

# DESIGN AND IMPLEMENTATION OF ONTOLOGY BASED ON SEMANTIC ANALYSIS FOR GIS APPLICATION

Dr. S.S Mantha<sup>1</sup>, Mrs. Madhuri Rao <sup>2</sup>, Ms. Naina Saita<sup>3</sup>

<sup>1</sup>Chairman AICTE, Professor, CAD/CAM, Robotics, VJTI, Mumbai, India

ssmantha@vjti.org.in

<sup>2</sup>Asst Prof, Department of Information Technology, TSEC, Mumbai University, India

my\_rao@yahoo.com

<sup>3</sup>Student, Department of Information Technology, TSEC, Mumbai University, India

nainasaita@gmail.com

## ABSTRACT

*The Agricultural Census information is a leading source of facts and figures about a country's agricultural development. Such information is used by many who provide services to farmers and rural communities including federal, state and local governments, agribusinesses etc. Also such information when integrated with other agricultural surveys and statistics can help in monitoring progress towards the achievement of Millennium Development Goals (MDGs) of a country. But such huge volumes of census data are available at various geo-spatial portals either in proprietary formats like shape files, .dat files etc or in form of database tables, word documents, PDF's etc. In order to do analysis or to just see the progress of a particular area such huge datasheets have to be scanned. This paper provides solutions to various problems related to Geo-spatial Agricultural Census data in three aspects: (1) Storage / Organization of census data using enhanced methods such as ontologies. (2) Visualization of data using Google Maps and Column Charts. (3) Analysis of data using interactive methods like Column Charts.*

## KEYWORDS

*Geo-spatial Ontology, Enhanced Visualization, Google Maps, Column Charts, Geo-spatial Agricultural Census data.*

## 1. INTRODUCTION

The Agricultural Census is a leading source of facts and figures about a country's agricultural development. Conducted every five or ten years, the Census provides detailed information about a country's farms and ranches and the people who operate them. It is the only reliable source of uniform, accurate and comprehensive agricultural data for every state and county in the country. Agricultural Census information is used by many who provide services to farmers and rural communities including federal, state and local governments, agribusinesses etc. The uses of agricultural census data [1] in a variety of economic and social areas are shown below in Figure 1. This agricultural census data is mainly used in monitoring progress towards the achievement of Millennium Development Goals (MDGs) and in analyzing poverty, food security and gender issues. An agricultural census could provide a range of data related to the MDGs as explained below:

David Bracewell, et al. (Eds): AIAA 2011, CS & IT 03, pp. 107–119, 2011.

© CS & IT-CSCP 2011

DOI : 10.5121/csit.2011.1310

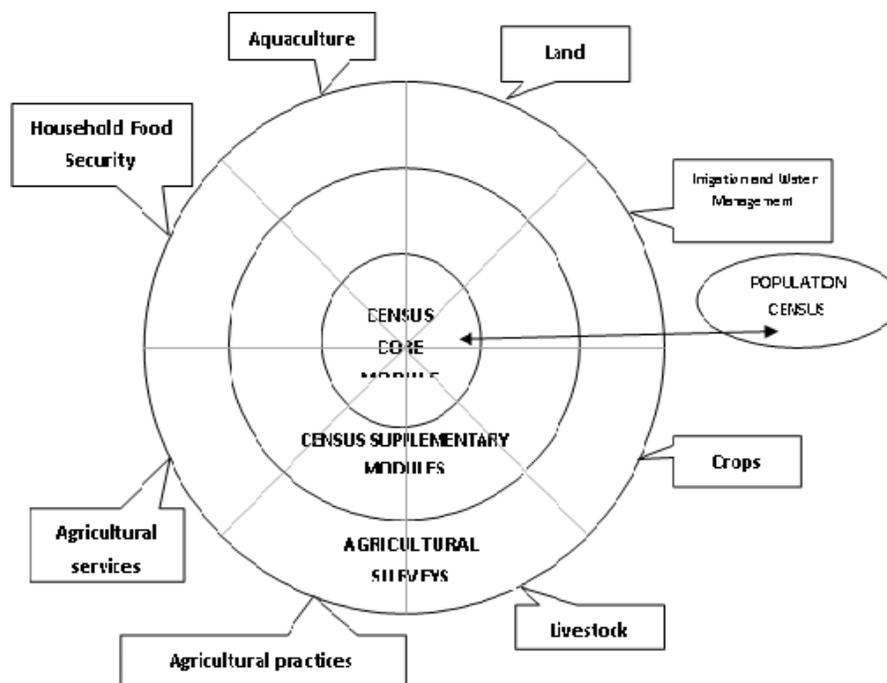


Figure 1 : The Agricultural Census in the framework of the system of integrated agricultural censuses and surveys [1]

### 1.1. Poverty monitoring and analysis

Shortage of land is one of the main reasons of poverty and agricultural census data related to farm size, types of cropping systems etc can help to understand whether farmers have sufficient land to support their needs. Also crop data can highlight the potential for raising farm incomes through crop diversification and investing in high value crops.

