Mining Online GIS for Crime Rate and Models based on Frequent Pattern Analysis

John David Elijah Sandig, Ruby Mae Somoba, Ma. Beth Concepcion and Bobby D. Gerardo, *Member, IAENG*

***Abstract*— The researchers proposed to develop a mechanism in mining an Online Graphical Information System (GIS) for crime rate and models using frequent pattern analysis. It is a web-based system that includes GIS for robbery, homicide and physical injury incidents within Iloilo City. This system can help police identify where and what time crime frequently happen. It generates who are the frequent victims in the community. The main focus of the proposed system is the analysis of frequent crime patterns and its association rules and crime hot spots. Results show the hot spots which displays clustered crimes in the Google Map. It uses the Google Heat Map layer in clustering plotted crimes. These hot spots are showing the density of crime. Furthermore, the system is able to generate frequent patterns of crimes. This uses Apriori algorithm in generating the results.**

***Index Terms*— Data Mining, GIS, Apriori algorithm, frequent pattern, Association rules, clustering, cloud computing, crime mapping, hot places, hot prey; hot spot**

I. Introduction

There is a direct correlation between poverty and crimes nowadays, that people's life is not getting any better. Crimes can be everywhere. People, even police are not even aware that a specific crime is already rampant. Police departments take time analyzing records just to know the current crime rate. It is observed that they only identify crime rates every end of the month.

In the advent of technology, provision for crimes continues to improve. One of the priorities is crime reduction and avoidance. Police awareness is the most relevant factor that may lead to reduction of crime rate or fast response to crimes. Police department cannot easily visualize where or what time certain crimes frequently happen.

Data mining is the process of discovering patterns in the data [1]. Using data mining in crime mapping can help authorities identify frequent crime patterns and hotspots that can help in crime intelligence.

The main objective of this study is to develop an algorithm that will generate frequent crime pattern and occurrence using Apriori Algorithm. The proposed system will aid police in identifying crime incident rate status of Iloilo City. The web-based system can generate patterns within a certain month, year or specific period using Classification Algorithm. It can also generate crime hot spot showing the density of crime in an area using a Google heat map.

Manuscript received August 16, 2013.

John David Elijah B. Sandig and Ruby Mae A. Somoba, belong to the Institute of Information and Communications Technology, West Visayas State University, Iloilo City, Philippines.

Prof. Ma. Beth S. Concepcion, is the College Secretary, Institute of Information and Communications Technology, West Visayas State University, Iloilo City, Philippines.

Prof. Bobby D. Gerardo, Ph.D. is an IAENG Member and currently a Full-Fledged Professor at Institute of Information and Communications Technology, West Visayas State University, Iloilo City, Philippines.

II. Related Systems

The Abeline Police Department uses a computer-based crime mapping program called RAIDS (Regional Analysis and Information Data Sharing) Online to display crime data to the public using Google Maps' well known, intuitive mapping interface [2].

The user is able to search a city or address and then select what crimes they would like to see and the date range of these crimes. A map will come up with pin points of these crimes. The user can click on the Analytic tab and this will pull up several charts. These charts show the selected crime in different ways. The user can see the crime by day of the week and even by the hour of the day. There is a feature that shows crime in the form of a density map, where the colors are hotter when there is more crime in that area [2].

On the other hand there is a web-based tool called GENECODIS that integrates different sources of information to search for annotations that frequently co-occur in a set of genes and rank them by statistical significance. The analysis of concurrent annotations provides significant information for the biologic interpretation of high-throughput experiments and may outperform the results of standard methods for the functional analysis of gene lists [3]. The Apriori algorithm is used to find sets of annotations that frequently co-occur in the input list [3].

In the study Cluster analysis for identifying sub-groups and selecting potential discriminatory variables in human encephalitis, cluster analysis is implemented with the aim of identifying sub-groups in human encephalitis [4]. Patient information were used to perform clustering of variables which include: (1) symptom variables such as fever, lethargy and confusion, (2) exposure variables such as water exposure, tick and mosquito bites, (3) diagnostic/laboratory measurements from brain scans and measurements from cerebrospinal fluids to identify sets of variables that are clustered in some way [4]. This allows users to explore similarities and differences between the groups of variables.

ISBN: 978-988-19253-1-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

III. SYSTEM ARCHITECTURE

The system uses data mining techniques to generate useful results for police purposes. The system analyzes the frequent patterns of crimes and clusters of it. The results are presented using Google applications such as map, line graph, bar graph and pie chart. The system also generates the crime rates, trends and hot places in Iloilo City.

