國立台東大學資訊管理學系

碩士論文

應用無線定位系統於博物館觀眾行為分析之研究

Applying Wireless Locating Systems to Analyze Museum Visitor Behavior

研究生：李紹永　撰

指導教授：鍾青萍博士

謝明哲博士

中華民國九十八年七月

國立台東大學
學位論文考試委員審定書

系所別：資訊管理學系

本 班 李 紹 永 君

所 提 之 論 文 應 用 無 線 定 位 系 統 於 博 物 館 觀 眾 行 為 分 析 之 研 究

業 經 本 委 員 會 通 過 合 於 碩 士 學 位 論 文 條 件

論 文 學 位 考 試 委 員 會 ：

（學位考試委員會主席）

（指導教授）

論 文 學 位 考 試 日 期 ： 98 年 7 月 6 日

國 立 台 東 大 學

博碩士論文電子檔案上網授權書

（提供授權人裝訂於紙本論文書名頁之次頁用）

本授權書所授權之論文為授權人在 國立臺東大學 資訊管理學系碩士班 97 學年度第 二 學期取得 碩士 學位之論文。

論文題目： 應用無線定位系統於博物館觀眾行為分析之研究

指導教授： 鍾青萍 謝明哲

茲同意將授權人擁有著作權之上列論文全文（含摘要），非專屬、無償授權國家圖書館及本人畢業學校圖書館，不限地域、時間與次數，以微縮、光碟或其他各種數位化方式將上列論文重製，並得將數位化之上列論文及論文電子檔以上載網路方式，提供讀者基於個人非營利性質之線上檢索、閱覽、下載或列印。

●讀者基非營利性質之線上檢索、閱覽、下載或列印上列論文，應依著作權法相關規定辦理。

授權人：李紹永

簽 名：李紹永 中華民國 98 年 07 月 27 日

博碩士論文授權書

本授權書所授權之論文為本人在 國立臺東大學資訊管理_____系(所)
97 學年度第 二 學期取得 碩士 士學位之論文。
論文名稱：應用無線定位系統於博物館觀眾行為分析之研究

本人具有著作財產權之論文全文資料，授權予下列單位：

同意	不同意	單　位
■	□	國家圖書館
■	□	本人畢業學校圖書館
■	□	與本人畢業學校圖書館簽訂合作協議之資料庫業者

得不限地域、時間與次數以微縮、光碟或其他各種數位化方式重製後散布發行或上載網站，藉由網路傳輸，提供讀者基於個人非營利性質之線上檢索、閱覽、下載或列印。

■同意 □不同意 本人畢業學校圖書館基於學術傳播之目的，在上述範圍內得再授權第三人進行資料重製。

本論文為本人向經濟部智慧財產局申請專利(未申請者本條款請不予理會)的附件之一，申請文號為：＿＿＿＿＿＿＿＿，請將全文資料延後半年再公開。

公開時程

立即公開	一年後公開	二年後公開	三年後公開
		✓	

上述授權內容均無須訂立讓與及授權契約書。依本授權之發行權為非專屬性發行權利。依本授權所為之收錄、重製、發行及學術研發利用均為無償。上述同意與不同意之欄位若未勾選，本人同意視同授權。

指導教授姓名：(親筆簽名)

研究生簽名：李紹詠 (親筆正楷)

學　　號：9601303 (務必填寫)

日　　期：中華民國 98 年 7 月 27 日

謝誌

終於，來到結束碩士生活的時刻，寫完這段誌謝辭也是時候得離開這美麗的台東了。2007 年 5 月，研究所考試放榜。很開心自己能考上台東大學，也很猶豫是否要就讀，畢竟自己已經在業界工作了一陣子，還適合到學校念書嗎？抱著不確定的心情，看著學校網頁上專任老師的介紹一鍾青萍老師，研究的方向似乎與我的興趣很相近，就這樣，進行了與鍾老師的第一次會面。鍾老師的學識淵博與詼諧風趣，讓我對就讀研究所有了更多的興趣，也因此開啓了我的碩士生活。回顧這 700 多個日子，該感謝的人太多了，希望能藉由這段誌謝辭，表達心中最深的感謝。

首先當然得感謝引領我進入研究殿堂的指導教授，鍾青萍老師與謝明哲老師。在兩位老師的引導之下，順利的完成於博物館的專案與研究。鍾老師對問題分析精闢，從多種角度切入核心，開拓我研究的視野，增進分析問題的能力。謝老師在資訊科技的專精以及對人因的研究，讓我在鑽研資訊技術時也不忘以人類福祉爲本的研究精神。兩位指導教授對這個研究的全力支持，是我完成這份論文最重要的助力。

此外，也要感謝黃正魁老師細心的教導，爲我開啓資料探勘研究的大門；感謝許功明教授擔任論文計畫書審查委員，提供寶貴的意見；感謝辛信興老師擔任口試委員，細心檢視論文的完整性；感謝國立史前文化博物館提供研究所需的協助，讓此研究得以順利進行。

再者，感謝在這兩年陪伴我一起成長的同學與學弟妹們，有各位的陪伴，讓我在遇到瓶頸時不覺孤單，使我有動力突破面臨的難關。與小洪一起同甘共苦的生活、算ㄏ天真而邪惡的本性、宜澤的性感火辣、阿宅的純樸善良、佳慧的豪爽及對事物的堅持，都讓我印象深刻、銘記在心。與各位相處的點點滴滴，所有的酸甜苦辣，都是我最寶貴的回憶。

另外也特別感謝龍富、聖熙、慧文與千千在口試前的鼓勵與幫忙，讓我順利的完成口試。當然也要感謝提供英文撰寫建議及最後幫我校稿跟的小鴨、小餅、大砲、安迪、育霖，因爲你們讓我的論文更加完整。最後我要感謝全力支持我來就讀研究所的家人，一路走來，因爲有你們，讓我可以沒有後顧之憂，全力追求我的理想。

Chinese Abstract

無線射頻辨識 (RFID) 及類似的無線定位技術，在近幾年間逐漸的成熟，也發展出相當多的應用。以往相關的研究中，大部分以物料管理、產品管理及物流管理等為主要應用範圍，很少應用在消費者 (遊客) 行為的研究上。本研究應用無線網路定位技術，設計並實作一套遊客位置紀錄系統 (Visitor Location Record System) 於國立台灣史前文化博物館，用以紀錄遊客 (*who*) 甚麼時候 (*when*) 出現在哪個展區 (*where*)。並且利用程式從大量的位置資料中萃取出遊客參觀的熱點 (hotspots)、路徑(paths)及熱門路徑(hot-paths)，增加資料的可用性。基於所發現的結果中，提出能改善博物館展覽及服務的建議。

關鍵詞：觀眾行為研究、無線定位系統、熱點分析

Abstract

Radio frequency identification (RFID) and similar wireless locating systems have gained momentum in the last few years, and many applications have sprung up with the maturation of the technology. However, the applications seem to remain focused on material/product logistics such as product tracking and tracing; and there are few studies devoted on applications on consumer/visitor behavior. In this paper, wireless locating technology was applied to design and implement a Visitor Location Record System (VLRS) to record the time (*when*) and position (*where*) of specific visitors (*who*) during the tour. A program was developed to extract the hotspots and paths of visitors, and in order to increase data usability, a hot-paths mining algorithm was developed as well. Based on the findings of this study, additional applications and implications are proposed for improving the exhibitions and services.

Keyword: consumer (visitor) behavior study, wireless locating system, hotspots analysis.

Table of Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1. Background and Motivation

Radio frequency identification (RFID) and similar wireless locating systems have gained momentum in the last few years, and many applications have sprung up with the maturation of the technology. However, the applications seem to remain focused on material/product logistics such as product tracking and tracing; and there are few studies devoted on applying wireless locating systems on consumer behaviors -- for example, to study the behavior of those visitors in large exhibition environments. However, since these environments, in a broader sense, may include department stores and supermarkets, the implication could be expansive in the field of marketing.

The original function of a museum is to preserve artifacts or collect special objects, and these functions have expanded to include education, research, and tourism. In terms of education, museums attempt to enhance visitor knowledge through a variety of exhibits; in terms of research, museums provide the subjects and objects as well as a collection of previous research data/results; and in terms of tourism, with its entertaining displays museums are the best places for tourists to spend some meaningful time, especially with their children.

By the end of the 20th Century, there are more than 25,000 museums around the world, and more than half of them were built after 1950 (Chang, 1994). This rapid increase in the number of museums can be attributed to several reasons, including the maturity of archaeological excavation, the increase in the number of newly discovered artifacts, and the increasing need of diversity in the types of museums (including small and personal museum), to name a few.

As the standard of living improves, people yearn for a life beyond material things. Culture and history becomes important as people look into their spiritual (includes but not limited to religions) lives. Thus, museums may assist in enlightening and promoting psychological well-beings through the cultural and historical artifacts they keep; this, in turn, opens up additional commercial opportunities for the adjacent communities in the form of tourism – such as in-depth experiential touring packages, or in-depth knowledge-based touring packages. This type of tourism growth is clearly shown in Taiwan and also appears in developed countries. The growth of the number of museums provides additional places for people to visit and thus becomes an important part of the tourist industry.

The National Museum of Prehistory sets its mission on exposing the general public to Taiwan's pre-historical – especially those of aboriginal – culture and natural ecology. It is essential for the museum to assess the number of visitors and the depth of their knowledge after the visit, since this information would be used to justify the museum's existence and continuation. Visitor satisfaction affects the future number of visitors through "returning" and "referral," (Anderson and Sullivan, 1993) and it would be important to study how the layout of exhibits may enhance visitor interest and knowledge. Through analyzing the amount of time each visitor spends on each exhibit – in other words, visitor behaviors – may help researchers understand how such allocation of time may be related to or affect this visitor's perception of museum. The research works presented in this thesis illustrate the design and implementation of a wireless locating system for capturing data related to visitors' whereabouts in museum, and the subsequent analyses of the captured data.

1.2. Objectives

The primary function of wireless locating (or location) system is to provide the information of the time (when) and location (where) of a particular item (what), and there are two major technological approaches currently exist in the market for exhibition environments.

Hsi and Fait (2005) points out that the RFID technology can improve visitors' learning and experience in museum. They used RFID tags and readers to establish a system which was named "eXspot". That system can record visitors' exploration through RFID system -- like photos captured at the "Heat Camera" exhibit -- and provides additional resources; all of which can be browsed at anytime on Web. In this application, eXspot uses the RFID system's sensing capability to know visitors' position and identify their location. Also, eXspot uses this information to improve visitor exploration experience.

Wireless locating system locates wireless devices (and thus their carriers) through access points (APs) or sensing devices at any time in the area covered by wireless signals. Once a wireless locating system is set up in a museum, the position of every visitor (i.e., those who carry wireless device in that setting) can be identified. A concept similar to time and motion study is adopted to model visitor behavior. Visitor behavior is divided into measurable element according to exhibition zones, and a visitor location data record system was designed and implemented to expedite the process. The visitor location data includes who and when appear at where, and these location data can be re-organized from points to lines. These data were analyzed to identify visitor hotspots, paths, and hot-paths. According to the result of analysis, Recommendations then can be made to improve services and management in museum, which in turn could increase service quality and visitor satisfaction, and visitors would be more willing to visit again and/or recommend the museum to others.

Most museums use either sensor-based systems or the traditional button-based tour guide machines in their applications. Although the wireless locating system is powerful, they have not been used in museums. The induction system can only know the position of a visitor who is near a sensing device, and it cannot record the visitor positions continuously or to differentiate the position when its signal is received by two or more receivers (signal readers). The wireless locating system can locate visitors anywhere in the museum that is covered by wireless signal. The data it can record are much richer than induction system. It is believed that the wireless locating system can provide better guide service than older systems and also expand the possibilities in visitor behavior researches. The main objectives of this research are list as follow:

(1) Apply wireless locating technology to design and implement a system to record a particular visitor's (*who*) location in the exhibition hall (*where*) at any given time (*when*) during a visitor's tour.

(2) Analyze the above data and identify *hotspots* in museum and make recommendation for improving exhibitions and services.

1.3. Research Methodology

The research process is depicted in Figure 1-1. It includes four parts.

Step 1: Consider the features of experimental environment and choose a wireless locating system production from software providers.

Step 2: Design and implement the system which can record durations of visitors at each exhibit with the chosen wireless locating system.

Step 3: Collect data with the system and analyze the data to identify hotspots and patterns of visitor behavior.

Step 4: Propose suggestions for the museum according to the patterns and hotspots.

