e-Census Based Decision Making for Employment Planning

Ajay Dutta[1*], M. Syamala Devi[2], Manish Arora [3]

[1]Department of Computer Science and Applications, Panjab University, Chandigarh, India
[2] Department of Computer Science and Applications, Panjab University, Chandigarh, India
[3] National Institute of Electronics and Information Technology, Chandigarh, India

*Corresponding Author: ajaydutta13@pu.ac.in, Tel.: +91-9876240638

Abstract— e-Governance refers to the delivery of government information and services using Information and Communications Technology (ICT) to citizens or businesses or government. Census is a government activity of citizen socio-economic data collection process for the preparation of the citizen database. Census data is used for various activities planning like increasing literacy rate, removal of unemployment and healthcare issues etc. The aim of the paper is to present web based spatial dimensional modeling of the job seekers for assisting in taking high level decisions for government institutions like National Institution for Transforming India (NITI) Aayog and State Governments etc. The proposed system employs Online Analytical Processing (OLAP) and Data mining techniques for finding the job seeker population correlations among multiple fields like gender, the age structure of the population, geographical area and educational qualification in census database. OLAP techniques like roll-up, drill-down, slice and dice are implemented for dimensional modeling of the proposed system. Microsoft Visual Studio is used as an integrated environment for prototype development of a web based application with c# as the programming language. Microsoft SQL server is used as a backend relational database management system. It is concluded that the census database can be effective and efficient in achieving sustainable development goals of government as it can provide accurate and timely information within a reasonable amount of time.

Keywords—Unemployment, Dimensional modeling, Census, e-Governance

I. INTRODUCTION

e-Governance refers to the delivery of State and Central Government information and services using Information and Communication Technology (ICT) to citizens, businesses and Government departments. ICT provides access to information through communication mediums like the Internet, wireless networks and cell phones etc. Census is of great importance to the political, social and economic planning of a country. An online census is the web based system which permits the users to enter the census data via the ICT. An online census database can help the government to make decisions related to need of skill development centre, educational institutions etc. In this paper, the Unemployment problem is taken because it is the root of various major problems. Unemployed workers are those who are jobless, seeking a job and ready to work if they are given a job. India is the second most populous and seventh major country in the world. The government needs basic information on its Job Seeker Population for job policy planning and improvement of the education. Good planning is possible if the reliable, accurate and detailed information is

available. This information makes it possible to plan better services and solve existing problems like unemployment etc. A detailed data of information in the census database contains hidden information of strategic importance. Data mining is a key element in finding this hidden information and relationships that can help organizations, governments, and businesses to take efficient and effective decisions. This suggests census can provide required information and knowledge for decision making in unemployment policy planning. This crucial information and knowledge are made available using OLAP and Data Mining techniques. The proposed model will make available the detail of unemployed population like a number of unemployed males, a number of unemployed females, their age, and literacy level. Census data mining can also be used with government organization for taking decisions such as opening new educational institutions and granting funds on the basis of data analysis using parameters like literacy rate and gender the rate of a specific location. Section II contains the related work of e-Governance and census. Section III contains the dimensional modeling of job seeker data, Section IV contains

the architecture design of process and database and Section V. contains the conclusion and scope for future work.

II. LITERATURE REVIEW

The literature is reviewed related to e-Governance applications, decision support systems, OLAP and data mining techniques. Dimensional modeling is the most appropriate approach to design a data warehouse for the purpose of analytical data mining [1]. The authors discussed the e-Governance services, infrastructure, and technologies in the implementation of electronic governance in India [2]. They are discussed e-services provide better delivery of government services to citizens, less corruption, increased transparency, greater convenience, citizen empowerment through access to information, a decrease in time and effort, revenue growth and cost reductions. In [3], the authors explained the selection of an appropriate data schema like Snowflake, Star, Star Cluster etc. and its Impact on performance and usability of the designed data warehouse. Data schema represents the arrangement of the fact table and dimension tables and the relations between them. In [4], the authors described the major objectives of dimensional modeling are to design database structures that are easy for end users to understand and write queries against and also maximize the efficiency of queries. In [5], the authors apply OLAP and data mining approaches to ERP data in order to generate reports, charts and identify new knowledge to support decision makers. . In [6], the authors described the three phases (information, interaction, and transformation) for government circulating information and services among citizens, business and other departments. The majority of multidimensional queries deal with high levels of data as dimensions, levels of hierarchies and attributes [7]. ICT today helps in enabling the transformation of the most expensive public services such as education, healthcare and also advanced low income countries' economies such as agriculture and transportation. In [8], the authors present the modeling and simulation of the decision support system for population census in Nigeria. A structural framework is presented with a logical model for the classification of population parameters such as related to the head of household, age, gender and literary etc. In [9], the authors presented the Government Information Sharing System (GISS) implemented by the Public Administration Information Sharing Center in Korea. The System facilitates with sharing of citizens' records among different government, educational, public and financial organizations to serve citizens with government services.

