Intelligent Decision Support Systems

Research in Computing Science

Series Editorial Board

Research in Computing Science es una publicación trimestral, de circulación internacional, editada por el Centro de Investigación en Computación del IPN, para dar a conocer los avances de investigación científica y desarrollo tecnológico de la comunidad científica internacional. **Volumen 109,** abril de 2016. Tiraje: 500 ejemplares. *Certificado de Reserva de Derechos al Uso Exclusivo del Título* No. : 04-2005-121611550100-102, expedido por el Instituto Nacional de Derecho de Autor. *Certificado de Licitud de Título* No. 12897, *Certificado de licitud de Contenido* No. 10470, expedidos por la Comisión Calificadora de Publicaciones y Revistas Ilustradas. El contenido de los artículos es responsabilidad exclusiva de sus respectivos autores. Impreso en la Ciudad de México, en los Talleres Gráficos del IPN – Dirección de Publicaciones, Tres Guerras 27, Centro Histórico, México, D.F. Distribuida por el Centro de Investigación en Computación, Av. Juan de Dios Bátiz S/N, Esq. Av. Miguel Othón de Mendizábal, Col. Nueva Industrial Vallejo, C.P. 07738, México, D.F. Tel. 57 29 60 00, ext. 56571.

Editor responsable: *Grigori Sidorov, RFC SIGR651028L69*

Research in Computing Science is published by the Center for Computing Research of IPN. **Volume 109**, April 2016. Printing 500. The authors are responsible for the contents of their articles. Printed in Mexico City, in the IPN Graphic Workshop – Publication Office.

Volume 109

Intelligent Decision Support Systems

Cuauhtémoc Sánchez-Ramírez,
Giner Alor-Hernández,
Jorge Luis García-Alcaraz (eds.)

Instituto Politécnico Nacional, Centro de Investigación en Computación
México 2016

ISSN: 1870-4069

Instituto Politécnico Nacional (IPN)
Centro de Investigación en Computación (CIC)
Av. Juan de Dios Bátiz s/n esq. M. Othón de Mendizábal
Unidad Profesional "Adolfo López Mateos", Zacatenco
07738, México D.F., México

http://www.rcs.cic.ipn.mx
http://www.ipn.mx
http://www.cic.ipn.mx

Indexed in LATINDEX, DBLP, and Periodica / Indexada en LATINDEX, DBLP y Periódica

Printing: 500 / Tiraje: 500

Printed in Mexico / Impreso en México

Editorial

Decision-making has always been part of human's life and evolution. Humans have developed many techniques to analyze different scenarios and environments in order to improve such decision-making processes. Some of these techniques make it possible to simultaneously analyze multiple criteria, while some others allow decision groups to make evaluations by integrating the opinions of all members.

Within industrial contexts, managers are usually the top decision makers. They hold permanent contracts and have high incomes. Nowadays, advanced information and communication technologies support and assist these administrators in their decisions. They enable to consider several variables – including time costs – and perform calculations with greater precision. Such systems integrated into business technologies are usually called Decision Support Systems (DSS).

As a result of their usefulness, DSS are increasingly gaining popularity in different domains, not only for businesses. They have also been implemented in engineering, the military, and medicine. Moreover, they are particularly valuable in situations in which the amount of available information is prohibitive for the intuition of an unaided human decision maker and when precision and optimality are of importance. Similarly, DSS can support choice among well-defined alternatives and build on formal approaches, such as the methods of engineering economics, operations research, statistics, and decision theory. Finally, they can use artificial intelligence methods to heuristically address problems that are intractable by formal techniques.

Therefore, it can be stated that the proper application of decision-making tools increases productivity, efficiency, and effectiveness. Furthermore, these systems offer many businesses a comparative advantage over their competitors, which allows them to make optimal choices in technological processes and their parameters, planning business operations, logistics, or investments.

This special issue collects nine papers, some of them are best papers presented at the First International Workshop on Intelligent Decision Support Systems (DSS) for Industry Application, co-located with the 14th Mexican International Conference on Artificial Intelligence (MICAI 2015). It aims to publish developments from practitioners in the research areas of industrial engineering, computer science, management science, systems engineering, operations research, optimization process, software engineering, computational engineering, innovation systems, and logistics engineering, among others. All these areas at some point focus on the use of Decision Support Systems for the industry. Similarly, this special collection addresses topics on the implementation of conceptual frameworks, strategies, techniques, methodologies, informatics platforms, and models for developing Intelligent Decision Support Systems.

Finally, editors would like to express their gratitude to the reviewers who kindly contributed to the evaluation of papers at all stages of the editing process. They equally thank the Editor-in-Chief, Prof. Grigori Sidorov, for the opportunity offered to edit this special issue and for providing his valuable comments to improve the selection of research works.

Acknowledgments. This special issue was partially supported by the Tecnológico Nacional de México (TecNM), the National Council of Science and Technology (CONACYT), and the Secretariat of Public Education (SEP) through PRODEP.

Cuauhtémoc Sánchez-Ramírez (Instituto Tecnológico de Orizaba)
Giner Alor-Hernández (Instituto Tecnológico de Orizaba)
Jorge Luis García-Alcaraz (Universidad Autónoma de Ciudad Juárez)

April 2016

Table of Contents

Page

Heuristic for Multi-objective Solution of the Periodic Vehicle Routing Problem

Angélica Tenahua[1], Elías Olivares-Benitez[1], Jorge Esparza[2]

[1] Universidad Popular Autónoma del Estado de Puebla,
Mexico

[2] Universidad Anáhuac, Puebla,
Mexico

{angelicamaria.tenahua, elias.olivares}@upaep.mx, jorge.esparza@anahuac.mx

Abstract. This paper presents a methodology for solving the Periodic Vehicle Routing Problem (PVRP). Customer locations each with a certain daily demand are known, as well as the capacity of vehicles. In this work, the problem is solved by means of a two-phase heuristic. The first phase is the itinerary assignment to visit each customer, and in the second phase the routing is done for each day of the itinerary. The itinerary assignment is made randomly and the routing is solved using the Savings Heuristic. The problem is treated as a bi-objective problem that minimizes the total distance travelled in the time horizon, minimizing also the maximum number of routes in any day. Three instances were solved identifying the non-dominated solutions and presenting them in a graph of the Pareto front. Computational results show that the proposed approach can yield good quality solutions within reasonable running times.

Keywords: Vehicle routing problem, non-dominated solutions, multi-objective, savings heuristic.

1 Introduction

The Vehicle Routing Problem (VRP) is a classic combinatorial optimization problem, widely studied. The first publications were in 1959 [1], and the problem is to generate routes at minimal cost, for a given set of vehicles with homogeneous capacity [2]. In order to include features to the real world, variants of the VRP have been studied. Some of them are, vehicles with heterogeneous fleet, time windows, periodical visits, more than one depots, etc. The main variants of the VRP can be found in publication [3]. The PVRP (Periodic Vehicle Routing Problem) can be considered as a generalization of the conventional VRP, which looks for determining an optimal set of daily routes for a planning horizon of time. Customers should be visited on different days during the planning horizon. According to the number of visits required by the customer, there are combinations of these visits, named itineraries. The solution of the PVRP implies solving two entangled problems: an assignment problem and the vehicle routing problem. In the assignment problem it is necessary to determine a valid itinerary (visiting days) for each customer within the planning horizon. The other is the vehicle

routing problem for each day. The PVRP is considered a NP-hard problem [4], so the techniques with more efficient solution for these problems are the heuristics. The most common heuristics used for this problem are: Tabu Search [5, 6], GRASP [7], Ant Colony Optimization (ACO) [8], Variable Neighbourhood Search (VNS) [9] and a hybrid heuristic based on joint coverage [10].

The total cost to minimize includes the costs associated with the distances travelled, vehicle capacity, transit time, fuel, etc. In this paper the distance is directly related to the cost. For more information about costs, the reader can review the publication in [11] that made a review of fixed and variable costs of transportation. In this work, we are interested in giving solutions to problems that may happen in real situations for the companies. This paper considers the transports (delivery trucks) with the same capacity, known as homogeneous fleet. The planning horizon is of six days, which may result in a week from Monday to Saturday. The solution of PVRP in this work is made first with assignment of itinerary to each customer based on number of visits required during the planning horizon. After that routing is performed by the Savings Method [12].

The paper is structured as follows. Section 2 is the Literature Review where the major papers have served as a reference and analysis for this study. Problem description is made in Section 3. The development of the methodology is shown in Section 4, which states how the PVRP results are obtained. Section 5 describes the implementation details of the method proposed. The results obtained are shown in Section 6. Finally in Section 7 the conclusions of this work are drawn.

2 Literature Review

This section presents the literature review of the papers that have served as a reference and analysis of this study. Each of them is mentioned in a very general way presenting the problem, the solution method and the results obtained.

Mendez et al. [13] solved the infectious waste collection problem as a Periodic Vehicle Routing Problem, using a hybrid technique that involves an evolutionary algorithm for optimal planning of the collection days, and a local search algorithm to determine the set of optimal routes every day. They divide the problem in the same way as we propose since it is natural to construct that hierarchy. Gaudioso and Paletta [14] describe a model for the optimal management of a commodity. The goal is to schedule the deliveries according to feasible combinations of delivery days and to determine the scheduling and routing policies over the planning horizon for a maximum number of vehicles simultaneously employed, i.e., the fleet size. Heuristic algorithms were proposed: An implicit enumeration algorithm for the feasible combinations of delivery to each customer, a heuristic algorithm for assignment of customers to routes, and Bin Packing algorithm for scheduling of routes. In this paper, the fleet size is unknown, but the objective is to get the minimum. Wen et al. [15] consider the Dynamic Multi-Period Vehicle Routing Problem, which deals with the distribution of orders from a depot to a set of customers over a multi-period time horizon. Customer orders and their feasible service periods are dynamically revealed over time. The objectives were to minimize total travel costs and customer waiting, and to balance the daily workload over the planning horizon. This problem originates from a large distributor operating in Sweden. It is modelled as a mixed integer linear program, and solved by means of a three-phase

heuristic that works over a rolling planning horizon. The multi-objective aspect of the problem is handled through a scalar technique approach. Computational results show that the proposed approach can yield high quality solutions within reasonable running times. In this paper although the model is multi-objective, demands and visit requirements are fixed and not dynamic as in the work of Wen et al. (2009), which is a more general problem, however, has served as a guide considering fixed visits.