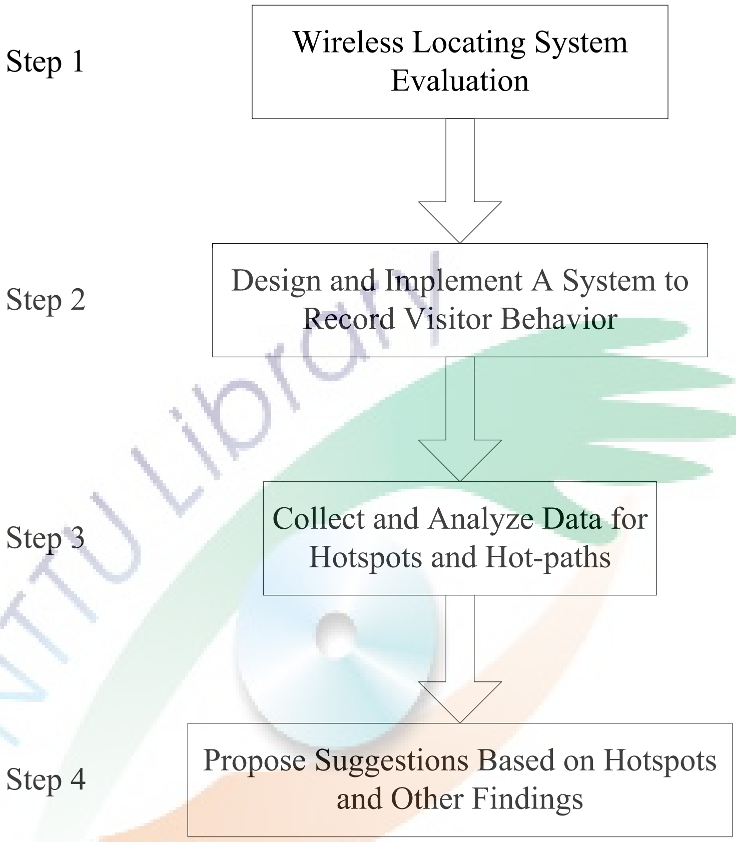

Figure 1-1: Research Process

1.4. Research Scope and Limitations

Since the primary purpose of this research is to design, implement, and validate a wireless locating system in the museum setting and not in exploring developments in wireless technology, most of the locating devices (and software) used are commercial products. The scope of this research does not include improving the accuracy of wireless locating system but to identify the strengths and weaknesses in applying such system. On the one hand, only the features that are “proven” to be useful and feasible can be used to develop new applications for such systems. As long as locating accuracy of wireless locating system becomes better in the future, the result will be more precise.

This research was completely conducted at National Museum of Prehistory in Taiwan. Though externality is not tested in this research, this methodology should also be feasible in other museums. Future researches are encouraged to apply this methodology to other museums.

1.5. Organization of the Thesis

The rest of this paper is organized as follows. Chapter 2 is the literature review in which various related topics such as different wireless locating technologies, time-and-motion study, and hotspot identification and usage are discussed. Chapter 3 describes the research methodology, the processes of designing/implementing the system, and the subsequent data collection. Chapter 4 analyzes the data for hotspots and paths, and proposes suggestions for the museum based on the findings. Chapter 5 describes the contributions and future works.

CHAPTER 2

LITERATURE REVIEW

2.1. Wireless Locating System

Since 1970s, the America government starts to develop Global Positioning System (GPS) (Shi, 1989). Many people notice that the real time locating system is a high commercial value and potential tool. Until now (2009), there are lots of different kinds of wireless locating technology including GPS, GSM (or CDMA), Wi-Fi, RFID, infrared ray and supersonic wave. But not all of these technologies are fit to use in museum. Because most museums are covered by roof, GPS and GSM signals are not easy to send in or out the buildings. For this reason, GPS and GSM locating system are not fit to use in museum. Besides, most of the display areas are divided by walls in the museum. Infrared rays and supersonic waves are not easy to pass through walls, so they are not fit to use in museum neither. Consider the properties of museum and wireless locating technologies discussed above, GPS, GSM (or CDMA), infrared rays and supersonic waves are excluded. RFID and Wi-Fi Systems will be discussed and compared below.

2.1.1 RFID System

There is no global public RFID standard now. Every country set their own rules for RFID and every manufacture set their own protocol for their products. General speaking, a RFID system includes two components: tags and readers. Readers can receive radio frequency sent by tags and identify the tags. RFID tags contain at least two parts. One is an integrated circuit for storing and processing information, modulating and demodulating a radio frequency (RF) signal, and other specialized functions. The second is an antenna for receiving and transmitting the signal. There are generally two types of RFID tags: active RFID tags, which contain a battery and thus can transmit its signal autonomously, and passive RFID tags, which have no battery and require an external source to initiate signal transmission.

2.1.2 Wi-Fi System

In IEEE's standard (IEEE 802.11, 2007) for wireless local area network (WLAN) which is also called Wi-Fi, there are two different structure of wireless network: Ad-hoc and infrastructure. Ad-hoc network means one node is able to communicate with another node. Stations can directly communicate with each other on a peer to peer level sharing a given cell coverage area. Infrastructure network needs an AP (Access Point), which is used as a bridge between wireless and wired LANs. All the clients transfer data to each other through the APs. Infrastructure network can be described as the network which is formed by clients and APs. Wi-Fi terminal/station is the most basic component of the infrastructure network. The terminal can be any kind of device that contains the functionality of the 802.11 protocol such as laptops, mobile devices, or APs, and typically the 802.11 functions are implemented in the hardware and software of a network interface card (NIC). Terminals can be mobile, portable, or fixed and all terminals have to support the 802.11 terminal services such as authentication, de-authentication, privacy, and data transformation.

2.1.3 Comparison of RFID and Wi-Fi System

Wi-Fi and RFID technologies seem the best solutions in this research. The features of these two technologies are compared with each other next (see Table 2-1). Wi-Fi access points (AP) have wider signal coverage than RFID readers. It means that less AP will be used to set up Wi-Fi system in the same area. It saves more devices and work than setting up RFID system. Although the locating accuracy of RFID system is better than Wi-Fi system, 15cm (Fact Sheet of Ubisense Series 7000 Sensor, 2008) and 1m (Ekahau RTLS Brochure, 2008) are not different in this case. The proportion of floor which stood by an adult is close to $1m^2$. So, there is no difference between 15cm and 1m locating accuracy. Data transfer speed of Wi-Fi system is much higher than RFID system. Normally, they also set up Wi-Fi system for data transfer in RFID application. Besides, Wi-Fi system is easier to connect to internet and integrate with other systems. So, it is more convenient to enhance Wi-Fi system than RFID system.

Table 2-1: Comparison of RFID and Wi-Fi

Properties	RFID	Wi-Fi
Max signal coverage (radius)	About 150m.	About 400m.
Locating accuracy	About 15cm.	About 1m.
Tag size	Small.	Big.
Environment setting	Need to set up reader.	Can use the access points already set up in museum.
Setting cost	Need more readers when set up in the same area.(more works and devices needed)	Need less access points when set up in the same area. (less works and devices needed)
Data transfer speed	About 10Kbps.	About 54Mbps~300Mbps.
Applications	Major in sensibility.	Rich applications on internet.

Consider the discussions above; it is believed that Wi-Fi locating system is the best solution in this case. After survey the wireless locating system production in the market, Ekahua Real Time Location System (ERTLS) is chosen for this research. The ERTLS is a powerful wireless real time locating system. The ERTLS claims to be able to simultaneously locate 50,000 mobile devices, provides visual management interface to monitor all of the mobile devices, has high speed of searching mobile devices (600 devices per second), has high locating accuracy (1-3m), and provides APIs for applications development. It completely satisfies our requirements. Chapter 3 will introduce the system infrastructure which adopts the ERTLS as the kernel and how to set up the system.

2.2. Time and Motion Study

Recalling the three fundamental data attributes of *when*, *where*, and *what* in wireless locating system, we may combine the *where* and *what* into the concept of "*motion*" and explore how time-and-motion study may be used to study visitor behaviors.

Time and motion study is the origin of scientific management and has been widely applied to other research areas. In order to analyze and improve working efficiency, Taylor (1911) proceeds to start time-and-motion study in factory. He used a clock to record how much time spent in different motions of a job or series of jobs and evaluated the performance of each motion. Time-and-motion study can not only check the efficiency of workers and equipments but also aid to standardize the work procedure. It is helpful to improve work by subdividing a job into measurable elements (motions).

Bergman et al. (1966) observed four pediatricians with stopwatch for 18 days to gain the profile of how their working days were spent. They pointed out that there are not enough pediatricians to care the critical ill children. The lack of pediatricians in United States was serious at that time.

McDonald and Dzwonczyk (1988) used time and motion study of anesthetists during 32 surgical procedures. They pointed out that care of patient in surgery can be improved by increasing anesthetist's ability to focus full attention on patient.

Daviaud and Chopra (2008) quantified staff requirements in the primary health care facilities in South Africa. Their study shows the lack of medical resources and huge variants between districts.

Baker and Kizil (2009) applied time and motion study to determine the key components of coal mining cycle. They identified bolting and shuttle car away time of the cycle have great impact on cutting time of mining and were the main inhibiters to the overall operating rate due to the large variance in these component times. They also pointed out the potential for improvement.

Gillespie et al. (2008) applied time and motion study to analyzing labor requirements and profitability in choosing grazing strategy for beef production. They explain low adoption rates of rotational grazing in Gulf Coast region.

The time and motion studies which were discussed above are summarized in Table 2-2.

Table 2-2: Summary of Time and Motion Studies

Researchers	Year	Research Areas	Main objective
Taylor	1911	Work efficiency	Analyze the jobs in factory and improve work efficiency.
Bergman et al.	1966	Medical	Analyze pediatricians in U.S.A.
McDonald and Dzwonczyk	1988	Medical	Analyze anesthetist behavior in surgical procedures.
Gillespie et al.	2008	Grazing	Analyze labor requirements and profitability in choosing grazing strategy for beef production.
Daviaud and Chopra	2008	Medical	Quantified staff requirements in primary health care facilities in South Africa.
Baker and Kizil	2009	Coal mining	Determine the key components of coal mining cycle.

2.3. Hotspots

Searching for hotspots is helpful for problem solving and decision making. Hsu (2006) analyzed the commercial hot points to assist users to choose their addresses of convenience stores. Li (2002) analyzed the criminal hot points to assist to deploy police and prevent crimes. Hotspots may be found using Pareto principle (80/20 principle) and data mining techniques.

According to Pareto principle, handle the "significant few" causes of problems then solve most of the problems. It shows that identifying the hotspots of problems is helpful to solve the problems and can yields twice the result with half the effort. Chen et al. (2001) provide a theoretical foundation for Pareto principle through establishing an analytical model to depict the skew distribution; and the result of this study encourages researchers to find the "significant few" of their problems.

Approaches of data mining for discovering the association rules and sequential patterns are based on the user behaviors expressed in terms of frequency. Thus the frequent association rules and sequential patterns can be seen as the hotspots of user behaviors. Clustering is a grouping method to divide data samples into groups by calculating distance between samples. In other words, clustering is a method to find areas which lots of data samples gathered together. From this point of view, clustering is also an approach to identify hotspots. These approaches are widely adopted in various research areas including finance, accounting, medical and so forth.

Wal-Mart mined the sale data and found that the most frequent item buy with diaper was beer. After market survey, they found lots of wives asked their husbands to buy diapers and husbands usually took some beers with diapers in the mart. Base on this pattern, Wal-Mart put the beer next to the diapers. Then sales of both of them were increased. This is a famous application of frequent association rule. It shows the benefits of applying frequent association rules to decision making.

In this research, the hotspots of exhibits are the "significant few" of visitor's tour. To identify the hotspots of each exhibit by analyzing visitor's location in exhibits will be helpful to reallocate the exhibits, design new exhibits and redesign the guide service content. Besides, the information of hotspots can also use in product promotion and museum marketing. A system will be proposed to record visitor behaviors in Chapter 3 and identify the hotspots of exhibits out in Chapter 4.

2.4. Visitor Behavior and Time and Motion Study

Robison (1928) was funded by The America Association of Museum to execute a project to observe visitor's behavior in the museums including what they did in front of the exhibits, how long they stayed in those locations, what they looked at, etc. He is the first scholar who proposes the concept of *holding power* and uses a quantitative attribute *average time* as the operational definition to measure it. Thus, holding power measures the attractiveness of an exhibit which keeps a visitor to *hold* his or her position in front of exhibit. This concept has significant impacted the scholars who followed.

Bitgood et al. (1988) observed selected exhibits in 13 zoos throughout America. They record the percentage of visitors who stopped and the duration of visitor viewing time at each exhibit of various species. They assumed the viewing time on various species can represent the attraction of various species. Visitor behavior was found to correlate with both the characteristics of animals and the architectural characteristics of the exhibits. These finding were helpful for designing zoo exhibits.

According to these researches, it is obviously that time and motion study is useful and widely adopted. It also fits to analyze visitor behavior. Chen (2001) adopted that Bitgood et al. (1988) assumed viewing time can represent attraction of exhibits to research visitor behavior in The World of Calcite Special Exhibition. Lin (2002) also adopted their assumption to study visitor behavior at Children Environmental Education

Exhibition in Taroko National Park. In our research, we will carry out a time and motion study to know how the time of visitors spent in National Museum of Prehistory, and identify the hotspots of exhibits.