Figure 1 shows the main architectural design of the system. Crime data are processed to generate crime rates, trends, hot spots, hot places, hot preys and frequent patterns of crimes. Clustering is a data mining technique used to analyze data objects without consulting a known class label [5]. Clustering analysis together with Google heat map layer is used in clustering crimes to determine crime hot spots. Frequent patterns have found broad application in the area like association rule mining, and clustering [6],[7]. In this paper, frequent pattern analysis is used to generate frequent crime patterns and its associations, and hot preys.

Figure 2 shows the detailed processes of frequent pattern analysis in determining frequent crime patterns and hot preys. Apriori algorithm is the best-known pattern mining algorithm whose aim is to find frequent itemsets [8]. Apriori algorithm was used to generate frequent crime patterns and as basis to generate hot preys. Apriori algorithm was also used to extract frequent crime patterns and its association rules through five processes.

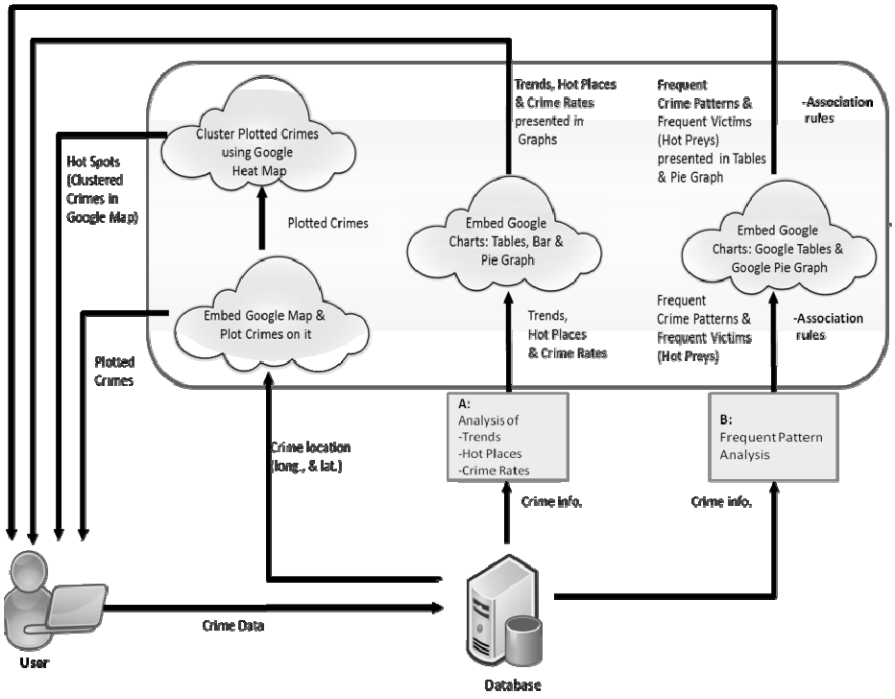


Fig. 1 System architecture of the proposed system

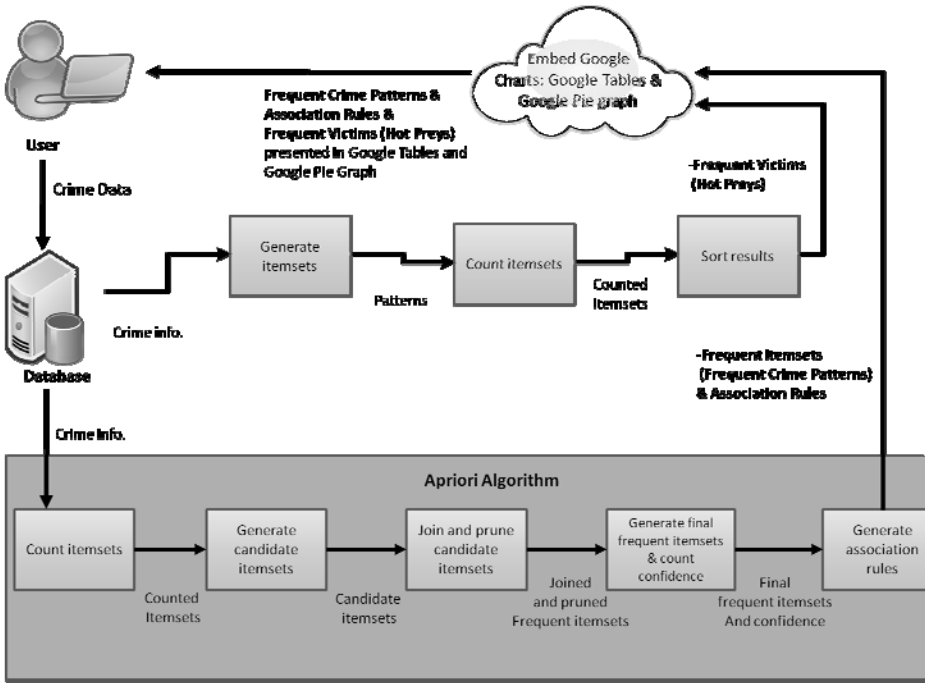


Fig. 2 System architecture for frequent pattern analysis

ISBN: 978-988-19253-1-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

IV. PROPOSED SYSTEM

The proposed system aims to alert police on the current crimes status and crime trends and to give them insights on when and where crimes frequently happen. These will guide police on their investigation and help them prevent future crimes to happen.

It is a web-based system that uses cloud computing. Google applications were utilized as a service provider and for the overall presentation of the system functions specifically in the entry of data in the system. Google Chart was for displaying the crime rates, trends and frequent crime patterns. Google Heat Map was used for clustering crimes. Google Map was utilized to plot the crimes.

The system was developed using PHP, CSS, HTML, Json, Ajax, Jquery and Javascript as its programming languages. It is anchored on a client and server framework. The server holds the database. It uses MySql as its server scripting language. The client is any computer that access to the internet.

The main users of the system are the police in each precinct who are authorized to use the system. The system requires them to input victim's(s) data, suspect's(s) data and crime data. In the crime data, the user is required to plot the exact location of the crime in the embedded Google Map.

The system has four major components, specifically, the "Crime Rate", the "Trends", the "Maps and Analysis" and the "Search".