In [11], cube operations, dimension hierarchies, function applications, and navigation are introduced. The administrative setup of the country has a hierarchical structure. The country is divided into 29 States and 7 Union Territories (UTs) which are further divided into several districts where every district is further divided into tehsils

and villages at the lower level [14]. In India, job seeker population data varies from state to state. For example number of job seekers per thousand in Gujarat State is 9 and in Tripura State 197. Based on a review of the above literature, Web based Geographical dimensional modeling of the Unemployment data can also help to indicate the overall balancing between the supply of workforce and demand for workforce within a specific time.

III. DIMENSIONAL MODELING OF JOB SEEKER

Dimensional modeling refers to different views of the data from different dimensions. It is widely adopted design technique to support end-user queries. It uses the concepts of Facts, Dimensions, Attribute, Hierarchy and Fact Table. Facts are numeric values that can be grouped. Dimensions are groups of hierarchies and define the facts. Hierarchy is the description of levels that represents the relationship between different attributes within a dimension. A location hierarchy describes the organizational structure and logical parent child relationship within the census data. For example, the village is at the lowest level and the nation is at the highest level in location hierarchy depicted below. At the national level, complete data is compiled. At next level aggregate data distribution is limited to state level, next state to district level, district to tehsil level and last the village level. The summarized collection of data attributes make information retrieval more efficient. Diagrammatically one such scenario has been shown in Figure 1.
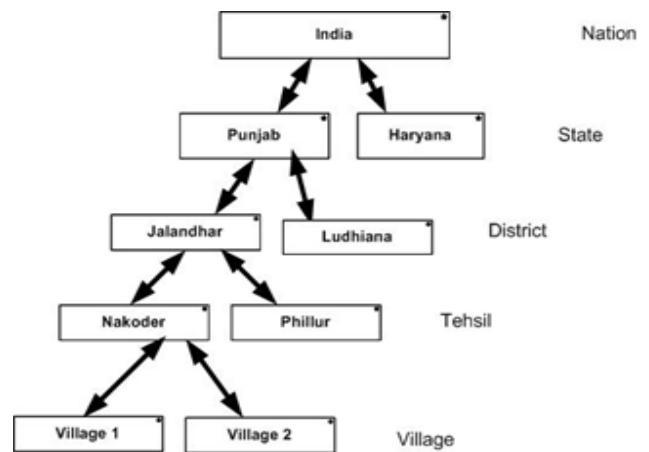

Figure 1: Hierarchical View of Geographical Area

The proposed model can provide a count of cross-classification of dimensions like Job Seeker, age, gender, literacy level and the area in the various dimensions. Dimensions like 0-Dimension (0-D), 1-Dimension (1-D), 2-Dimension (2-D) and 3-Dimension (3-D) have been represented in Figure 2.
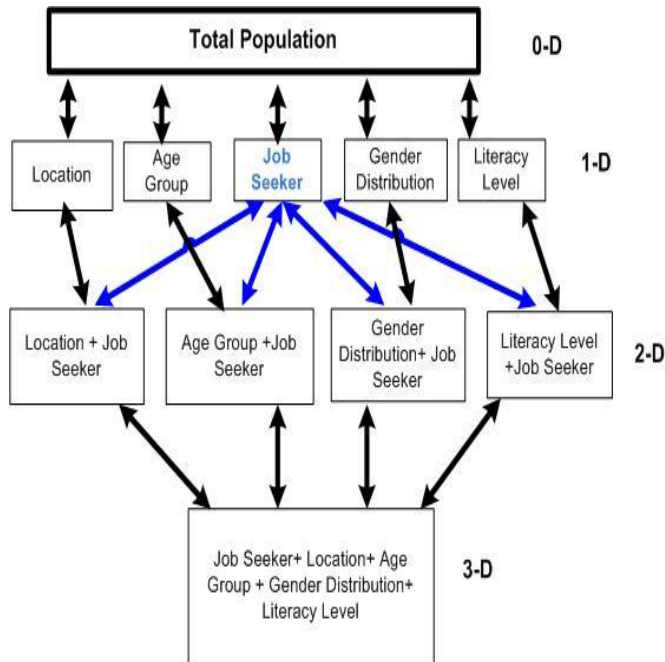
Figure 2: Types of dimensional views in the cube

The aggregated information of facts can be viewed as the 0-D level which is top most hierarchical level. The same information can be viewed at a 1-D level in the context of location or Age Group or Job Seeker or Gender distribution or Literacy level. Further, the combination of two dimensions can be analyzed at the 2-D level and a combination of consolidated four dimensions at the 3-D level.