In the past, in order to collect durations of visitors in exhibition, researchers must send lots of observers to follow visitors or deploy observers at each display areas. This method takes lots of manpower and costs very much. Another method is questionnaire. Using questionnaire can gather durations of each exhibit with less cost but there are still some problems. The accuracy of duration data collect by questionnaire is not reliable. Visitors will not remember all the details about how their time spent in each exhibit even if you remind them to do this before.

2.5. Business Intelligence

Gorry and Scott-Morton (1971) first articulated the main concept of decision support system. They defined decision support system (DSS) as "interactive computer-based systems, which help decision makers utilize data and models to solve unstructured problems". Keen and Scott-Morton provided another classic definition of decision support system as follow:

> Decision support systems couple the intellectual resources of individuals with the capabilities of computer to improve the quality of decisions. It is a computer-based support system for management decision makers who deal with semistructured problems

The term "decision support system" is a content-free expression just like management information system (MIS) or other terms in the information technology field. Turban et al. (2007) drew the architecture of DSS as Figure 2-1. The data which come from many sources is the first component of DSS architecture. Data are manipulated by using different models to solve various problems. The models are the second component of the DSS architecture. Some systems have an intelligence component (or knowledge component). Knowledge can be extracted from data directly. It also can be extracted from data by models. Intelligence is the third component of DSS architecture. Users are the fourth component of the architecture and the user interface which is between users and system is the fifth component of the architecture. Results from studying consumer/visitor behavior in a large exhibition environment, thus, can become part of business intelligence to enhance a company's marketing capability and performance.

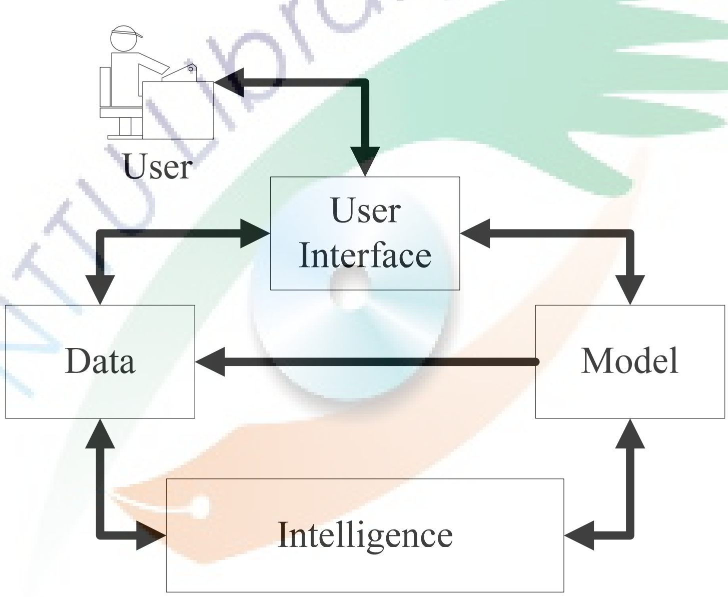

Figure 2-1: Architecture of Decision Support System

CHAPTER 3

RESEARCH METHOD

The system described in this thesis was designed and implemented in the National Museum of Prehistory, Taitung, Taiwan, and Section 3.1 will provide the background for the selection of this location. As discussed in section 2.1, ERTLS is used to locate visitor positions, and the complete visitor location record system will be introduced in section 3.2. Section 3.3 describes the process of data collection. Section 3.4 introduces the data analysis approaches and applications. Section 3.5 describes the concept of modeling visitor behavior. Section 3.6 evaluates the feasibility of this research method. The research process is shown in Figure 1-1.

3.1. National Museum of Prehistory

Through a joint research project on museum visitor satisfaction assessment, this research team injects the concept of using "time spent on exhibit" as the operational definition of "interest level on that exhibit" in the study. The museum collaborated with the research team to install or replace equipments such as additional AP, and the research team operates.

The National Museum of Prehistory is a huge concrete building. There are two floors in the main exhibition area. Our system is set in this area. The floor plan is shown as Figure 3-1 and 3-2. There is a spin corridor which connects B1 and 2F, corridors around the centre area, exhibition areas around the corridors. In this research, each exhibition area was given a unique code (see Figure 3-1 and 3-2) to identify itself. The 1st temporary exhibition room is set at another side and it does not be considered in this research. Table 3.1 shows the contents of each exhibition area. More details of geographic information can be found on the official website of Taiwan National Museum of Prehistory.

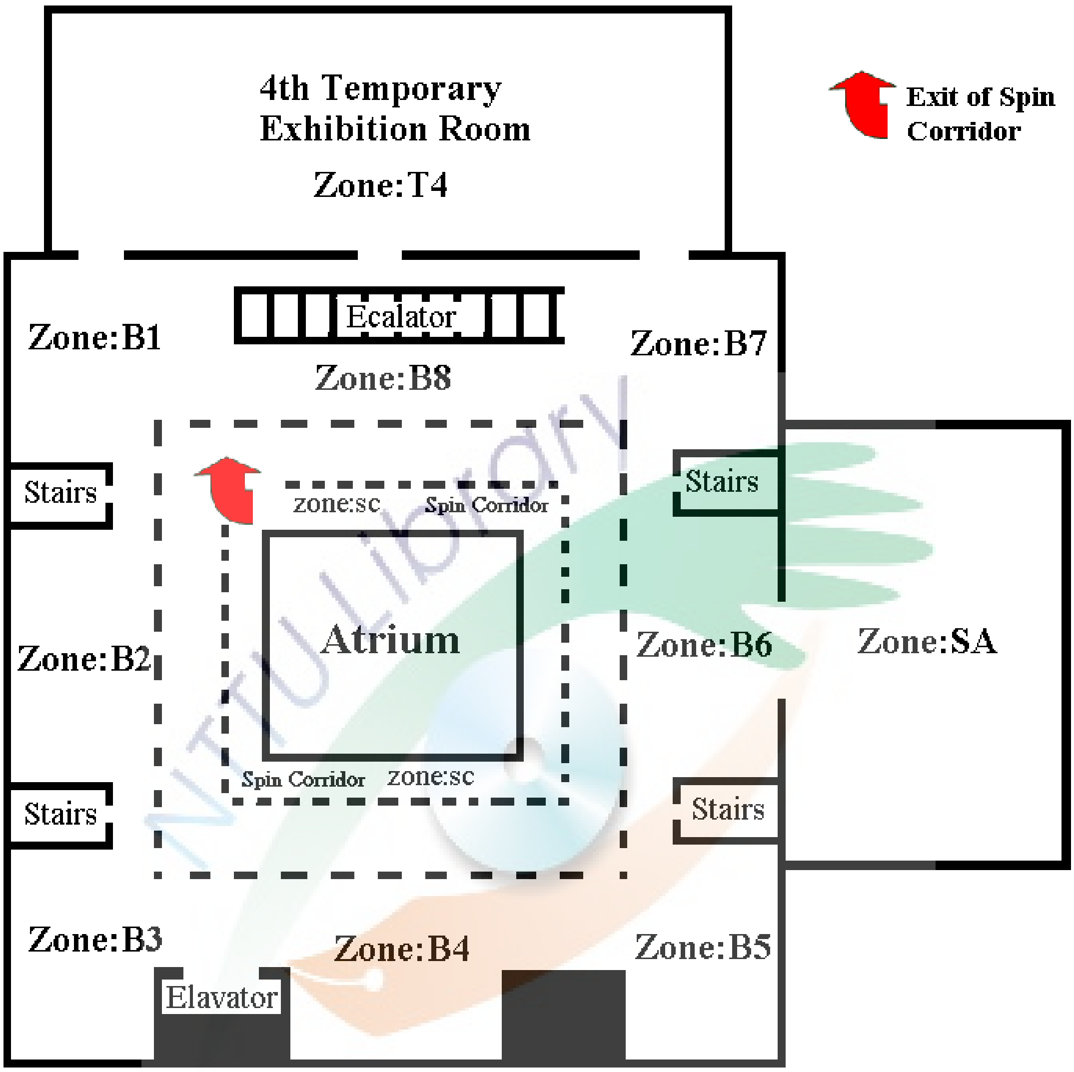

Figure 3-1: Floor Plan of B1

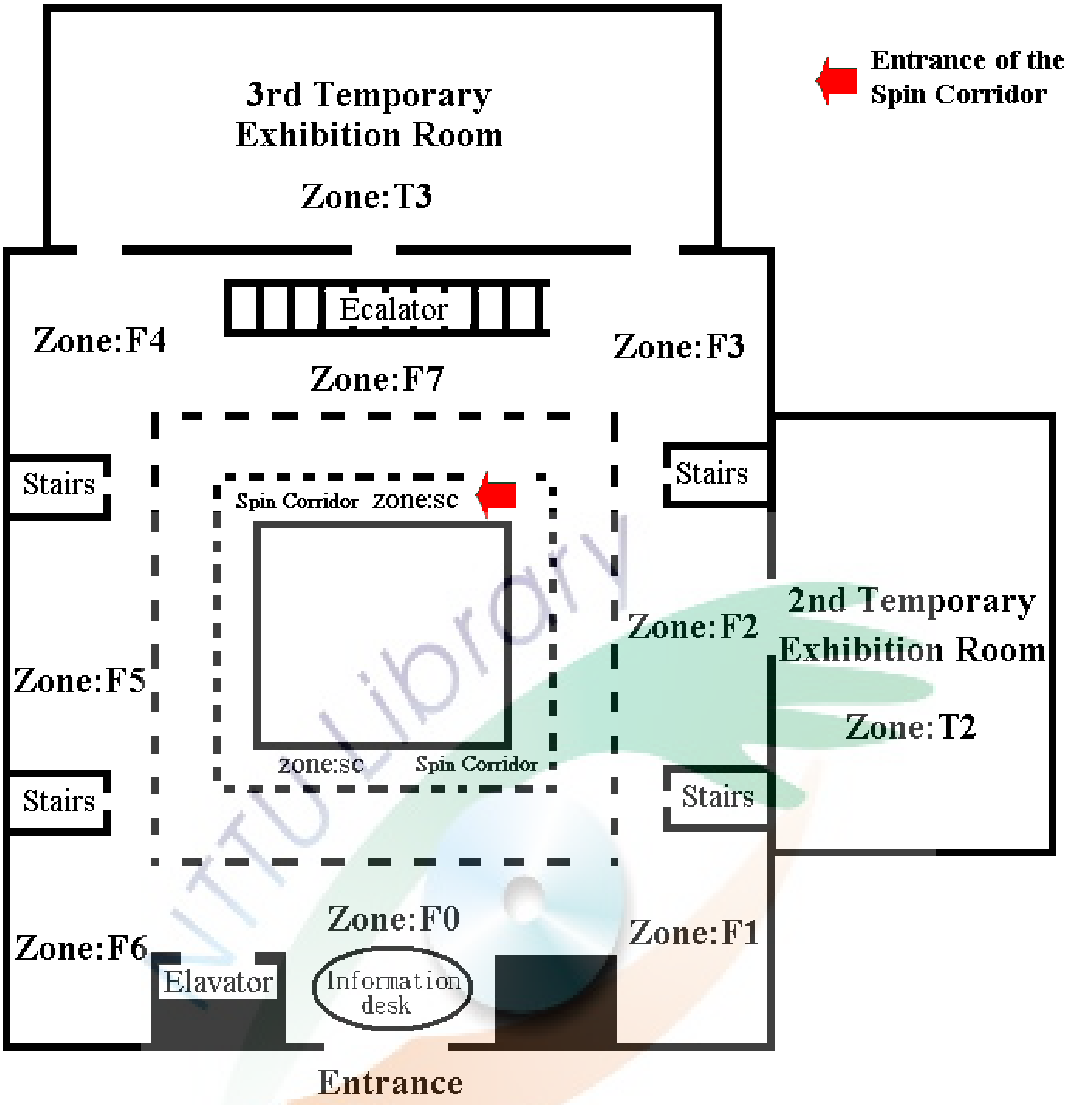

Figure 3-2: Floor Plan of 2F

Table 3-1: Introduction of Each Zone

Zone	Introduction
F0	Lobby
F1	Birth of natural history of Taiwan.
F2	Ice age of natural history of Taiwan.
F3	New age of natural history of Taiwan.
F4	Social relationship of the Austronesian language speaking people.
F5	Technology, subsistence and social organization of the Austronesian language speaking people.
F6	Ceremony and spiritual concept of the Austronesian language speaking people.
SC	Spin corridor between B1 and F2 with exhibits about human evolution.
B1	The beginning of Taiwan's prehistory.
B2	How Taiwan's prehistoric people lived.
B3	Pottery making in Taiwan's prehistoric age.
B4	Taiwan's prehistoric people and ocean.
B5	Pei-Nan Culture.
B6	Megaliths and rituals.
B7	Prehistoric stoneworking and iron age in Taiwan.
B8	Pottery game and some introductions of prehistoric relics.
SA	Introduction of scientific archaeology.
T2	Second temporary exhibition room. Only open for special exhibitions.
T3	Third temporary exhibition room. Only open for special exhibitions.
T4	Fourth temporary exhibition room. Only open for special exhibitions.
null	Undefined area. Most of the null area are rest areas and corridors

3.2. Visitor Location Record System (VLRS)

VLRS is used to analyze the wireless signals for locating visitor and store the location data into database. At the onset of this research, there did not exist a VLRS in the museum. The shortcomings of existing VLRS, as pointed out by previous studies, include high cost and its lack of precision. To offset these problems, the system designed and implemented in this thesis utilizes ERTLS to locate visitors, a database for storing data, a program to capture location data from the positioning engine, and access points for communication linkage. The structure of VLRS is constructed by all of the components discussed above and shows in Figure 3-3. The mobile devices will communicate with positioning engine through APs and send the information of signals including strength of signals to the positioning engine for locating themselves. The APIs provided by software manufacture are applied to write a location data capture program (LDCP) to capture the position data of visitor from positioning engine and store into database server when the positioning engine located the mobile devices.