Francis and Smilowitz [16] presented a continuous approximation model for the Period Vehicle Routing Problem with Service Choice (PVRP-SC). The PVRP-SC is a variant of the Period Vehicle Routing Problem in which the visit frequency to nodes is a decision of the model. This variation can result in more efficient vehicle tours and/or greater service benefit to customers. The continuous approximation model can facilitate strategic and tactical planning of periodic distribution systems and evaluate the value of service choice. Francis and Smilowitz suggested making a parametric analysis of future conditions, which may be useful to test the sensitivity of the solutions for future scenarios before committing to the highest service levels. For this work, no time windows are considered, but the mathematical model from Smilowitz and Francis [17] proposed, is a great help in this paper considering large windows of time. Griffis and Bell [18] made a comparison between results obtained from five different instances of the type: random, cluster, Sector, Rural and Urban Coastal, through the metaheuristic Ant Colony Optimization (ACO), and Clark and Wright heuristic. The best results are Sector type. For future work related to this article would make an interesting analysis with the five types of population distribution and identify if any of them has better solutions. Ballou and Agarwal [19] made a comparison between the methods of Savings, Cluster and sweep under five types of population distribution, random, cluster, Sector, Urban-rural and Coastal. The results reflect an advantage of the Savings method vs. Cluster and sweep methods with different types of distribution. In this paper, savings method is used to find routes each day of the period.

3 Problem Description

Consider the graph $G=(N, A)$ and period np days (planning horizon), where N is the set of nodes representing customers and A is the set of arcs connecting the nodes. Each customer requires the collection task with a frequency f_i measuring the number of times the customer i must be visited with $1 \leq f_i \leq np$. A combination of visit times (days) into the planning horizon is named an itinerary. The basic PVRP is to select an itinerary for node i and solve a vehicle routing problem for each day of the selected itineraries, in order to minimize the total cost of travel. As mentioned previously, it is a NP-hard problem because it includes the VRP with single period as a special case [13].

Christofides and Beasley [20] proposed the PVRP model as follows. Let S_i be the set of allowable itineraries for customer i. Let $x_{ik} = 1$ if the *kth* itinerary is chosen for i, 0 otherwise. Let n be the total number of customers and T the number of days in the period. Let $a_{kt} = 1$ if day t is in itinerary k, 0 otherwise; q_i be the demand of customer i (for each delivery); c_{ij} be the travel time from customer i to customer j ; $N = (i|i = 1, 2, \ldots, n)$ be the set of customers; Q, be the capacity of vehicle r ; D_r be the total allowable driving time of vehicle r; and R_t be the given set of available vehicles for day t . Let $v_{it} = 1$ if customer i is visited on day t, 0 otherwise; with the depot

represented by customer 0 define $v_{0t} = 1, t = 1, \ldots, T$, and let $u_{ijtr} = 1$ if vehicle $r \in R_t$ goes from *i* to *j* on day *t*, 0 otherwise; then the program is to

Minimize

$$\sum_{t=1}^{T} \sum_{i=0}^{n} \sum_{j=0}^{n} \sum_{r \in R_t} c_{ij}\, u_{ijtr} \tag{1}$$

Such that

$$\sum_{k \in S_i} x_{ik} = 1, \forall i\ (i \neq 0) \tag{2}$$

$$v_{it} = \sum_{k \in S_i} x_{ik} a_{kt}, \forall t, \forall i\ (i \neq 0) \tag{3}$$

$$\sum_{r \in R_t} u_{ijtr} \leq \frac{v_{it} + v_{jt}}{2}, \forall i, j, t\ (i \neq j), \tag{4}$$

$$\sum_{i=0}^{n} u_{iptr} = \sum_{j=0}^{n} u_{pjtr}, \forall p, t, r \in R_t, \tag{5}$$

$$\sum_{r \in R_t} \sum_{i=0}^{n} u_{ijtr} = v_{jt}, \forall j, t(\neq 0) = |R_t| \forall t(j = 0) \tag{6}$$

$$\sum_{i \in W} \sum_{j \in W}^{n} u_{ijtr} < |W| - 1, \forall t, r \in R_t, \forall W \subseteq N, \tag{7}$$

$$\sum_{j=1}^{n} u_{0jtr} < 1, \forall t, r \in R_t, \tag{8}$$

$$\sum_{i=1}^{n} q_i \left(\sum_{j=0}^{n} u_{ijtr} \right) < Q_r, \forall t, r \in R_t, \tag{9}$$

$$x_{ik} \in (0,1), \forall i, k \in S_{i,} \tag{10}$$

$$u_{ijtr} \in (0,1), \forall i, j, t, r \in R_t. \tag{11}$$

Eq. (1) is the objective function to evaluate the total travel cost. Eq. (2) ensures that only one delivery combination is chosen for each customer, Eq. (3) ensures that a customer is only visited on a particular day if the delivery combination chosen has a delivery on that day, Eq. (4) ensures that no vehicle can go between two customers on a particular day unless they are both scheduled for delivery on that day, Eq. (5) ensures that if a vehicle visits a customer it also leaves that customer, Eq. (6) ensures that each customer is visited on the days that it is scheduled for delivery. Eq. (7) is the set of subtour elimination constraints. Eq. (8) ensures that a vehicle can only be used at most once. Eq. (9) is the vehicle capacity constraint.

4 Methodology

It is considered a planning horizon of six days, so $np = 6$. Each customer has an associated frequency of visits f_i. It is possible to represent itineraries in Table 1. If a customer is visited on day t, the value is 1, 0 otherwise. The pattern of visits can be represented by a set P of binary vectors x. Each element has a value of 1 if the customer is visited on day t, 0 otherwise. Some restrictions are implemented to prevent consecutive visits. Assignment of itineraries is random according between all possible itineraries that meet f_i of each customer i. Figure 1 presents graphically the methodology explained.

Table 1. Itineraries for one, two or three visits.

Itinerary	Number of visits	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6
1	1	1	0	0	0	0	0
2	1	0	1	0	0	0	0
…							
7	2	1	0	1	0	0	0
8	2	1	0	0	1	0	0
…							
16	3	1	1	0	1	0	0
17	3	1	1	0	0	1	0
…							

Table 2 presents an example of generated routes each day, for a planning horizon of 6 days. For example observe customers 4, 14 and 9 they require to be visited 1, 2 and 3 times a week respectively. For customer 4, the set P is P = {100000, 010000, 001000, 000100, 000010, 000001}. The element of the set that was assigned is the last element, which means customer 4 is visited on day 6 of the planning horizon. For customer 14, it is visited on day 3 and 5 of the period. Customer 9 is visited on days 1, 4 and 6.

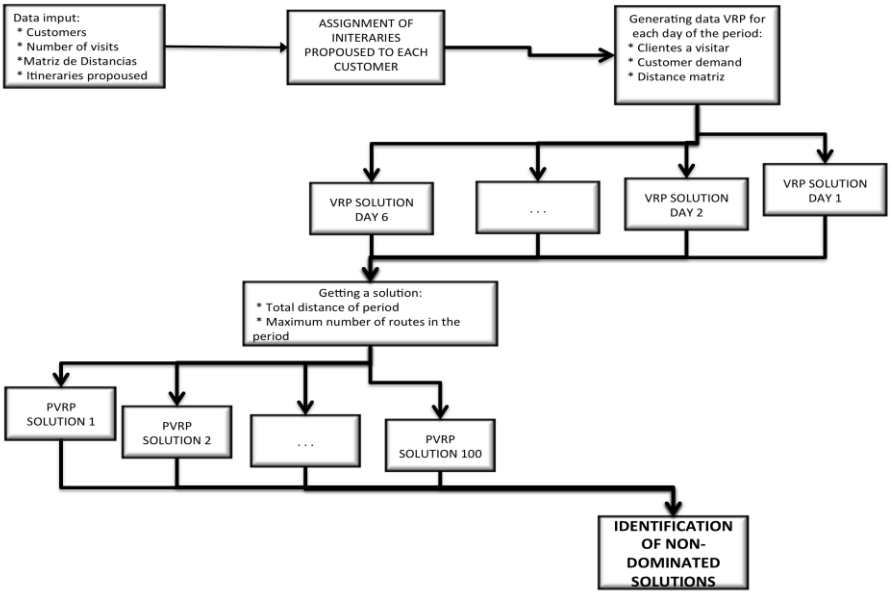


Fig. 1. Methodology diagram.

Table 2. Example of routes generated on each day of a planning horizon.

Day	Sequences of nodes in the routes for each day
1	0 9 15 10 0 3 12 0
2	0 2 6 5 0 8 12 0
3	0 15 6 14 0
4	0 15 1 5 0 9 11 0 3 12 0
5	0 8 2 14 0 3 13 0
6	0 6 7 9 0 4 11 0

5 Details of Implementation

The algorithm was coded in C++ programming language to solve three instances. Details of instances are presented in Table 3. The computer is a Mac running OS X Processor 2.4 GHz Intel Core 2 Duo and Memory 5 GB 1067 MHz DDR3. Average runtime was 2725 seconds.

Table 3. Data of the instances worked.

Instance	Total Customers	Average demand	Stantandar deviation of demand
1	15	389.33	54.95
2	15	296.13	174.17
3	25	2.72	1.21

According to the number of visits required by each costumer, an itinerary is assigned randomly. Once every customer has an itinerary assigned, the total costumers to be visited each day is known. With the last information it is possible to generate a distance matrix for each day of the period, for routing.

The nodes to visit each day on the planning horizon, is solved as a VRP. The Savings Method obtains the solutions of VRP. The solution on a planning horizon has the following information: the number of nodes to be visited every day on the planning horizon, the nodes to be visited every day of the planning horizon, the number of routes that are generated each day, routes, the distance travelled on each route, the distance travelled each day on the period, and the total distance travelled throughout the period. To obtain solutions for different assignments of itineraries, one hundred solutions are created and evaluated for each instance. Subsequently the non-dominated solutions are identified to construct an approximate Pareto front. To define non-dominated solutions in this paper, we consider two objective functions: the total distance and the maximum number of routes between all days in a period. An example of solution of a period is presented in Table 2, in which the routes are indicated in each day. Note day 1, in which two routes were generated. The routes are separated by node 0 which is the depot. In the same example, it can be observed that on day 3 only one route is generated, while on day 4, three routes are generated.

