Available online at www.sciencedirect.com

ScienceDirect

Procedia Environmental Sciences 24 (2015) 277 – 284

Procedia Environmental Sciences

The 1st International Symposium on LAPAN-IPB Satellite for Food Security and Environmental Monitoring

Spatial online analytical processing for hotspots distribution based on socio-economic factors in Riau Province Indonesia

Putri Thariqa*, Imas Sukaesih Sitanggang

Department of Computer Science, Bogor Agricultural University, Jl. Meranti Wing 20 Level 5 - 6P, Kampus IPB Darmaga, Bogor 16680, Indonesia

Abstract

Hotspot data analysis will generate more useful information if the data are combined with socio-economics condition. The relation between hotspots and socio-economic conditions is important to be explored in order to identify the human factors that may cause forest fires. In the previous research, a spatial data warehouse and Spatial Online Analytical Processing (SOLAP) system for hotspots distribution were developed. However, the system does not provide other useful information such as socio-economy to analyse the cause of forest fires. The purpose of this study is to create socio-economy dimensions in the spatial data warehouse system.

Selection and peer-review under responsibility of the LISAT-FSEM Symposium Committee

Keywords: land and forest fires; Riau Province; socio-economic condition; spatial OLAP; hotspot

1. Introduction

Land and forest fire is a serious problem which affects the condition of the forest. One activity to detect the occurrence of forest fire is performed by observing the hotspots. A hotspot is an indicator of forest fire occurrence which has a relatively higher temperature than the surrounding temperature [1]. An area will be detected as a hotspot if it exceeds the threshold temperature of 46.85°C [2].

The hotspots provide less information if it is not supported by further analysis and interpretation. Hotspot data will become more useful when it is combined with the regional characteristics such as socio-economic conditions.

* Corresponding author. Tel.: :+62-251-8625-584.
E-mail address: putri.thariqa@gmail.com.

Selection and peer-review under responsibility of the LISAT-FSEM Symposium Committee
doi:10.1016/j.proenv.2015.03.036

Reference [3] states that "the socio-economic factors and ignorance of the population is the main reason for the occurrence of forest fires". Socio-economic factors include population growth, the level of formal education, and employment opportunities. Each of the socio-economic condition is relevant to the occurrence of forest fires. For instance, the relation between forest fires with the increase of population was explained by Mangandar [3], in this research the growing number of local population has caused the logging of forest for opening new land and later on, selling this land to the migrants. Furthermore, there is a relation between forest fires with conditions of employment opportunities where the difficulty in obtaining job while forests provide the opportunity for people to work, especially for the case of forest clearing, logging, land clearing, and others. Lastly, there is a relation between forest fires and the level of formal education where a low level of formal education, especially public understanding to the use of forests as a conservation area could cause of forest fires [4].

Hotspot data which has a large number of objects can be stored into a data warehouse so that the summary of the hotspot data based on the characteristics of the area where the hotspot occurred can be easily obtained. Data warehouses generally include the concept of online analytical processing (OLAP). Displaying data on the map to compare different phenomenon from year to year and combining maps with statistical tables and graphs allows one to obtain better information to handle spatial data. This can be done by combining Geographic Information Systems (GISs) with On-Line Analytical Processing (OLAP) resulting in "SOLAP" (Spatial OLAP) [5]. Spatial OLAP is provided for analyzing large data in a spatial data warehouse, because spatial OLAP (SOLAP) allows the visualization of data through cartography (maps) and non-cartographic displays (e.g., tables) [6]. Thus the analysis of the spatial distribution of hotspots can be performed easily which makes it easier to analyze the causes of forest fires.

