

Ain Shams University  
Faculty Of Computer  
& Information Science  
Computer Science Department

---

# Using Spatio-Temporal Relationships For Data Fusion And Mining Of GIS

---

A Thesis Submitted To The Department Of Computer Science, Faculty Of Computer  
& Information Science, Ain Shams University, In Partial Fulfillment Of The Requirement For  
The Degree Of Master Of Computer And Information Sciences

By :

Ahmed Abd El Rahman Abd El Lateef Osman

B.Sc. In Computer Science

Faculty Of Computer & Informatics, Suez Canal University.

Demonstrator At Information System Department,

Faculty Of Computer & Informatics,

Suez Canal University- Ismailia - Egypt

Supervised By:

Prof.Dr./Ahmed Hamad

Prof .Dr/ Mohammed

Hashem

The dean, Faculty Of Computer  
& Information Science. October  
University

IS Dept., Faculty Of Computer and  
Information Sciences.  
Ain Shams University

Prof.Dr./M.Helmy A.Mahran

The Previous Dean, Faculty Of Computer And Informatics, Suez Canal  
University, Professor Qassim University – Saudai Arabia

Faculty Of Computer and Information Science

Ain Shams University

Cairo □□□□

## **Abstract**

Many different database locations have to be searched to get complete geographical data about an object. The process of collecting geographical data from all resources and stores them in only one location is called Data Fusion. Similarity search algorithm is the most known algorithm for performing this kind of data fusion. This algorithm is based on searching and comparing in all resources sequentially which increase its complexity and wastes a lot of time, effort and space.

In this thesis, we propose a proposed algorithm named tree algorithm to minimize the fusion complexity of GIS data. The proposed algorithm is based on the simplicity of searching and comparing sorted trees. The resulting database can enhance the GIS description and reduce the search complexity.

This research introduces an evaluation of the proposed algorithm and compares it with the similarity search algorithm.

The fusion process may results in large database after combining different databases in only one. Mining these large databases to extract knowledge from it is a complex process. The most popular data mining technique -association rules technique- has proved very efficient in mining theses large databases. However, if the database to be mined has descriptive fields, as is the case in GIS databases, the technique may lead to wrong results. Some GIS fields have a special nature that prevents their categorization (the process which the association rules technique depends on), and thus are not amenable for manipulation by the

normal association rules technique. To overcome this deficiency, we propose in this research a modification to this technique.

The modification is a new algorithm that depends on skipping the categorization steps in the normal association rules technique. The algorithm uses instead some statistical methods, such as the correlation coefficient.

The modified technique is applied to a real GIS database---the demographical database of the USA. When compared with the expected results, the results obtained by the modified technique have proved twice as accurate as those obtained by the normal technique.

To give the results some practical value, they have been used to develop a tiny Decision Support Systems (DSS). Namely, the extracted rules are incorporated in the decision making of GIS-related problems.

Decision support systems play a very important role in our every day life, since no one can live without taking decisions at home, work even in street. These decisions must be based on perfect background and experience and previous studies in order to be valuable and correct. Decisions with no pre-calculation and studies may lead to bad results. Different data have to be searched to get complete view about any object which consumes a lot of time. Data fusion is the process that combines the data found in many different locations in one location so that all required data could be found in one place.

Data mining process is most used to get perfect data and knowledge about some thing, usually in the process of pre-

studying before the decision making process. Association rules technique is the most famous technique for data mining as it generates strong rules from large databases. These generated rules can then be used as pre-study knowledge that help in the decision support process.

In this thesis different GIS databases are fused in only one. Association rules technique is then applied in order to generate strong rules which are then combined in the decision support rules to induct new rules to be used in the decision making process. These extracted rules are then compared to those from experts systems.

