274 Measuring The Quality Of E-Services And Its Impact On Students Satisfaction At Jordanian Universities
Hareth Alshamayleh, Rasha Aljaafreh, Ali Aljaafreh, Dara Albadayneh

285 Key Determinants Of Customer Satisfaction: Evidence From Malaysia Grocery Stores
Mohammed Al-Ali, Nor Erne Nazira Bazin, Siti Maryam Shamsuddin

300 Efficient Techniques For Predicting Suppliers Churn Tendency In E-Commerce Based On Website Access Data
Veronica S. Moertini, Niko Ibrahim, Lionov

310 Review Of Quality Of Service In Routing Protocols For Wireless Sensor Networks
Abdul Bamatraf, Mohammad Shafie Bin Abd Latiff, Yahaya Coulibaly, Ahmad M. Khasawneh

321 Intensive Fixed Chunking (IFC) De-Duplication For Space Optimization In Private Cloud Storage Backup
M. Shyamala Devi, V. Vimal Khanna, M. Shaheen Shah

329 Microwave Imaging For The Detection And Localization Of Breast Cancer Using Artificial Neural Network
Abdelfetah Miraoui, Lotfi Merad, Sidi Mohamed Meriah

336 An Assessing Approach Based On Fmea Methodology To Evaluate Security Of A Third Party Cloud Provider
Bentajer Ahmed, Abouelmehdi Karim, Dali Loubna, El-Fezazi Said, Hedabou Mustapha, El Amrani Fatimaezzahra

345 Proposed It Financial Management Process Using Itil (IT Infrastructure Library) For Port Company In Indonesia
Dedy Rahman Wijaya

355 An Enhancement Algorithm Using Gabor Filter For Fingerprint Recognition
Einas Almarghni Azzoubi, Rosziati Bint Ibrahim

364 Carrier Based PWM Technique For Harmonic Reduction In Cascaded Multilevel Inverters
S. Leela, S. S. Dash

374 Portfolio Selection Using The Cat Swarm Optimization
Hamza Kamili, Mohammed Essaid Riffi

381 New Architecture Of IDS Based Interaction On Mobile Agents And Data Mining
Chaimae Saadi, Rachid Cherkaoui, Habiba Chaoui, Hassan Erguig

386 IDFP-Tree: An Efficient Tree For Interactive Mining Of Frequent Subgraph Patterns
Mohammad H. Nadimi-Shahraki, Maryam Taki, Mohammad Naderi

391 A Low Power Phase Frequency Detector For Delay-Locked Loop
Lau Weng Loon, Mamun Bin Ibne Reaz, Khairun Nisa Minhad, Noorfazila Kamal, Wan Mimi Diyana Wan Zaki
Welcome to Journal of Theoretical and Applied Information Technology

Journal of Theoretical and Applied Information Technology
April 2015 | Vol. 74 No.3

Title: MEASURING THE QUALITY OF E-SERVICES AND ITS IMPACT ON STUDENTS SATISFACTION AT JORDANIAN UNIVERSITIES

Author: HARETH ALSHAMAYLEH, RASHA ALJAAAFREH, ALI ALJAAAFREH, DARA ALBADAYNEH

Abstract: This research aims at measuring the quality of E-Services and its impact on student satisfaction at Jordanian universities. Toward this aim, a research model was developed and tested within the context of Jordanian universities. A questionnaire consisting of 40 (Likert) type items were distributed to a convenience sample of 755 respondents. Primary data were collected from 703 students with a response rate of 90.7%. SPSS package was used to determine if the quality of E-services dimensions (Website design, reliability, efficiency, responsiveness, ease of use, availability and privacy) has an impact on students satisfaction. The results show that all of the E-services dimensions have an impact on students satisfaction expect for the dimension of ease of use. In addition, the researchers present some recommendations for universities to apply e-services quality dimensions through websites, and some suggestions for future research.

Keywords: Students Satisfaction, Electronic Services Quality, Jordanian Universities.

Source: Journal of Theoretical and Applied Information Technology

Full Text

Title: KEY DETERMINANTS OF CUSTOMER SATISFACTION: EVIDENCE FROM MALAYSIA GROCERY STORES

Author: MOHammed Al-ali, Nor Erne Nazira Bazin, Siti Maryam Shamsuddin

Abstract: Customer satisfaction is critical to retail success. Concentrating on customer satisfaction has become a major goal in retailing industry, especially in grocery retail. This work examines the key determinants of customer satisfaction in grocery retailing and measure the link between store attributes and customer satisfaction. In addition, it aims to find out the effect size of these determinants on overall customer satisfaction in an emerging market, namely Malaysia. For this purpose, an extensive dataset from 313 shoppers who had carried out their purchase in different type of grocery stores has been analyzed. Using Partial Least Squares-SEM (PLS-SEM) analysis method, we show that three determinants (monetary value, service and convenience, and store quality image) have a direct impact on customer satisfaction. However, the weight that each factor shows is different. Results may help managers of grocery stores in Malaysia to develop and implement more successful relationship marketing strategies.

Keywords: Customer Satisfaction, Grocery Stores, Store Attributes, PLS-SEM

Source: Journal of Theoretical and Applied Information Technology

Full Text

Title: EFFICIENT TECHNIQUES FOR PREDICTING SUPPLIERS CHURN TENDENCY IN E-COMMERCE BASED ON WEBSITE ACCESS DATA

Author: VERONICA S. MOERTINI, NIKO IbraHim, LIONOV

Abstract: Electronic supplier relationship management (e-SRM) is important in order to maintain strong, long lasting and beneficial relationship between e-commerce firms and their suppliers. One important function of e-SRM is to predict suppliers who tend to churn such that early treatment can be given. In the e-commerce systems that involve suppliers as the websites users, predicting suppliers churn tendency can be based on analyzing their frequencies in accessing the e-commerce websites. Our proposed techniques include data warehouse design (supporting the data collection and pre-processing) and unsupervised algorithms that analyze the preprocessed bitmaps of time series data representing suppliers website access from time to time. Having bitmaps as inputs, our proposed algorithms are efficient (the time complexity is O(n)) as proven with our experiments. In experimenting with real world data of an e-commerce system selling hotel rooms, our techniques produce output of supplier segments where each segment has certain churn level tendency and need specific treatment.