### 1.2. Food security monitoring and analysis

Cropping patterns can be analysed along with information on the use of irrigation, variety of improved seeds used, high value crops, farm machinery to help develop programmes for increasing food production. The agricultural census also helps to tackle issues related to food access. For example, data on the location and availability of agricultural produce markets and other infrastructure in the community can help to assess the effectiveness and accessibility of the food distribution system.

### 1.3. Measuring the role of women in agriculture

The agricultural census provides indirect measures related to the role of women in agricultural production activities and the participation of rural women in non-farm economic activities. The identification of the agricultural holder provides the basis for comparing the characteristics of holdings operated by men and women. Study of aspects such as area of holding, cropping patterns, use of different agricultural practices can help to focus on the problems faced by women in operating agricultural holdings. For example, if women tend to be responsible for managing livestock, the census would provide information on the number of women performing similar roles, type of livestock they manage, their demographic characteristics, main occupation, and the amount of time they spend in working on and off the holding.

#### **1.4. Ensure environmental sustainability**

This emphasizes on “integrating the principles of sustainable development into country policies and programmes and reducing the loss of environmental resources”. An agricultural census collects a large amount of environmental information related to irrigation, soil degradation, use of mineral fertilizers, use of pesticides, fungicides etc.

#### **1.5. Agricultural planning and policy-making**

An agricultural census can be used to analyze the characteristics of agricultural holdings and their agricultural production activities and thereby acts as an aid in helping the government and other regulatory bodies in effective planning and policy-making.

#### **1.6. Providing data for the private sector**

Private sector basically requires such data for making commercial decisions. A food processing company could use data such as number of growers and area for specific crops in each district to help identify target sites for its processing plants. An input supplier could use data such as input use for each crop by district to better identify the market opportunities. Farm machinery suppliers use data such as area of each type of crop grown and the number of growers to identify the potential demand for their products. A company planning to setup a business in a particular location could use such data to assess the availability of labour and the variety of skills available in that location.

Agriculture surveys are presently carried out throughout the nation in order to collect information and statistics on crops, rangeland, livestock and other related agricultural resources. This data is collected on district, county and state and also at the local level. But to gather such strategic and highly useful data, various problems are faced [2]. Such data when integrated with other census data and surveys and the resulting if analyzed and interpreted properly can help for agricultural development planning and formulation of national agricultural policies. Huge volumes of census data are available at various geo-spatial portals in formats like database tables, propriety formats like shape files, or simple formats like Word documents, PDF's etc. But retrieval of such information from such sources is difficult. Also it is cumbersome to analyze such high volumes of information for strategic planning purposes.

This paper tackles problems related to Agricultural Census data in three aspects-

(1) Storage / Organization of census data- Rather than storing data in relational databases or propriety formats like shape files, it stores data using improved knowledge and data modeling techniques such as ontologies [3].

(2) Visualization of data- Rather than pondering over multiple .dbf, .xls, or pdf files for analyzing and interrelating it for strategic planning and decision making purposes, the paper demonstrates an interactive way of projecting such valuable information using Google Maps and Column Charts.

(3) Analysis- Interactive methods like column charts etc have been used to compare growth, variances, improvements etc among various states and counties. Also such data can be integrated with other surveys and analysis to solve various problems.

As participation in the Census is required by law, and the same law protects the confidentiality and integrity of all individual responses. So many large geo-spatial data portals store and publish only statistical aggregates and no detailed information from individual operations are ever disclosed for security purposes. Based on such statistical aggregates collected, a semantic application is developed to analyze such data and provide such crucial data in a timely manner thereby improving and shortening the planning and decision making process. The application proposed by this paper involves developing OWL Ontology for agriculture domain thereby

modelling US AGRICULTURAL 2007 CENSUS DATA [4]. Depending on the input query, the data is analyzed and extracted dynamically from the ontology structure. Finally the information is projected on the Map and displayed to the user. The analysis results are interactively displayed on Column Charts and Bar Charts.