The "Crime Rate" generates the current and annual crime rate status of Iloilo City. It is based on the current population of the city of Iloilo by a ratio of number of crimes per 100,000 persons. Google Charts, specifically, Google Bar Graph is used to represent the results.

The "Trends" generates trends of crimes in Iloilo City. It generates what crime is high or low in a specific time, thus, the police can identify seasonal occurrence of crimes. Google Line Graph is used to display the crime trends.

The "Maps and Analysis" has five sub-components specifically, the 'Plotted Crimes', the 'Hot Spots', the 'Hot Places', the 'Hot Preys' and the 'Patterns'. These sub-components are categorized into two, namely: Maps and Charts. The 'Hot Spots' and the 'Plotted Crimes' are under the Maps category while the 'Hot Places', the 'Hot Preys', and the 'Pattern' are under the Charts category. In these components, a classification algorithm was used to classify the time. The user sets the range of time he or she wants the results will be based. The algorithm uses 'from and to what time' as time classifier.

In the Map category, users can view crimes in Iloilo City plotted in Google Map and the clustered crimes on it. They can customize the settings of the map based on what they want to see. The system can generate specific results provided the users will choose the parameters like crimes on what district, what kind of crime and from and to what time the specific crime happened. The changes made by the user will only apply until the user logs-out or closes the browser.

The 'Plotted Crimes' displays the locations of crimes pin pointed in the Google Map. Crimes are presented in different colors correspondingly.

The 'Hot Spots' displays clustered crimes in the Google Map. It uses the Google Heat Map layer in clustering plotted crimes. The users can set the standard of radius and intensity of crimes to be clustered.

In the Charts category, users can view graphs of places where crimes are high or low, victim patterns, and frequent patterns of crimes. They can set specific results by choosing on what district the results may come from, from and to what time crimes happened, what type of crimes they wanted to analyze, order of the result and the number of results they wanted to view.

The 'Hot Places' displays the number and percentage of crimes in each district and barangays represented in Google pie graph.

The 'Hot Prey' generates the patterns of victims attacked. The researchers developed an algorithm generating these results based on Apriori algorithm. The system generates patterns and counts each itemsets. It then determines frequent patterns of victims. Also, the results are represented in Google Pie Graph and Tables.