Keywords: Churn Prediction In E-Commerce, Supplier Relationship Management, Web Usage Mining

Source: Journal of Theoretical and Applied Information Technology
<table>
<thead>
<tr>
<th>Title:</th>
<th>REVIEW OF QUALITY OF SERVICE IN ROUTING PROTOCOLS FOR WIRELESS SENSOR NETWORKS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author:</td>
<td>ABDULLAH BAMATRAF, MOHAMMAD SHAHIE BIN ABD LATIFF, YAHAYA COULIBALY, AHMAD M. KHASAWNEH</td>
</tr>
<tr>
<td>Abstract:</td>
<td>Rapid innovative improvements in wireless communication technology have revolutionized wireless sensor networks (WSNs). A WSN is comprised of self-ruling sensors that are distributed spatially to monitor physical or environmental conditions, such as temperature, sound, vibration, pressure, motion, or pollutants, and to pass this information through the network to a main area. Sensor nodes in wireless sensor networks experience the ill effects of resource constraints, such as energy deficits, buffers, and bandwidth issues. The expanding demand for real-time services in WSN applications means that interest in quality of service (QoS)-based routing has risen. Offering consistent QoS in sensor networks creates considerable challenges. In real time applications, it is important to deliver data as soon as it is sensed. If the network has multiple real and non-real-time applications, its ability to manage them will be challenging due to different requirements. In this study, we investigate QoS-based routing protocols for WSNs. Metrics of analysis are end-to-end delay, congestion, energy efficiency, and reliability. The aim of the study is to identify the limitations of relevant papers and show research direction in routing. This will not only help new comers to the field of WSN but also will ease the tasks of WSN researchers in developing appropriate routing solutions for WSNs.</td>
</tr>
<tr>
<td>Keywords:</td>
<td>QoS, WSN, End-to-end Delay, Congestion, Energy Efficiency, Reliability, Heterogeneity.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Title:</th>
<th>INTENSIVE FIXED CHUNKING (IFC) DE-DUPLICATION FOR SPACE OPTIMIZATION IN PRIVATE CLOUD STORAGE BACKUP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author:</td>
<td>M.SHYAMALA DEVI, V.VIMAL KHANNA, M.SHAHEEN SHAH</td>
</tr>
<tr>
<td>Abstract:</td>
<td>Cloud Storage provides users with abundant storage space and make user friendly for immediate data access. But there is a lack of analysis on optimizing cloud storage for effective data access. With the development of storage and technology, digital data has occupied more and more space. According to statistics, 60% of digital data is redundant, and the data compression can only eliminate intra-file redundancy. In order to solve these problems, De-Duplication has been proposed. Many organizations have set up private cloud storage with their unused resources for resource utilization. Since private cloud storage has limited amount of hardware resources, they need to optimally utilize the space to hold maximum data. In this paper, we are going to discuss the flaws in the existing de-duplication methods and introduce new methods for Data De-Duplication. Our proposed method namely Intensive Fixed Chunking (IFC) De-duplication which is the enhanced File level de-duplication that provides dynamic space optimization in private cloud storage backup as well as increase the throughput and de-duplication efficiency</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Cloud Computing, Private Storage Cloud, Cloud Backup, Data De-Duplication, Chunking, Redundancy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Title:</th>
<th>MICROWAVE IMAGING FOR THE DETECTION AND LOCALIZATION OF BREAST CANCER USING ARTIFICIAL NEURAL NETWORK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author:</td>
<td>ABDELFETTAH MIRAOU, LOTFI MERAD, SIDI 1MOHAMED MERIAH</td>
</tr>
<tr>
<td>Abstract:</td>
<td>Localization and Reconstruction of homogenous objects from electromagnetic scattered fields have been shown to be of great importance, because of their various applications in many areas such as medicine, biology, geophysics and other sciences. In this paper, we propose a computational method for detection and localization of the object for medical application (breast cancer). The proposed technique is based on the use of artificial neural network ANN. A spherical tumor was created and at arbitrary locations in a breast model using an EM simulator. Bow-tie antennas were used to transmit and receive Ultra-Wide Band (UWB) signals at 4GHz. A training and validation sets were constructed to train and test the ANN. A very optimistic results have been observed for early received signal components with the ANN model. Hence, the proposed model is very potential for early tumor detection to save human lives in the future.</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Microwave Imaging, Artificial Neural Network (ANN), Computer-Aided Diagnosis, Breast Cancer.</td>
</tr>
</tbody>
</table>