Developing a web-based OLAP that has been integrated with the data warehouse and also implements spatial dimensions has been done by Hayardisi and Sitanggang [7] for mapping the distribution of hotspots in Indonesia. The data warehouse implementation adopted the three-tier architecture that consisted of the data warehouse (DBMS), web servers (OLAP server) and web browser. In this system, the data were represented in multi-dimensional models using the star scheme in a data cube that consists of two dimensions namely time and location, and the number of hotspots as a measure. Furthermore, OLAP operations, including roll-up and drill-down were applied to show the summary of hotspot data and to visualize it in the form of tables and graphs [7]. This research was continued by Sitanggang *et al.* [8] which visualized the clustering results from the data taken in the study conducted by Hayardisi [8], to help users for getting the distribution of hotspots groups for every region in Indonesia. However, this system still has some limitations, such as, it is only displaying the data based on time and hotspot areas, and it is not presenting other information that can be used to analyze the cause of hotspot occurrences.

This research aims to create new dimensions to the spatial data in the SOLAP cube so that the system can improve the previous system that was developed by Sitanggang *et al.* [8].The socio-economic dimensions included in this research are population density, school density, and income source of community in Riau Province. These socio-economy dimensions will help users to see the causes of forest fires that were usually caused by human activities. This research expects that SOLAP which presents more complete information may help users to analyze the cause of hotspot occurrences in Riau Province which is essential in forest fire prevention.

2. Data and Method

This study used hotspots data from 2006 to 2008 in the Riau province. The hotspot data were obtained from MODIS FIRMS Fire/Hotspot, NASA/University of Maryland. The socio-economic data including population density, school density, and income source were taken from village data of Riau Province in 2008. The data were collected from Indonesia Statistical Agency(BPS) in the form of text files. Additional data are districts delineation borders in Riau province in vector format that were taken from BPS.

This study was performed in several stages. First, the creation of a new dimension in the data warehouse scheme where the new dimension was added to the data cube using the Schema Workbench software. Furthermore, the data cube scheme was stored in the form of XML file that will be read by the GeoMondrian OLAP server to execute queries requested by the user. The structure of multi-dimensional data cube is adopting the snowflake schema. Later on, the second stage is the system testing. In this stage, testing was performed to determine whether the application complies with the requirements and to evaluate whether the basic OLAP operations were successfully implemented. The testing was done in two parts namely OLAP operations and data queries. The OLAP operations tested are rollup, drill down, and pivoting, while the query testing was done for queries in MDX functions that can handle multi-dimensional data structures. The last stage of the methods is analysis of the test results. The test is considered to be successful the entire function can run properly. However, if there are some functions that do not perform well then causes of the problem are investigated so that the flaw of the system can be corrected later on. The software that was used in this study is as follows:

- Operating system Windows 7 Professional,
- Apache Tomcat 6.0 as the web server,
- Spasialytics as spatial OLAP framework,
- GeoMondrian as spatial OLAP server,
- GeoServer 2.1 as a web map server,
- Geokettle as a data migration tool,
- OpenLayers 2.8 as a JavaScript library for displaying map,
- PostgreSQL 9.1 as a database server with PostGIS extension to manage spatial data, and
- Schema Workbench 1.0 to create multidimensional data cubes.

3. Results

This work adopts the snowflake scheme for creating multi-dimensional data cube (Fig. 2). Snowflake scheme consists of a fact table and six dimensional tables. Fact_forest fire table is the act table at the center of the datacube, this table has a single measure that is the number of hot spots area captured by satellite within a certain time. Dimension table consists of a time, satellite, location, school density, population density, and income source dimension. Time dimension is written on the_time table which has three levels, namely the year, quartiles, and month. Satellite dimension is written on the_satellite table that has one level, which is the name of the satellite that detects the occurrence of the hot spots areas. Location dimension are written in the geo hotspot table which has two hierarchies such as, district table connected with geo hotspot table and province table connected with district table. Income source dimension is written on in source table, this table contains the types of income source of the people. School density dimension is written on the school table, this table contains the classification of the level of school density such as, low (school density <= 0.1), medium (0.1 <school density <= 0.2), and high (school density >0.2). Population density dimension is written on population table which in this table contains a population density class that is low (population density <= 50), medium (50 <population density <= 100), and high (population density>100).