## **2. DESIGN APPROACH**

### **2.1. Develop Ontology for Agriculture Domain**

Ontology models knowledge for US Agricultural Census and is based on 4 viewpoints:-

- Farm Land Features
- Land Features
- Operator Features
- Crop Production Features

The main entities of interest are:

- a. Farm Land-This class models information related to farms in a county. It models information such as “% of farms with sales less than \$10,000”, “Average size of the farms” etc. This class is further sub-divided into Cropland and Orchard land. Also it has other sub-classes such as Production Expenses, Machinery, and Agricultural Products. Such data can be used for poverty monitoring and analysis and can be very useful to private sector as stated earlier.
- b. Crop Land-This class models information related to a crop lands under a farm of a particular county. It models information such as “acres of cropland in the farm” etc. Such data can be used for food security monitoring and analysis as stated earlier.
- c. Annual Crop Production-This class models detailed information related to all annual crops grown under a particular county. Information such as pesticide used, fertilizer used, improved seeds used, acres harvested etc is modeled in this class. Such data can be used to ensure environmental sustainability and agricultural planning and policy making as stated earlier.
- d. Perennial Crop Production-This class models detailed information related to all perennial crops grown under a particular county. Information such as pesticide used, fertilizer used, improved seeds used, acres harvested etc is modeled in this class. Such data can be used to ensure environmental sustainability and agricultural planning and policy making.
- e. Operator-This class models information related to operators of the farm. It models important information such as “% of farms operated by family”, “% of farms operated by female principal operators” etc. Such data can be used to measure the role of women in agriculture as state earlier.

### **2.2. Associate the ontology concepts and relations with the geographic data**

The Ontology created is represented in XML (Extensible Mark-up Language) format that supports RDF (Resource Description Framework) schema and OWL (Web ontology language). The Resource Description Framework (RDF) is a family of World Wide Web Consortium (W3C) specifications originally designed as a metadata data model. The Web Ontology Language (OWL) is a family of knowledge representation languages for authoring ontologies, and is endorsed by the World Wide Web Consortium. OWL ontologies are most commonly serialized using RDF/XML syntax.



<http://www.infoplease.com/atlas/latitude-longitude.html>,  
[http://www.worldatlas.com/aatlas/latitude\\_and\\_longitude\\_finder.htm](http://www.worldatlas.com/aatlas/latitude_and_longitude_finder.htm)

Then the available information from these sites is in degree, minutes and seconds but Google Maps accepts decimal degrees format. So the degree, minutes and seconds format data was converted to decimal degrees using the following site <http://www.satsig.net/degrees-minutes-seconds-calculator.htm>

```

Attribute:

Attribute_Label: M003_07
Attribute_Definition:
    The average size of farms in acres. A value of zero may represent a zero data value or may mean that no data are available for the
    county.
Attribute_Definition_Source: National Atlas of the United States
Attribute_Domain_Values:

    Range_Domain:

        Range_Domain_Minimum: 0
        Range_Domain_Maximum: 47421

Attribute_Domain_Values:

    Enumerated_Domain:

        Enumerated_Domain_Value: -9999
        Enumerated_Domain_Value_Definition:
            Data withheld to avoid disclosing information for individual farms.
        Enumerated_Domain_Value_Definition_Source:
            U.S. Department of Agriculture, National Agricultural Statistics Service

Attribute:

Attribute_Label: M009_07
Attribute_Definition:
    The percentage of farms with sales less than $10,000. A value of zero may represent a zero data value or may mean that no data
    are available for the county.
Attribute_Definition_Source: National Atlas of the United States
Attribute_Domain_Values:

    Range_Domain:

        Range_Domain_Minimum: 0.00
        Range_Domain_Maximum: 100.00
  
```

Figure 3: Metadata snapshot

### 3.1.1 Final Attributes

- Farmland Characteristics
  - a. Average size of farms in acres
  - b. % of farms with sales less than \$10,000
  - c. % of farms with sales between \$10,000 and \$249,000
  - d. Average value of agricultural products sold per farm
  - e. Average value of all machinery and equipments per farm
  - f. Average total farm production expenses per farm
- Operator Characteristics
  - a. % of farms operated by family
  - b. % of farms with female principal operators
  - c. Average age of principal farm operator
- Land Characteristics
  - a. Area of total cropland as % of land area in acres
  - b. Total area of land in orchards
  - c. Irrigated Land in orchards
- Crop Characteristics
  - a. Crop name
  - b. Crop Code
  - c. Soil preparation method
  - d. Irrigation Used
  - e. Fertilizer Used

- f. Pesticide Used
- g. Harvesting Method
- h. Threshing Method
- i. Improved Seeds Used
- j. Area harvested
- k. Quantity harvested
- l. Mostly sold to

### 3.2. Ontology development

Ontology for the US agricultural census data is developed using Protégé OWL version 4.1 [5] software as shown in the Figure 4 below. Ontology is then populated (individuals are created) with data from shape files. The final ontology is saved in xml format.