6 Results

Outcomes of interest are the total distance and the maximum number of routes during the period. Table 4 presents the results and shows the non-dominated solutions of the three instances with average execution time of 2725 seconds. Figure 3 is the graph of the solution with instance 1, identifying with dark and bigger square the non-dominated solution.

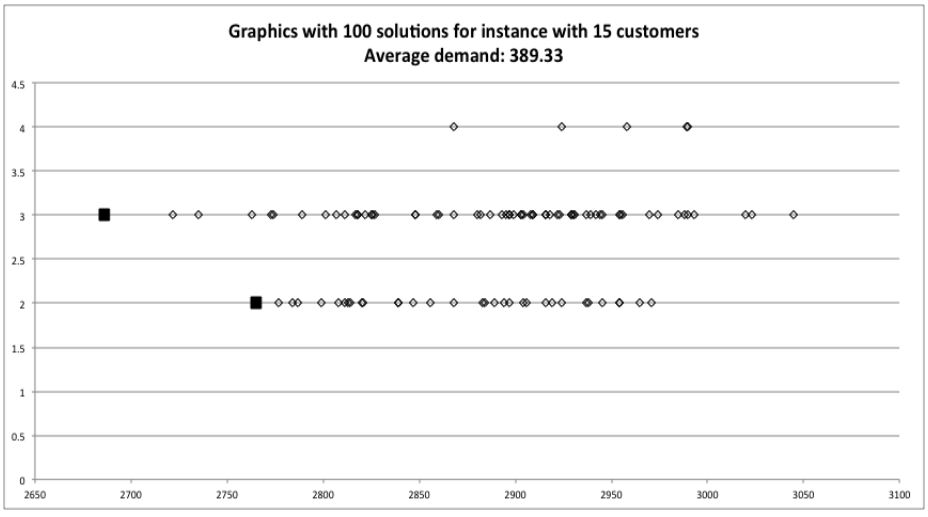


Fig. 3. Shows the graphic with 100 solutions of the PVRP for Instance 1.

Table 4. Non-dominated solutions from three instances.

Instance	Number of solution	Total distance travelled during the period	Maximum number of routes during the period
1	12	2686.00	3
1	99	2765.00	2
2	82	3295.00	2
3	55	130.04	3
3	79	132.01	2

7 Conclusions

One hundred solutions for each instance, allow finding one or two non-dominated solutions. It indicates that there is a high correlation between distance and maximum number of routes per period. The assignment is random so two analyses are proposed to address this for future works, the first is in which the assignation has certain criteria that could improve and verify the solutions proposed improvement. The second proposal is to do intra-route and inter- routes changes. A further opportunity is to find another trade- off to identify non-dominated solutions.

References

1. Dantzig, G., Ramser, J.: The Truck dispatching problem. Management Science, Vol. 6, No. 1, pp. 80–91 (1959)
2. Abreu, R., Arroyo, J.: Aplicação da heurística ILS para o Problema de Roteamento de Veículos Periódico com Frota Heterogênea e Custos Fixos (2014)
3. Toth, P., Vigo, D.: The Vehicle Routing Problem (2014)
4. Chao, I., Golden, B., Wasil, E.: An improved heuristic for the period vehicle routing problem. Networks, Vol. 26, Issue 1, pp. 25–44 (1995)
5. Cordeau, J., Gendreau, M., Laporte, G.: A tabu search heuristic for period and multi-depot vehicle routing problems. Networks, Vol. 30, pp. 105–119 (1997)
6. Cordeau, J., Laporte, G., Mercier, A.: A unified tabu search heuristic for vehicle routing problems with time windows. Journal of the Operational research society, Vol. 52, No. 8, pp. 928–936 (2001)
7. Pacheco, J., Álvarez, A., García, I., Angel-Bello, F.: Optimizing vehicle routes in a bakery company allowing flexibility in delivery dates. Journal of the Operational Research Society, Vol. 63, No. 5, pp. 569–581 (2012)
8. Bell, J., Griffis, S.: Swarm intelligence: Aplication of the ant colony optimization algorithm to logistics-oriented vehicle routing problems. Journal of Business Logistics, Vol. 31, No. 2, pp. 157–175 (2010)
9. Hemmelmayr, V., Doerner, K., Harlt, R.: A variable neighborhood search heuristic for periodic routing problems. European Journal of Operational Research, Vol. 195, No. 3, pp. 791–802 (2009)
10. Cacchiani, V., Hemmelmayr, V., Tricoire, F.: A set-covering based heuristic algorithm for a periodic vehicle routing problem. Discrete Applied Mathematics, Vol. 163, Parte 1, pp. 53–64 (2014)
11. Da Silva, I., Reis, R., Gomes, M.: Custos e otimização de rotas no transporte de leite a latão e a granel: um estudo de caso. Organizações Rurais & Agroindustriais, Vol. 2, No. 1 (2011)

12. Clarke, G. Wright, J.: Scheduling of Vehicles from a Central Depot to a Number of Delivery Points. Operations Research, Vol. 12, Issue 4, pp. 568–581 (1964)
13. Méndez, A., Pontin, M., Ziletti, M., Chávez, L.: Heurísticas para la resolución de un Problema de Ruteo de Vehículos Periódico. Mecánica Computacional, Vol. 24, pp. 2951–2960 (2005)
14. Gaudioso, M., Paletta, G.: A Heuristic for the Periodic Vehicle Routing Problem. Transportation Science, Vol. 26, No. 2, pp. 86–92 (1992)
15. Wen, M., Cordeau, J.F., Laporte, G., Larsen, J.: The Dynamic Multi-Period Vehicle Routing Problem. Computers & Operations Research, Vol. 37, Issue 9, pp. 1615–1623 (2010)
16. Francis, P., Smilowitz, K.: Modeling Techniques for Periodic Vehicle Routing Problems. Transportation Research Part B Methodological, Vol. 40, Issue 10, pp. 872–884 (2006)
17. Christofides, N., Beasley, J.E.: The Period Routing Problem. Networks, Vol. 14, Issue 2, pp. 237–256 (1984)

Knowledge Management Systems for Supporting Enterprise Wide Optimization and Modeling: Strategic and Tactical decisions

Edrisi Muñoz Mata[1,2], Elisabet Capon García[2], Jose M. Lainez Aguirre[3], Konrad Hungerbuehle[2], Antonio Espuña Camarasa[4], Luis Puigjaner Corbella[4]

[1] Centro de Investigacion en Matematicas, Guanajuato, Mexico

[2] ETH Zürich, Department of Chemistry and Applied Biosciences, Zürich, Switzerland

[3] University at Buffalo, Department of Industrial and Systems Engineering, Amherst, NY, USA

[4] Universitat Politecnica de Catalunya, Department of Chemical Engineering, Barcelona, Spain

emunoz@cimat.mx, {elisabet.capon, konrad.hungerbuehler}@chem.ethz.ch, jmlainez@gmail.com, {antonio.espuna, luis.puigjaner}@upc.edu

Abstract. Industries constitute a highly complex process network including multiple business and process units, which interact with each other. Enterprise-wide optimization addresses the optimization of supply operations, manufacturing and distribution in process industries. This work presents the evolution of knowledge management approaches for supporting enterprise wide optimization in process system engineering.

Keywords: Knowledge management systems, enterprise wide optimization.

1 Introduction

The chemical process industry is the base for producing a multitude of goods worldwide. Industries constitute a highly complex process network including multiple business and process units, which interact with each other. In order to tackle the aforementioned complexity, collaboration among different scientific disciplines, namely chemical engineering, operations research and computer science is of utmost importance. It is important to underscore that process system engineering (PSE) is a well-established discipline of chemical engineering which covers a set of methods and tools to support decision-making for the creation and operation of the process supply chain constituting the discovery, design, manufacturing and distribution of chemical products from a holistic approach. In this sense, the so-called enterprise-wide optimization (EWO) addresses the optimization of supply operations, manufacturing and distribution in process industries [1]. Specifically, EWO aims to develop flexible modeling environments for the chemical supply chain representation, which is the main

basis for reaching efficient decision-making. An important challenge in this arena is the integration and management of data, information and decisions along the various hierarchical levels of the company and the whole supply chain. Recent efforts have been devoted to reach a better transactional systems integration in order to improve the use of data for analytical models.

This work presents the evolution of knowledge management approaches for supporting EWO in PSE. In order to deal with the problem complexity, the proposed approaches decouple the system across a hierarchy of appropriately chosen levels. Here, we will focus on the strategic and tactical levels. The knowledge management solution that we proposed has been continuously broadening its scope. We will chronologically summarize such extension. The knowledge management systems was originally created to deal with batch process, however its functionality was later extended to include supply chain planning, life cycle assessment, mathematical model and operation research modeling.

2 Decision-Making and Modeling at the Chemical Process Industry

For years, companies have been developing information systems to help people exploiting data, models and information with the purpose of supporting decision-making. Nowadays, global competition has made decisions related to certain manufacturing characteristics such economic efficiency, product quality, flexibility, and reliability essential for the viability of the enterprise [2].

Decision support systems (DSS) are information technology solutions that can be used to support complex decision-making. DSS are defined as "aid computer systems at the management company level that combine data and sophisticated analytic models to support decision-making" [3]. A Classic DSS comprises different components, which interact along functionalities at different levels of development and implementation. The first functionality relates to sophisticated database management capabilities with access to internal and external data, information, and knowledge. The next functionality concerns to the modeling functions accessed by a model management system. Other functionality regards to user interfaces design that enable interactive queries, reporting, and graphing functions. Finally, the optimization functionality is found which is supported by mathematic algorithms and/or intuition/knowledge [4].

One of the major challenges of DDS concerns to functionality modeling, which attempt of devising an approximate representation of a system with the goal of providing predictions of its performance. Such a representation is called a model. A model formalizes the relationship between various flows of information and can adopt different forms which range from spreadsheets to complex mathematical programs, neural networks, and expert systems among others. Furthermore, a model is designed to capture certain behavioral aspects of the modeled system – those that are of interest to the analyst – in order to gain knowledge and insight into the system's behavior [5].