To add the socio-economic dimensions and display it in the existing SOLAP system requires several steps. First, prepare the table dimensions and import the data in the shapefile format into database. Furthermore, the spatial join was done to retrieve hotspot data which have the socio-economic information. The final step is to update the fact table.

Data presented in this system consists of two modules which are Jpivot module and map module. Jpivot module displays tables and graphs of the data from the query execution while the map module displays cartographic visualization of the data that consists of Riau map layer and the layer of hot spots.

SOLAP work flow system was described in Fig. 1. Based on Figure 1 the flow SOLAP system can be described through the following stages:

(a) To run the system, the users submit a query input from the MDX query editor in Spatialytics client. This query is received and handled by olap4js.
(b) Olap4js send MDX queries that the user entered into Spatialytics server. While spatialytics olap4js server receives the query execution results in the form of JSON to be checked in the multi-dimensional data structures, such as hierarchies, members, levels, elements, and dimensions.
(c) Queries in spatialytics server are forwarded to the Geomondrian for in execution. After the query process is executed, the result will be returned to the server in the form of Javascript Object Notation (JSON).
(d) GeoMondrian connects with the JDBC Driver that acts as a connector interface with PostgreSQL. While the data is retrieved through the JDBC driver sent to Geomondrian.
(e) JDBC driver retrieves data from PostgreSQL, and PostgreSQL restores the required data to the JDBC driver.
(f) After checking, the JSON data will be sent to the module of Featurizer to do the parsing process. This parsing process is used to obtain spatial features of the object or member.
(g) The object features obtained were sent to SOLAP Context to be collected with other data that will be needed to display a map.

(h) SOLAP Context sends data that has been collected to Open Layers to be visualized.
(i) GeoServer connects with the JDBC driver to take a basic map from the data base while JDBC driver will return a map that takes from GeoServer.
(j) Base map that has been obtained is sent to Open Layers to be visualized.
(k) The results of the query execution GeoMondrian will be sent directly to the module JPivot.