| Title:         | AN ASSESSING APPROACH BASED ON FMECA METHODOLOGY TO EVALUATE SECURITY OF A THIRD PARTY CLOUD PROVIDER |
| Title: | AN ENHANCEMENT ALGORITHM USING GABOR FILTER FOR FINGERPRINT RECOGNITION |
| Author: | EINAS ALMARGHINI AZZOUBI, ROSZIATI BINT IBRAHIM |
| Abstract: | Fingerprint recognition is being widely applied in the personal identification for the purpose of high degree of security by matching processes between two human fingerprints. Many different techniques have been proposed to have a satisfactory fingerprint identification. The widely used minutiae-based representation does not utilize a significant component of the rich discriminatory information available in the fingerprints. Local ridge structures cannot be completely characterized by minutiae. Further, minutiae-based matching has difficulty in quickly matching two fingerprint images containing different numbers of unregistered minutiae points. We introduced an enhancement algorithm using Gabor filter based matching to capture both local and global details in a fingerprint as a compact fixed-length Fingercode. The improved filtering used to feather extraction in our proposed algorithm with eight different directions. The last step of our proposed algorithm is fingerprint matching, which is based on the Euclidean Distance (ED) between the two corresponding Fingercodes. The proposed algorithm can be personalized according to the value of Euclidean Distance (ED) and threshold (TH). If the ED is less than TH or equal to zero, it means that the two fingerprint images came from the same person. We compared our proposed algorithm (enhanced) with algorithm (without enhanced). The values of false acceptance rate (FAR), false reject rate (FRR) and equal error rate (EER) are lower than the algorithm (without enhanced). Also, the result based on proposed algorithm are also presented and based on the result, it gives higher accuracy and recognition rate. |
| Keywords: | Fingerprint Recognition; Gabor Filter; Fingercode; Matching; Euclidean Distance |
| Full Text: | [
| Title: | PROPOSED IT FINANCIAL MANAGEMENT PROCESS USING ITIL (IT INFRASTRUCTURE LIBRARY) FOR PORT COMPANY IN INDONESIA |
| Author: | DEDY RAHMAN WIJAYA |
| Abstract: | ASEAN Single Window (ASW) implementation ensures compatibility of Member States National Single Windows (NSWs) with international open communication standards while also ensuring that each of those Member States can then exchange data securely and reliably with any trading partners that use international open standards. Simpler and faster processing time, and a more transparent way of doing business. In order to facing of ASW, Indonesian Port Companies have to improve their services. Based on this fact, some Indonesia Port Companies try to invest Information Technology to achieve their goals and improve competitiveness. However, frequently processes related with IT investment do not managed properly. Port Company in Indonesia often ignore linkage analysis between IT investment, business objectives, and business impact. Therefore, the value of IT service not been well defined. This work has raised this issue and try to propose systematic approach for IT Financial Management process based on ITIL best practice standard. Research case study in one of the largest government-owned port company in Indonesia that implements Port Integrated IT services. This work start with planning, analysis, and design to be produced proposed IT Financial Management process. Business case analysis perform in planning phase to make sure the IT investment related with business objectives and has measurable impact. Assessment perform in analysis phase to measure the IT Financial Management process readiness. Total Service Valuation (TSV) also formulated to know the value of Port Integrated IT services. The last is design phase which is defined complete procedures, role and responsibilities for IT Financial Management process implementation. The result of this work is systematic approach to implementation IT Financial Management process in Indonesian Port Company. |
| Keywords: | IT Financial Management, ITIL, Port Company |
| Full Text: | [
| Title: | CARRIER BASED PWM TECHNIQUE FOR HARMONIC REDUCTION IN CASCADED MULTILEVEL INVERTERS |
| Author: | Bentajer Ahmed, AbouElMehdi Karim, Dali Loubna, EL-Fezazi Said, Hedabou Mustapha, El Amrani FatimaEzzahra |
| Abstract: | Cloud Computing is recognized as a great eliminator of the hefty costs and complex processes that come with evaluating, purchasing, configuring or managing software and hardware essentials that are necessary for enterprise applications. However it presents a significant security concerns that need to be addressed when moving to the cloud that should be well studied and quantified for more visibility. In this article we will study how to quantify a risk associated with a cloud service/deployment model and use FMECA methodology to audit a third party cloud provider risks. |
| Keywords: | Trust Computing, Fmeca, Cloud Computing, Predictive Analysis, Confidentiality |
| Full Text: | [Full Text]
### Title: PORTFOLIO SELECTION USING THE CAT SWARM OPTIMIZATION

**Author:** HAMZA KAMILI, MOHAMMED ESSAID RIFFI

**Abstract:** The portfolio selection is a discipline in finance interested in the optimization of the investment represented by a mixed quadratic programming problem. The approach of this paper to studying the portfolio selection problem is the implementation of the metaheuristic cat swarm optimization, a method inspired from the behavior of different felines and characterized by two modes: the seeking and the tracing mode; the seeking mode is when a cat is at rest observing its environment, the tracing mode is when the cat is hunting. In this article, we have adapted this method to the cardinality constrained efficient frontier (CCEF) compared to the data of mean return and risk obtained by the unconstrained efficient frontier (UEF) for five indexes markets and we have obtained efficient results.

**Keywords:** Portfolio Selection Problem, Metaheuristic, Cat Swarm Optimization, Efficient Frontier, Cardinality Constrained Efficient Frontier.

**Source:** Journal of Theoretical and Applied Information Technology 30th April 2015 – Vol. 74. No. 3 – 2015

### Title: NEW ARCHITECTURE OF IDS BASED INTERACTION ON MOBILE AGENTS AND DATA MINING

**Author:** CHAIMAE SAADI, RACHID CHERKAOUI, HABIBA CHAOUI, HASSAN ERGUIS

**Abstract:** Intrusion detection system (IDS) is a very useful tool for the defence of a network against attacks. Nowadays, intruders use a complex attacks towards a target. This make the IDS unable to detect all intrusions and the latter generate two frequent false alarms. The purpose of this paper is to propose a new architecture of an IDS allowing to increase the rate of intrusion detection and to minimize the false positive rate. The proposed architecture is based on mobile agents and data mining algorithms.

**Keywords:** Intrusion, Intrusion Detection System, Mobile Agent, Data Mining Algorithms.

**Source:** Journal of Theoretical and Applied Information Technology 30th April 2015 – Vol. 74. No. 3 – 2015

### Title: IDFP-TREE: AN EFFICIENT TREE FOR INTERACTIVE MINING OF FREQUENT SUBGRAPH PATTERNS

**Author:** MOHAMMAD H. NADIMI-SHAHRAKI, MARYAM TAKI, MOHAMMAD NADERI

**Abstract:** Recently, knowledge extraction from transactional graph databases by mining frequent subgraph patterns has become an interesting research topic. One of the important challenges in this topic is the situation called interactive mining in which the minimum support threshold can be frequently changed to find proper frequent subgraph patterns. Obviously, running the mining method from scratch is very costly. Therefore, in this paper, an efficient tree called IDFP-tree is proposed to develop our previous two-layer model for interactive mining of frequent subgraph patterns from transactional graph databases. The proposed tree constructs the mining model separated from the mining process. Thus, when the minimum support threshold is changed, there is no need to reconstruct the mining model, and only the mining process must be rerun. The experimental results show when the mining model is constructed by IDFP-tree, it can be frequently used and the total runtime of interactive mining of frequent subgraph patterns can be reduced.