Theoretically, the distributed structure system has better efficiency. But consider the mobility of the system, convenience to data collection, and the environment of National Museum of Prehistory, we choose to use a laptop with Intel Pentium M 1.86GHz CPU and 2G RAM as main server which will not overload. The Ekahua Real Time Location System (ERTLS) is installed on the laptop for locating and set up MySQL5.0 on it for storing data. In other case, researchers could consider to separate the ERTLS and MySQL5.0 into two machines depend on their requirements.

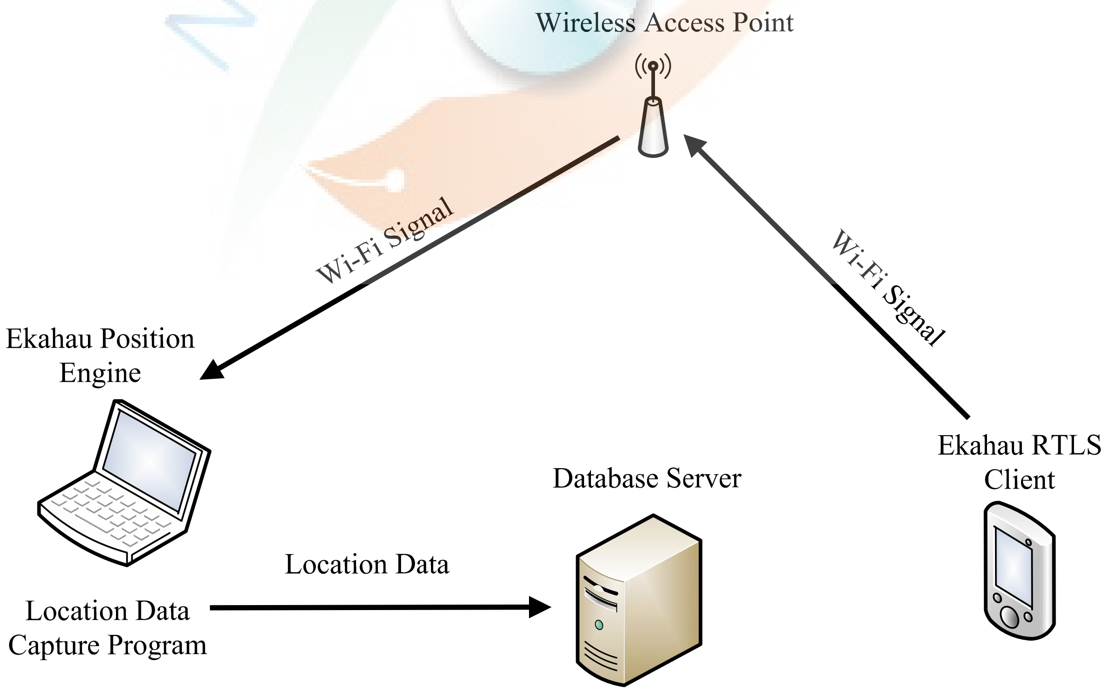

Figure 3-3: Structure of Visitor Location Record System

3.2.1 Access points

Before this research proceeds, there is already a Wi-Fi environment in National Museum of Prehistory. But the APs are too old to provide stable and strong signals. We suggest the museum to change the old APs for new ones. After the old APs were replaced with new ones, the signals become more stable and stronger. Besides, all channels of APs are the same originally. The signals of APs interfere with each others. In order to avoid interferences between APs, the channels of APs is set to be different with their neighbors. The distribution of wireless access points in display areas is shown as the red points in Figure 3-4 and Figure 3-5. The signal around all of the display areas is surveyed, and the signal is ensured to cover all of the display areas fully.

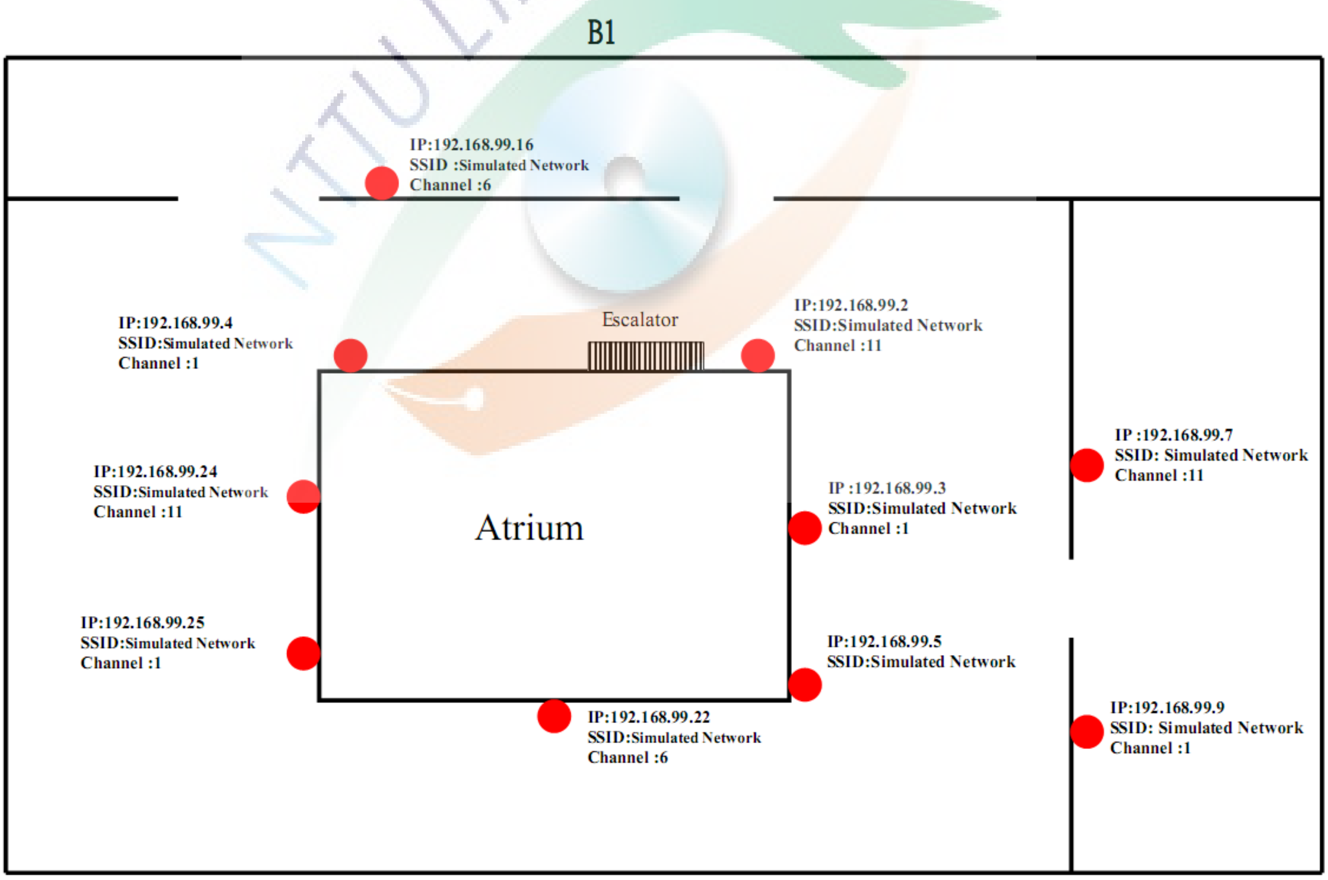

Figure 3-4: Wireless Access Points Distribution at B1

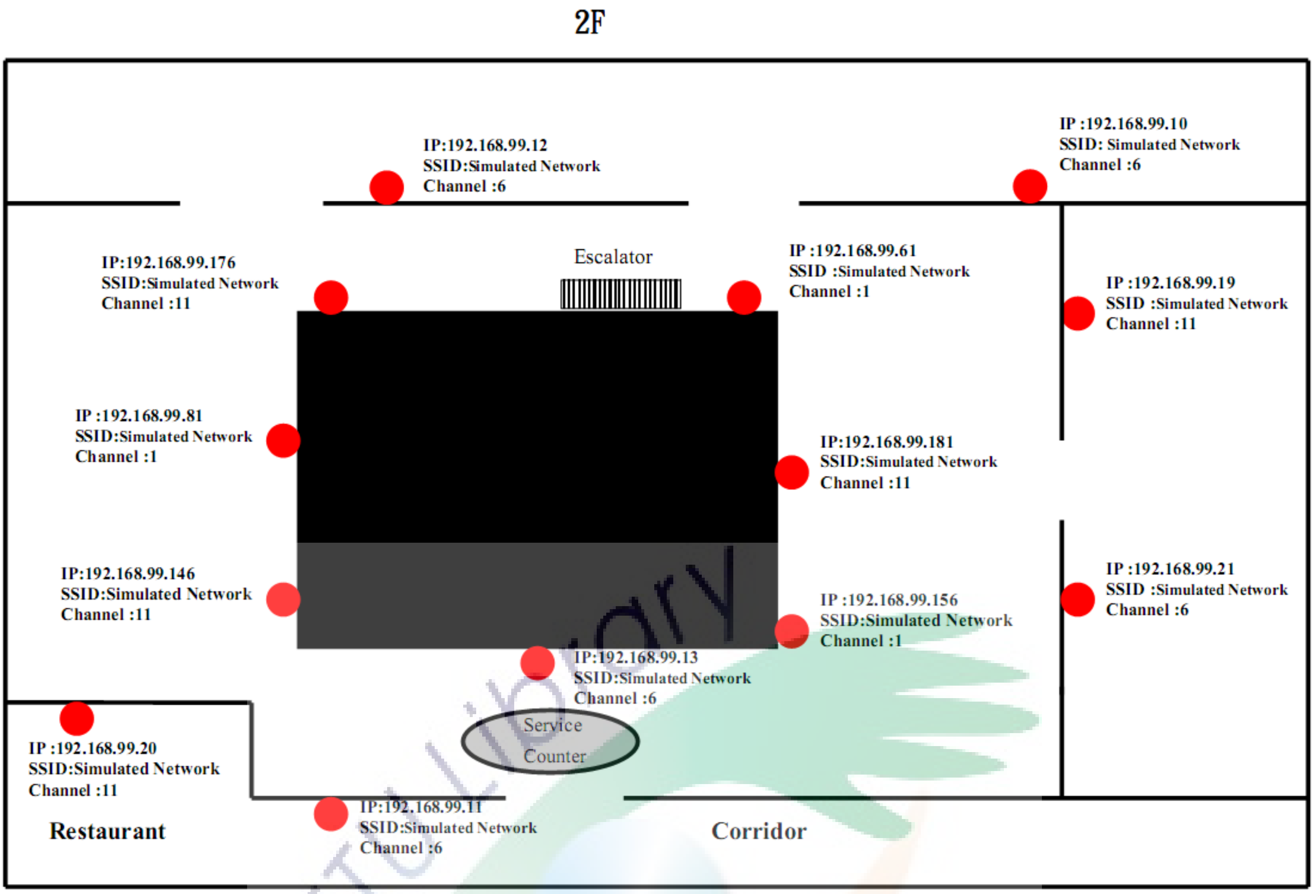

Figure 3-5: Wireless Access Points Distribution at 2F

Before starting to use the ERTLS, it is necessary to survey the signals around whole display areas to build a *positioning model*. The positioning model stores the information of signals of the display areas which be surveyed. Although the locating algorithm of ERTLS is not introduced in the ERTLS user manual, it is clearly that the positioning model influences the locating accuracy in our experiments. The positioning engine compares the signals which come from the mobile devices with the positioning model kept in the position engine too determine the position of mobile devices.

After several times of trials and errors, two major factors influence the quality of positioning model are found. First, walk speed and stability. When survey the signals, operator have to walk slow and steady as documented in the manual. It is suggested that walk one step per second and keep steady when survey the signals or used a wheelbarrow to assist the laptop be moved steady. Second, divide the whole survey into small segments. Because human beings are not easy to move steady for long distance, to reduce the distance of each segment is helpful for moving slow and steady. Follow these two principle, a position model with best locating accuracy close to 1M and worst locating accuracy about 5M is built. The model is precise enough to distinguish the visitors in different display areas and we adopt this model in this research.

