/~liuj/teaching/CS685_s18/06-Clustering-hierarchical-demo.pdf)
- [2]. Cooney-Gam, A. R. (2010): An Overview of the Greater Port Harcourt City Master Plan and Opportunities in Building a World Class City Over the Next 20 years. September, 2010. Extracted from [http://gphcity.com/media/publications/Presentation %20%20An%20overview%20of%20GPHC%20Master%20Plan.pdf](http://gphcity.com/media/publications/Presentation%20%20An%20overview%20of%20GPHC%20Master%20Plan.pdf)

- [3]. Fung, G. (2001) A Comprehensive Overview of Basic Clustering Algorithms. A Technical Report, University of Wisconsin, Madison.
- [4]. Jain, A. K. & Dubes, R. C. (1988). Algorithm for Clustering Data. Prentice Hall Inc., Eagle Cliff, New Jersey. ISBN: 0-13-0222278-X
- [5]. Jain, A., Murty, M., & Flynn, P. (2001). Data Clustering: A Review, ACM Computing Surveys (CSUR), 31(3), 264–323.
- [6]. Manning, C. D., Raghavan, P. & Schütze H. (2009): An Introduction to Information Retrieval. Cambridge University Press. April 1, 2009. From: <http://www.informationretrieval.org/>
- [7]. Luxburg, U. (2007): A Tutorial on Spectral Clustering. Journal of Statistics and Computing, 17(4), 395-416.
- [8]. Manning, C. D., Raghavan, P. & Schütze, H. (2009): An Introduction to Information Retrieval. Cambridge University. Extracted from <http://www.informationretrieval.org>
- [9]. Rahmat W. S., Jasni M. Z. & Abdullah E. (2010): A Comparative Agglomerative Hierarchical Clustering Method to Cluster Implemented Course. Journal of Computing, 2(12), 1-6.
- [10]. Sasirekha, K. & Baby, P. (2013): Agglomerative Hierarchical Clustering Algorithm—A Review. International Journal of Scientific and Research Publications, 3(3), 1-3.