# Table of Contents

|                        |     |
|------------------------|-----|
| Acknowledgment .....   | II  |
| Publications.....      | III |
| Abstract.....          | IV  |
| Table of Contents..... | VII |
| List of Figures.....   | X   |
| List of Tables.....    | XII |

| CHAPTER                                        | Page     |
|------------------------------------------------|----------|
| <b>1 Introduction</b>                          | <b>1</b> |
| 1.1 Thesis Objective.....                      | 1        |
| 1.2 Thesis Organization.....                   | 1        |
| <b>2 Literature Survey</b>                     | <b>2</b> |
| 2.1 Geographical Information System (GIS)..... | 2        |
| 2.1.1 Components of GIS .....                  | 2        |
| 2.1.2 Data In a GIS .....                      | 2        |
| 2.1.3 Functions of GIS .....                   | 2        |
| 2.1.4 Planning for GIS .....                   | 2        |
| 2.1.5 GIS Operations.....                      | 2        |
| 2.1.6 GIS Software Architecture .....          | 2        |
| 2.1.7 Implications of GIS .....                | 2        |
| 2.1.8 Applications of GIS.....                 | 2        |
| 2.2 Data Fusion .....                          | 2        |
| 2.2.1 Data Fusion Steps .....                  | 2        |
| 2.2.2 Data Fusion Techniques .....             | 2        |
| 2.2.3 Application of Data Fusion .....         | 2        |
| 2.3 Data Mining.....                           | 2        |
| 2.3.1 Component of Data mining.....            | 2        |
| 2.3.2 Data mining functions.....               | 2        |
| 2.3.3 Association Rules.....                   | 2        |

|     |                                                                                               |     |
|-----|-----------------------------------------------------------------------------------------------|-----|
|     | □ □ □ □ - Association rule techniques.....                                                    | □ □ |
|     | □ □ □ □ - Data Mining Applications.....                                                       | □ □ |
| ۳   | <b>Proposed Data Fusion Technique Of GIS Data</b>                                             | ۴۴  |
| □ □ | Introduction.....                                                                             | □ □ |
| □ □ | Tree Search Algorithm .....                                                                   | □ □ |
| □ □ | Experiments and Results                                                                       | □ □ |
|     | □ □ □ □ Test Databases                                                                        | □ □ |
|     | □ □ □ □ Measurement Metrics                                                                   | □ □ |
|     | □ □ □ □ Fusion Algorithm.....                                                                 | □ □ |
|     | □ □ □ □ .□ □ □ Similarity Search Algorithm                                                    | □ □ |
|     | □ □ □ □ .□ □ □ Tree Algorithm                                                                 | □ □ |
| ۴   | <b>Decision Support System For GIS Data Using Proposed Association Rule Mining Technique.</b> | ۴۹  |
| □ □ | Introduction.....                                                                             | □ □ |
| □ □ | Related Work.....                                                                             | □ □ |
| □ □ | Proposed Model.....                                                                           | □ □ |
|     | □ □ □ □ Preprocessing Module.....                                                             | □ □ |
|     | □ □ □ □ Association Rule mining.....                                                          | □ □ |
|     | □ □ □ □ DSS Module.....                                                                       | □ □ |
| ۵   | <b>Proposed Mining Framework For DSS Using Fused GIS Data</b>                                 | ۷۳  |
| □ □ | Introduction .....                                                                            | □ □ |
|     | □ □ □ □ Overview.....                                                                         | □ □ |
|     | □ □ □ □ Numerical Association rule Algorithm.....                                             | □ □ |
| □ □ | The proposed framework.....                                                                   | □ □ |
|     | □ □ □ □ Data Fusion module.....                                                               | □ □ |

|                                                          |             |
|----------------------------------------------------------|-------------|
| ❑ ❑ ❑ - Preprocessing Module.....                        | ❑ ❑         |
| ❑ ❑ ❑ Numerical Association Rule Mining<br>Module (NARM) | ❑ ❑         |
| ❑ ❑ ❑ Clustering module.....                             | ❑ ❑         |
| ❑ ❑ ❑ -The Decision Support Module.....                  | ❑ ❑         |
| <b>❧ Experiments and Results</b>                         | ❑ ❑         |
| ❑ ❑ Test Databases                                       | ❑ ❑         |
| ❑ ❑ Measurement Metrics                                  | ❑ ❑         |
| ❑ .❑ Modified Numerical Association Rule                 | ❑ ❑         |
| ❑ .❑ .❑ -First experiment                                | ❑ ❑         |
| ❑ .❑ .❑ - Second experiment                              | ❑ ❑         |
| ❑ .❑ .❑ - Third experiment                               | ❑ ❑         |
| ❑ .❑ Proposed Mining Framework For DSS Using             | ❑ ❑         |
| ❑ .❑ .❑ - Testing Categories                             | ❑ ❑         |
| ❑ .❑ .❑ - DSS Categories                                 | ❑ ❑         |
| ❑ .❑ .❑ -Results                                         | ❑ ❑         |
| ❑ .❑ .❑ - Decision support GIS system                    | ❑ ❑ ❑       |
| ❑ .❑ .❑ - Comparisons                                    | ❑ ❑ ❑       |
| <b>Conclusions.....</b>                                  | <b>❧ .❧</b> |
| <b>Future Work .....</b>                                 | <b>❧ .❧</b> |
| <b>References.....</b>                                   | <b>❧ .❧</b> |
| <b>Arabic Abstract.....</b>                              | <b>❧</b>    |