3.2.2 Mobile devices

In order to access the positions of visitors, it is needed to request them to carry mobile devices for position locating. Because there are already some PDAs for digital guiding in National Museum of Prehistory, the ERTLS client is installed on PDAs for position locating instead Wi-Fi tags. There are two advantages. First, do not have to buy more mobile devices, it can save some cost. Second, the signal of PDAs is stronger the signal of Wi-Fi tags. It will be more precise to locate PDA than Wi-Fi tags. Although the size of PDA is a little larger than Wi-Fi tags, it may reduce visitors' will of carrying mobile device. In this research, it is attempted to find volunteers of visitors to assist data collection and discards this problem.

3.2.3 Data Format

The data format used for storing visitor location data is shown in Table 3-2. The fields include AUTO, QUESNUM, X, Y, ZONE, SEQ, and TIMESTAMP. AUTO (auto-numbering) is a serial number for ensuring that uniqueness of each record and is used as the primary key. QUESNUM (questionnaire number) is the serial number of the questionnaire; and this number is used to link the questionnaire data with the movement data of a specific visitor (*who*, operationally defined as who carries the mobile device).

Although QUESNUM can be used as the primary key without AUTO, AUTO preserves the original sequence of the questionnaire distribution in the case of responses deletion and maintains the integrity of the database contents. Furthermore, AUTO serves as a system-checking provision during the data entry. That is, not only a mistaken QUESNUM during the manual data entry would not halt the process, but each entry may be verified against a system-provided number to help prevent entry errors. Furthermore, in this study QUESNUM also doubles up as the identity of the visitor.

X and Y are the coordinates (*where*) used for expressing the visitor location captured by the positioning engine. The museum's display areas are divided into *zones*, with each zone basically contains one exhibition hall. Each hall/zone is assigned a code, and each hall's position is expressed by its coordinates (i.e., X and Y values). The location data capture program receives the hall's code from a visitor's device (via positioning engine) to assess the visitor's whereabouts and stores this code into the data field ZONE. The temporal sequence of a visitor's record is kept in the field SEQ (sequence), and the last SEQ value of a visitor also represents the number of signals that have been sent by the positioning engine. TIMESTAMP contains the date and time (*when*) at which the record is stored into database according to the system clock of the database server.

Table 3-2: Sample of Data Records

AUTO	QUESNUM	X	Y	ZONE	SEQ	TIMESTAMP
1	20825001	0	0	F1	0	2008/8/25 10:14:41
2	20825001	310	124	F1	1	2008/8/25 10:14:42
3	20825001	300	126	F1	2	2008/8/25 10:14:45
4	20825001	287	127	F1	3	2008/8/25 10:14:48
5	20825001	278	127	F1	4	2008/8/25 10:14:52
6	20825001	229	127	null	5	2008/8/25 10:14:55
7	20825001	226	127	null	6	2008/8/25 10:14:58
8	20825001	226	127	null	7	2008/8/25 10:15:01
9	20825001	227	127	null	8	2008/8/25 10:15:04
10	20825001	154	117	F2	9	2008/8/25 10:15:07
11	20825001	150	111	F2	10	2008/8/25 10:15:10
12	20825001	102	52	F2	11	2008/8/25 10:15:15
13	20825001	105	54	F2	12	2008/8/25 10:15:18
14	20825001	119	61	F2	13	2008/8/25 10:15:21
15	20825001	118	59	F2	14	2008/8/25 10:15:24

3.3. Data Collection

In this research, data is retrieved from the visitor satisfaction survey result of National Museum of Prehistory (Hsieh et. al, 2008). It is attempted to ask all visitors who come to National Museum of Prehistory at the entrance of exhibition as possible as they can and give PDAs to the volunteers. There are lots of visitors who do not come alone. In fact, most of them visit the museum with companies. Undoubtedly, a group of visitors usually stick together during their tour. In order to maximize the effectiveness of limited number of PDAs to get more kinds of data, only one PDA is given to one of the member of the group. And the volunteers are told to follow their will to discover the museum and reclaim PDAs at the exit. The system will record their position after the volunteers get PDAs. When the volunteers walk out, they will be requested to fill questionnaire which design by the National Museum of Prehistory satisfaction survey team.

Turban et al. (2007) suggest that the decision support includes the steps: *intelligence*, *design*, *choice*, and *implement* (see Figure 3-6). An efficient system is designed to require duration data with less cost and more precise result. This approach which was proposed in this research can help us do the work of intelligence phase easier and faster. It is an innovation of intelligence phase in the decision support process in museum.

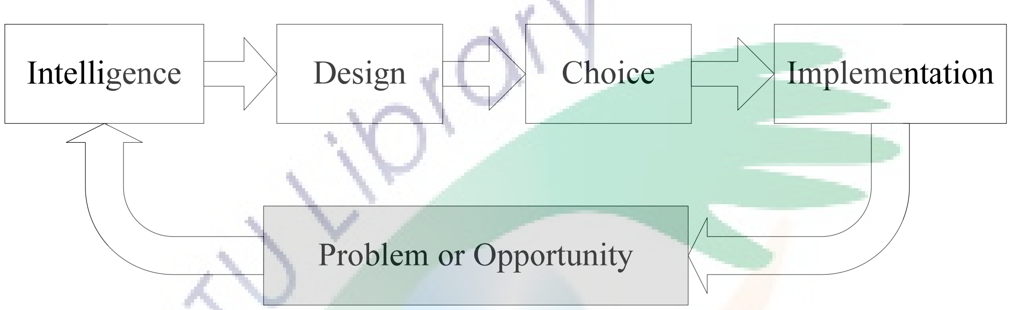

Figure 3-6: Four Steps of Decision Support

All of the questionnaires will be keyed into an Excel file then import into database for analyzing. After data collection phase is complete, two kinds of data can be gained from database server including visitor location data and questionnaire data. For convenience to analyze these data, this data is imported into MSSQL server 2005. It will be easier to manage these data by SQL language. Basically, descriptive statistics approaches are used to analyze these data. After this process, the hotspots with the exhibits with high average durations can be found in the museum.

3.4. Data Analysis Structure and Applications

The visitor location record system can be a data source of decision support system in museum. Because visitor location data is large, it will be more efficient to sort out the data before using in DSS. The DSS can require visitor location data from the visitor duration data warehouse (see Figure 3-7). This data could be transformed to business intelligence and keep in the organization. And it is helpful for decision makers when they face some problems in museum.

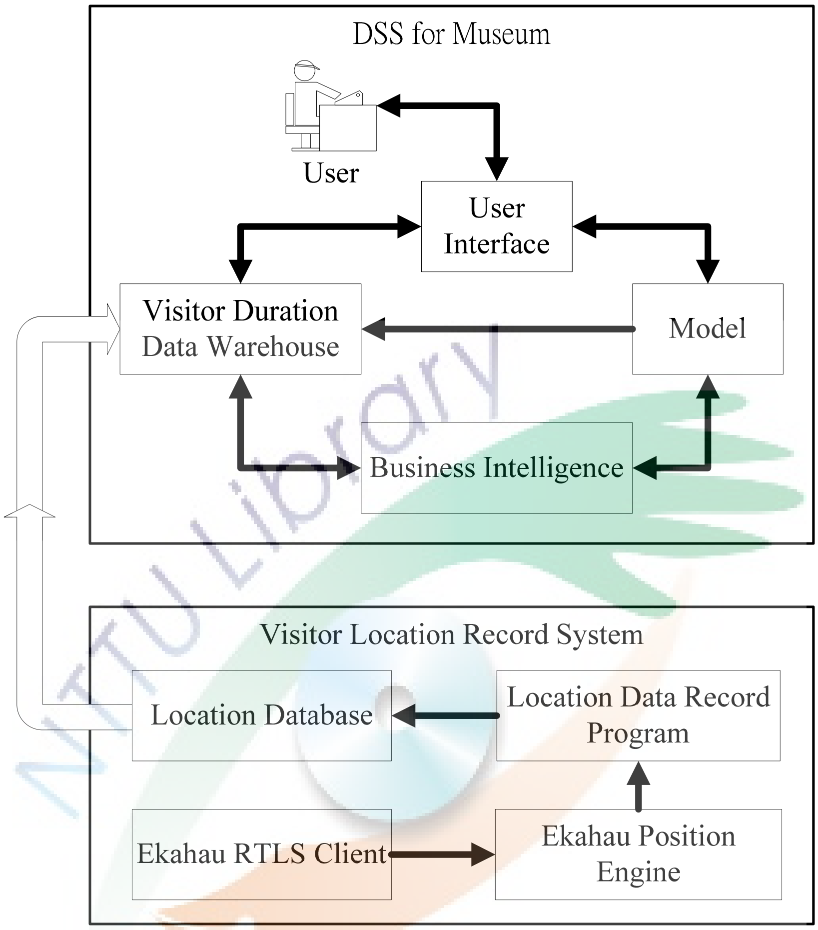

Figure 3-7: Framework for Decision Support System Integration

3.5. Concept of Modeling Visitor Behavior

In traditional RFID research work, tags were attached to merchandizes to track their movements along, say, the assembly line. In this research, visitors are "treated as objects," that is, they are modeled as if they were merchandizes that go through assembly lines with RFID tags attached to them, in this case by carrying PDAs. Thus, we may separate objective positions from subject humanity such as feeling, emotion, and thought of visitors is discarded. The VLRS can be used to record location data of visitor's tour automatically just like the traditional applications of materials/products management and pattern extracting techniques are used to analyze the data systematically. This approach can avoid processing complex problem of humanity. After the patterns are identified, researchers can re-present visitor's behavior by analyzing the patterns with the viewpoint of human beings. Figure 3-8 show the concept of this approach.

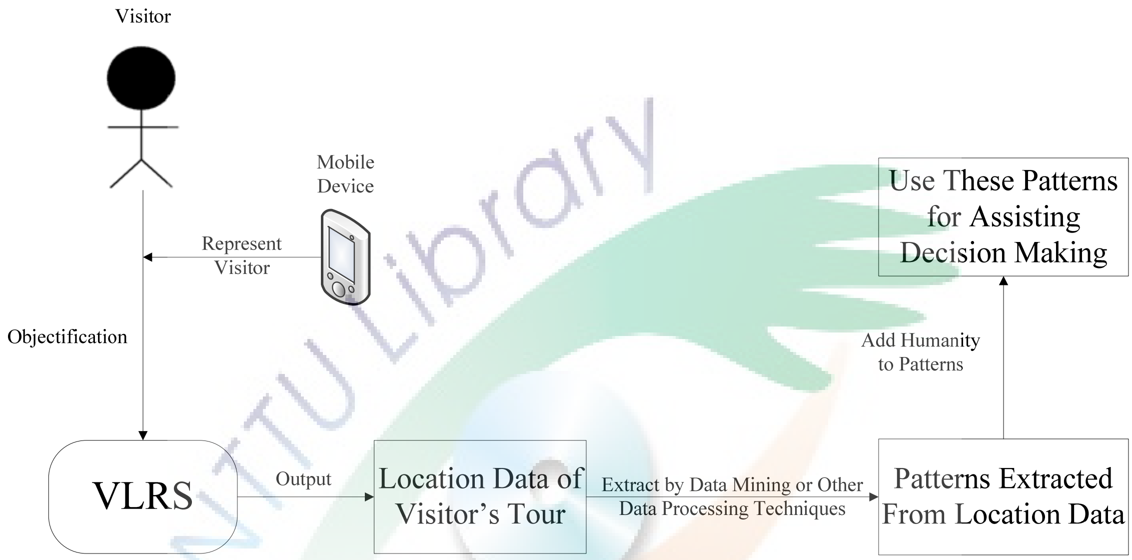

Figure 3-8: Concept of Modeling Visitor Behavior

3.6. Feasibility

Compare with the traditional approach and new approach proposed in this thesis, the new one is more *efficient* and *effective* in long period observation of visitor behavior. Table 3-3 shows the comparison of these approaches. Traditionally, researchers have to send observers to follow visitors or use questionnaires to access their behavior in the tour. A system will be constructed to record visitor behavior in new approach. Observation of visitor's tour takes the same time in all kinds of approach because it depends on how much time a visitor spent in his/her tour. But the cost of observation is different between tradition approach and new one. The most expensive part of traditional approach is manpower. The unit cost of recording per visitor's tour in traditional approach is fixed. The more data needed, the more manpower needed. The cost will increase with the number of data needed. Once the system is set completely, the unit cost of recording per visitor's tour will decrease with numbers of data recorded. The new approach will be more *efficient* (cost lower) to observe visitor behavior for long period of observation. Since the new approach records the visitor behavior automatically, it will produce fewer human-caused errors. Results will be more objective and precise. In a word, the new approach is more *effective*.