**Keywords:** Interactive Mining, Graph Mining, Transactional Graph, Frequent Subgraph Pattern

**Source:** Journal of Theoretical and Applied Information Technology 30th April 2015 – Vol. 74. No. 3 – 2015

### Title: A LOW POWER PHASE FREQUENCY DETECTOR FOR DELAY-LOCKED LOOP

**Author:** S.LEELA, S.S.DASH

**Abstract:** Cascaded multilevel inverters have received more attention due to their ability to generate high quality output waveforms with low switching frequency. This paper deals with a novel analysis of a carrier based PWM method for cascaded multilevel inverters. Its effect on the harmonic spectrum is analysed. The voltage source inverters are modelled and the same is used for simulation studies. The test results verify the effectiveness of the proposed strategy in terms of computational efficiency as well as the capability of the inverter to produce very low distorted voltage with low switching losses.

**Keywords:** Carrier Based PWM technique, Cascaded Multilevel Inverter, Matlab Simulink, Total Harmonic Distortion.

**Source:** Journal of Theoretical and Applied Information Technology 30th April 2015 – Vol. 74. No. 3 – 2015
Author: LAU WENG LOON, MAMUN BIN IBNE REAZ, KHAIRUN NISA MINHAD, NOORFAZILA KAMAL, WAN MIMI DIYANA WAN ZAKI

Abstract: High performance phase frequency detector (PFD) is one of the key modules in high speed delay-locked loop (DLL). The operation of DLL depends on the performance of its detector. The demand for the reduction of power dissipation in CMOS design is a challenge in order to optimize circuit power consumption. A low power dynamic pseudo-PMOS PFD is proposed to make DLL system more reliable. In this work NOR gate of typical TSPC PFD is replaced with a low power dissipation pseudo-PMOS AND gate built of 3 PMOS transistors. Pseudo-PMOS AND integrated into proposed TSPC PFD to run maximum frequency at 1G Hz with 1.8 V input power supply. This proposed PFD has been implemented in Mentor Graphics 0.18 μm CMOS process technology and consumed 163.36 μm2 active layout area with 206 nW total power dissipation will further trim down the total cost of the DLL.

Keywords: DLL, Dynamic PFD, Low Power PFD, Low Noise PFD, Pseudo-PMOS

Source: Journal of Theoretical and Applied Information Technology

Full Text
EFFICIENT TECHNIQUES FOR PREDICTING SUPPLIERS
CHURN TENDENCY IN E-COMMERCE
BASED ON WEBSITE ACCESS DATA

VERONICA S. MOERTINI, NIKO IBRAHIM, LIONOV
1Informatics Department, Parahyangan Catholic University
2Information System Department, Maranatha Christian University
Bandung, Indonesia
E-mail: moertini@unpar.ac.id, niko.ibrahim@it.maranatha.edu, lionov@unpar.ac.id

ABSTRACT

Electronic supplier relationship management (e-SRM) is important in order to maintain strong, long lasting and beneficial relationship between e-commerce firms and their suppliers. One important function of e-SRM is to predict suppliers who tend to churn such that early “treatment” can be given. In the e-commerce systems that involve suppliers as the websites users, predicting suppliers’ churn tendency can be based on analyzing their frequencies in accessing the e-commerce websites.

Our proposed techniques include data warehouse design (supporting the data collection and preprocessing) and unsupervised algorithms that analyze the preprocessed bitmaps of time series data representing suppliers website access from time to time. Having bitmaps as inputs, our proposed algorithms are efficient (the time complexity is $O(n)$) as proven with our experiments. In experimenting with real world data of an e-commerce system selling hotel rooms, our techniques produce output of supplier segments where each segment has certain churn level tendency and need specific treatment.

Keywords: Churn Prediction In E-Commerce, Supplier Relationship Management, Web Usage Mining

1. INTRODUCTION

E-commerce systems come in several business models. Few of the models - such as B2C Transaction Broker, B2C Market Creator, B2B E-procurement and Private Industrial Networks - involve customers (buyers) and suppliers as the websites users [1]. For those models, having electronic supplier relationship management (e-SRM) is important in order to maintain strong, long lasting and beneficial relationship between the e-commerce firms and their suppliers. One of the important functions of e-SRM is to predict suppliers who tend to churn such that early "treatment" can be given to prevent them from churning. To reduce cost and maximize effectiveness, churn prediction techniques have to be fast and accurate to ensure that the right limited suppliers are being targeted for retention treatment.

While electronic customer relationship management (e-CRM) concept has been matured and discussed widely in literatures, e-SRM is a relatively new concept [2]. In e-CRM, many methods for predicting customers’ behavior of churning have been developed. The methods adopt traditional methods as well as soft computing [3], where many of these are developed based on recency, frequency and monetary (see Subsection 2.2 for more discussions) ([4], [5]). Among those methods, we have not found any churn prediction method that is based on the recency and frequency of customers website access. This seems to be justified as e-commerce customers may just visit and shop at an e-commerce website intermittently or occasionally, not regularly, just based on their impulsive need or desire. Contrarily to this fact, in e-commerce systems that involve suppliers as the websites users, the suppliers are expected to continuously manage their profile, products, services or campaigns from time to time. Thus, one approach for predicting suppliers’ churn behavior is by adopting web usage mining techniques, specifically by analyzing suppliers recency and frequency in accessing the websites.

In mining web usage, raw dataset must be collected and then preprocessed with certain techniques such that the transformed dataset is ready to be fed into data mining algorithms [6]. In the case of mining
suppliers' web usage, we view that the better approach is to develop a data warehouse that is used to store the preprocessed historical data such that the data needed can be added as well as analyzed periodically. Another important concept that we adopt is bitmap index in DBMS technology, which is known for its efficiency in processing queries involving bit operations, such as AND, OR, or COUNT [7]. In order to gain speedy computations, in our techniques, data representing the recency and frequency of the suppliers are transformed into bitmaps of time series data.

In this research, we intend to contribute in developing novel techniques used to predict suppliers' behavior towards churning based on recency and frequency in accessing the websites. The techniques include data warehouse design (specifically to support the data collection and preprocessing) and unsupervised algorithms for predicting the level of suppliers tendency towards churning by taking inputs of bitmaps representing the recency and frequency of website usage. The use of bitmaps is intended to gain time complexity of $O(n)$ of the proposed algorithms.