The Pattern generates frequent patterns of crimes. It uses Apriori algorithm in generating the results. First, the system counts the k-itemsets from the database. Then, it generates candidate k-itemsets where the counts of k-itemsets are equal or higher than the minimum support count required by the system. These candidate k-itemsets are then considered as frequent itemsets. These frequent k-itemsets will undergo joining and pruning to generate frequent k+1 itemsets. Frequent k-itemsets is combined in itself and removes infrequent itemsets and those which have infrequent subsets. Frequent K+1 itemsets are then generated. The processes are repeated until the frequent itemsets become null. The algorithm is terminated having found the global frequent itemsets and generation of Association rules starts.

Confidences of frequent itemsets are determined using the formula presented in equation (1).

$$Confidence(AtoB) = P(B \mid A) = \frac{\sup\_count(A \cup B)}{\sup\_countA} \quad (1)$$

If the confidence of the frequent itemset is greater than or equal to the minimum confidence threshold, the frequent itemset is considered under the association rules and becomes the final output of the Pattern. These results are represented in Google Pie Graph and Tables.

V. SIMULATION AND RESULTS

Figures 3-8 show the important input and output of the system. These are the generated results of the different analysis and processes in the system.

Figure 3 shows the trends of crime accumulated in a particular year depicting the various crimes such as physical injuries, robbery and homicide. On the other hand Figure 4 shows the plotted crimes on the map while Figure 5 shows the map option demonstrating filters for use for generating patterns of crimes.

One output of the simulation results is the generated crime hotspots overlaid on the Google heat map as depicted in Figure 6. And finally Figures 7 and 8 demonstrated the pie charts for the crime hot places and frequent crime patterns, respectively.