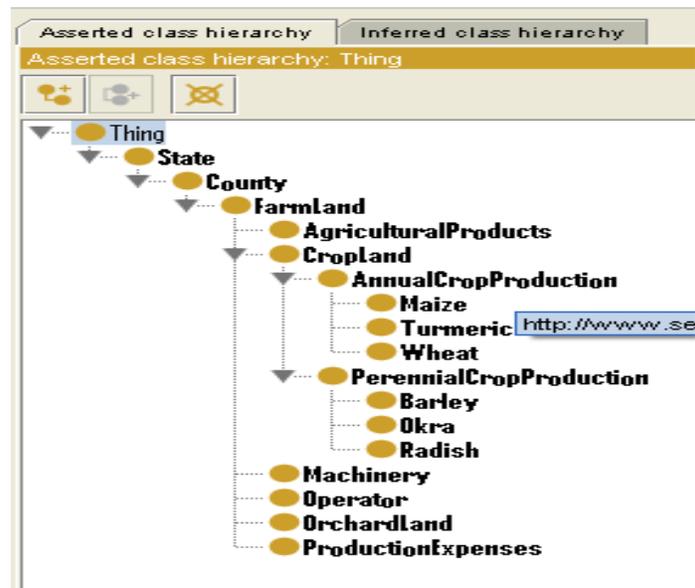


Figure 4: Asserted Class Hierarchy

### 3.2. Modelling the Agricultural census data

In the Agricultural Census Data page as shown in Figure 5, initially the State and County dropdown boxes are displayed. As soon as the page loads, the State dropdown box is filled with options which are dynamically extracted from the ontology. After the user selects a particular state, the county dropdown box is loaded dynamically with its options that are extracted from the ontology. County dropdown box is loaded with all the counties only under the selected state. Also at the same time all the counties of that state are plotted on Google Map as shown in Figure 6. Data related to all characteristics i.e. Farmland, Operator and Land is displayed in the marker plotted on Map for each and every county under than state.

If the user selects a particular county, then data for all characteristics (Farmland, Operator and Land) only for that particular selected county is displayed on the marker plotted on Google Map as shown in Figure 7. Once a county is selected, all 3 dropdown boxes i.e. Farmland characteristics, Operator characteristics and Land characteristics are filled with options dynamically extracted from the ontology.

If a user selected a particular option from any of these 3 drop-down boxes then data for that selected options is displayed in the marker plotted on the map for the selected county as shown in Figure 8. Implementation of above functionality has been done using JavaScript [6], AJAX [7] and Google Maps API version 3 [8].