This paper is organized as follows: Introduction, literature review depicting relevant concepts, proposed systems and techniques, experiments and conclusions.

2. LITERATURE REVIEW

2.1. e-Commerce Models Involving Online Suppliers

E-commerce comes in several business models. There are few models that involve customers (buyers) and suppliers as e-commerce websites users, such as [1]:

(a) B2C Transaction Broker: The e-commerce firms act as brokers who sell suppliers’ products or services via the websites and obtain transaction fees. Here, the suppliers and customers do not interact directly. Industries using this model includes financial services, travel services, job placement services.

(b) B2C Market Creator: The owner firms uses the websites to create markets that bring buyers and sellers together. Hence, the suppliers and customers do not interact directly.

(c) B2B E-procurement, Exchanges and Industry Consortia: The websites act as a digital marketplace where suppliers and purchasers conduct transactions.

(c) Private Industrial Networks: The websites are designed to coordinate flow of communication among firms engaged in business together.

The users of customer access those websites mainly to search and purchase products, which can be incidentally, impulsively or infrequently. On the other hand, the suppliers use the websites to maintain company profiles, update products/services offered, create campaigns, and so on, and are expected to be frequent at all times in conducting these activities.

2.2. CRM Techniques

Recency, Frequency, Monetary (RFM) analysis is a traditional approach to analyzing customer behavior in the retailing and have been widely adopted. This analysis divides customers into groups, based on how recently they have made a purchase, how frequently they make purchases, and how much money they have spent [14], [15].

The RFM analysis uses:

(a) Recency: The time (in units such as days/months/years) since the most recent purchase transaction or shopping visit;

(b) Frequency: The total number of purchase transactions or shopping visits in the period examined;

(c) Monetary: The total value of the purchases within the period examined or the average value (e.g., monthly average value) per time unit. This analysis divides each of the three dimensions of RFM into equal sized chunks and places customers in the corresponding chunk along each dimension.

It is reported in [3] that popular methods for churn predictions in CRM are: (a) Traditional methods: Decision Trees, Regression Analysis, Support vector machine (SVM), Naive Bayes, k-nearest neighbor (KNN); (b) Soft computing: fuzzy logic, neural networks, and genetic algorithms. Hadden [3] develops supplier churn prediction model based on Neural Networks, Regression Trees and Linear Regression. The data used in developing the models are customer profiles and various data related to customer complaints and repair.

2.3. Web Usage Mining

Clickstream is foundational data. The data can be used to measure pages and campaigns and analyze all kinds of behavior, such as visits, visitors, time on site, page view, bounce rate, and so on [8]. Clickstream and associated data, which are generated from user interactions with websites, can also be analyzed to discover patterns [6]. In general, the process consists of two phases:
(a) Data preparation: Data of web application logs database are selected as needed, cleaned, identified, parsed, integrated, transformed and then stored in another database. The data analyzed can be categorized into four primary groups, which are usage, content, structure and user data.

(b) Pattern discovery phase: The transformed data are fed into pattern discovery algorithms, the output are analyzed for selecting the useful ones. There are several techniques for analyzing the data used in pattern discovery phase, which are clustering, classification, association and correlation analysis, and so on. The useful patterns discovered can then be used to personalize services (for examples of such services, see [9] and [4]).

In this research, we choose clustering approach. Clustering is a data mining technique that groups together a set of items having similar characteristics. In the web usage domain, there are two kinds of interesting clusters that can be discovered, which are user clusters and page clusters. Clustering of user transactions records is one of the most commonly used tasks in Web usage mining [6].

2.4. Data Warehouse

Data warehouse is a subject-oriented, integrated, time-variant (historical) and non-volatile collection of data supporting fast and systematic data analysis needed in decision making process. It is populated from the enterprise databases as well as other sources by ETLCL (extract, transform, clean and load) functions [10]. The data warehouse system architecture can be single, two or three layer. Three-layer design consist of source, reconciled and data warehouse layer (see Figure 1). The reconciled layer is a temporary database used to store the selected data from sources, which may need to be cleaned and transformed before loaded into the data warehouse schema. To facilitate the storing of historical websites access data, three-layer architecture is needed.

The basic schema employed in data warehouse and data mart is star schema. It consists of fact and dimension tables. The fact table, which is usually the center of a star schema, contains facts that are linked through their dimensions. Hence, it has primary keys of its linked dimensions and measure attribute(s) that represent specific business aspect or activity. The dimensions provide descriptive characteristics about the facts through their attributes.

2.5. Research Results of SRM

[11] reviews supplier selection and order allocation models based on an extensive search in the literature (170 paper during 2000-2010) and show the contributions to supply chain management (SCM). The techniques found in the research results are grouped into Multi Attribute Decision Making (MADM), mathematical programming, Artificial Intelligence methods (case-base reasoning, neural network, decision tree, association rules, cluster analysis), statistical models, fuzzy set theory, and hybrid models. Among the results of selecting (segmenting) suppliers models/techniques, the most common used criteria are material/products quality, lead time or delivery time and price.

3. PROPOSED SYSTEM AND TECHNIQUES

Based on our survey and knowledge we conclude that the frequency and recency of supplier activities (accessing the website for updating their profile, products catalog, creating campaigns, etc.) can be used to predict their tendency of churning. That is, if the frequency of online activities is low or declining during some period of observation, it can be interpreted that the suppliers are less interested (in conducting business activities), which may lead to churn. Hence, the system and techniques that we propose are intended to analyze
the supplier log activities stored in operational databases.

3.1. Data Warehouse System Architecture

In e-commerce systems, data warehouses can also be designed to support e-CRM and e-SRM. To provide data needed for suppliers' churn prediction based on the website access frequency, the data warehouse should store the summarized data of suppliers websites access data from time to time. Then, time series data representing the recency and frequency of suppliers' activities can be generated from the data mart (created from the data warehouse) and fed into the prediction algorithms. (Note: Time series data is a set of data about individual/agent behavior over time, where the time unit of observation can be day, week, month, or year [12]).