## List of Figures

|                 |                                                                      |     |
|-----------------|----------------------------------------------------------------------|-----|
| <b>Fig. ٢,١</b> | Components of a GIS include more than just computer technology ..... | □   |
| <b>Fig. ٢,٢</b> | The concept of a GIS life cycle .....                                | □ □ |
| <b>Fig. ٢,٣</b> | Data Fusion process .....                                            | □ □ |
| <b>Fig. ٢,٤</b> | Data Fusion Steps .....                                              | □ □ |
| <b>Fig. ٢,٥</b> | Principal Components Transformation (PCT) technique                  | □ □ |
| <b>Fig. ٢,٦</b> | Similarity search Algorithm .....                                    | □ □ |
| <b>Fig. ٢,٧</b> | Apriori Algorithm .....                                              | □ □ |
| <b>Fig. ٢,٨</b> | The Numerical Association rule Algorithm .....                       | □ □ |
| <b>Fig. ٣,١</b> | Example of tree algorithm.....                                       | □ □ |
| <b>Fig. ٣,٢</b> | Tree search algorithm .....                                          | □ □ |
| <b>Fig. ٣,٣</b> | Map of Cities table before fusion .....                              | □ □ |
| <b>Fig. ٣,٤</b> | Map of Cities□ table before fusion                                   | □ □ |
| <b>Fig. ٣,٥</b> | Map of Cities□ table Continued before fusion                         | □ □ |
| <b>Fig. ٣,٦</b> | Cities □ table before fusion .....                                   | □ □ |
| <b>Fig. ٣,٧</b> | Cities□ table Continued before fusion                                | □ □ |
| <b>Fig. ٣,٨</b> | Roads table before fusion .....                                      | □ □ |
| <b>Fig. ٣,٩</b> | Roads□ table before fusion.....                                      | □ □ |
| <b>Fig.</b>     | Similarity Search algorithm time                                     | □ □ |
| <b>Fig.</b>     | Tree algorithm time                                                  | □ □ |
| <b>Fig.</b>     | Comparison between Tree and Similarity Search No. of                 | □ □ |
| <b>Fig.</b>     | Comparison between Tree and Similarity Search No. of                 | □ □ |
| <b>Fig.</b>     | Comparison between Tree and Similarity Search Time                   | □ □ |
| <b>Fig.</b>     | Comparison between Tree and Similarity Search Time                   | □ □ |
| <b>Fig.</b>     | Map of Fused Cities .....                                            | □ □ |
| <b>Fig.</b>     | Map of Fused Cities(Continued).....                                  | □ □ |
| <b>Fig.</b>     | Map of Fused                                                         | □ □ |
| <b>Fig.</b>     | Map of Fused Roads .....                                             | □ □ |
| <b>Fig. ٤,١</b> | Numerical Association Rule Algorithm                                 | □ □ |
| <b>Fig. ٤,٢</b> | Block Diagram of the Proposed model                                  | □ □ |
| <b>Fig. ٤,٣</b> | the Proposed Algorithm Module                                        | □ □ |
| <b>Fig. ٤,٤</b> | DSS module .....                                                     | □ □ |
| <b>Fig. ٤,٥</b> | overall system model .....                                           | □ □ |
| <b>Fig. ٥,١</b> | Numerical association rule mining algorithm                          | □ □ |
| <b>Fig. ٥,٢</b> | Block Diagram of the Proposed Framework.....                         | □ □ |