Figure 5: Agricultural Census Data Page

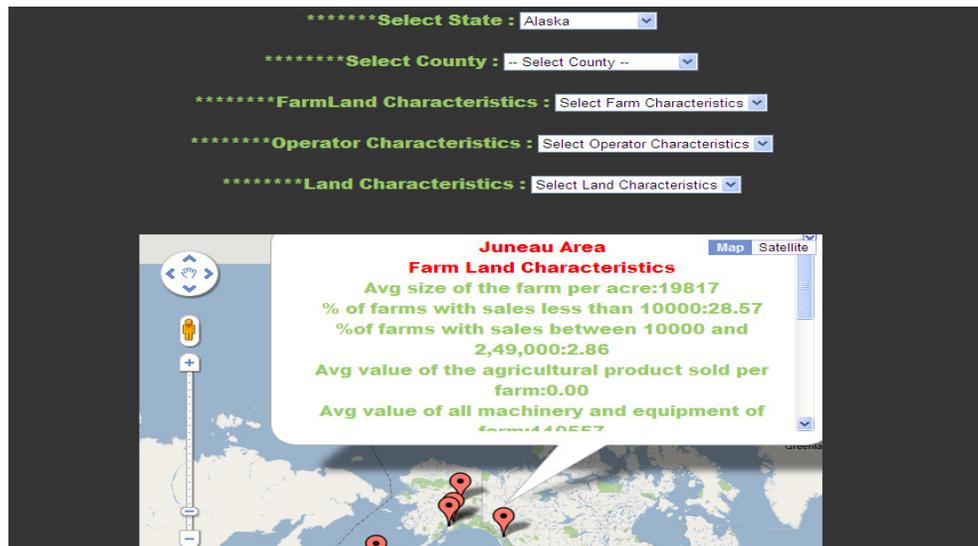


Figure 6: Agricultural Census Data Page – State Selected

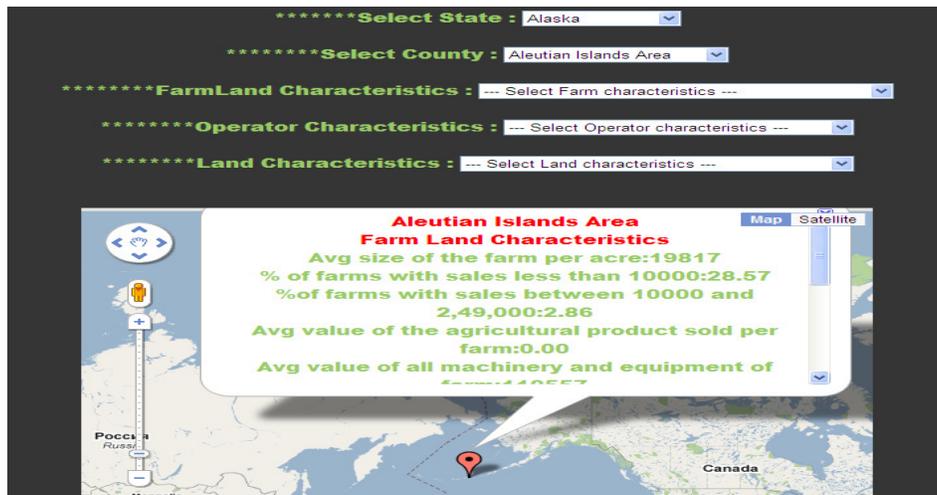


Figure 7: Agricultural Census Data Page – County Selected

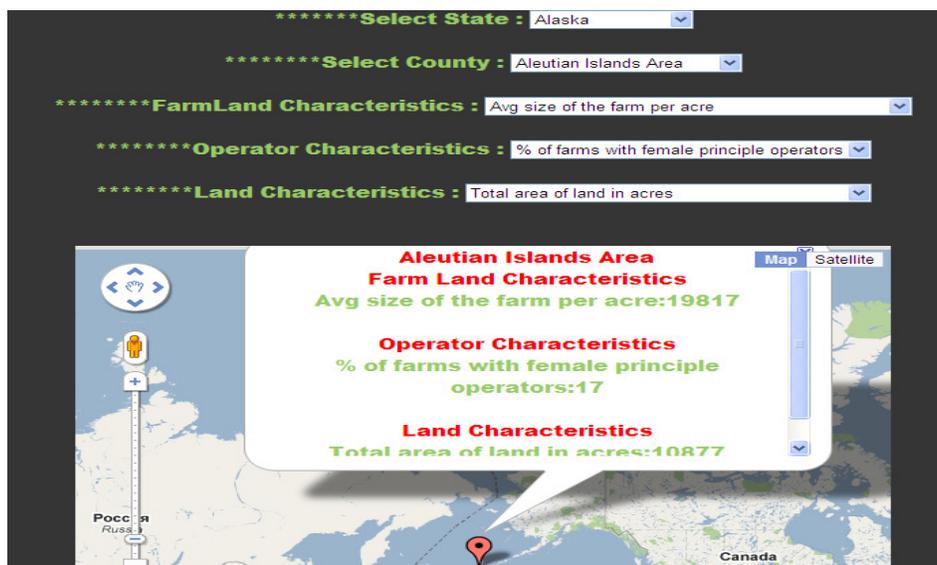


Figure 8: Agricultural Census Data Page – Characteristics Selected

### 3.3. Modelling the Crop production characteristics

In the Crop Production Characteristics Page, initially the State and County drop-down boxes are displayed. As soon as the page loads, the State dropdown box is filled with options which are dynamically extracted from the ontology. After the user selects a particular state, the county dropdown box is loaded dynamically with its options that are extracted from the ontology. County dropdown box is loaded with all the counties only under the selected state. Also at the same time all the counties of that state are plotted on Google Map. All the Annual and Perennial Crops grown for all the counties under than state are listed in the place marker plotted on Map for each and every county under than state.

If the user selects a particular county, then detailed data such as irrigation used, fertilizer used etc for both annual and perennial crops only for that particular selected county is displayed on the

marker plotted on Google Map. Once a county is selected, annual and perennial dropdown boxes are filled with options dynamically extracted from the ontology.