IV. E-CENSUS BASED DECISION MAKING

Census database uses the multi-dimensional cube to store the data. Every cube is particularly its own value because all these make possible to roll-up and drill-down operation with other cubes. Front-end for the proposed model is designed using OLAP operations such as roll-up, roll-down, dicing and slicing to enable users to navigate data, analyze data at different granularities and visualize results in different structures. The roll-up operation aggregates the job seeker data from village to country by location hierarchy. This is done when one moves up in the hierarchy or reduces a dimension, i.e. one aggregate to all level. The Drill-down operation reduces the level of aggregation moving from one general level to a detailed level in a hierarchy The Slicing operation selects one dimension of a cube that results in a sub-cube. The Dice operation selects two or more dimensions of a cube that result in a sub-cube. Web based Graphical User Interface (GUI) is designed for analysis of census data from high level to a low level such as from National, to State, to District, to Tehsil and finally to Village level.

A. *Database Design*

The multidimensional census database is implemented by Relational online analytical processing (ROLAP) Server. The multidimensional census database operations have to be mapped into relations and SQL queries. Multidimensional census models represent data with a "cube" structure, making more compatible logical data representation with OLAP data analysis. Census data can be directly queried by passing complex database queries by using various combinations of dimensions. There are two types of schema - Star schema and Snowflake schema that are used to design a census database. If all dimensions are directly joined to the fact table then it called Star schema If the dimensions are not directly joined to the fact table but through the other dimension, this is called snowflake schema. For example, the fact table (i.e. tbl_insertpriaryinfo) stores numeric value on location is based on the foreign keys StateID (for storage of state information) and DistrictID (for storage of district information) etc. This makes a start type of structure around the fact table and is known as Star Schema and it is represented as shown in Figure 3.
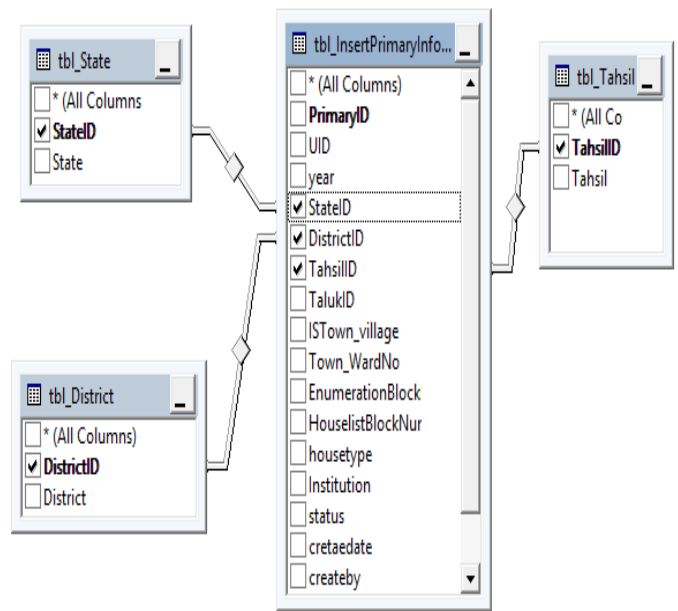