Table 3-3: Comparison of Visitor Behavior Observing Approaches

Approach	Time	Cost	Manpower	Precision	Result
System	Depends on visitors.	Decrease with numbers of data recorded.	Low.	High.	Objective
Questionnaire	Depends on visitors.	Medium.	Low.	Low.	Subjective.
Observing by Person	Depends on visitors.	Increase with numbers of data needed	High.	High.	Subjective.

CHAPTER 4

DATA PROCESSING AND ANALYSIS

In this chapter, missing data rate is calculated to evaluate the data quality and use average viewing time to define the hotspot and identify them. Besides data mining approaches are used to extract the potential maximum visit path in the museum.

4.1. Data Quality Control and Processing

369,006 location data records are collected from 370 visitors. During data collection, T2 and T3 did not open for visitor. After delete the invalid data, there are still 365,902 location data records from 367 visitors. During the data collection phase, it is found that visitors may play the PDAs and cause the ERTLS client work improperly. So, the missing data rate of each visitor is calculated to evaluate the data quality.

According to the settings of ERTLS, PDAs report their location every 3 seconds. Since the start time and end time of each visitor are recorded, the real total can be got time by calculating end time minus start time and calculate theoretical location data reporting times.

$$TTT = ET - ST, \quad (1)$$

$$TLDRF = TTT/3, \text{ and} \quad (2)$$

$$MDR = [1 - (ALDRF / TLDRF)] * 100\% \quad (3)$$

where

TTT is the total touring time (in seconds),

ET is the ending time (to the seconds),

ST is the starting time (to the seconds),

TLDRF is the theoretical location data reporting frequency (the number of data submission), MDR is the missing data rate (in %), and

ALDRF is the actual location data reporting frequency (the number of data submission).

The partial result of missing data rate is show in Table 4.1. In order to ensure data quality, the data with missing data rate higher the 25% were discarded. The missing data rate level could be modified according to the research needs; for example, if higher quality data are needed, the missing data rate level can be raised. After such "scrubbing," 189,677 location data records from 145 visitors are used in this research.

Table 4-1: Partial Result of Missing Data Rate

QUESNUM	STARTTIME	ENDTIME	MISSINGDATARATE
20828001	2008/8/28 09:16:47	2008/8/28 12:03:30	98%
20828002	2008/8/29 09:22:09	2008/8/28 10:42:14	6%
20828003	2008/8/28 09:29:41	2008/8/28 11:13:06	45%
20828004	2008/8/28 09:37:13	2008/8/28 10:15:42	26%
20828005	2008/8/28 09:44:23	2008/8/28 09:48:45	78%
20828006	2008/8/28 09:59:51	2008/8/28 10:55:18	57%
20828007	2008/8/28 10:01:55	2008/8/28 12:08:15	18%
20828008	2008/8/28 10:07:43	2008/8/28 11:23:35	28%
20828009	2008/8/28 10:12:43	2008/8/28 11:51:25	84%
20828010	2008/8/28 10:26:28	2008/8/28 11:27:51	49%

These data are saved in a database, maintained using Microsoft SQL Server 2005. Once they are stored, a simple SQL statement:

"select zone,count(quesnum) from TRACKTRANS where quesnum in(select quesnum from HOTSPOTPATH where convert(int,missingdata)<=25) group by zone"

can be used to calculate the total viewing time visitors spent at each zone. The total number of visitors of those zones can be obtained using the following SQL statement:

“select count(quesnum) from HOTSPOTPATH where convert(int,missingdata)<=25”

It can be seen that when a visitor returns to a zone for additional viewing, this revisiting time would be included in the total viewing time for the analysis. However, the number of returns and their effects are not analyzed for the time being since they are beyond the scope of current research.

4.2. Hotspots

As described in Section 2.4, many scholars use average viewing time to assess the visitor’s holding power and thus the attractiveness of the exhibit. In this research, *average viewing time* of all visitors is used to assess the location of hotspots, and the data needed for this simple measurement can be obtained from the SQL statements described in Section 4.1.

Hotspots is defined to be the exhibition areas with the highest holding power -- in other words, the highest average viewing time – in the museum. Table 4.2 shows the average viewing time of exhibition zones in the museum.

Table 4-2: Average Viewing Time of Each Zone

Zone	Average viewing time (Min.)	Zone	Average viewing time (Min.)
F0	2.47	B1	3.31
F1	5.16	B2	2.67
F2	5.01	B3	1.53
F3	4.44	B4	1.61
F4	2.03	B5	2.67
F5	0.93	B6	1.95
F6	0.88	B7	2.45
T4	4.54	B8	2.60
SC	6.92	SA	6.59
null	7.64	--	--

From Table 4-2, it is easy to identify that the top five hotspots are null (7.64), SC (6.92), SA (6.59), F1 (5.16), and F2 (5.01); where the numbers in the parentheses are their corresponding average viewing time. Since most of the null areas are rest areas and corridors and not the point of exhibition interest, they are excluded from this analysis. Thus, the top five hotspots are identified to be exhibitions SC (6.92), SA (6.59), F1 (5.16), F2 (5.01), and T4 (4.54); and these should correspond to the favorite exhibitions for the museum visitors.

The managers may use these data as starting points to conduct in-depth analyses for assess the "attractiveness" of exhibitions. For example, zone SC contains the main spiral corridor that connects floors and it does take some time just to walk through the length of the zone. This walking time (estimated to be about 2.5 minutes) should be excluded from its total to better represent the average viewing time. Once these hotspots are identified, museum personnel may check against their preconceived notion to find discrepancies. These discrepancies may be used to point out some "hidden" treasures or problems that have escaped notice.

These hotspots naturally represent high-traffic areas in the museum, and to increase the viewership of otherwise-overlooked exhibits, the museum may place important or new items around these hotspots. The museum may also take advantage of these hotspots for promoting souvenir, merchandise items, as well as cross-marketing future exhibits or other partners. Since "coldspots" represent low-traffic areas, once they are identified the museum may be able to find the causes and make improvement. According to the usage pattern, the museum may be able to arrange exhibits to balance the "load" – which may allow visitor flow more smoothly and quickly from exhibit to exhibit.

4.3. Hot-paths

4.3.1. Hot-paths Extraction

Since the *sequence* of zone visited is recorded, paths – that is, the movement from zone to zone – can be organized, and naturally the most frequently used passageways may be considered as *hot-paths*. Common sense dictates that many of such paths would be determined by the natural layout of the building, yet many variations could exist in radial-designed buildings with “planks” which allow visitors to jump from zone to zone. Using music media as an analogy, the sequential layout would be similar to that of a cassette tape while the latter a CD. Based on the data format shown in Table 3.1, an algorithm is designed to organize the spots into paths.

Because the error range of ERTLS is around 3 meters in radius, a gray zone (see Figure 4-1) exists between two exhibition zones in which a visitor is hard to be located accurately. The gray zone of the partial data shown in Table 4-3 is an example of such messy data.

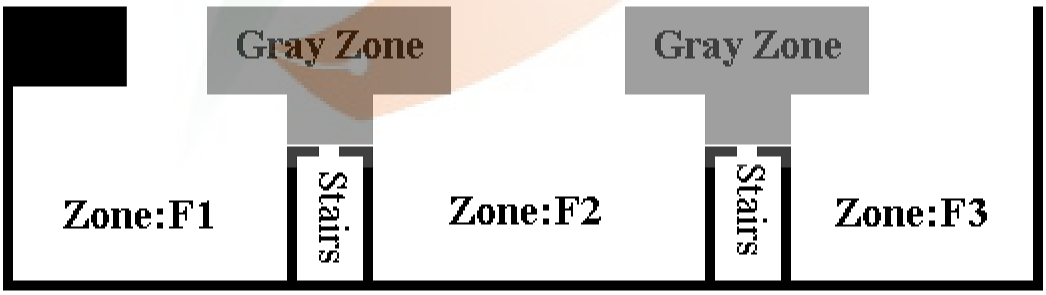

Figure 4-1: Framework of Decision Support System Integration

Table 4-3: Partial Data Collected in Gray Zone

AUTO	NUM	X	Y	ZONE	SEQ	TIMESTAMP
10534	20826001	667	620	F4	914	2008/8/26 10:27:23
10539	20826001	658	620	F4	915	2008/8/26 10:27:26
10546	20826001	643	620	F4	916	2008/8/26 10:27:29
10549	20826001	647	620	F4	917	2008/8/26 10:27:32
10555	20826001	647	620	F4	918	2008/8/26 10:27:35
10561	20826001	651	620	F4	919	2008/8/26 10:27:38
10571	20826001	650	620	F4	920	2008/8/26 10:27:44
10577	20826001	636	620	F4	921	2008/8/26 10:27:47
10582	20826001	636	620	F4	922	2008/8/26 10:27:50
10639	20826001	631	366	F5	923	2008/8/26 10:28:22
10646	20826001	576	397	null	924	2008/8/26 10:28:26
10654	20826001	576	469	null	925	2008/8/26 10:28:29
10663	20826001	679	468	F5	926	2008/8/26 10:28:32
10671	20826001	576	475	null	927	2008/8/26 10:28:35
10683	20826001	656	470	F5	928	2008/8/26 10:28:40
10691	20826001	671	471	F5	929	2008/8/26 10:28:43
10696	20826001	672	471	F5	930	2008/8/26 10:28:46
10703	20826001	667	471	F5	931	2008/8/26 10:28:49
10716	20826001	576	279	null	932	2008/8/26 10:28:52
10721	20826001	576	455	null	933	2008/8/26 10:28:56
10728	20826001	576	449	null	934	2008/8/26 10:28:59
10736	20826001	576	398	null	935	2008/8/26 10:29:02
10742	20826001	576	386	null	936	2008/8/26 10:29:05
10751	20826001	576	385	null	937	2008/8/26 10:29:10
10758	20826001	631	381	F5	938	2008/8/26 10:29:13
10765	20826001	638	288	F5	939	2008/8/26 10:29:16
10780	20826001	716	309	F5	940	2008/8/26 10:29:22
10788	20826001	632	289	F5	941	2008/8/26 10:29:26
10794	20826001	635	288	F5	942	2008/8/26 10:29:29
10799	20826001	629	292	F5	943	2008/8/26 10:29:32
10806	20826001	628	283	F5	944	2008/8/26 10:29:35
10812	20826001	631	287	F5	945	2008/8/26 10:29:38
10820	20826001	629	289	F5	946	2008/8/26 10:29:43

It is obvious that the data from SEQ 923 to SEQ 937 is a mess. In order to overcome this problem, we set a parameter α as the lowest staying time level. If ZONE do not hold the same value continually cause that the staying time level is not higher than α, it's will be considered that the visitor do not really stay in the area. In order to reduce the data, the records which are next to each other and have the same ZONE value will be merged. According to these two rules, a program is written to extract paths from the data which is qualified by the data evaluating procedure and store into another table in the database. Table 4-4 is the partial result of the paths extraction program. As it shows in Table 4-4, the symbol "=>" is used to divide each zone and the number record how many records with the same ZONE value after the comma. These paths can easily show that the specific visitor stayed in which exhibition zones, how long he/she stayed in each exhibition zone, and the sequential relationship between these zones in a tour.

Table 4-4: Partial Result of the Paths Extraction

QUESNUM	PATH
20825003	F1,48=>F2,32=>F3,63=>SC,85=>B1,39=>B2,55=>B5,53=>B8,118=>F4,34
20827016	F1,176=>F2,142=>F3,30=>SC,147=>B8,20=>B1,53=>B7,41
20827020	F1,123=>F2,26=>F3,26=>SC,29
20828013	F1,133=>F2,56=>F3,22=>F3,20=>SC,77=>SC,38=>B8,22=>B7,56=>B6,78
20829003	F1,26=>F2,53=>F3,84=>SC,200=>B1,44=>B2,30=>B8,24=>T4,192
20830008	F1,81=>F2,64=>F3,80=>F3,20=>SC,129=>B8,22=>B6,24=>B6,42=>SC,20
20830012	F1,24=>F1,55=>F2,116=>F3,79=>SC,35=>B6,33
20831009	F0,26=>F1,64=>F2,77=>F3,43=>SC,73=>B3,55=>B5,29=>T4,121
20831011	F0,118=>F0,21=>B1,117=>B2,92=>B3,39=>B3,56=>F0,714
20831020	F1,116=>F2,102=>F3,83=>F3,26=>F4,66=>F5,28
20902007	F2,34=>F3,24=>SC,31=>B1,51=>B2,29=>B5,22=>SA,77=>T4,142
20903004	F1,44=>F2,59=>F2,52=>F3,20=>F4,209=>F5,56=>F6,31=>F6,23
20903005	F1,53=>F2,44=>F3,34=>SC,21=>B5,41=>SA,25=>SA,56
20903010	F1,62=>F2,105=>F4,64=>F6,22
20903011	F1,21=>T4,369=>T4,196=>F4,164=>F5,23=>F6,90

4.3.2. Hot-paths Mining

Although each visitor's tour paths can be extracted as shown in Section 4.3.2., it is still difficult to recognize the patterns through visual inspection. Agrawal and Srikant (1995) propose *sequential pattern mining* to extract patterns in such situations in consumer purchasing behavior. Han et al. (2000) propose a "Freespan" algorithm to improve the efficiency of sequential pattern mining; Pei et al. (2001) extended these research works and improve the efficiency through a "Prefixspan" algorithm. Lee and Wang (2003) apply this sequential pattern mining technique to mine the frequent calling paths in GSM networks.