If a user selected a particular crop from any of this 2 drop-down boxes then detailed data for that selected crop is displayed in the marker plotted on the map for the selected county as shown in Figure 6. Implementation of above functionality has been done using JavaScript and Google Maps API version 3.



Figure 9: Crop Production Characteristics Page

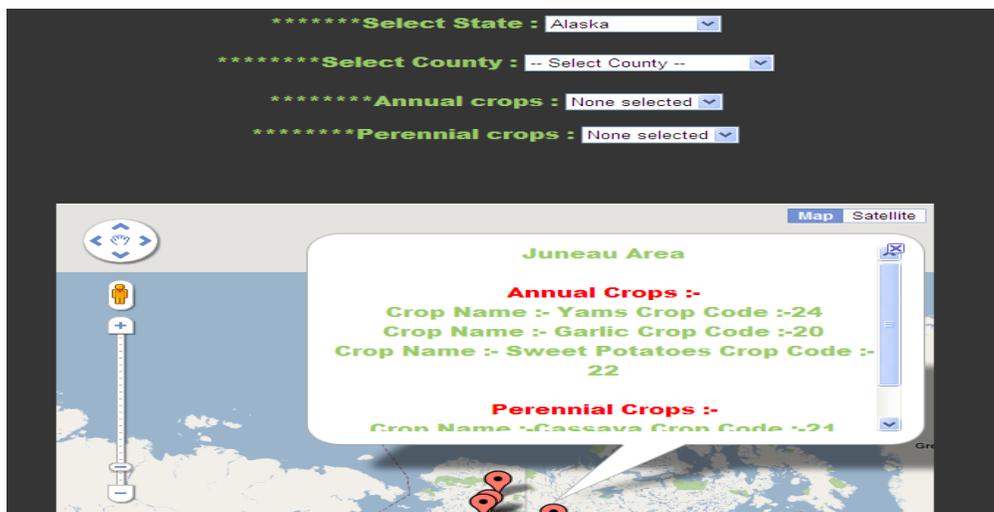


Figure 10: Crop Production Characteristics Page – State Selected

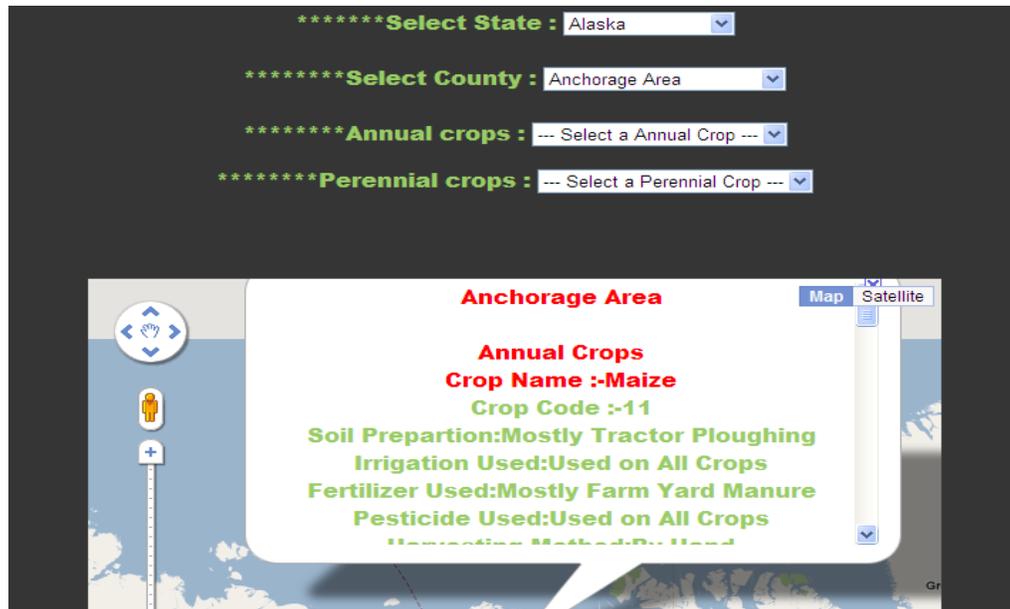


Figure 11: Crop Production Characteristics Page – County Selected

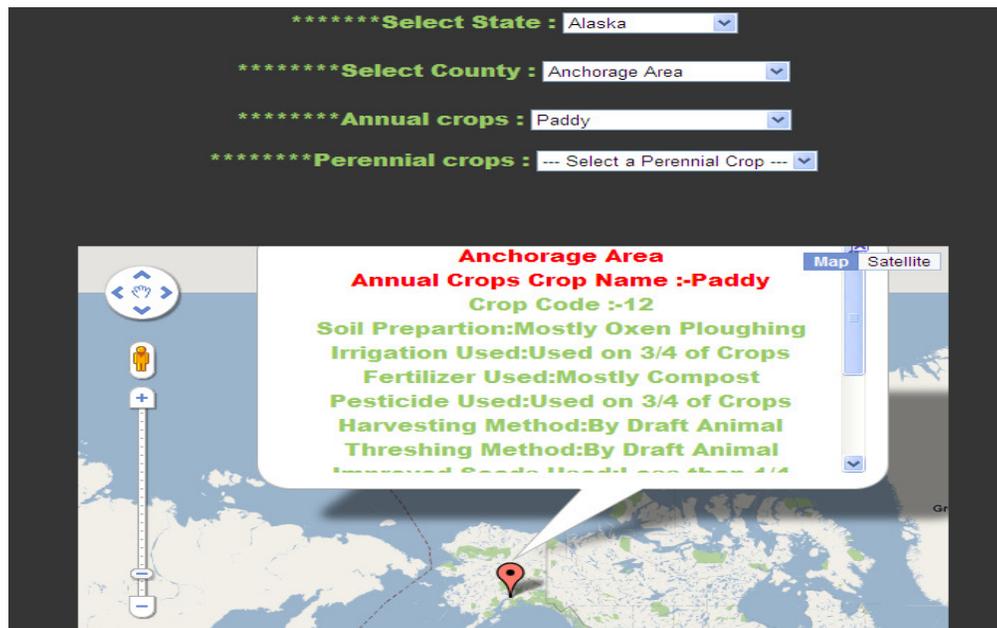


Figure 12: Crop Production Characteristics Page – Crop Selected

### 3.4. Comparison between 2002 and 2007 agricultural census data

This page does comparison between 2002 and 2007 US agricultural census data. Initially as the page loads, State and 'Select a characteristics' dropdown box are displayed. All options for State dropdown box are loaded dynamically from the ontology. For comparison, user needs to select a

particular state and a characteristic for analysis and then when display button is clicked the result of the analysis is displayed interactively using Google Column Charts API as shown in Figure 13. The column chart does comparison for the selected parameter between 2002 and 2007 data for all counties under the selected state.

Example – if a particular state such as “Alaska” is selected and a parameter such as “% of farms with sales less than \$10,000”, then the column chart does comparison between all counties under Alaska for the parameter “% of farms with sales less than \$10,000” for 2002 and 2007 data as shown in Figure 13. So the column chart serves 2 purposes i.e. it provides comparison between all counties under than state and also comparison between 2002 and 2007 data for a particular county. Implementation of the above functionality has been done using JavaScript and Google Column Charts of Google Visualization API [9].