Figure 3: Star Schema

For example, the fact table (i.e. tbl_inserthouseinfo) as shown in Figure 4 stores numeric value on the foreign keys like MaritalStatusID (for storage of Marital information) and UID (for storage of UID information) etc. UID information comes from table name (i.e. tbl_Insertprimaryinfo) that is already linked with other tables. This makes a Snowflake Schema type of structure that is represented in Figure 4.
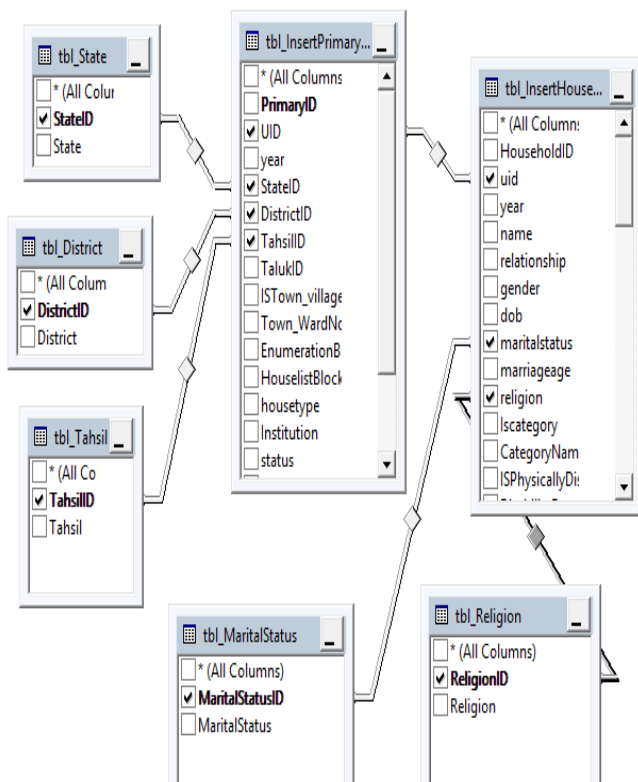
Figure 4: Snowflake Schema

B. *Process Design*
The citizen employment data analysis system provides information on a number of job seekers at a specific location depending on education level. The education level can be matric passed, undergraduate, graduate or post-graduate etc. It will help government organizations to analyse the job seeker information at the state level or district level or tehsil level as shown in Figure 5.

Following steps are performed for employment data analysis

Step 1 − Start
Step 2 − declare seven variables State, District, Tahsil, age group, gender, literacy, and output
Step 3 − define values of variable State, District, Tahsil, age group, gender, and literacy status
Step 4 − select values of variable State, District, Tahsil, age group, gender, and literacy status
Step 5 − Count values in the database to corresponding to input variable State, District, Tahsil, age group, gender, and literacy status
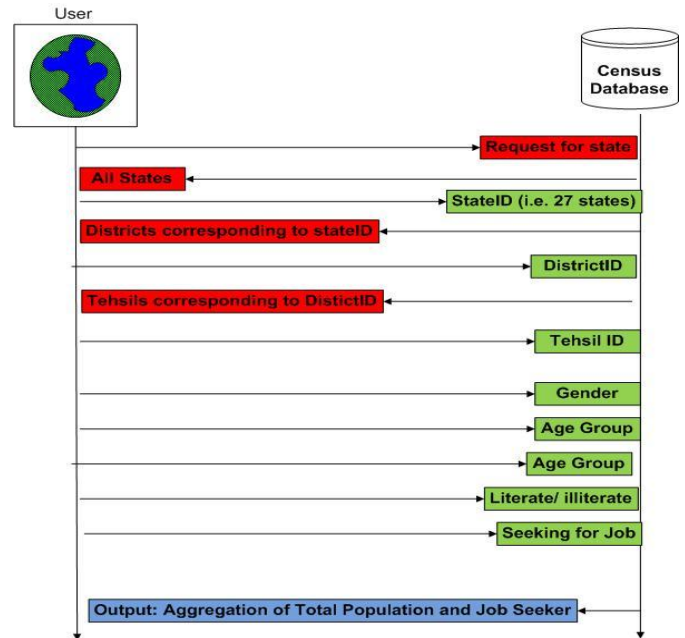Step 6 – store result of step 5 to output
Step 7 – display output
Step 8 – Stop

Figure 5: Activity Diagram for Job Seeker Data Analysis

V. CONCLUSION AND SCOPE FOR FUTURE WORK

The proposed web based geographical dimensional modeling of the Unemployment data is useful for assisting in taking high level decisions like trends of unemployment, need of opening skill development institutions, planning for female suitable skill development institutions. It can be helpful at both State and Central Government level employment planning. The proposed multidimensional modeling is reliable, flexible and accurate. In future scope, it can be integrated with various departments for circulation of employment information based on geographical area, gender, the Age structure of the population, literacy level etc. It can also be linked with mobile based Short Message Service (SMS) technology.

REFERENCES

[1] Ljubljana, Slovenia, "Multi-relational data mining: an introduction," ACM SIGKDD Explore, Vol. 5, pp. 1–16, 2003. DOI: http://doi.acm.org/10.1145/959242.959245,.
[2] A. Dutta and M. S. Devi, "E-Governance Status in India," International Journal of Computer Sciences and Engineering, Vol 3, Issue-7, pp. 27–31, 2015.
[3] R. Rupnik and M. Kukar, "Data Mining Based Decision Support System to Support Association Rules," Electro Technical, Vol. 74, pp. 195–200, 2007.
[4] M. H. Peyravi, "A Schema Selection Framework for Data Warehouse Design," International Journal of Machine Learning and Computing, Vol. 2, No. 3, June 2012.
[5] T. Abdellatif, M. Abo Elsoud, and H. Arafat Ali, "Comparing Online Analytical Processing and Data Mining Tasks In Enterprise Resource Planning Systems," International Journal of Computer Science Issues, Vol. 8, Issue 6, No 2, pp 161-174, November 2011.