A general taxonomy for modeling systems distinguishes between transactional and analytical modeling approaches. Transactional systems are concerned with the acquisition, processing and communication of data over the enterprise. Analytical techniques introduce some reasoning to evaluate the problems, and are further classified

into descriptive and normative models. Descriptive models can be used to analyze a system, but not to improve it, and provide a better understanding of internal and external functional relationships in the enterprise. Forecasting models, cost relationships, resource utilization relationships, and simulation models fall into this category. On the other hand, optimization or normative models are developed as decision-support systems to assist managers in the identification of efficient and improved decisions.

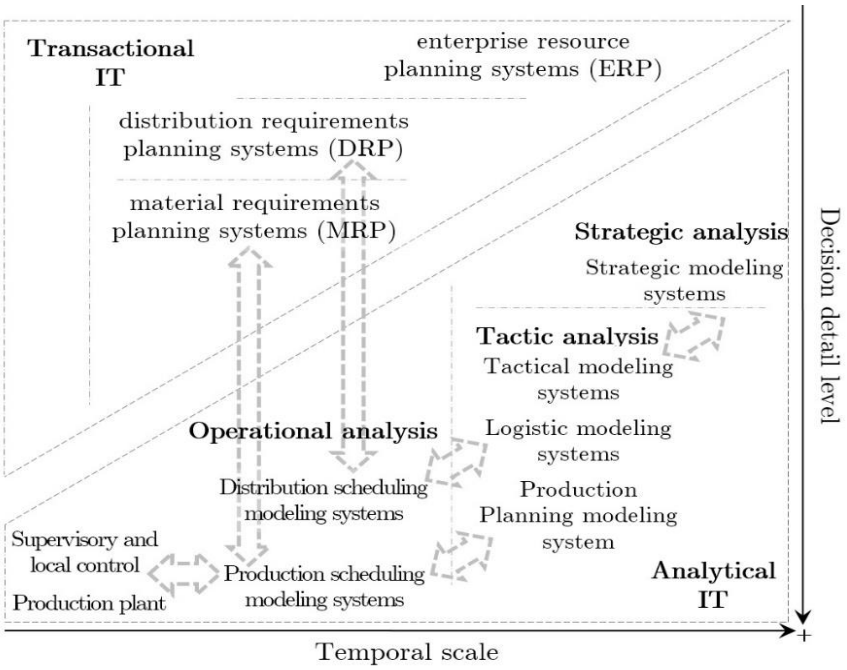


Fig. 1. Hierarchical scheme of modeling systems within organization.

In general, descriptive and optimization algorithms can be broadly classified into equation-oriented or procedure-oriented approaches. Equation-oriented approaches involve rigorous mathematical programs, either deterministic or stochastic, constraint programming and graph theory. Procedure-oriented approaches comprise rule-based techniques, heuristics, and meta-heuristics such as simulated annealing, genetic algorithms (GA), or taboo search. These techniques are based on generic principles and schemes and attempt to improve a given solution time effectively, but the optimality and convergence are difficult to assess. In addition, there is no systematic procedure for obtaining good bounds on the attainable optimum values of the objective function [6].

The basis for solving a systems problem is to formulate a representation in an adequate model that captures the features relevant to efficiently support decision-making. Precisely, decision making in process industries results in a highly challenging task. Figure 1 represents a hierarchy of modeling systems that can be found in process system engineering. They are organized based on the temporal scale and the level of decision [7].

The need to integrate the different modeling approaches in a hierarchical decision-support system makes necessary the use of consistent terminology and concepts to

improve the communication and collaboration tasks over the entire system. We believe that a knowledge management system is a suitable tool for this task.

3 Knowledge Management

Knowledge management is about leveraging corporate knowledge for greater productivity, value, and competitiveness [8]. How to manage this knowledge has become an important issue in the past few decades, and the knowledge management community has developed a wide range of technologies and applications for both academic research and practical applications. Recent technological advances have made knowledge management expand in a rapidly manner. Enterprises are not apart from this trend.

Along this line, ontology is the knowledge integration of different representations of knowledge at different levels of formalization. The experts who participate in the ontology process are allowed to use their own terminology, facilitating knowledge integration with cooperative tools [9]. Ontologies are increasingly seen as a key semantic technology for addressing heterogeneities and mitigating the problems they create and for enabling semantics-driven knowledge processing. Ontologies are formal structures enabling acquiring, maintaining, accessing, sharing and reusing information [10]. Knowledge management systems benefit from ontologies that semantically enrich information and precisely define the meaning of various information artifacts.

From the essence of enterprise modeling, we could say that ontologies play a critical role in this integration, enabling better analysis of their performance, and management of their operations. An enterprise model is a computational representation of the structure, activities, processes, information, resources, people, behavior, goals, and constraints of a business, government, or other enterprise. The role of an enterprise model is to achieve model-driven enterprise design, analysis, and operation [11, 12]. In summary, ontologies provide a shared and common understanding of a domain, and an explicit conceptualization that describes the semantics of the enterprise's data.

4 Knowledge management approaches

This section will present in a brief manner four knowledge approaches developed to improve knowledge management in the PSE domain. These knowledge systems address areas that range from batch process modeling to supply chain domain models. As the reader will notice, different enterprise functions have been covered in order to improve the effectiveness in the decision-making process. In the next sections, we briefly describe the batch process ontology, enterprise ontology project, ontological math representation and operations research ontology in the chronological order of development.

4.1 Batch Process Ontology (BaPrOn)

A crucial step for batch process improvement and optimization is to develop information structures which (i) streamline data gathering, and even more, (ii) are

capable of integrating transactional systems with the analytical tools used in an enterprise. The batch process ontology (BaPrOn) [13] is a generic ontology following and ANS/ISA88 standards [14, 15, 16, 17] which allows creating an infrastructure that should be general enough to be applied to any batch system. Additionally, this ontology may be used as a straightforward guideline for standardizing batch process management and control.

BaPrOn integrates different concepts regarding batch processes which are categorized, and the relationships between them are examined. As aforementioned, these concepts are structured in accordance with the ANS/ISA-88 standards. Furthermore, this ontology is also aiming at system integration, improving accessing to information quality by the knowledge description, and thus allowed an improved decision-making process. This was possible using common terminology and manufacturing models which simplifies communication between customers and suppliers in the system.

4.2 Enterprise Ontology Project (EOP)

Enterprises comprise several functions, such as production, marketing, sales, human resources, logistics, safety and environment, which interact with each other. Specifically, environmental management is closely related to several levels in the enterprise structure, since they share a large amount of data and information and has been a topic that has received attention in the last years for a firm to operate sustainably. Enterprise ontology project (EOP) [18] has been developed as the technology for information and knowledge models sharing for the environmental assessment of the enterprise. The model provides an enterprise decision-making supporting tool by combining different information systems, which adapts and recognizes the different elements associated with the enterprise functions, and facilitates assessing the environmental performance of enterprises. Besides, this semantic model considers the environmental system representation within the various SC decision levels. It shows how the link between transactional and analytical systems and semantic and quantitative models can be exploited to propose environmentally and economically sustainable solutions for the design and operation of SCs. As a result, more environmentally conscious supply chains and production processes can be automatically proposed to support the decision-making.

4.3 Ontological Math Representation (OMR)

The basis of decision-making in the enterprise consists in formally representing the system and its sub-systems in models which adequately capture those features which are necessary to reach consistent decisions. Ontological math representation (OMR) [19] aims to semantically representing mathematical models in engineering domain. The ontology encompasses: (i) the mathematical entities themselves and their behavior (i.e. decisions, parameters, constraints, indicators); (ii) the relation-ships among these entities to build mathematical models; and (iii) the properties which allow to relate the mathematical entities to the semantic abstraction of the elements that they represent. Therefore, the entities of the mathematical models can be directly associated with the engineering concepts, which are unified/standardized into the classes, properties, and

axioms of an existing ontology in the enterprise and process domain. As a result, the links between the mathematical elements and the abstraction of the reality are explicitly formalized in the ontological framework. Therefore, although the modeler is still responsible for establishing such links, this framework provides a tool for formalizing them, thus clarifying and unveiling the assumptions in the modeling process and the relations of the mathematical model to the real system. This approach is advantageous because it provides the possibility to integrate mathematical models from different decision levels into a single platform, and to easily incorporate other system aspects, such as the identification of mathematical elements (e.g. variables, performance indicators) that refer to the same engineering entity.

4.4 Operations Research Ontology (ORO)

Finally, we would like to mention an operations research ontology (ORO) semantic model that has been developed as a step forward in capturing the nature of problems and technologies for decision making in the enterprise. Specifically, the whole process of decision making and the creation and classification of equations according to their structure are qualitatively represented in terms of OR principles. This OR ontology is integrated with other two semantic models previously developed, (EOP) and (OMR), thus enhancing the functionalities of the original ontological framework. The scope of these models comprises the representation of the real system for EOP, the mathematical representation domain for OMR, and finally the problem design representation for ORO.

5 Conclusions

The ontological frameworks provides a powerful tool to support computational optimization models for strategical and tactic decision-making, allowing the comprehensive application of enterprise wide optimization throughout the process industry. This work contributes to improve communication within plant process environment, and represents a step forward to support the integration of analytical software tools, resulting into an enhancement of the supply chain functions.

Thus, ontological models enhances the way for achieving a successful decision-making supporting tool which adapts and recognizes the different elements found recipe models. Moreover, a general semantic framework is proposed, which is able to model any plant layout, proving its re-usability. As a whole, the main contributions of this environment and the model behind are re-usability, usability, higher efficiency in communication and coordination procedures. In addition, it has been proved the adequacy of an ontology as a means for sharing information about a general model for different problem representations. As a result, it solves the problem of integration, standardization and compatibility of heterogeneous modeling systems. Even more, the response time for decision making task could be reduced and better decisions adopted owing to faster availability of higher quality data and the improved visibility of the existing relationships between the scheduling function and other hierarchical levels functions.

Acknowledgements. Authors would like to acknowledge the Spanish Ministerio de Economía y Competitividad and the European Regional Development Fund for supporting the present research by projects EHMAN (DPI2009-09386) and SIGERA (DPI2012-37154-C02-01). Finally, the financial support received from CIMAT México is also fully acknowledged.