Mining the hot-paths in museum faces similar problems in mining the frequent calling paths in GSM networks; therefore this research adopted, with modification, Lee and Wang's algorithm in hot-paths mining. In GSM network, because of the characteristics of hexagonal cell, each cell is surrounded by at most six neighboring cells; consequently, each user can only have six possible path selections when moving into neighboring cells. In the museum case, the path selection is not structured at all; a visitor can move to any zone including circling back to itself by passing through the gray zones. Before presenting the algorithm of mining hot-paths, some terminologies which will be used later are defined below.

Definition 1. A visit path P, denoted by $v_1 \Rightarrow v_2 \Rightarrow \ldots \Rightarrow v_n$, $n \geq 2$, is the sequence of visited zones during a tour of specific visitor where $v_1, v_2, \ldots, v_n$ are codes of zones.

Definition 2. The *support* of a visit path P is the ratio of visit paths containing P to all of the visit paths in the database D. The support of P is defined as $sup(P) = |P| / |D|$

where |P| denotes the number of visit paths containing P in D, and |D| denotes the number of all visit paths in D.

Definition 3. A visit path P is said to be a *hot-path* if sup(P) which is termed the *minimum support* is not less than a user-specified threshold. If all the paths in graph *G* are hot-paths, *G* is called a *frequent visit graph*.

Definition 4. A *visit path graph G* is consisted of *vertexes* (v_1, v_2,…, v_n) and *edges* (e_1, e_2,…, e_n). An edge is the link of two vertexes. An *in-edge* of vertex *v* in a visit path graph is an edge ending at *v*. An *out-edge* of vertex *v* in a visit path graph is an edge starting at *v*. If a visit path graph G is not connected, it must be consisted of G_1, G_2,…, G_n. G_1, G_2,…, G_n are called *sub-graphs* of graph G

This research proposes the following algorithm to process this special case of "move back to itself" pattern to visitors to move between zones more freely. This algorithm includes 6 steps:

Step 1: Divide all paths into edges to construct a visit path graph;

Step 2: Delete the edges that do not meet the support level needed;

Step 3: Find the special "move back to itself" paths and store them in a collection of frequent paths;

Step 4: Find the nodes which have no in-edge as the start-vertexes;

Step 5: Trace the graph from each *start-vertex* and store the mined paths in the frequent paths collection until all start-vertexes are traced; and

Step 6: Trace rest of the cyclic sub-graphs from any untraced node and store the mined path in frequent paths collection until all nodes are traced.

In step 1, all paths are divided into edges to construct a visit graph *G*. The count of each edge will be stored (see Figure 4-2). Step 2 removes the paths which do not meet minimum support and G becomes a frequent visit graph. Because the special "move back to itself" paths are not easy to process with other paths, they are identified and removed in step 3. G may not be connected and it may be consisted of two kinds of sub-graph including cyclic sub-graphs and non-cyclic sub-graphs (see Figure 4-3). In Figure 4-3 the cyclic graph contains a cyclic path v_2=>v_4=>v_5=>v_3=>v_2 and a non-cyclic path v_1=>v_2, the non-cyclic graph contains two paths v_1=>v_3=>v_4 and v_1=>v_2=>v_5. In step 4, start-vertexes which have no in-path of these vertexes are identified and will be used to be the beginning of tracing the frequent visit graph. All non-cyclic sub-graphs and the cyclic sub-graphs which connect with non-cyclic sub-graphs with will be traced in step 5. For example, v_1 will the beginning of tracing the sub-graph and the result will be the path v_1=>v_2=>v_4=>v_5=>v_3=>v_2. In step 6, rest of the cyclic sub-graphs will be traced.

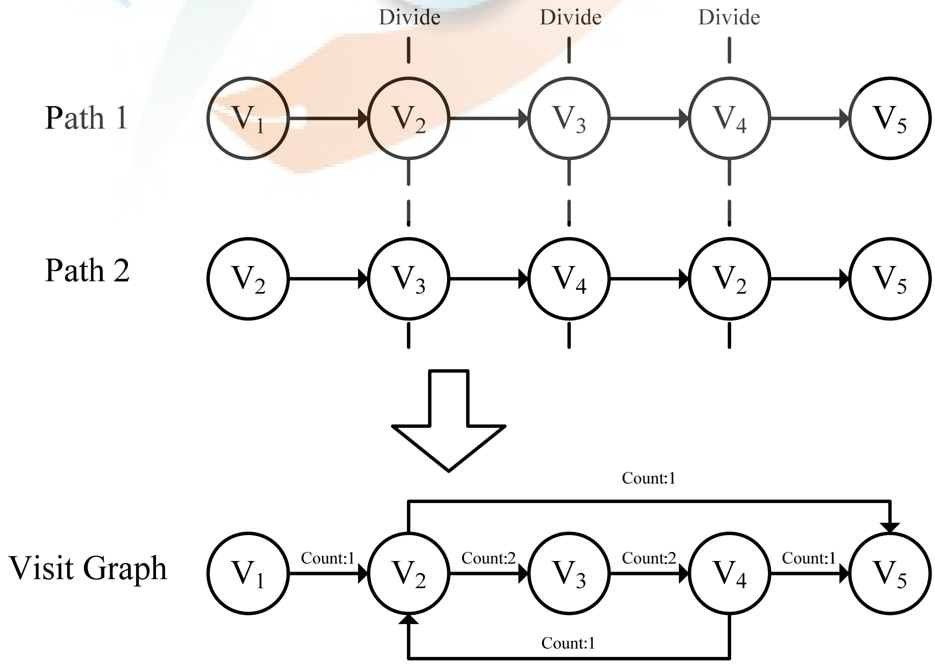

Figure 4-2: Visit Graph Construction

Cyclic Sub-graphs Non-cyclic Sub-graphs

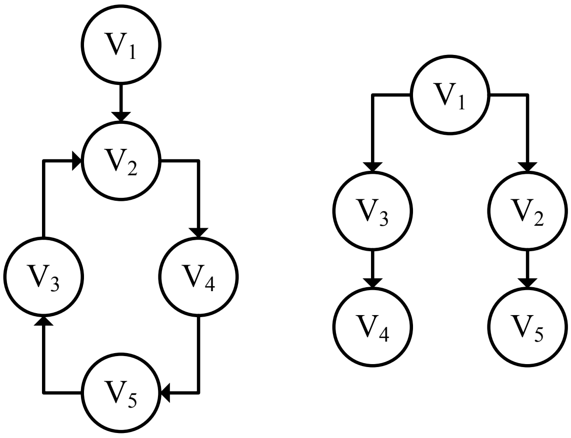

Figure 4-3: Cyclic and Non-cyclic Sub-graphs

Using the algorithm above, a computer program (see Appendix) is written to process the positioning data. The minimum support can be modified according to the researchers' needs. Figure 4-2 shows the result of hot-paths mining with 30% minimum support. Five hot-paths are identified: F1=>F1, SA=>SA, F3=>F3, B4=>B5=>B6=>SA, and F1=> F2=>F3=>SC=>B1=>B2=>B3.

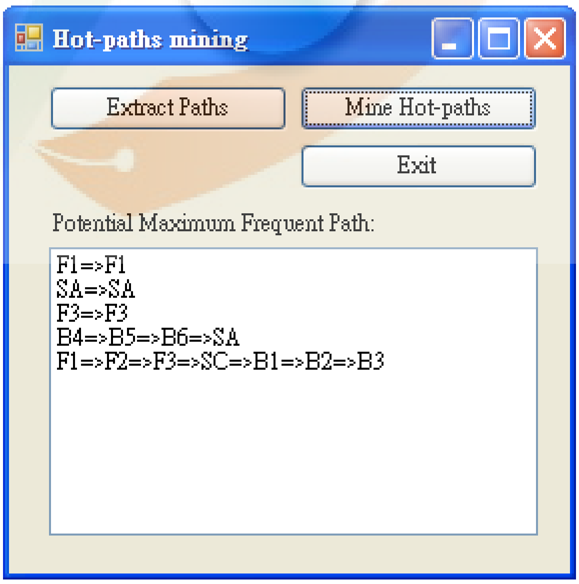

Figure 4-4: Result of Hot-paths Mining

Strictly speaking, the special “move back to itself” paths are not paths, after excluding these special paths, we obtain two hot-paths: B4=>B5=>B6=>SA and F1=> F2=>F3=>SC=>B1=>B2=>B3.

According to the museum tour guides, a comprehensive visitation would take 8 hours to cover all exhibitions, but most visitors have far less time to spend. For the convenience to visitors, based on these hot-paths managers may design recommended touring paths with different time requirements for visitors to choose. In our example here, there may be two recommended touring paths. One path is B4=>B5=>B6=>SA which includes part of prehistory exhibitions and scientific archaeology exhibition and takes about 1.5 hours to complete. The other is F1=>F2=>F3=>SC=>B1=>B2=>B3, and it includes nature history of Taiwan, part of prehistory exhibitions, and human evolution exhibits; and this path takes about 2 hours.

CHAPTER 5

CONCLUSION AND DISSCUSION

In this research, we have developed a new application for wireless locating technology that is different from the traditional "product positioning," "product tracking," and "product history." In the past, the wireless locating technology such as RFID and Wi-Fi locating technology is used in logistic management, warehouse management, material/product management – just to name a few; but none of them is focused on consumer behavior research. In our research we first treat visitors as "objects" so we may take advantage of the traditional applications such as VLRS to collect the whereabouts of visitor's tour and then analyze the collected data to assess their behaviors – in other words, add the "humanity" back to these "objects". This approach and concept can greatly expand the applications in psychology, sociology, and especially consumer marketing researches.

The other objective of this thesis is to develop a system to make previous infeasible research concepts work. This research designed and implemented the VLRS system to record the time (*when*) and location (*where*) of specific visitor's (*who*) tour. This system is more efficient and cost less than old approaches to collect the data of visitor behavior. This methodology is also easier to be applied in large scale exhibitions. Lot of things which researches want to understand in the past may be measured, indentified, or found by the objective data provided by this system. This system can observe the visitors continually. It is hard to accomplish in the past, especially in large exhibitions and long time observation, because it is too complex and cost too much manpower. Time, location, and the object (visitor) are the most important data which wireless locating system can provide. In fact, these three kinds of data are the only things what wireless locating system can provide. Based on these data, the hotspots are identified by calculating average viewing time as done in section 4.2, paths can be extracted by the approach proposed in section 4.3.1, and hot-paths are identified by the modified algorithm introduced in section 4.3.2. All objective information of visitor behavior can combine the result of questionnaire (subjective information) of satisfaction survey team of National Museum of Prehistory and increase the dimension of data analysis of satisfaction study. For example, the average viewing of each zone can compare with the most impressed exhibitions of visitor surveyed by questionnaire. The objective data may explain the result of questionnaire or create conflict between them.

Because the defects of software and hardware in the ERTLS, it exists an error range of the result of the locating system. During setting up the system, we found that the old APs cannot provide stable and strong signal for wireless locating engine. After replacing the APs with new model ones, the accuracy of locating system is improved. Such as this kind of hardware problems are solvable. Even the triangular locating algorithm is possible to redesign and improve. In a word, it is believed that the accuracy of the locating system can be improved, but it not the objective in this research as mentioned in section 1.5. Though the accuracy of the locating system affects the correctness of the result and creates the gray zone problem. Once the wireless locating technology is improved, the result will also be improved. In the mean time, it only has to modify few details of implement of the system (i.e., data schema and size of zones) to ensure the system can work as well as which does in this research. The scalability was considered when designing the system.

The gray zone which causes by the inaccuracy of locating system may affect the hotspots and hot-paths. However, we did not study how the gray zone affects the hotspots and hot-paths in depth. When extracting the paths, the parameter α is used in section 4.3.1 decide the lowest staying level. Is there an approach to set the best value of α? How the gray zones affect setting the value of α? It is encouraged to study these topics in depth.

VLRS can integrate with touring paths recommend system, active guide system, and questionnaire system. Objective data can be collected from VLRS and subjective data can be collect from questionnaire system. Since both of these two data are retrieved, it is possible to find out who is interested in which exhibitions. These patterns could be used to construct touring paths recommend system and provide the motion trend needed in active guide system. It will be helpful for visitors and raise the satisfactions of visitors.

Reference

[1] Agrawal, R. and Srikant, R., (1995), "Mining Sequential Patterns," *Proceedings of the 11th International Conference on Data Engineering*, Taipei, Taiwan, March 6-10, 1995, pp. 3-14.

[2] Anderson, W. and Sullivan, M., (1993), "The Antecedents and Consequences of Customer Satisfaction for Firms," *Marketing Science*, Vol. 12, pp. 125-143.