ISBN: 978-988-19253-1-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

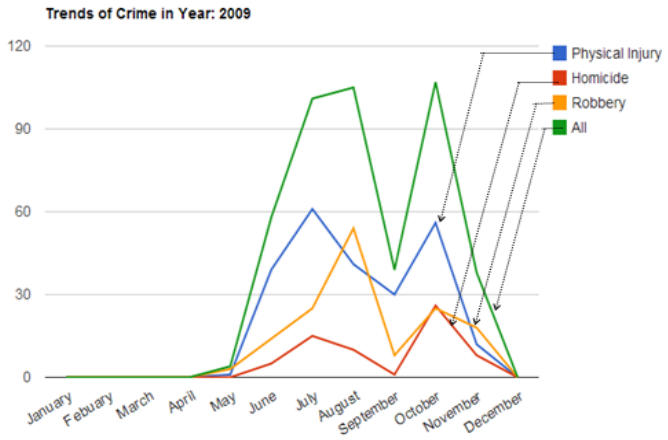


Fig. 3 Monthly Crime trends of different sample crimes

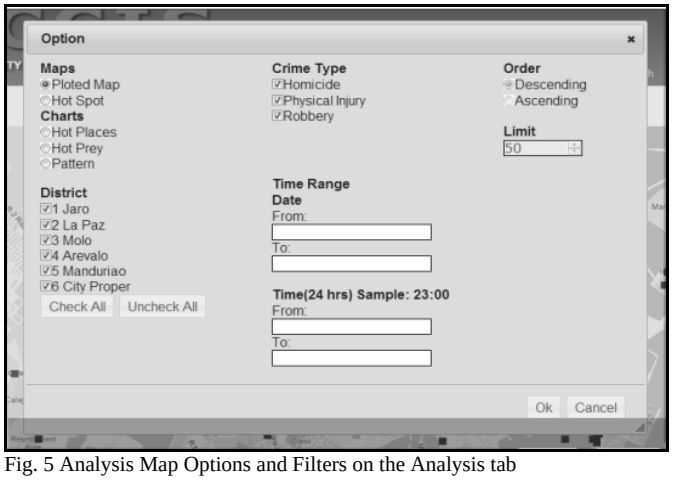


Fig. 5 Analysis Map Options and Filters on the Analysis tab

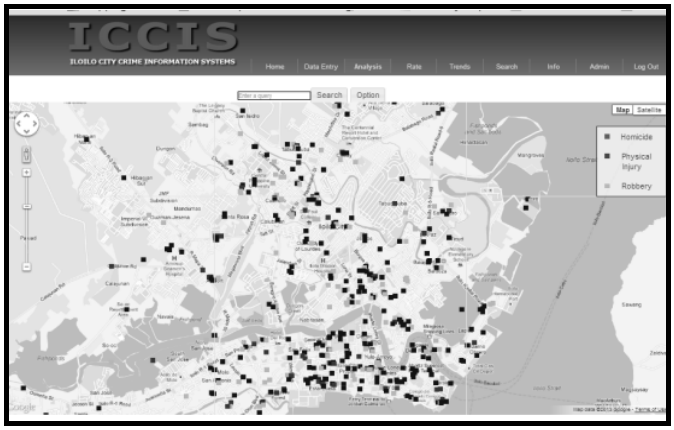


Fig. 4 Default map on the Analysis page that shows plotted crimes

Fig. 6 Generated Crime hotspots using the Google Heat Map Layer

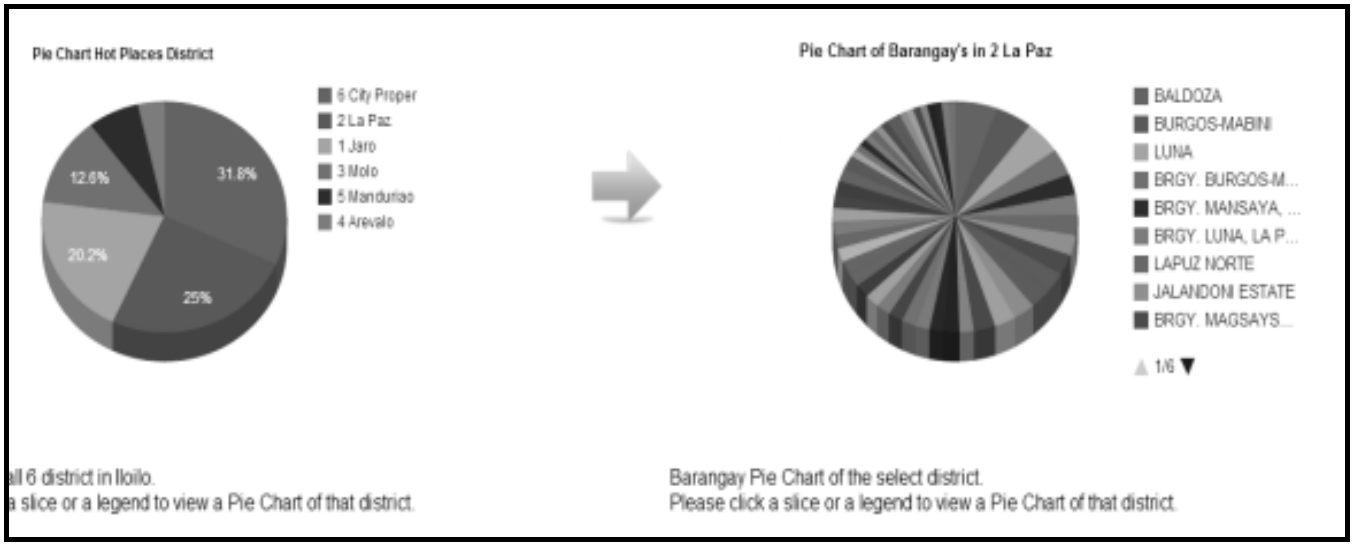
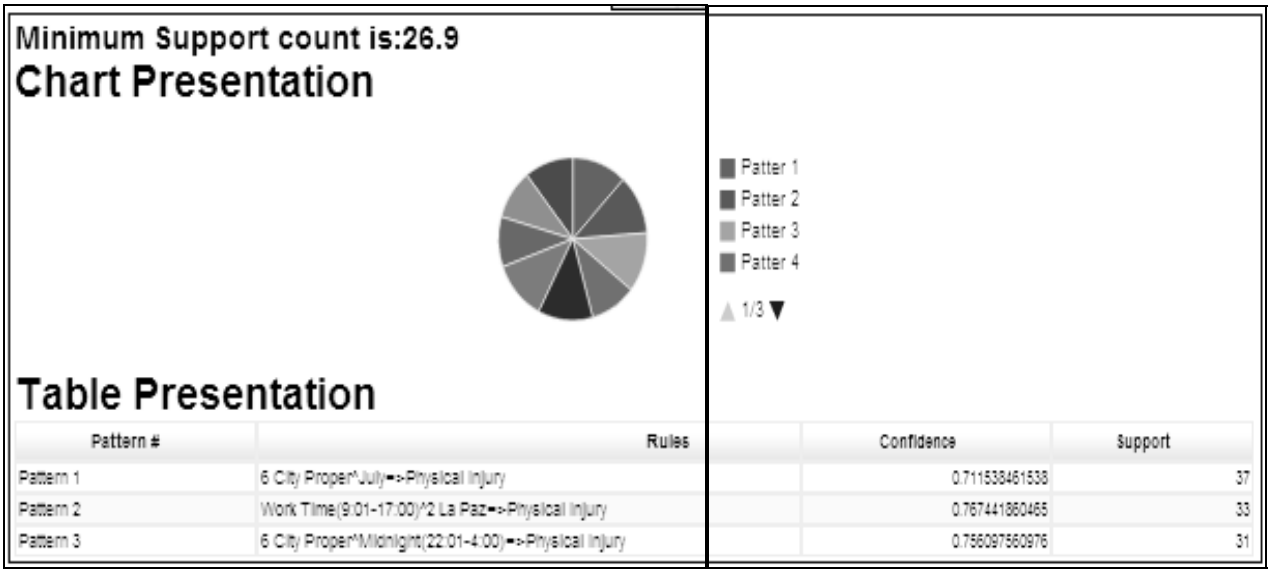