References

1. Grossmann, I.: Enterprise-wide optimization: A new frontier in process systems engineering. Aiche Journal, Vol. 51, No. 7, pp. 1846–1857 (2005)
2. Venkatasubramanian, V., Zhao, C., Joglekar, G., Jain, A., Hailemariam, L., Suresh, P., Akkisetty, P., Morris, K., Reklaitis, G.: Ontological informatics infrastructure for pharmaceutical product development and manufacturing. Computer and Chemical Engineering 30 (July), pp. 1482–1496 (2006)
3. Simon, F., Murray, T.: Decision support systems. Commun. ACM, Vol. 50, No. 3, pp. 39–40 (2007)
4. Shim, J.P., Warkentin, M., Courtney, F.J., Power, J.D., Sharda, R., Carlsson, C.: Past, present, and future of decision support technology. Decision Support Systems, Vol. 33, No. 2, pp. 111–126 (2002)
5. Morris, W.T.: On the arts of modelling. Management Science, Vol. 13, pp. 707–717 (1967)
6. Pekny, J.F., Reklaitis, G.V.: Towards the Convergence of Theory and Practice: A Technology Guide for Scheduling/Planning Methodology. Foundations of Computer Aided Process Operations (1998)
7. Harjunkoski, I., Nystrom, R., Horch, A.: Integration of scheduling and control—Theory or practice? Computers & Chemical Engineering, Vol. 33, No. 12, pp. 1909–1918, Pergamon (2009)
8. Schreiber, G., Akkermans, H., Anjewierden, A., de Hoog, R., Shadbolt, N., Van de Velde, W., Wielinga, B.: Knowledge Engineering and Management: The Common KADS Methodology. Cambridge, MA: MIT Press (2000)
9. Fernandez-Breis, J.T. Martinez-Bejar, R.: A cooperative tool for facilitating knowledge management. Expert Systems with Applications, Vol. 18, No. 4, pp. 315–330 (2000)
10. Gruber, T.R.: A translation approach to portable ontology specifications. Knowledge Acquisition, Vol. 5, No. 2, pp. 199–220 (1993)
11. Fox, M.S., Barbuceanu, M., Gruninger, M., Lin, J.: An organization ontology for enterprise modelling. In Modeling, In: International Conference on Enterprise Integration Model-ling Technology 97, Springer (1997)
12. Fox, M.S., Gruninger, M.: Enterprise modeling. AI Magazine, Vol. 19, No. 3, pp. 109–121 (1998)
13. Muñoz, E., Espuña, A., Puigjaner, L.: Towards an ontological infrastructure for chemical batch process management. Comput. Chem. Eng., Vol. 34, pp. 668–682 (2010)
14. International Society for Measurement and Control (1995). Batch control part 1 models and terminology.
15. International Society for Measurement and Control. Control batch parte 4 registros de producción de lote (2006)
16. International Society for Measurement and Control. Batch control part 5 automated equipment control models & terminology (2007)
17. International Society for Measurement and Control. Isa-88/95 technical report: Using isa-88 and isa-95 together. Technical report, ISA The Instrumentation, Systems, and Automation Society (2007)

18. Muñoz, E., Capón-García, E., Lainez, J.M., Espuña, A., Puigjaner, L.: Considering environmental assessment in an ontological framework for enterprise sustainability. Journal of Cleaner Production, Vol. 47, pp. 149–164 (2013)
19. Muñoz, E., Capón-García, E., Lainez, J.M., Espuña, A., Puigjaner, L.: Using mathematical knowledge management to support integrated decision-making in the enterprise. Computers and Chemical Engineering. Vol.66, pp.139–150 (2014)

Application of Business Intelligence to the Power System Process Security

N. Jacome-Grajales, G. Escobedo-Briones, G. Arroyo-Figueroa

Instituto de Investigaciones Electricas, Gerencia de Sistemas Informaticos, Cuernavaca, Mexico

{nejacome, gescobedo, garroyo}@iie.org.mx

Abstract. Nowadays, the companies have a growing need for timely and reliable information for operative and strategic decision-making. Companies have understood the importance of enforcing achievements of the goals defined by their business strategies through business intelligence concepts. This paper describes the design and development of business intelligence tools applied to industrial security. The goal is to have available and timely information to make better decisions, to reduce the number of accidents and incidents. The dashboards show a clear and simple way the main indicators of security process. Also, shows the relationship between indicators to determine which may be the attitude possible that caused the accident. The BI tool can be applied in future to other Electric Power Utility processes.

Keywords: Information management, decision support systems, business intelligence, industrial security, power system.

1 Introduction

In a competitive world, information management becomes strategic. The management information aids the modern executive in the making strategic decision, providing quality information taken from the organization's large volumes of past and present data in a summarized and timely manner. Comprehensive and timely information and knowledge are crucial in improve business operations [1]. Management information and business intelligence play a central role in producing up-to-date information for operative and strategic decision-making.

The conceptualization, planning, implementation, development and putting into operation of enterprise information system and business intelligence, in a Company are not a trivial matter [2]. It brings about countless challenges to organization, administration, training and integration, as well as the resolution of technological and cultural problems. Some of the main problems can be [3]:

There exists a large amount of data but it is not easily accessible. Generally, an executive must turn to other people to obtain data. The time required for a middle- or high-level company manager to obtain the required information to be able to make a decision is long, sometimes days.

There are same data with different value, the same data from two or more different sources, in general they do not agree. The same data can follow different paths through each of the departments or areas in the organization, from its origin.

There is a lot of data. But what is important is often unknown. The databases tend to grow indiscriminately and as they do, it becomes difficult to access the relevant information for a certain decision making context.

There are some data not very precise. The data can have an error, such as a measurement error; also, the data can be input more than once throughout its different paths, each input can be a potential source of error.

There are large volume of data that is generated daily in the different technical and administrative processes of production and control related to the generation, transmission, transformation and distribution of electric energy, as well as the associated aspects of administration and finances.

There are variety of platforms, operating systems, programming languages and methods of communication between the operational systems. The age of certain operational systems developed with obsolete techniques.

There are lack of methodologies and ways of working, lack of updated information or the lack of documentation of the operational systems, lack of knowledge about the computer information technology culture based on new paradigms.

In this complex scenario, the application of business intelligence technology can be a big task. Business intelligence requires reliable and timely information. Business Intelligence can be defined as a collection of decision support technologies for gathering, providing access to, and analyzing data for the purpose of helping enterprise users (executives, managers and analysts) make better and faster business decisions [4].

This paper present the application of business intelligence tools for providing comprehensive and timely information of security in an Electric Power Utility (EPU).

2 Business Intelligence

Business Intelligence is a collection of decision support technologies for gathering, providing access to, and analyzing data for the purpose of helping enterprise users (executives, managers and analysts) make better and faster business decisions [5, 6]. The term implies having a comprehensive knowledge of all of the factors that affect the business. It is imperative that companies have an in depth knowledge about factors such as the customers, competitors, business partners, economic environment, and internal operations to make effective and good quality business decisions. Business intelligence enables firms to make these kinds of decisions. Enterprise aimed at enabling knowledge executives, managers and analysts to make better and faster decisions.

The typical components of Business Intelligence architecture for an Enterprise are the following:

Data Sources

Data sources can be operational databases, historical data, external data for example, from market research companies or from the Internet), or information from the already existing data warehouse environment. The data sources can be relational databases or any other data structure that supports the line of business applications. They also can

reside on many different platforms and can contain structured information or unstructured information. Thus, the problems of integrating, cleansing, and standardizing data in preparation for BI tasks can be rather challenging.

Extractions, transformation and loading (ETL)

Extract-Transform-Load (ETL) refers to a collection of tools that play a crucial role in helping discover and correct data quality issues and efficiently load large volumes of data into the warehouse.

Data Warehouse and data marts

The data warehouse is the significant component of business intelligence. It is subject oriented, integrated. The data warehouse supports the physical propagation of data by handling the numerous enterprise records for integration, cleansing, aggregation and query tasks. A data mart is a collection of subject areas organized for decision support based on the needs of a given department. The key difference is that the creation of a data mart is predicated on a specific, predefined need for a certain grouping and configuration of select data [7].

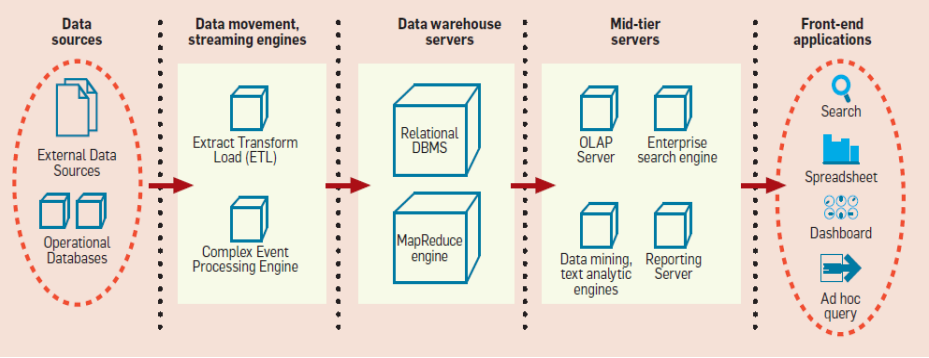


Fig. 1. Typical BI architecture [8].

Data analysis tools

BI includes several tools for the data analysis. It refers to the way in which business users can slice and dice their way through data using sophisticated tools that allow analytical processing and advance analytics. Online analytic processing (OLAP) provides multidimensional, summarized views of business data and is used for reporting, analysis, modeling and planning for optimizing the business. OLAP tools provide the common BI operations such as filtering, aggregation, drill-down and pivoting. Advanced analytics is referred to as data mining, text analytics, forecasting or predictive analytics and artificial intelligence algorithms, this takes advantage of statistical analysis and artificial intelligence techniques to predict or provide certainty measures on facts.

Front-end-applications

There are several popular frontend applications through which users perform BI tasks: spreadsheets, enterprise portals for searching, performance management applications

that enable decision makers to track key performance indicators of the business using visual dashboards, tools that allow users to pose ad hoc queries, viewers for data mining models, and so on. Rapid, ad hoc visualization of data can enable dynamic exploration of patterns, outliers and help uncover relevant facts for BI.

Business intelligence integrates various technologies, such as local and extensive networks, database managers, data visualization, decision support tools, artificial intelligence algorithms and others [9, 10]. Its use depends on the company's specific line business, for this reason each business intelligence application has very particular characteristics for each organization. What makes the concept important is that today the implementation is perfectly feasible given the technological advances in hardware and software that have occurred in the last ten years. These advances include the following among others:

The arrival of On Line Analytical Processing (OLAP) and its application beyond the traditional areas of marketing and finances.