The architecture of the proposed system is shown in Figure 2. It is basically a three-layer data warehouse system (see Subsection 2.4). The supplier web access logs and other data are extracted from data sources (operational databases), which are then cleaned, transformed, and load onto data warehouse. A data mart containing time series data of frequency and statuses of suppliers' web access is created from the data warehouse, where its content is used to feed the proposed churn prediction algorithms.

![Figure 2: The System Architecture.](image)

The data mart design is shown in Figure 3. In DimTime table, UnitTimeOfYear attribute can be number day of year (having values of 1 to 362), week of year (1 to 52), month of year (1 to 12) or other unit as needed. In FactAccess table, AccessFreq is the frequency of accessing the e-commerce website, while AccessStatus is the status (having value '1' if the frequency is greater than 0 and '0' if otherwise).

![Figure 3: Some Part Of The Data Warehouse Schema.](image)

From the schema, if UnitTimeOfYear depicts week, some example of the time series data created in the last 24 weeks that can be queried from the data warehouse is shown in Table 1.

<table>
<thead>
<tr>
<th>IdSup</th>
<th>Year</th>
<th>IF</th>
<th>IS</th>
<th>IF</th>
<th>IF</th>
<th>IF</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>2014</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2014</td>
<td>2014</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2014</td>
<td>2014</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Note: IF = frequency of week-i, IS = status of week-i

Data Sources and Transform Functions Design:
Among the available data sources in the e-commerce systems, in this web usage mining (see Subsection 2.3), the data needed is usage data. Usage data is the log data collected automatically by the website and application servers representing the behavior of visitors [6]. To populate the designed data mart (see Figure 3), the operational data needed include the logs of suppliers' online transactions. Typical transaction logs tables usually have the following attributes: ID of parties/users (such as suppliers) involved, ID of the related transaction and timestamp.

As known, ETCL functions is run periodically (daily, weekly or others) to add data stored in a data warehouse. The algorithm for transforming data stored in transaction logs into the staging tables is:

**Algorithm:** Tranform-log

**Table accessed:**
- Transaction log tables
- StagingTable(IdSup, Year, UnitTimeOfYear, AccessStatus, AccessFreq)

**Steps:**
3.4. Input of Bitmap Time Series Data

A bitmap is defined as an array of bits. In DBMS technology, bitmap indexing is known for its efficiency (speed) in querying tables indexed on attributes with small unique values, as querying involve bitmap operations with bits operation (COUNT, AND, OR and NOT) [7]. In creating indices, a bitmap is created for each distinct value of attributes. For attribute value v, the bit for a table record is 1 if the record has the value v, and is 0 if otherwise. Bitmap indexing is commonly used in data warehouses that manage large volume of data. The proposed algorithms can take the advantages of bitmap index.

The proposed algorithms take bitmap time series data as inputs. Some examples of bitmap time series (selected from Table 1, where attribute of IdSupplier and Wi_Status are chosen) fed into the proposed algorithms are shown in Table 2 or Table 3. In Table 3, the statuses of each supplier are concatenated into a single bitmap string.

### Table 2: Some example of time series dataset.

<table>
<thead>
<tr>
<th>IdSupplier</th>
<th>15</th>
<th>25</th>
<th>...</th>
<th>24S</th>
</tr>
</thead>
<tbody>
<tr>
<td>IdS1</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>IdS7</td>
<td>1</td>
<td>1</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>IdS23</td>
<td>0</td>
<td>1</td>
<td>...</td>
<td>1</td>
</tr>
</tbody>
</table>

### Table 3: Time series data with 24-bit bitmap string.

<table>
<thead>
<tr>
<th>IdSupplier</th>
<th>Statuses (24 bits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IdS1</td>
<td>101011111101001111111110</td>
</tr>
<tr>
<td>IdS7</td>
<td>1101011110110111011111110</td>
</tr>
<tr>
<td>IdS23</td>
<td>01110011111010000111101111</td>
</tr>
</tbody>
</table>

3.5. Proposed Algorithms

As discussed in Section 2.3, clustering of user transactions records is one of the most commonly used tasks in web usage mining. We design algorithms to cluster suppliers giving results where each supplier is labeled with its predicted level of churning. In this section, we discuss the data structure (array) that supports efficient computation and the algorithms.

Array of Count (Frequency) of Bit 1: The arrays are designed to store the count of bit 1 in the bitmaps. That is, an array is created for a bitmap with certain length (8, 16, 24 bits, and so on). The bitmap decimal value is used as the index while its bit 1 count is stored as its element value (see Figure 4). The operation of getting an element of an array with its index (look up table), i.e. arrayName(i), is O(1), which is very efficient computation.

The following are some examples of bitmaps and their corresponding decimal value (given in (1)):
- 8-bits: 00000000 (0), 00000001 (1), 00000010 (2), ..., 11111111 (254), 11111111 (255);
- 16-bits: 0000000000000000 (0), 000000000000001 (1), 0000000000000010 (2), ..., 1111111111111111 (65534), 1111111111111111 (65535);
- 24-bits: 000000000000000000000000 (0), 000000000000000000000001 (1), ..., 111111111111111111111111 (16777214), 111111111111111111111111 (16777215).

For bitmap of 32-bits and 64-bits, the maximum value is 4294967295 and 18446744073709551615.

For the arrays depicted in Figure 4, for instance, the operation for obtaining the count of 1 in bitmap “11111110” is simply countOnes8b[254], while “11111111111111111111111111110” is countOnes24b[16777214].

![Figure 4: Array of bit 1 count of 8-bit and 24-bit.](image-url)
Algorithms: Our algorithms are designed with the assumption that the patterns of suppliers' frequency in conducting activities (during some period of observation time) used to define predicted churn levels are known in advance. For example:

1. If in the last period of observation of 24 weeks (6 months), a supplier is less active, i.e., dormant during k% x 24 weeks (for instance, k = 25) or more, then this supplier may churn in the future.

2. If in the last 6 months, a supplier's frequency in conducting activities decline from month to month, then this supplier is very likely to churn. If the frequency is declining sometime in the middle of the time only (then up again sometime later), then then this supplier may be churn.