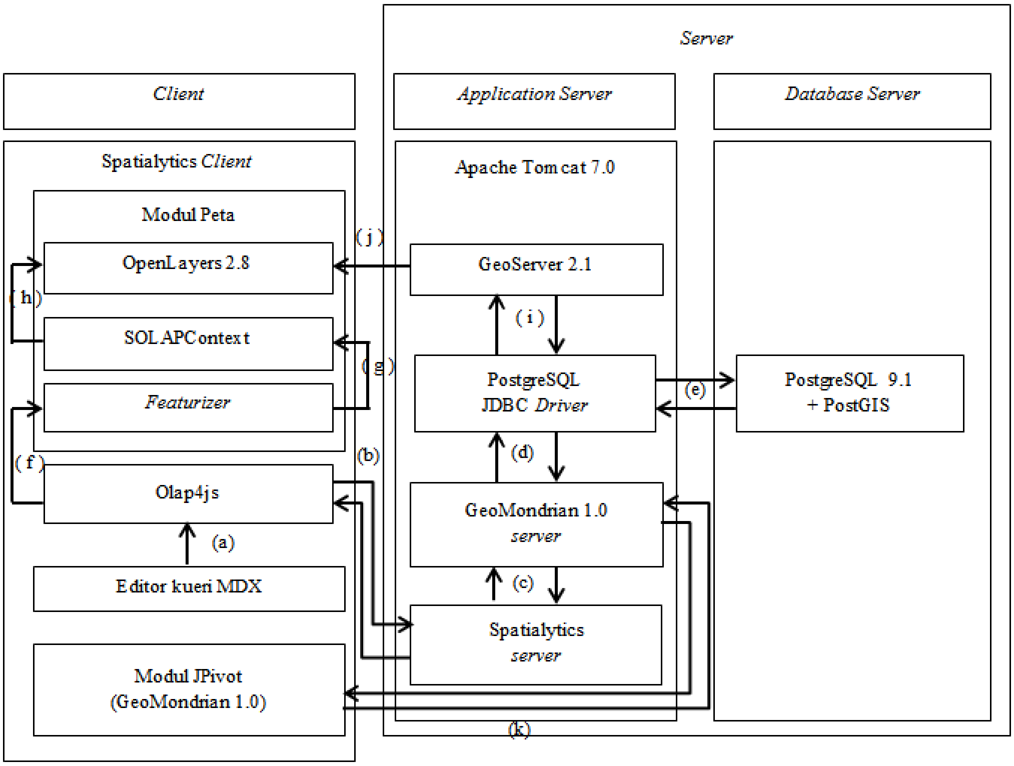

Fig.1. Workflow of SOLAP system

In order to evaluate the SOLAP, queries testing and basic OLAP operations testing were conducted. These procedures executed several queries to retrieve hotspot data based on the socio-economic conditions, time and locations.

3.1. Query Retrieving Location of Hotspots Based on Population Density

The following query retrieves number of hotspots in the district Pelalawan, Kuantan Singingi, Indragiri Hulu, and Indragiri Hilirin 2006 with the high population density.

```
SELECT
{[Measures].[Hotspot_Total]} ON COLUMNS,
NON EMPTY Crossjoin(
{[location].[All Riau].[RIAU].[PELALAWAN].Children, [location].[All
Riau].[RIAU].[KUANTAN SINGINGI].Children,
[location].[All Riau].[RIAU].[INDRAGIRI HULU].Children, [location].[All
Riau].[RIAU].[INDRAGIRI HILIR].Children},
{[population].[Population Class].[High]}) ON ROWS
FROM [geohotspot]
WHERE [Time].[AllTime].[2006]
```

The number of hotspots in 2006 for the fourth district is 4121 and the query process was resulted in 349 hotspots. SOLAP views and JPivot (Geomondrian) that display the query results can be seen in Fig. 3.