The growth of client/server computing that has created hardware and software servers that are more powerful and sophisticated than ever. Today, the serves compete with the mainframes of yesterday and offer technologically superior memory architectures, with high-speed processor and massive storage capacities.

The appearance of modern database management systems-DBMS's that provide greater support for complex data structures.

The generalized use of internet and intranet accompanied by new and more powerful application.

The advances in data storage, query processing, enterprise search, visualization, data mining, text analytics, artificial intelligence and cloud data services have driven the development of feasible applications.

3 Industrial Security System

The management of industrial security has as main objective to preserve the physical integrity, health and welfare of staff. Also, manage health and safety risks at work associated with the processes of the company. An important aspect of security is the prevention and reduction of risks, this is achieved through an existing hazard identified, assess the magnitude of risk and establish controls for inspection, monitoring and implementation of preventive measures for compliance existing regulations. Other action for prevention and reduction of risk is the learning and training for the operation and maintenance of high risk tasks.

To reduce the number and severity of accidents is necessary to have information on the different risks. In order to support safety management, industrial safety system was developed. The industrial safety system, called as *Sistema Integral de Seguridad y Salud en el Trabajo* (SISST) [11], manages three areas of industrial safety: accident management, infrastructure safety and health protection, safety management. The figure 2 shows the main interface of the system.

Accident module

The accident module manages of information on accidents, incidents, illnesses, accident cause, management indicators of frequency, severity and degree of risk.

Infrastructure safety and health protection module

The accident module manages of information on accidents, incidents, illnesses, accident cause, management indicators of frequency, severity and degree of risk.

Safety management module

Safety management module manages of information on program evaluations security assessment and compliance with legal requirements, hazard identification, risk assessment and control, and occupational health and preventive-control remedial measures.

Fig. 2. Security information system.

Although the system handles various standard reports and statistics of accidents to inform at different levels of aggregation, is necessary to relate data for discovery new knowledge about accident and incident. The goal is to design analytical tools that will enable the reduction of accidents. With Business Intelligence tools could devote their efforts to prevent, rather than react.

4 Business Intelligence Implementation

The goal of Business Intelligence is to provide new knowledge to the company, from the automated exploitation of historical information for business actions are taken to be better supported. The results obtained by applying techniques of BI to the database of industrial safety of Mexican Electric Utility are shown in this section.

The benefits it will bring to the implementation of business intelligence technologies are mainly:

a) Accessing security information in a timely and reliable, which allow reducing the time in making and decisions, creating more effective decisions to have the information available.
b) Display detailed information of the security process, making further analysis as a result of having consolidated historical information and current information.
c) Allow delivery of data in a flexible, dynamic and in many cases to solve unplanned queries.
d) As result of above: having a decrease in the number of injured or dead; having a decrease in the economic impact caused by accidents and Decrease the number of days lost due to accidents.

Today there are many tools that offer similar products to both large and small organizations. BI vendors propose solutions both horizontal and vertical and the best choice will depend on the specific need of each organization. With horizontal solutions from scratch by an application tailored to the need. Vertical solutions are aimed at an industry already developed components and only fit specific needs. In this case the implementation of BI tools for the industrial security was done by development proprietary tools under Windows platform and a solution horizontal. For the design and development of BI tools take in count the following queries:

Goal alignment queries: the application of BI tools has the aim to reduce the number of accidents and incidents.

Baseline queries: the source of information is generated by security information system (in Spanish, *Sistema de Integral de Seguridad y Salud en el Trabajo* –SISST). The system manages three areas of industrial safety for power system processes: accident, infrastructure safety and health protection; and safety management.

Customer and stakeholder queries: There three kinds of users: operative, analytical and executive.

Metric-related queries: the metrics was defined by the security expert. The metrics includes: security indicators, and related variables. A success factor for the development of BI applications is the definition of metrics.

Measurement queries. The operational variables are load by the security information system. There are procedures and methodologies defined by the company for the variables and its load frequency.

The methodology used for the definition of the BI architecture considers the construction of a single enterprise data warehouse and from it will emerge Data Marts (small units of analysis), with an overall vision of the company. The figure 3 shows the BI-SISST architecture.

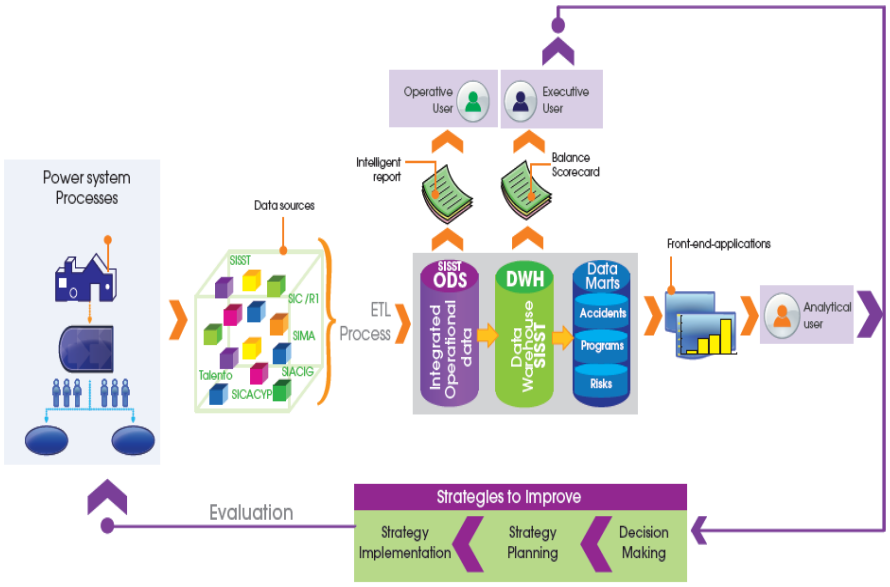


Fig. 3. BI-SISST proposed architecture.

The primary tasks include gathering, preparing and analyzing data. From operational database of industrial security system, extract the most relevant data to the user through an ETL process, which loads and makes the necessary transformations and data cleansing. These are the integrated operational data (ODS-SISST) of the security system. The data itself must be of high quality.

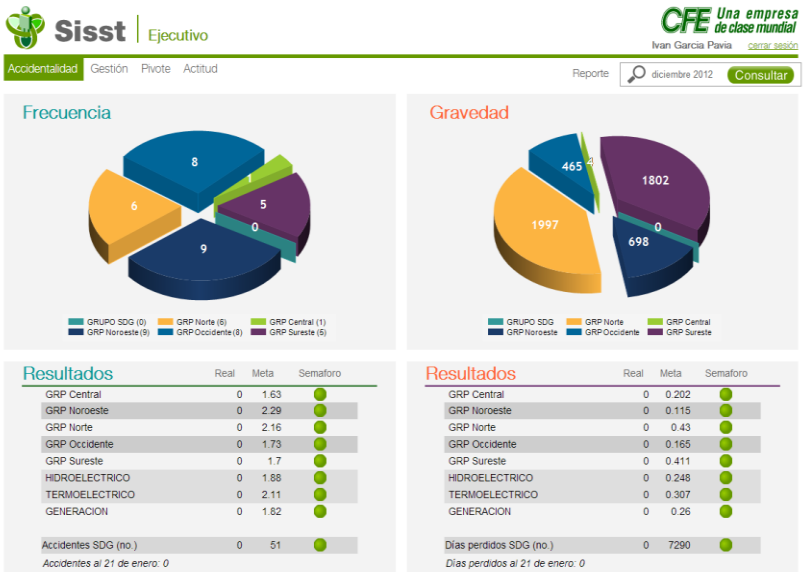
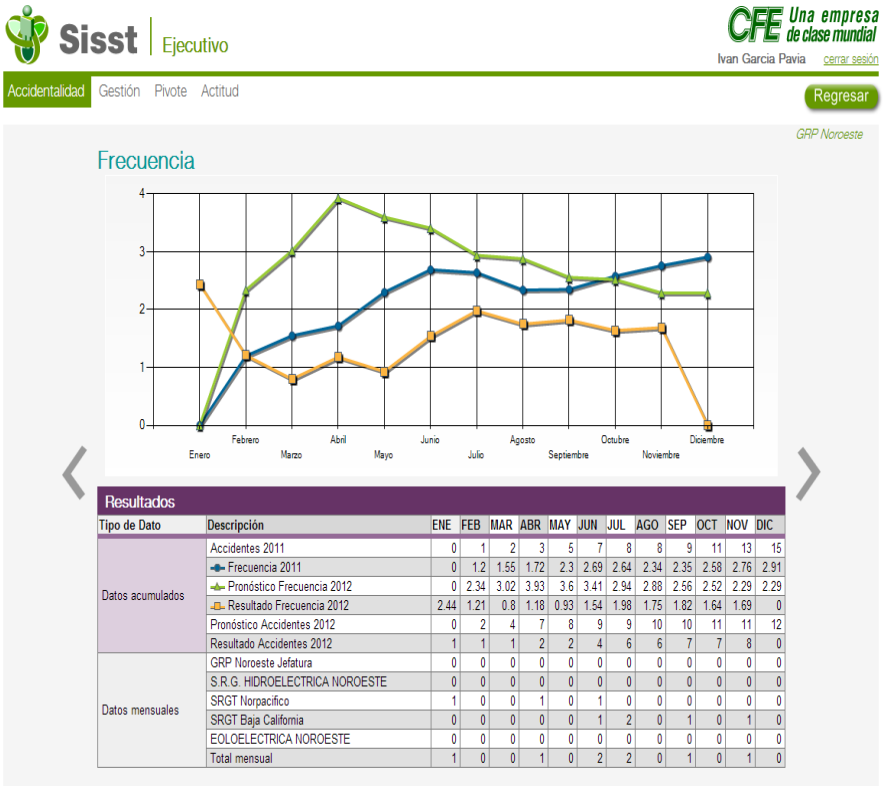


Fig. 4. Accident frequency and severity.

Once the data is complete and clean, they pass through another ETL process to data warehouse (DWH-SISST). Finally, data marts are created for accidents, safety programs and hazards. From these data were carried out the data analysis, through multidimensional analysis and consulted manager dashboards.

The results of the BI application are the front-end-applications through of dashboards. The Dashboard should provide the executive with a tool for navigating through company's information. The figure 4 shows the accident frequency by area with pie contribution and table of results with goal and indicator state. Also shows the severity by area.