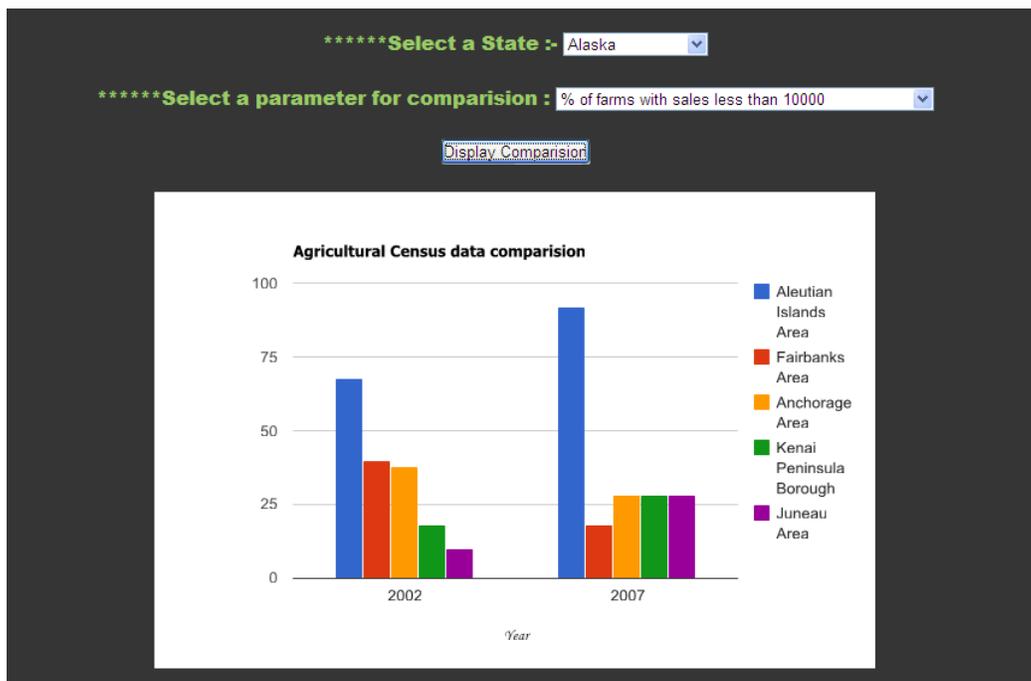


Figure 13: Comparison of 2002 and 2007 Census Data Page

#### 4. CONCLUSIONS

Thus the proposed application in this paper successfully solves various problems related to storage, retrieval and visualization of agricultural census data as explained below.

- Organization/ Storage of Geo-spatial data -Data stored in ontology rather than PDFS, RDBMS or proprietary formats such as shape files.
- Visualization- Data visualized using user-friendly methods such as Google Maps and Column charts. Rather than using ontology as a semantic knowledge modelling tool we have used it to develop Ontology Enhanced User Interface.
- Analysis- Easy to use, fast, user-friendly methods for census data comparison such as Column Charts have been used. Also they serve as baseline data and can be integrated with other agricultural surveys and statistics. Also the various attributes selected help in monitoring the various MDGs as –

- Poverty monitoring and analysis – Attributes such as Average size of farms in acres, % of farms with sales less than \$10,000, % of farms with sales between \$10,000 and \$249,000, Average total farm production expenses per farm, Average value of agricultural products sold per farm etc selected in this application can help to monitor various aspects of poverty.
- Food security monitoring and analysis –Parameters related to food availability, food access, stability of food supplies etc can help to monitor food conditions for a particular area. Attributes such as Area of total cropland as % of land area in acres, Improved Seeds used for that area, Area harvested, and Quantity harvested etc selected in this application can help in food security monitoring.
- Role of women in agriculture – Attributes such as % of farms operated by family, % of farms with female principal operators, Average age of principal farm operator etc selected in this application help to determine the level of participation of women in agriculture and also help to understand the various problems faced by women in operating the agricultural holdings.
- Ensure environment sustainability – Range of parameters such as irrigation used, pesticides used, and minerals used fertilizers used etc can help to determine the condition of environmental resources of a particular area.
- Provide data to the private sector – Attributes such as produce mostly sold to, Area harvested, Quantity harvested, Average value of agricultural products sold per farm, Average value of all machinery and equipments per farm etc selected in this application provide important information to the private sector and help them in making commercial decisions and plan up their various marketing strategies.
- Agricultural planning and policy - making – Study of a specific crop – classification between annual and perennial crops, Structure of agriculture in a particular area etc help in planning various agricultural policies and integrate such information with other areas.
- Providing baseline data for monitoring agricultural development projects – 2002 and 2007 Agricultural Census Data Comparison page is an example of the above application.

## REFERENCES

- [1] Food and Agriculture Organization of the United Nations Rome (2005) - A System of Integrated Agricultural Censuses and Surveys, Volume 1- World Programme for the Census of Agriculture 2010
- [2] ERSI. "GIS Best practices: GIS for Agriculture"- <http://www.esri.com/library/bestpractices/gis-for-agriculture.pdf>, June 2009, [July. 20, 2010].
- [3] Ontology Development 101: A Guide to Creating Your First Ontology - [http://protege.stanford.edu/publications/ontology\\_development/ontology101.html](http://protege.stanford.edu/publications/ontology_development/ontology101.html)
- [4] US Agricultural Census Data - <http://www.nationalatlas.gov/mld/agcen07.html>
- [5] Understanding protégé software - [http://owl.cs.manchester.ac.uk/tutorials/protegeowl/tutorial/resources/ProtegeOWLTutorialP4\\_v1\\_2.pdf](http://owl.cs.manchester.ac.uk/tutorials/protegeowl/tutorial/resources/ProtegeOWLTutorialP4_v1_2.pdf).
- [6] JavaScript and AJAX - [http://en.wikipedia.org/wiki/Ajax\\_\(programming\)](http://en.wikipedia.org/wiki/Ajax_(programming))
- [7] XML DOM - [http://www.w3schools.com/dom/dom\\_intro.asp](http://www.w3schools.com/dom/dom_intro.asp)
- [8] Google Maps API v3 - <http://code.google.com/apis/maps/documentation/javascript/tutorial.html>
- [9] Google Visualization API - <http://code.google.com/apis/visualization/documentation/>