We design 2 algorithms, PredictChurn_1 and PredictChurn_2, with those 2 assumptions and discussed as follows.

PredictChurn_1: The predicted churn level of every supplier is determined by how frequent each supplier conduct online activity (update product catalog, create campaigns, and so on) during the period of observation. Using some defined threshold frequency values of low, medium and high, the algorithm predicts the churn level of every supplier accordingly.

Algorithm: PredictChurn_1
Input: (a) File containing records of IdSupplier and its bitmap of statuses (see Table 3), fninput, (b) File containing count of 1 (see Figure 4), fCountOne; (c) Value of threshold1, threshold2, threshold3
Output: array of IdSupplier who predicted to churn, IdSupChurn[], and its churn level, ChurnLevel[] where the level l = low or 2 = medium or 3 = high

Descriptions: Predicting supplier’s churn behavior based on count (frequency) of bit 1: If the count < threshold1 then cLevel = 3, else if threshold1 <= count < threshold2 then cLevel = 2, else if threshold2 <= count < threshold3 then cLevel = 1

Steps:
1. Read fninput, store IdSuppliers in IdSup[] and bitmaps in bitStr[]
2. Read fCountOne and store the array in countOnes[]
3. Initialize IdSupChurn[]; Initialize ChurnLevel[]
4. bitmapSize ← length of bitStr[0]
5. th1 ← threshold1 x bitmapSize;
6. th2 ← threshold2 x bitmapSize;
7. th3 ← threshold3 x bitmapSize
8. j ← 0
9. for i=0 until size of IdSup[]-1 // evaluate every supplier
   10. cLevel ← 0 //default value: not churn
   11. decVal ← decimal value of bitStr[i]
   12. ctOnes ← countOnes[decVal] // look up count
   13. if ctOnes < th1 then cLevel = 3 // high
   14. else if th1 <= ctOnes < th2 then cLevel = 2 // medium
   15. else if th2 <= ctOnes < th3 then cLevel = 1 // low
   16. IdSupChurn[j] = IdSup[i]
   17. ChurnLevel[j] = cLevel
   18. j ← j + 1
19. return IdSupChurn[], ChurnLevel[]

PredictChurn_2: The predicted churn level of every supplier is determined by the frequency changes from “window time” to “window time”. Here, the overall bitmap will be divided into windows by masking it with the defined masks. Some example of the bitmap where divided into 4 windows and the corresponding 4 masks are as follows:

<table>
<thead>
<tr>
<th>1011</th>
<th>0111</th>
<th>1011</th>
<th>1110</th>
</tr>
</thead>
<tbody>
<tr>
<td>window1</td>
<td>window2</td>
<td>window3</td>
<td>window4</td>
</tr>
</tbody>
</table>

mask1 = 1111000000000000;
mask2 = 0000111111000000;
mask3 = 0000000011111000;
mask4 = 0000000000001111;

The number of windows and masks can be designed as needed (based on the known knowledge). In the detailed algorithm below, we provide some example of known patterns (for predicting churn level) with 4 windows.

Algorithm: PredictChurn_2
Input: (a) File containing records of IdSupplier and its bitmap of statuses (see Table 3), fnInput, (b) File containing count of 1 (see Figure 4), fCountOne; (c) Masking bitmaps, mask1, mask2, mask3, mask4
Output: array of IdSupplier who predicted to churn, IdSupChurn[], and its churn level, 1, 2, or 3

Descriptions: Predicting supplier’s churn behavior based on the change of count (frequency) of bit 1 in the windows.

Known rules (for examples): If the frequency of web access:
- is zero then churn level = 4
- is fully declining then churn level = 3
- is declining at the beginning of 2 windows then churn level = 2
- is declining at the beginning of 1 window then churn level = 1
- is declining at the end of 2 windows then churn level = 2

Steps:
1. Read fninput store IdSuppliers in IdSup[] and bitmap strings in bitStr[]
2. Read fCountOne store in count of 1 in countOnes[]
3. Initialize IdSupChurn[]; Initialize ChurnLevel[]
bitmapSize ← length of bitStr[0]
threshold ← threshold x bitmapSize.
j ← 0
for i = 0 until size of IdSup[i] //evaluate every supplier in IdSup[]
    CL = 0 //default value: not churn
    //Mask the bitmaps and get the decimal values
    dV1 ← decimal of (binary of bitStr[i] AND mask1);
    dV2 ← decimal of (binary of bitStr[i] AND mask2);
    dV3 ← decimal of (binary of bitStr[i] AND mask3);
    dV4 ← decimal of (binary of bitStr[i] AND mask4);
    //Get the count of bit 1
    ct1 ← countOnes[dV1]; ct2 ← countOnes[dV2];
    ct3 ← countOnes[dV3]; ct4 ← countOnes[dV4];
    if ct = 0 then CL = 4
    if ct = 1 then CL = 2
    if ct = 2 then CL = 1
    if ct = 3 then CL = 0
    if CL = 0
        IdSupChurn[j] = IdSup[i]
        ChurnLevel[j] = CL
        j ← j + 1
    return IdSupChurn[j], ChurnLevel[j]

Algorithms Complexity: The complexity of PredictChurn-1 and PredictChurn-2 is O(n), where n denotes the number of suppliers. Hence, the algorithms are efficient.

The Use and Implementation of the Algorithms: The algorithms can be implemented and used "standalone" for predicting suppliers based on the frequency of conducting activities as well as combined with other algorithms based on monetary analysis. Those algorithms can be implemented:

- Of database: The algorithms are implemented is a program module that read bitmaps stored in the database or files populated with bitmaps read from the data warehouse.
- In the database: The algorithms are implemented as stored procedures in the DBMS which access tables of bitmaps, which can be indexed with bitmap indexing.

4. EXPERIMENT

The experiments were performed in a personal computer with Intel(R) Pentium 4, 3.0 GHz of CPU, 2 GB of RAM, and 32-bit Windows operating system, where bitmaps data are stored in files.