The figure 5 shows the historical, real and forecast accident bye year and by moth. This information is obtained to accident module.



Tipo de Dato	Descripción	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SEP	OCT	NOV	DIC
Datos acumulados	Accidentes 2011	0	1	2	3	5	7	8	8	9	11	13	15
	Frecuencia 2011	0	1.2	1.55	1.72	2.3	2.69	2.64	2.34	2.35	2.58	2.76	2.91
	Pronóstico Frecuencia 2012	0	2.34	3.02	3.93	3.6	3.41	2.94	2.88	2.56	2.52	2.29	2.29
	Resultado Frecuencia 2012	2.44	1.21	0.8	1.18	0.93	1.54	1.98	1.75	1.82	1.64	1.69	0
	Pronóstico Accidentes 2012	0	2	4	7	8	9	9	10	10	11	11	12
	Resultado Accidentes 2012	1	1	1	2	2	4	6	6	7	7	8	0
Datos mensuales	GRP Noroeste Jefatura	0	0	0	0	0	0	0	0	0	0	0	0
	S.R.G. HIDROELECTRICA NOROESTE	0	0	0	0	0	0	0	0	0	0	0	0
	SRGT Norpacifico	1	0	0	1	0	1	0	0	0	0	0	0
	SRGT Baja California	0	0	0	0	0	1	2	0	1	0	1	0
	EOLOELECTRICA NOROESTE	0	0	0	0	0	0	0	0	0	0	0	0
	Total mensual	1	0	0	1	0	2	2	0	1	0	1	0

Fig. 5. Accident frequency: historical, real and forecast.

The main indicators of the security have presented in the figure 6. The definition of indicators is an important task to ensure the success of business intelligence. This dashboard shows the security maturity in terms of identification of hazards, legal requirements, compliance of security programs, safety forms, incidents and acts of government.

With the business intelligence is possible to combine different information. For example it is possible to relate the accident with the attitudes. For this case, the figure 7 presents the relations between the accidents with attitudes.

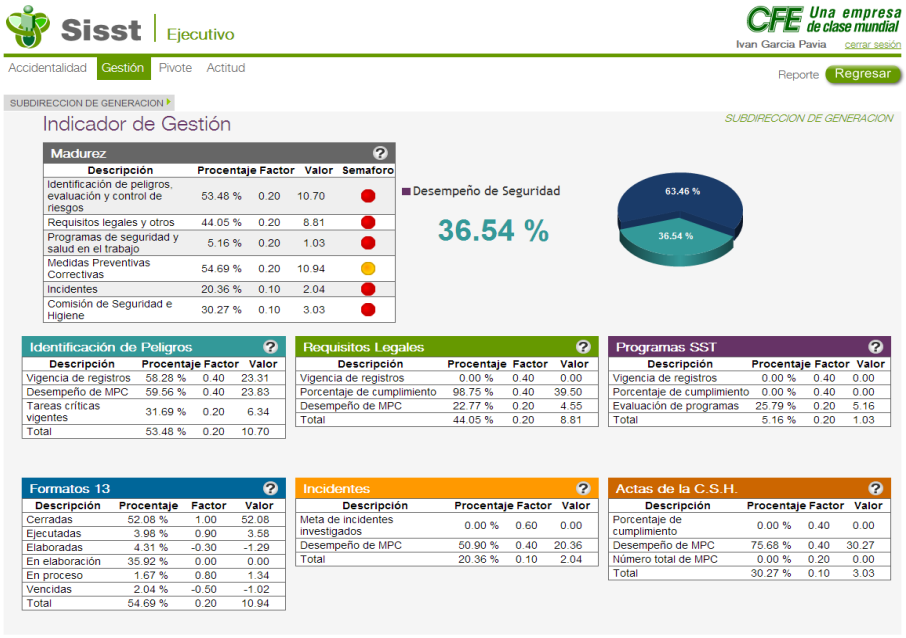


Fig. 6. Management indicators.

Fig. 7. Relationship between accidents and incidents with attitudes.

Also, it is possible to classify the kind of attitudes: learning, knowledge and responsibility. This means that the an accident occurs can determine if the accident occurred due to lack of training, lack of knowledge or lack of responsibility. For the case of incidents also it is possible determinate if occurred by lack of training, knowledge or responsibility.

The capabilities of business intelligence include decision support, online analytical processing, statistical analysis, forecasting and data mining. BI-SISST system is a traditional business intelligence application with back-end database and front-end user interface, software that processes the information and reporting systems.

5 Conclusions

Business Intelligence as a concept is becoming more common in everyday business life. BI incorporates people, process, and also knowledge as an end product. The implementation of business intelligence (BI) system is a complex undertaking requiring considerable resources.

An important factor to build BI applications is the information management. BI requires reliable and timely information and generates summary information for the operative and strategic decision making. In addition, the implementation of a BI system is often associated with the following challenges: underlying original back-end systems and processes which were not adapted for BI applications; poor data quality derived from source systems that can often go unnoticed until cross-systems analysis is conducted; and the maintenance process that tends to be vague and ill-defined

To attack this problem is necessary to implement enterprise architecture with its two main components: business architecture and technological architecture can help ensure that the data source will be reliable.

The business intelligence tools developed for the industrial security have had good results. The information displayed through dashboards make career choices have led to the decrease of accidents. In particular, the relationship between accidents and attitudes has been a great help to generate preventive actions to avoid accidents. Also, it indicates if the accident occurred due to lack of training, knowledge, or lack of responsibility.

Acknowledgment. The authors wish to thank Israel Paredes Rivera, Department Head of Technical Services Unit of CFE for their important work in supporting, organizing and promoting the project. We would like to convey special thanks to all the operators who selflessly shared their reliable knowledge and experience to make this project successful.

References

1. Chaudhuri, S., Dayal, U., Narasayya, V.: An Overview of Business Intelligence Technology. Communications of the ACM, Vol. 54, No. 8, pp: 88–98 (2011)
2. Hannula, M., Pirttimaki, V.: Business intelligence Emperical Study on to the 50 Finnish Companies. Journal of American Academy of Business, Vol. 2, No. 2 (2003)

3. Mejia-Lavalle, M., Arroyo-Figueroa, G.: Construction of a Corporative Information System for an Electric Power Company. IEEE Conference on Intelligent Engineering Systems Proceedings, pp. 177–182 (2010)
4. Ranjan, J.: Business Intelligence: concepts, components, techniques and benefits. Journal of Theoretical and Applied Information Technology, pp. 60–69 (2007)
5. Turban, E., Aronson, J., Liang, T., Shard, R.: Decision support and business intelligence systems. Pretince Hall (2005)
6. McKay, L.: Businees Intelligence comes out of the back office. CRM Magazine (2008)
7. Rizzi, S., Abelló, A., Lechtenborger, J., Trujillo, J.: Research in Data Warehouse Modeling and Design: Dead or Alive? In 9th International Workshop on Data Warehousing and OLAP (DOLAP), Arlington (USA), ACM Press, pp. 3–10 (2006)
8. Watson, H.J., Wixom, B.H.: The current state of business intelligence. Computer, Vol. 40, No. 9, pp. 96–99 (2007)
9. Olszak, C., Ziemba, E.: Approach to Building and Implementing Business Intelligence Systems. Interdisciplinary Journal of Information, Knowledge, and Management, Vol. 2, pp: 135–148 (2007)
10. Wang, H., Wang, S.: A Knowledge Management Approach to Data Mining Process for Business Intelligence. Industrial Management & Data Systems, Vol. 108, No. 5, pp. 622–634 (2008)
11. Jacome-Grajales, N., Escobedo-Briones, G., Guadarrama Villa, E.: Inteligencia de Negocios en el área de seguridad de la CFE. Congreso Internacional sobre Innovación y Desarrollo Tecnológico, pp. 677–685 (2011)

Impact of the Work Culture and Suppliers of Equipment on the Benefits of TPM

José Roberto Diaz Reza[1], Deysi Guadalupe Marquez Gayosso[2], Valeria Martínez Loya[2], Julio Blanco Fernandez[1], Emilio Jimenez-Macias[1]

[1] University of La Rioja, Spain

[2] Autonomous University of Ciudad Juarez, Mexico

roberto.diaz.reza@gmail.com, {emilio.jimenez, julio.blanco}@unirioja.es, {marquezdeysi, valeriie_05}@hotmail.com

Abstract. Today's markets are uncertain and highly competitive, surviving in them becomes a challenge for companies. Therefore, companies should be at the forefront having its equipment in optimal conditions to avoid failures, in addition the updating of personnel in order to ensure that their processes provide quality products. For this reason, it is necessary to perform maintenance of the machines regularly. In that sense, a tool that facilitates this process is Total Productive Maintenance (TPM), which has played an important role within companies in recent times. Because of this, the objective of this research is to quantify the impact of the human factor in the benefits that TPM provides. The Structural Equation Modeling technique is considered in the development of this research which considers three latent variables related to Total Productive Maintenance (TPM), Suppliers, Work Culture and the TPM's benefits.

Keywords: Structural equation modeling, total productive maintenance (TPM).

1 Introduction

Manufacturing companies operating in the fast changing market and highly competitive in the last two decades have adopted the principles of Lean thinking. In doing so, they are reorganized into cells, and value stream to improve the quality, flexibility and customer response time in their manufacturing processes [1]. A key point that all authors note regarding Lean Manufacturing is not about reducing headcount, but rather the improvement of productivity in the broadest sense of the concept and costs are seen as a result of practices, systems and processes, which means that the cost reduction occurs when you manage to improve them [2].

Lean Manufacturing concepts were developed mostly in Japanese industries, especially Toyota, but today Lean Manufacturing tools are used for waste reduction as suggested by many authors, [1, 3-5], but in practice, Lean Manufacturing maximizes product value through the minimization of waste [5]. Lean principles define the value of the product / service as perceived by the customer and then make the flow in line

with customer attraction and pursuit of perfection through continuous improvement to eliminate waste by classification activities that add value (VA) and activities that do not add value (NVA) [5].

When a company is Lean simultaneously minimum inventory is obtained in the form of raw materials, work in process, finished products, minimum product nonconformance, rework, rejects and returns, losses minimal production through unplanned downtime and planned change and transition time, fee reductions and short stops, and quality problems, minimum cycle time of the system, minimum lag times between processes, minimal variability in the rates of production and processes, minimum unit cost production, excellent performance of on-time delivery, customer satisfaction, and gross profit, etc. [2].