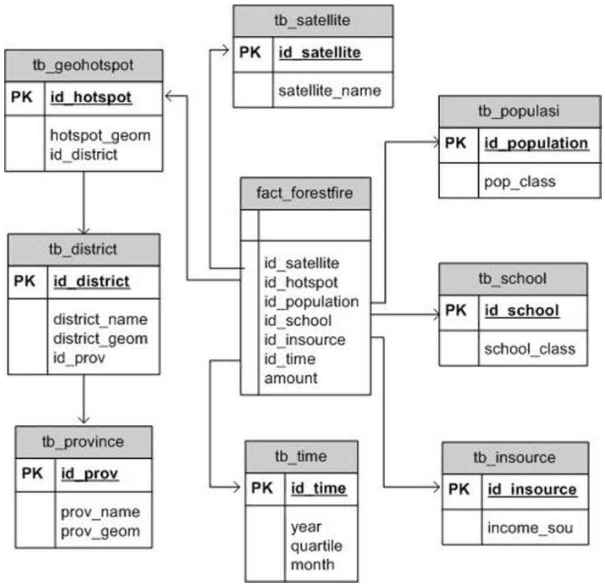

Fig. 2. Scheme of multi dimensional data cube

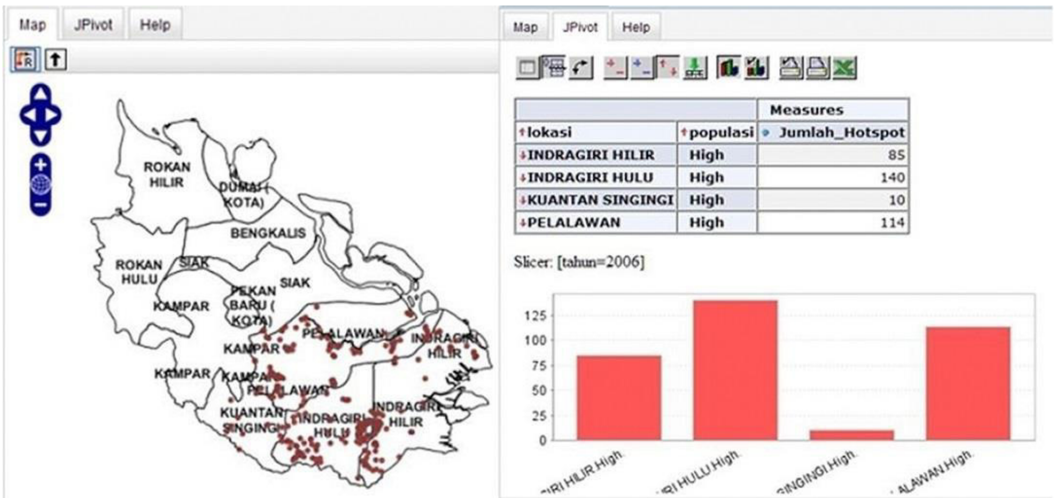

Fig. 3. SOLAP and J-pivot output that display hotspots in the area with high population density

3.2. Query Retrieving Location of Hotspots Based on School Density

Testing of this query is similar to the previous query testing. In this query test data used is the hot spot data of 2007 in the district/city Bengkalis, Pelalawan, Dumai (City), Rokan Hilir, Rokan Hulu, Siak, and Indragiri Hulu. A Socio-economic condition is shown is a low density schools.

```
SELECT
{[Measures].[Jumlah_Hotspot]} ON COLUMNS,
NON EMPTY Crossjoin(
{[location].[All Riau].[RIAU].[BENGKALIS].Children, [location].[All
Riau].[RIAU].[PELALAWAN].Children, [location].[All Riau].[RIAU].[DUMAI
(KOTA)].Children, [location].[All Riau].[RIAU].[ROKAN HILIR].Children,
[location].[All Riau].[RIAU].[ROKAN HULU].Children, [location].[All
Riau].[RIAU].[SIAK].Children,
[location].[All Riau].[RIAU].[INDRAGIRI HULU].Children}, {[school].[School
Class].[Low]}) ON ROWS
FROM [geohotspot]
WHERE [Time].[AllTime].[2007]
```

For a low density schools generates 3124 hotspots, and the total number of hotspots in 2007 was 3449. SOLAP views and Jpivot that displays query results can be seen in Fig. 4.