4.1. Time Response Experiment

Experiment Method: In these experiments, we create 2 sets of file containing data simulation of 16-bits and 2 sets of file of 24-bits bitmaps. In the first set of files, the bitmaps are stored as unsigned integers while in the second files the bitmaps are stored as strings. One set include 11 files, where each stores 50K, 100K, 200K, ..., 900K or 1M bitmaps. We use the sets of files as the inputs of PredictChurn-1 and PredictChurn-2 algorithm, which are implemented as program modules. Each run of the module is repeated 10 times, where the CPU times for reading input file and computing are recorded in each run. The CPU times are then averaged and the results are discussed below.

In Figure 5, 6 and 7, the CPU time for processing:

- 16-bit bitmaps are presented in blue (for performing computation only) and black (for reading input file and performing computation);
- 24-bit bitmaps are presented in red (for performing computation only) and green (for reading input file and performing computation).

Bitmaps are represented as unsigned integers: As shown in Figure 5, the computation is linear and very efficient. For processing 1,000,000 bitmaps, the response time including for reading input files of PredictChurn-1 is less than 40 seconds. If the time for reading input files is omitted, the response times are less than 5 seconds. (Note: PredictChurn-2 cannot be experimented as the first bits of 0 in bitmaps will be omitted if bitmaps are stored as unsigned integers.)
Figure 5: The time responses of PredictChurn-1 in processing unsigned integer bitmaps.

Bitmaps are represented as strings: As shown in Figure 6 and 7, the computation is linear and also efficient. However, since the computation involve converting strings into unsigned integer values (for obtaining indices used to look up count one arrays), the execution times are approximately 4 times as long compared to the previous experiments. Here, for processing 1,000,000 bitmaps, the response time including for reading input files of both algorithms is less than 200 seconds. If the time for reading input files is omitted, the response times are less than 120 seconds.

Experiment results discussion: The time complexity of O(n) of both algorithms have been proven by the linearity of the execution times. If the computation does not involve conversion strings into unsigned integers, the execution times are very efficient. If the computation involves string conversion, three quarter of the times are needed for this operation.

Figure 6: The time responses of PredictChurn-1 in processing string bitmaps.

Figure 7: The time responses of PredictChurn-2 in processing string bitmaps.

4.2. Experiment with Case Study Data

The real data used for experiments is obtained from a transaction broker e-commerce selling hotel rooms in Indonesia, which is www.klikhotel.com. Klikhotel has been operating since August 2010 and in 2014 it has more than 1600 hotels (mostly of 3 and 4 star) who act as hotel rooms suppliers. Its customers are individual customers, travel agent and tour operators.

As discussed in [14], in our previous research we designed Supplier Relationship Managements (SRM) for Klikhotel and have implemented many of the SRM features used by the hotels. Few of the features include self-service hotel information management functions used for updating hotel profile, rooms information, and creating/updating campaigns or promotions. (Among those hotels, however, many have been integrated with an international hotel management system, which is also connected with Klikhotel system. Thus they do not use the SRM provide by Klikhotel.) As Klikhotel has been storing hotels transactions in its database, we can create and implement the data mart as depicted in Figure 3.

For this experiment, from the data mart we queried 6 month of weekly time series data (period of July to December 2013) that include of 1229 hotels. Among them, the total hotels being analyzed are 793 hotels as the rests do not use Klikhotel’s SRM. Each hotel activity frequency is represented by a bitmap consisting of 24 bits, where each bit represents the activity status of a week. After the bitmaps are processed with the proposed algorithms, the results are shown in Table 4.
Table 4: Clustering Results Of Supplier Hotels.

<table>
<thead>
<tr>
<th></th>
<th>Predict-Churn-1</th>
<th>Predict-Churn-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hotels with cl = 1</td>
<td>108</td>
<td>16</td>
</tr>
<tr>
<td>Hotels with cl = 2</td>
<td>173</td>
<td>28</td>
</tr>
<tr>
<td>Hotels with cl = 3</td>
<td>332</td>
<td>336</td>
</tr>
<tr>
<td>Total</td>
<td>613</td>
<td>380</td>
</tr>
<tr>
<td>Hotels with fine activity</td>
<td>180</td>
<td>413</td>
</tr>
<tr>
<td>TR1</td>
<td>0.74 sec</td>
<td>0.8 sec</td>
</tr>
<tr>
<td>TR2</td>
<td>0.24 sec</td>
<td>0.32 sec</td>
</tr>
</tbody>
</table>

cl.: churn level
TR1: time response including reading input file
TR2: time response without reading input file

As shown in Table 4, the output of algorithm PredictChurn-1 differ to PredictChurn-2, where the later predicts less hotels needing “treatments” (hotels with cl=2 and cl=3). However, both algorithms predicts that more than 330 hotels have low activity (churn level is 3), which should alert the KlikHotel supplier care division. These hotels can then be fed into an SRM module that generates email alert or other SRM module that predict the behavior of churning based on monetary variable. (The Klikhotel management has acknowledged that the results show consistency with their knowledge of the hotel members’ activity.) In the table, it is also shown that it takes less than 1 second to process the 1229 bitmaps data, which is very fast.

5. CONCLUSION

The suppliers’ online transaction logs data can be preprocessed into time series data that is then analyzed with efficient clustering algorithms to predict suppliers churn level tendency based on their frequency in accessing website. To perform such tasks, we propose efficient techniques that involve data mart design with its transform function needed to preprocessed raw log data into time series format, and two churn tendency clustering algorithms that process the time series bitmaps. Having time complexity of $O(m)$, the algorithms are proved to be very efficient.

Our proposed techniques are used to predict churn level tendency based on suppliers frequency in accessing a website only. For further works: Techniques processing other data, such as monetary and other necessary data that can be used to predict churn tendency, can be developed. These techniques are then combined with our proposed techniques to obtain better (more accurate) prediction techniques. Other future works: In the case that training and testing data set for classification models (bitmap time series data having a class attribute representing the suppliers churn level) can be obtained or generated, classification models used to predict suppliers churn tendency can be designed and then trained and tested with the data set.

ACKNOWLEDGMENT

We like to thank to the Directorate General of Higher Education, Ministry of Research and Technology and Higher Education, Indonesia who has funded this research in 2015 through Hibah Bersaing scheme.

REFERENCES