[6] Al-Hashmi and A. Darem, , "Understanding phases of E-government project," Emerg. Technol. E-Government, pp. 152–157, 2008.

[7] N. Padhy and R. Panigrahi, "Multi Relational Data Mining Approaches : A Data Mining Technique International Journal of Computer Applications,Vol 57, No.17, pp. 23–32, November 2012.

[8] O. C. Okeke and C. Boniface, "Using Data-Mining Technique for Census Analysis to Give Geo- Spatial Distribution of Nigeria .," Journal of Computer Engineering, Vol. 14, no. 2, pp. 1–5, 2013.

[9] J. Choi and S. A. Chun, "SecureGov: secure data sharing for government services", Proceedings of the 14th Annual International Conference on Digital Government Research, pp. 127–135, June 2013, DOI 10.1145/2479724.2479745

[10] V. R. Rao, "A Framework for e-Government Data Mining Applications (eGDMA) for Effective Citizen Services " -An Indian Perspective," International Journal of Computer Science and Information Technology Research, Vol. 2, no. 4, pp. 209–225, Dec 2014.

[11] J. Hareesh and V. V. Kumari, "Decision Tree Based Information System for Improving School Facilities," International Journal of Science and Research, Vol. 4, no. 10, pp. 2013–2016, Oct 2015.

[12] Siriginidi Subba Rao, "Achieving millennium development goals: Role of ICTS innovations in India", Telematics and Informatics 26, Vol 26, Issue 2, pp 127-143, February 2008

[13] A. Vijayaraj and C.-T. Nadu, "Design and Implementation of Census Data Collection System using PDA," vol. 9, no. 9, pp. 28–32, 2010.

[14] S. Mahendra Dev and M. Venkatanarayana, "Youth employment and unemployment in India," International Journal of Accounts, Economics & Business Management, Vol. 1 No. 1, pp 1-6, Sep. 2012.

Authors Profile

Mr. Ajay Dutta is pursuing Ph.D. from Department of computer science and Applications, Panjab University, Chandigarh, India. He has received MCA degree with distinction from Punjab Technical University, India in 2012. Before joining Ph.D., He has served various e-Governance projects like Punjab Land Record Computerization, Municipals Corporations Computerization, Grievance Management System, Punjab State Warehousing Corporation (PSWC) Computerization and NRI Punjab Department Grievance portal development. He has presented five research paper at national and International Level. In his credit four research papers related to e-Governance services are published in Journals. He is Google AdWords and Google Analytical certified. He is involved in various organization i.e. Computer Society of India (Life Member), Internet Society of Computers (Member), Open Web Application Security Project, Punjab Science Congress (Member), Special Interest Group on e-Governance under CSI (Member).

Dr. M. Syamala Devi is a professor in the department of Computer Science & Applications, Panjab University, Chandigarh. She received her Ph.D degree in Computer Science & Systems Engineering from Andhra University, Visakhapatnam and M.E. in Computer Science & Engineering, from NIT, Allahabad. She has completed M.Sc in Applied Mathematics from Andhra University, Visakhapatnam. Before joining Panjab University, she served Indian Space Research Organization, Sriharikota, and National Institute of Technical Teachers" Training & Research, Chandigarh. Her areas of expertise include algorithms, Image Processing, Distributed Artificial Intelligence and Educational Computing. She has guided (two students are awarded PhD) and has been guiding a number of student projects at Ph.D and post graduate level. She has published about fifty papers in national and international research journals, proceedings of various national and international seminars, and conferences. She has conducted various short-term courses, workshops, seminars and symposia. She has given expert talks in Computer Science and Information Technology related topics at various Colleges, Universities and Institutions.

Dr. Manish Arora has been working as Additional Director at National Institute of Electronics and Information Technology, Chandigarh since 1995. He completed MCA from Guru Nanak Dev University, Amritsar and obtained Ph.D. degree from Department of Computer Science and Applications, Panjab University, Chandigarh in the area of Multi-agent Systems. He has vast experience in software development and training on different platforms. His research interests include cloud computing, big data analytics and e-governance applications. He has published several research papers in various international journals and conferences.