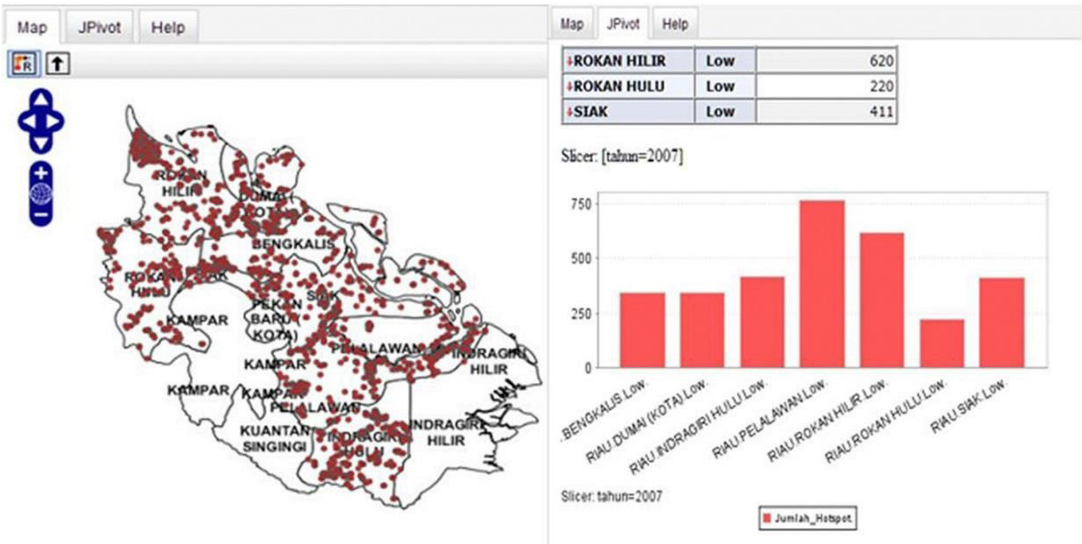

Fig. 4. SOLAP and Jpivot output that displays hotspots in the area with the low school density

3.3. Query Retrieving Location of Hotspots Based on Income Source

The third query testing using data of hotspots in 2008 in the district/city of Bengkalis, Pelalawan, Dumai (City), Rokan Hilir, Rokan Hulu, Siak, Indragiri Hulu, Indragiri Hilir, and Kuantan Singingi. Income source of a resident who was used to test this query only includes plantation, forestry, and services. The query that was used is as follows:

```
SELECT
{[Measures].[Jumlah_Hotspot]} ON COLUMNS,
NON EMPTY Crossjoin(
{[location].[All Riau].[RIAU].[BENGKALIS].Children,
[location].[All Riau].[RIAU].[PELALAWAN].Children,
[location].[All Riau].[RIAU].[DUMAI (KOTA)].Children, [location].[All
Riau].[RIAU].[ROKAN HILIR].Children, [location].[All Riau].[RIAU].[ROKAN
HULU].Children, [location].[All Riau].[RIAU].[SIAK].Children,
[location].[All Riau].[RIAU].[INDRAGIRI HULU].Children, [location].[All
Riau].[RIAU].[INDRAGIRI HILIR].Children, [location].[All Riau].[RIAU].[KUANTAN
```

```
SINGINGI].Children}, {[insource].[Income Source].[Plantation],
[insource].[Income Source].[Forestry],
[insource].[Income Source].[Services]}) ON ROWS
FROM [geohotspot]
WHERE [Time].[AllTime].[2008]
```

From the results of these queries, there are 150 hotspots with services source, 2278 hotspots with plantation resource, and there are 1834 hotspots with forestry resource. Jpivot views and SOLAP that displays query results can be seen in Fig. 5.

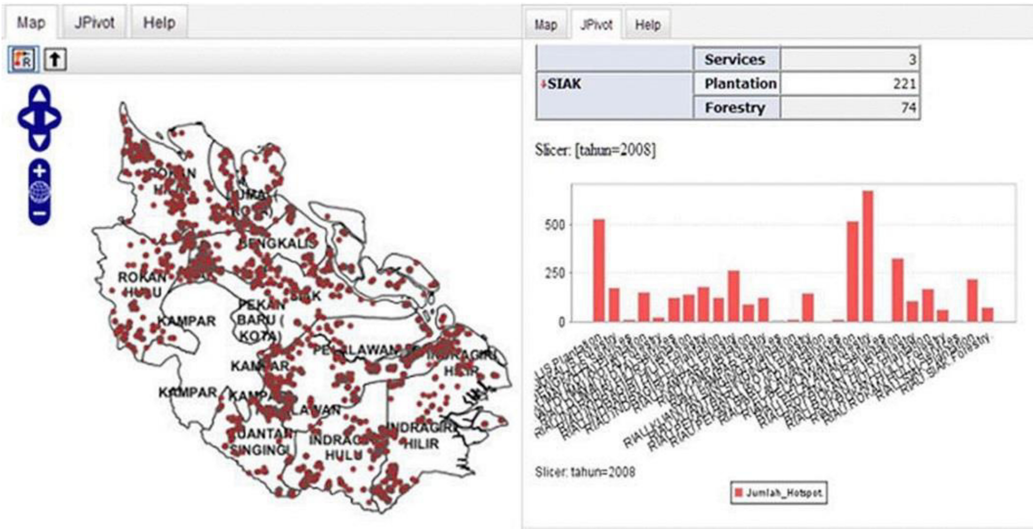

Fig. 5. SOLAP and Jpivot output that display hotspots in the area with multiple sources of income of a resident