|                  |                                                                            |     |
|------------------|----------------------------------------------------------------------------|-----|
| <b>Fig. ٥,٣</b>  | The cluster module algorithm                                               | □ □ |
| <b>Fig. ٥,٤</b>  | Decision Support Algorithm .....                                           | □ □ |
| <b>Fig. ٦,١</b>  | First experiment number of rules at different thresholds                   | □ □ |
| <b>Fig. ٦,٢</b>  | Proposed model metrics at different thresholds                             | □ □ |
| <b>Fig. ٦,٣</b>  | second experiment number of rules at different thresholds                  | □ □ |
| <b>Fig. ٦,٤</b>  | the proposed model metrics at different thresholds .....                   | □ □ |
| <b>Fig. ٦,٥</b>  | Proposed model number of rules at different thresholds                     | □ □ |
| <b>Fig. ٦,٦</b>  | NAR number of rules at different thresholds .....                          | □ □ |
| <b>Fig. ٦,٧</b>  | The proposed model metrics at different thresholds .....                   | □ □ |
| <b>Fig. ٦,٨</b>  | NAR metrics at different thresholds .....                                  | □ □ |
| <b>Fig. ٦,٩</b>  | Runtime for proposed and NAR algorithms .....                              | □ □ |
| <b>Fig. ٦,١٠</b> | Resulted rules from the proposed model .....                               | □ □ |
| <b>Fig. ٦,١١</b> | Expert DSS rules .....                                                     | □ □ |
| <b>Fig. ٦,١٢</b> | The inducted rules from augmented resulted rules with DSS rules .....      | □ □ |
| <b>Fig. ٦,١٣</b> | Examples Of DSS Rules .....                                                | □ □ |
| <b>Fig. ٦,١٤</b> | Examples Of The Inducted Rules From Augmented Resulted .....               | □ □ |
| <b>Fig. ٦,١٥</b> | The Proposed System Number Of Rules At Different Categories .....          | □ □ |
| <b>Fig. ٦,١٦</b> | The Expert DSS Number Of Rules At The Different Categories... ..           | □ □ |
| <b>Fig. ٦,١٧</b> | The Correct Ratio Of The Proposed System At The Different Categories ..... | □ □ |
| <b>Fig. ٦,١٨</b> | The Wrong Ratio Of The Proposed System At The Different Categories .....   | □ □ |
| <b>Fig. ٦,١٩</b> | The Miss Ratio Of The Proposed System At The Different Categories.....     | □ □ |

## List of Tables

|                  |                                                                                                |    |
|------------------|------------------------------------------------------------------------------------------------|----|
| <b>Table ٣.١</b> | Number Of Comparisons & Algorithm Time (Using Similarity Search Algorithm)                     | ١١ |
| <b>Table ٣.٢</b> | Number Of Comparison & Algorithm Time (Using TREE Search Algorithm)                            | ١١ |
| <b>Table ٣.٣</b> | Comparison Between Tree And Similarity Search<br>No. Of Comparisons For Cities && Roads Of USA | ١١ |
| <b>Table ٤.١</b> | Example of the data                                                                            | ١١ |
| <b>Table ٥.١</b> | Egypt Cities First View                                                                        | ١١ |
| <b>Table ٥.٢</b> | Egypt Cities Second View                                                                       | ١١ |
| <b>Table ٥.٣</b> | Egypt Cities Third View                                                                        | ١١ |
| <b>Table ٥.٤</b> | The Fused Table                                                                                | ١١ |
| <b>Table ٥.٥</b> | Egyptian Cities                                                                                | ١١ |
| <b>Table ٦.١</b> | USA Types Of Citizens .....                                                                    | ١١ |
| <b>Table ٦.٢</b> | USA Citizens Material Status .....                                                             | ١١ |
| <b>Table ٦.٣</b> | USA citizens type of houses                                                                    | ١١ |
| <b>Table ٦.٤</b> | Children Data                                                                                  | ١١ |
| <b>Table ٦.٥</b> | Type Of Building In A City                                                                     | ١١ |



# Acknowledgement

In the name of Allah, the Beneficent, the Merciful: “**And whatever of comfort you enjoy, it is from Allah...**” *Al-Nahl* (١٦). First and foremost, I humbly give my deep thanks to **Allah** for giving me the opportunity and the strength to accomplish this work.