[3] Barker, S. and Kizil, M., (2009), "System Management Approach to Improvement in Longwall Development," *The Proceedings of the 2009 Underground Coal Operators' Conference*, Gwynneville, NSW, Australia, February 13.

[4] Bergman, A.B., Dassel, S.W., and Wedgwood, R.J., (1966), "Time-Motion Study of Practicing Pediatricians," *Pediatrics*, Vol. 38, pp. 254-263.

[5] Bitgood, S., Patterson, D., and Benefield, A., (1988), "Empirical Relationship between Exhibit Design and Visitor Behavior," *Environmental and Behavior*, Vol. 20, No. 4.

[6] Chang, Y.T., (1994), *The Future of Museums in Globalization*, Taipei: Daw Shiang, pp. 219.
張譽騰，1994，「全球村中博物館的未來」，台北：稻鄉，頁 219。

[7] Chen, H.C., (2001), "The Research of Visitor Behavior: 'The World of Calcite Special Exhibition'," *Museology Quarterly*, Vol. 15, No. 3.
陳慧娟，2001，「碳酸鈣礦物展觀眾行為研究」，博物館學季刊，15(3)。

[8] Chen, J.C.H., Chong, P.P., and Chen, Y.S., (2001), "Decision Criteria Consolidation: A Theoretical Foundation of Pareto Principle to Porter's Competitive Forces," *Journal of Organizational Computing and Electronic Commerce*, Vol. 11, No. 1, pp. 1-14.

[9] Daviaud, E. and Chopra, M., (2008), "How Much is Not Enough? Human Resources Requirements for Primary Health Care: a Case Study from South Africa," *Bulletin of the World Health Organization*, Vol. 86, No. 1, pp. 46-51.

[10] Ekahau RTLS Brochure, http://www.ekahau.co m/file.php?id=99414, retrieve date: 2008/3/26

[11] Fact Sheet of Ubisense Series 7000 Sensor, http://www.ubisense.n et, retrieve date: 2008/9/26

[12] Gillespie, J.M., Wyatt, W., Venuto, B., Blouin, D., and Boucher, R., (2008), "The Roles of Labor and Profitability in Choosing a Grazing Strategy for Beef Production in the U.S. Gulf Coast Region," *Journal of Agricultural and Applied Economics*, Vol. 40, No. 1, pp. 301-313.

[13] Gorry , G.M., and Scott-Morton, M.S., (1971), "A Framework of Management Information Systems," *Sloan Management Review*, Vol. 13, No. 1, pp. 55-70.

[14] Han, J., Pei, J., Mortazavi-Asl, B., Chen, Q., Dayal, U., and Hsu, M. C., (2000), "FreeSpan: Frequent Pattern-Projected Sequential Pattern Mining," *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Boston, MA, August 20 -23, pp. 355-359.

[15] Hsi, S. and Fait, H., (2005), "RFID Enhances Visitors' Museum Experience at the Exploratorium," *Communications of the ACM,* Vol. 48, No. 9.

[16] Hsieh, M. C., Chong, P.P., Wang, W. C., and Shin, S.S., (2008), *Taiwan National Museum of Prehistory Visitor Satisfaction Survey 2008*, National Museum of Prehistory, Taitung, Taiwan.
謝明哲、鍾青萍、王文清、辛信興，2008，「國立臺灣史前文化博物館九十七年度遊客滿意度調查研究報告」，國立臺灣史前文化博物館。

[17] Hsu, T.P., (2006), "A Study of Applying 7-11 Location to Model the Section of Hot Point Mechanism Commercial," master's thesis, Graduate Institute of Information Management, Ming Chuan University, Taiwan.
徐子鵬 ，2006 ，「應用 7-11 超商成功模式於選點分析機制之研究」，銘傳大學資訊管理學系碩士在職專班碩士論文。

[18] IEEE Standard 802.11, http://standards.ieee.org/getieee802/download/802.11-2007.pdf, Retrieve Date: 2009/4/20.

[19] Keen, P.G.W. and Scott-Morton, M.S., (1978), *Decision Support System: An Organizational Perspective*, MA: Addison-Wesley.

[20] Lee, A.J.T., Wang, Y.T., (2003), "Efficient Data Mining for Calling Path Patterns in GSM Networks," *Information Systems*, Vol.28, No.8, pp. 929-948.

[21] Li, M.D., (2002), "Discuss the Police's Reaction from Criminal Hot Spots: a Case of Taipei County," master's thesis, Graduate Institute of Police Administration, Central Police University, Taiwan.
李明道，2002，「以犯罪熱點論警察因應作為--以台北縣為例」，中央警察大學行政警察研究所碩士論文。

[22] Lin, Y.C., (2002), "Visitor Behavior Study at Children Environmental Education Exhibition in Taroko National Park," master's thesis, Graduate Institute of Tourism & Recreation Management, National Dong Hwa University, Taiwan.
林宜君，2002，「太魯閣國家公園兒童環境教育館觀眾行為研究」，國立東華大學觀光暨遊憩管理研究所碩士論文。

[23] McDonald, J.S. and Dzwonczyk, R.R., (1988), "A Time and Motion Study of the Anaesthetist's Intraoperative Time," *British Journal of Anaesthetist*, Vol. 61, pp. 738-742.

[24] Pei, J., Han, J., Mortazavi-Asl, B., and Pinto, H., (2001), "PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-projected Pattern Growth," *Proceedings of the 17th International Conference on Data Engineering*, Heidelberg, Germany, April 2-6, pp. 106-115.

[25] Robinson, E.S., (1928), *The Behavior of the Museum Visitor*, Washington, DC: American Association of Museums.

[26] Shi, H.Y., (1989), “A Study of Global Positioning System,” master’s thesis, Institute of Maritime Technology, National Taiwan Ocean University, Taiwan.
石宏揚，1989，「GPS 衛星定位設計之研究」，國立海洋大學航運技術研究所碩士論文。

[27] Taylor, F.W., (1911), *The Principles of Scientific Management*, New York: Norton.

[28] Turban, E., Aronson, J.E., Liang, T.P., Sharda, R., (2007), *Decision Support and Business Intelligence Systems*, 8th Edition, Prentice Hall.

Appendix

The codes below are written in C# to implement the mining algorithm.

```
public partial class Form1 : Form
{
    private string ConnectString;
    private string _ParamA;          // Lowest Staying Level
    private string ParamB;          //Missing Data Rate
    private string _ParamC;          //Minimum Support
    private int ParamA, ParamC;
    private object strobj;
    private FHSP result;
    private LinkedList<string> VisitedNodes;
    private LinkedList<string> UnvisitedNodes;
    AppSettingsReader ASReader = new AppSettingsReader();

    public Form1()
    {
        InitializeComponent();
    }

    private void Form1_Load(object sender, EventArgs e)
    {
        ConnectString = ASReader.GetValue("ConnectString", typeof(string)).ToString();
        _ParamA = ASReader.GetValue("ParamA", typeof(string)).ToString();
        ParamB = ASReader.GetValue("ParamB", typeof(string)).ToString();
        _ParamC = ASReader.GetValue("ParamC", typeof(string)).ToString();
        ParamA = Convert.ToInt16(_ParamA);
        ParamC = Convert.ToInt16(_ParamC);
    }

    private void button4_Click(object sender, EventArgs e)
    {
        DataSet ph = new DataSet();
        FrequentGraph<String> FG=new FrequentGraph<string>();
        FHSP _FHSP = new FHSP();
        PathList<string> pl = new PathList<string>();
        string show = "";
        int LowestCount;
        string qSql = "select [path] from HOTSPOTPATH where convert(int,missingdata)<="+ParamB;
        ph = Select(qSql);

        for(int i=0;i<ph.Tables[0].Rows.Count;++i)
        {
```

```
                LinkedList<HotSpot>
llhs=StringToPath(ph.Tables[0].Rows[i]["path"].ToString());
                if(llhs.Count>1)
                    FG = addLinklist(FG, llhs);
            }

            LowestCount = (ph.Tables[0].Rows.Count * ParamC) / 100;

            foreach (Path _ph in FG.Path)
            {
                if (_ph.Times < LowestCount)
                    pl.Add(_ph);
            }

            foreach (Path _ph in pl)
            {
                if (_ph.Times < LowestCount)
                    FG.Path.Remove(_ph);
            }

            _FHSP = MineFHSP(FG);

            foreach (LinkedList<string> ll in _FHSP)
            {
                foreach(string str in ll)
                {
                    show += str + "=>";
                }
                show = show.Substring(0, show.Length - 2) + "\r\n";
            }
            textBox1.Text = show;
        }

        private    FrequentGraph<String>    addLinklist(FrequentGraph<String>    FG,
LinkedList<HotSpot> llhs)
        {
            string[] sllarr = new string[llhs.Count];
            int[] illarr = new int[llhs.Count];
            int i = 0;

            foreach (HotSpot hs in llhs)
            {
                sllarr[i] = hs.Zone;
                illarr[i] = hs.Times;
                ++i;
            }

            if (llhs.Count > 1)
```

```
        {
            for (int j = 0; j < llhs.Count - 1; j++)
            {
                if (FG.Path == null)
                {

                    FG.Path.Add(new Path(sllarr[j], sllarr[j + 1]));
                }
                else
                {
                    Path ph = new Path();
                    ph = FG.Path.FindByValue(sllarr[j], sllarr[j + 1]);

                    if (ph == null)
                    {
                        FG.Path.Add(new Path(sllarr[j], sllarr[j + 1]));
                    }
                    else
                    {
                        FG.Path.Remove(ph);
                        ph.AddTimes(1);
                        FG.Path.Add(ph);
                    }
                }
            }
        }
        return FG;
    }

    private FHSP MineFHSP(FrequentGraph<String> FG)
    {

        result=new FHSP();
        VisitedNodes = new LinkedList<string>();
        UnvisitedNodes = new LinkedList<string>();

        LinkedList<string> AllNodes = new LinkedList<string>();
        LinkedList<string> AllFromNodes=new LinkedList<string>();
        LinkedList<string> AllToNodes = new LinkedList<string>();
        LinkedList<string> StartVertex = new LinkedList<string>();
        LinkedList<string> CyclicVertex = new LinkedList<string>();
        LinkedList<string> currentPath = new LinkedList<string>();

        PathList<string> pl = new PathList<string>();

        foreach (Path ph in FG.Path)
        {
            if (ph.FromNode == ph.ToNode)
```

```
        pl.Add(ph);
}

foreach (Path ph in pl)
{
    LinkedList<string> tmp=new LinkedList<string>();

    FG.Path.Remove(ph);
    tmp.AddLast(ph.FromNode);
    tmp.AddLast(ph.ToNode);
    result.Add(tmp);
}

foreach (Path ph in FG.Path)
{
    if (!AllFromNodes.Contains(ph.FromNode))
        AllFromNodes.AddLast(ph.FromNode);

    if (!AllToNodes.Contains(ph.ToNode))
        AllToNodes.AddLast(ph.ToNode);
}

StartVertex = copyll(StartVertex, AllFromNodes);
AllNodes = copyll(AllNodes, AllFromNodes);

foreach (string str in AllToNodes)
{
    if (StartVertex.Contains(str))
    {
      StartVertex.Remove(str);
    }

    if (!AllNodes.Contains(str))
        AllNodes.AddLast(str);
}

UnvisitedNodes = copyll(UnvisitedNodes, AllNodes);

while (StartVertex.Count!=0)
{
    string str = StartVertex.First.Value;
    currentPath = new LinkedList<string>();

    UnvisitedNodes.Remove(str);
    VisitedNodes.AddLast(str);
    currentPath.AddLast(str);
    StartVertex.Remove(str);
    trace(currentPath, FG);
```

```
        }

        while (UnvisitedNodes.Count != 0)
        {
            currentPath = new LinkedList<string>();
            string str = UnvisitedNodes.First.Value;

            UnvisitedNodes.Remove(str);
            VisitedNodes.AddLast(str);
            currentPath.AddLast(str);
            trace(currentPath, FG);
        }
        return result;
    }

    private void trace(LinkedList<string> currentPath, FrequentGraph<String>
FG)
    {
        LinkedList<string> Unvisited = new LinkedList<string>();

        foreach (Path ph in FG.Path)
        {
            if (ph.FromNode == currentPath.Last.Value)
                Unvisited.AddLast(ph.ToNode);
        }

        if (Unvisited.Count == 0)
        {
            result.Add(currentPath);
        }
        else
        {
            while (Unvisited.Count!=0)
            {
                LinkedList<string> tmp =new LinkedList<string>();
                tmp=copyll(tmp, currentPath);
                string str = Unvisited.First.Value;

                if(VisitedNodes.Contains(str))
                {
                    Unvisited.Remove(str);
                    UnvisitedNodes.Remove(str);
                    tmp.AddLast(str);
                    result.Add(tmp);
                }
                else
                {
```

```
                Unvisited.Remove(str);
                UnvisitedNodes.Remove(str);
                VisitedNodes.AddLast(str);
                tmp.AddLast(str);
                trace(tmp, FG);
            }
        }
    }
}
```