Fig. 7 Generated Crime hot places presented using Google chart tools

ISBN: 978-988-19253-1-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)



Pattern #	Rules	Confidence	Support
Pattern 1	6 City Proper^July=>Physical Injury	0.711538461538	37
Pattern 2	Work Time(9:01-17:00)^2 La Paz=>Physical Injury	0.767441860465	33
Pattern 3	6 City Proper^Midnight(22:01-4:00)=>Physical Injury	0.756097560976	31

Fig. 8 Generated frequent crime patterns

VI. CONCLUSION

An Online GIS was developed to help authorities identify the current crime rate in Iloilo City. The system was able to generate annual and monthly crime rate using the Google chart tools.

The system was able generate frequent crime patterns using Aprioiri algorithm. The result was presented using the Google chart tools showing the patterns in both pie chart and tabular form. The user can limit the factors that can affect the generated pattern results using the filters in the system.

The researchers were able to develop a web-based system that is able to generate patterns within a certain month, year or specific period with the help of Classification algorithm.

The researchers were able to develop a system that can generate crime hot spots showing the density of crime in an area using a heat map with the aid of Google map heatmap layer. The user can therefore determine the maximum limit for the heatmap as well as its density radius. The system can generate a heatmap that shows three colors (red, yellow and green) to indicate the crime intensity.

VII. RECOMMENDATION

To further improve the effectiveness and efficiency of the system, the researchers recommend to create own map system for the greater control of the Online Graphical Information system environment. Using online services such as Google map may not be reliable and available all the time because any failure on their service may affect the reliability and efficiency of the results. Lastly, this can be improved into a system that will not only identify crime patterns, trends and hot spots but also predict crime occurrences and establish link between crimes.

ACKNOWLEDGMENT

The proponents wish to acknowledge the contribution of Vicente G. Albonian Jr. and Jimmel A. Samillano for their significant assistance in the development of the system. They would also like to thank their Professors, Ma. Beth S. Concepcion and Dr. Bobby D. Gerardo for their supervision, suggestion and support for the completion of this research.

REFERENCES

[1] Witten, I. H., & Frank, E., (2005). Data mining: Practical machine learning tools and techniques (2nd Edition ed.). Morgan Kaufmann.
[2] Abilene Police Department, “Abilene Police Department Information Release,” http://www.abileneaor.com/member_files/4/documents/254.pdf. January 18, 2011.
[3] Saez, M., et al.”GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists,” http://genomebiology.com/content/8/1/r3/. January 4, 2007.
[4] Hamid, J., et al.” Cluster analysis for identifying sub- groups and selecting potential discriminatory variables in human encephalitis.” http://www.biomedcentral.com/1471-2334/10/364. December 31, 2010.
[5] Han, J. & Kamber, M., (2006). Data mining: Concepts and techniques. Definition of data mining, association rules and clustering.
[6] Agrawal R., Imielinski T., and Swami A., (1993). Mining association rules between sets of items in large databases. In Proc. of SIGMOD, pages 207–216,.
[7] Wang K., Xu C., and Liu B., (1999). Clustering transactions using large items In Proc. of CIKM, pages 483–490,
[8] Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., & Verkamo, A. I., (1996). Fast discovery of association rules. In U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, & R. Uthurusamy (Eds.), Advances in knowledge discovery and data mining (p. 307-328). AAAI/MIT Press
[9] Dunham, MH, (2002) Data mining: Introductory and advanced topics. Prentice Hall, August 2002.
[10] Liu, B., Hsu, W., and Ma, Y., (1998) Integrating classification and association rule mining. In Proc. of KDD, pages 80–86