I would like to thank **Prof. / Mohamed Essam Khaleifa** - the dean of the faculty of Computer and Information Sciences, Ain Shams University- for his support. My great thanks to **Prof. / Hamed Nassar**, the dean of the faculty of Computers and Informatics, Suez Canal University, for his encouragement and support.

I would like to express my deep appreciation and thanks to my supervisors: **Prof. / Mohammed Hashim** and **prof. / Ahmed Hamad** and **prof. / Helmy Mahran**, for their great help and encouragement during the execution of this work. I would like to thank **Prof. / A. B. Salem**, **Prof. / M. E. Ghonemy** for their advice and support.

I would like to thank Eng./ **Ahmed Magdy Ghanem**, Eng./ **Amr Mostafa** and Eng/ **Mohamed Azab** for their help and advice. Finally, I would like to express my deep thanks to all my family and my colleague friends who were encouraging me strongly during the execution of this work.

# Chapter 1

## Introduction

The emergence of remote sensing, scientific simulation, telescope scanning, and other survey technologies has dramatically enhanced our capabilities to collect spatio-temporal data, however the explosive growth in data makes the management, analysis and use of data difficult and expensive. In decision support application with spatio-temporal data, it is important to study the temporal relationships of the parameters that influence the decision.

The geographical information system is a computer system capable of capturing, storing, analyzing, and displaying geographically referenced information; that is, data identified according to location. Practitioners also define a GIS as including the procedures, operating personnel, and spatial data that go into the system [1, 2, 3].

Data fusion is the process of combining data and knowledge from different locations in only one location with the aim of maximizing the useful information content. It improves reliability or discriminate capability while offering the opportunity to minimize the data retained.

The fusion process may results in large database after combining different databases in only one. Mining these large databases to extract knowledge from it is a complex process. Data mining combines techniques including statistical analysis, visualization, decision trees, and neural networks to explore large amounts of data and discover relationships and patterns that should solve problems.

In this thesis, we proposed three algorithms. The first one is the tree search algorithm for fusing different GIS databases in only one database. The second one is a modification to the normal association rules technique to mine the descriptive and special nature fields in the GIS databases. The third is a new frame work for making a decision support rules for the GIS data.

The three algorithms are testing using a real data – the GIS data of USA- which obtained from the internet. We choose the USA data because it is highly descriptive, correct and very real.

The three algorithms are implemented using C# programming language.

## 1.1 Thesis Objective

The first objective of this thesis is to make analysis for geographical data using data fusion by fuse geographical databases founded in diferent locations in only one location with a proposed technique which is then compared to the standard fusion technique. The second objective of this thesis is mining the fused geographical databases whether the special nature geographical data or the traditional geographical data. The last objective is to invest concept description and rule discovery for geographical information system data to aid in the decision making process.

## 1.2 Thesis Organization

This thesis is organized as follows:

**Chapter two** illustrates the different defenitions of the geographical information system, data fusion and the mining of gegraphical data. It produces the different algorithms and techniques which have been previously used in the proceses of fusion and mining of geographical data and its main problems.

**Chapter three** proposes an improved technique for making data fusion process of geographical data. This technique is called tree search algorithm as it based on storing data in binary tree.

This chapter continues by comparing the results of the tree search algorithm with the results of the old similarity search algorithm after applying them on the same data set. The proposed algorithm provides more suitable results using less number of comparisons and time than the old algorithm.

**Chapter four** introduces a new algorithm for modifying the numerical association rule mining technique to be suitable for using in mining the special cases geographical data by using some statistical measures such as correlation statistical measure to generate association rules describing these data.

**Chapter five** a new proposed framework for combining the process of fusion and mining of the geographical data. This framework is based on fusing the geographical data founded in different places in one place using the proposed algorithm in chapter three then use the numerical association rule algorithm on the fused data in order to generate association rules from these data which are then augmented in the decision support system to output new association rules which can then be used in the decision making process.

**Chapter six** illustrates the experiments which have been done and their output results. This chapter illustrates the data set which has been used.

This chapter continues by comparing the results of the modified association rule technique which proposed in chapter four with the results of the traditional association rule technique after applying them on the same data set, it is founded that the proposed technique provide a promising results with less miss ratio, wrong ratio and time than the traditional association rule technique.