The system testing by executing several queries shows that most hotspots are occurred in a low population density and low density schools. In addition, most hotspots are occurred in the areas where income source of the community is plantation. The maximum number of hotspots that can be shown in this study is 5571 points.

Sometimes the distribution of hotspots does not appear on the map while Jpivot was successfully executed the query requests. In this situation, queries and thematic style were not appropriately used. Children function was used to display the distribution of hotspots on the map because this function is useful to return a set of all nodes below the level of the member that is still in use [9]. Thus, in order to display the distribution of hotspots with certain socio-economic conditions, the cross join function is required. This function will return the combined of two or more dimensions [9].

4. Conclusion

This work has successfully created the socio-economic dimensions in addition to time and location dimensions in the data warehouse for hotspot distribution. The socio economic dimensions include population density, school density and income source in the villages in Riau Province. The data warehouse was integrated with the SOLAP (Spatial Online Analytical Processing) system that adopted the snowflake scheme as the structure of multi-dimensional data cube. The SOLAP system has several operations to create summary of hotspots based on the time, location and socio-economic dimensions. The operations are roll up, drill down, and pivoting that were run by performing the MDX query input. Summaries of the hotspots data are the results of these operations that were displayed in form of maps, tables, and charts. Based on this study, most of hotspots are occurred in a low population density and low density schools. In addition, most hotspots are occurred in the areas where income source of the community is plantation. By including these socio-economic dimensions, users will be able to analyze the cause of hotspots occurrences based on socio-economic conditions.

Reference

1. [MENHUT] Menteri Kehutanan. Peraturan Menteri Kehutanan Republik Indonesia Nomor. 12 Tahun 2009 tentang Pengendalian Kebakaran Hutan. Jakarta: Menhut; 2009.
2. [Dephut] Departemen Kehutanan. A guide to hotspots and forest fires in Sumatra [Online]. Available: http://www.dephut.go.id/INFORMASI/PHPA/FFPCP/articles/FFPCP%20%20Hotspot%20Guide.htm; 2013.
3. Mangandar. *Keterkaitan Sosial Masyarakat di Sekitar Hutan dengan Kebakaran Hutan: Studi Kasus di Propinsi Daerah Tingkat I Riau*. Thesis. Bogor: Program Pascasarjana, Institut Pertanian Bogor; 2000.
4. Hardjanto. *Dampak Sosial Ekonomi Kebakaran Hutan dan Lahan di Indonesia. Integrated Forest Fire Management In Indonesia Phase I: National Guidelines on the Protection of Forest Against Fir*". Bogor: Institut Pertanian Bogor; 1998.
5. Bédard Y, Rivest S, Proulx M-J. *Spatial On-Line Analytical Processing (SOLAP): Concepts, Architectures and Solutions from a Geomatics Engineering Perspective*. Canada: Laval University; 2005.
6. Viswanathan G, Schneider M. *On the Requirements for User-Centric Spatial Data Warehousing and SOLAP*. Florida: University of Florida; 2011.
7. Hayardisi G, Sitanggang IS, Syaufina L. Data warehouse and web-based OLAPfor hotspot distribution in Indonesia. In: Conference on Data Mining and Optimization; Malaysia, 27-28 Oktober 2009.
8. Sitanggang IS, Fuad T, Annisa. K-Means Clustering Visualization of Web-Based OLAP Operations for Hotspot Data; 2010.
9. Whitehorn M, Zare R, Pasumansky M. *Fast Track to MDX*. London: Springer; 2002.