One of the tools most used in lean manufacturing production systems is the Total Productive Maintenance (TPM), which essentially seeks to improve equipment performance, reduce variability, and shorten the period of the offer, but later, improved total operating performance became the main objective of TPM, even including the gradual improvement of the quality, maintenance, cleaning and order; new manufacturing technologies, and work on all employees with advanced technology equipment. [6].

Total Productive Maintenance (TPM) is defined as a philosophy designed to integrate maintenance of equipment in the manufacturing process. It is a system that maintains and improves the integrity of production systems and quality through the machines, equipment and processes that add value to the product [7]. TPM focuses on maintaining all equipment in good condition to avoid breakdowns and delays in the manufacturing process also strives to avoid minor stops and defects during the production process and provides a safe working environment [7].

Therefore, it appears that TPM provides a comprehensive enterprise-wide approach to the management of maintenance, which can be divided into elements in the short and long term. In the long term, efforts will focus on the design of new equipment and the elimination of downtime where the participation of the different areas of the organization is required. In the short term, TPM activities include autonomous maintenance program for the department of production and a planned maintenance program for the maintenance department [8].

In summary, based on the definition and activities of TPM, it can be concluded that the objective of TPM is to continually improve the availability and prevent degradation of equipment for maximum effectiveness thereof and to achieve this, strong support is required management as well as the continued use of work equipment and small group activities for improvement [9].

In [9] it mentions that operators and maintenance workers need to have a greater understanding of the functions performed by each and occasionally, to acquire new skills. For example, operators need to learn to anticipate problems and be able to solve minor adjustments and basic preventive maintenance such as checking routines, cleaning and lubrication, a more important role in multiple skills is viewed as an essential support.

To achieve its objective, the TPM requires a number of elements, and it would not be possible to make proper repairs on your machine if you do not have the parts or components required at the time they are needed, or that maintenance personnel not know the methods and procedures for carrying out such repairs. From this, we conclude

that the human factor is very important and so is the relationship with suppliers of machinery and components.

Therefore, the objective of this research is to investigate the impact of the human factor in obtaining the benefits TPM offers, so three latent variables and relationships, which are being studied: Suppliers, Culture and Employment Benefits TPM. The results are intended to provide from a quantitative point of view the impact between the variables, so that managers can determine those variables that are important from those that are trivial.

1.1 Work Culture

Charles Roger (2014) cited for [10] mentions that the culture of the organization is to the values, beliefs, attitudes and behaviors that employees share and use every day in their work. Culture determines how employees describe where they work, how they understand the business, and how they see themselves as part of the organization. Culture is also a driver of decisions, actions, and ultimately the overall performance of the organization.

The role of human resources, skills and abilities is considered the most important in the company and in all the operations it performs, but especially in the care we have with the materials and equipment operating in the production process [11] and it is why the management and senior management must work in generating a working culture for the company. Remember that employees are those planning activities in general terms about the company and therefore, are those who know of a better way its equipment, including strengths and weaknesses, so that the plans and maintenance programs should include [12].

1.2 Equipment Suppliers

With the globalization of the economy, market competition becomes more intense in order to maintain a competitive edge, companies try to form an alliance with partners, customers, suppliers, among others; for which it requires trust, loyalty of its members. Therefore, a good and stable supplier-buyer relationship is crucial to a strong partnership between a supplier of industrial equipment and the manufacturer [13].

Make a good selection of suppliers of equipment and industrial components significantly reduce purchasing costs and improve competitiveness of companies, it is why many experts believe that the selection of suppliers is the most important activity in any company for the acquisition your product and in this case, machinery and production equipment [14].

Yuqi Wang [15] Wang Yuqi mentions that it is important for manufacturers of large and complicated sure you have an effective after-sales service and fast, this in order that maintenance can be provided promptly in case of failure or falling equipment. It also mentions that the selection of suppliers of maintenance and repair of machinery of precision parts is very important in products after-sales services.

The maintenance service is an important part of product service and quality of service and attention during customer use of the products will have great influence on product branding, customer return rate, Gronroos [16] mentions that the service provider does not transfer value, but therefore, the value of the service is created by the

supplier and the customer together [17] and then it becomes crucial to seek to have a good relationship.

The provider of maintenance services can help companies manufacturing complex to choose the right service provider, thus reducing maintenance costs and improving the standard capacity maintenance management and servicing products and this is of great importance for improving the competitiveness of the products. [17].

As can be seen in paragraphs, it appears that the success of TPM depends largely on management to perform the administrative part with the supplier and therefore the following hypothesis is proposed:

H_1: Work Culture towards TPM has a direct and positive about suppliers used in the production process equipment impact.

1.3 Benefit Productivity

Having a good administration of TPM, you are allowed to increase productivity and to improve it in an industrial plant, it is necessary to selecting the tasks getting less consumption of resources, to achieve this it is necessary to combine many actions of organizational, management and technologies, and appropriate use of resources. In [18] a number of benefits arising from the implementation of TPM, such as improving the quality of the final product, the total productivity of the company is improved, the yield rate is improved, among others are reported.

Similarly, (Koelsch, 1993) cited by [19] provides a range of percentages of improvement in some areas of the companies that have implemented TPM, these percentages are provided by Constance Dyer, Director of Research and Product Development TPM, Productivity Inc. these percentages include: 50% reduction in failures work rates, 70% reduction in lost production, 50-90% reduction in set-up, 25-40% capacity increase, 50% increase in labor productivity, and 60% reduction in unit costs of maintenance.

According to [20] in the current industrial scenario, large losses / waste on the shop floor occur. These wastes are due to operators, maintenance personnel, tooling problems, the unavailability of components on time, etc., [20]. Therefore, management should raise awareness in employees in relation to the cost of having to have a detained without producing plant, which is achieved by creating an appropriate work culture. Therefore, the following hypothesis is proposed:

H_2: Work Culture has a direct and positive impact on the productivity benefits.

However, these benefits are obtained from the TPM and are reflected as indices of operating efficiency in the company, have several sources, one of which is the good relationship and collaboration we have with suppliers of equipment. It would not be possible to have high levels of machine efficiency if not delivered on time to repair components thereof. [21]. For example, it is reported that the supply of parts is crucial in highly integrated, such as the Chinese automotive industries [22], where efficiency aspects associated directly to providers.

Similarly, it has been associated with TPM processing quality final product [23], with the economic performance of the company [12] and the material flow along

production lines [24]. Therefore, given the importance of suppliers in obtaining productivity indices of the company, the following hypothesis is proposed:

H_3: Equipment suppliers have a direct and positive effect on the productivity benefits of the company.

A summary of the hypotheses is illustrated in Figure 1.

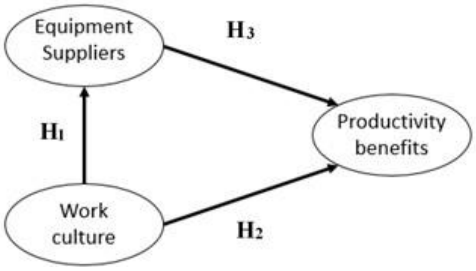


Fig. 1. Proposed Hypotheses.

2 Methodology

This section describes the methodology used for this research. This work was carried out by using a questionnaire that focus on the industrial sector in northern Mexico, specifically in Ciudad Juarez, Chihuahua.

2.1 Questionnaire Design

First, a questionnaire was designed, which consists of 23 items which are divided into four sections. The first section deals with demographics of the sample as the industrial sector to which the company belongs, position, years of experience and sex of the respondent (4 items), the second section is intended for "equipment suppliers", the third is focused the "work culture" and the fourth section is focused on the "productivity benefits", same as described below in Table 1 and which refer to authors who justify it becomes.

Table 1. Latent variable equipment suppliers.

Latent Variable	Item
Equipment Suppliers [25, 26, 27]	Do suppliers provide adequate maintenance manuals of the machines?
	Do suppliers provide technical assistance?
	Does the company signed some kind of agreement with suppliers?
	Are providers comply with the guarantees?
	Does the machine supplier provides training?
	Does the supplier provide training of the machinery in the buyer´s plant?
	The supplier is responsible for the installation and commissioning of machinery that is sold?

Latent Variable	Item
Work Culture [20, 18]	The company emphasizes in placing all the tools and accessories?
	Work places areas are clean and tidy?
	Does the company is kept clean at all times and in all departments?
	Do employees receive training to properly multitasking?
	Does the scheduling of maintenance is carried out in coordination with the production department?
Productivity Benefits [16, 19, 26]	Eliminate losses that affect plant productivity
	Improved reliability and availability of equipment
	Reducing maintenance costs
	Improving the quality of the final product
	Lower financial costs for spare parts
	Improve company's technology
	Increased responsiveness to market movements
	Create competitive capabilities from the factory

In order to answer the questions, likert scale with values from 1 to 5, the scale is a function of the frequency with which tasks are performed and the advantages are obtained during the process of implementing TPM is used. Table 2 shows the values and their meaning.

Table 2. Importance levels.

Scale	Description
1	Never
2	Sporadically
3	Regularly
4	Very often or almost always
5	Always

2.2 Data Collection

As mentioned above, the questionnaire was administered exclusively in the industrial sector in northern Mexico (Ciudad Juarez) and was answered by people involved in the maintenance area, and was applied at different hierarchical levels, such as managers, technicians, engineers, operators that perform or execute maintenance plans. The principle of inclusion should be that had an administrative position or had relationships with suppliers of equipment, since this variable is also being examined.

2.3 Capturing Information and Data Debbuging

After gathering information from the different questionnaires, the next step is to capture the information in the statistical software SPSS 21®, where each line represents a case

or completed survey and each column lists all the elements or variables that make up the questionnaire.

After the capture of information, we proceeded to make debugging database, which is to identify the missing values (unanswered questions) for each of the cases, if missing values are equal to or greater than 10 % this questionnaire is removed [28] , done this, the standard deviation of each of the cases is calculated, and those questionnaires are removed with, after that less than 0.5 standard deviation, and extreme outliers are identified, which may be due to data entry errors data. Missing values are filled with the median, which because it has an ordinal scale and extreme values are replaced by this measure of central tendency.

Once the database was refined, the following was the validation of all latent variables, some indexes for this validation were used. Index Cronbach's alpha was used to measure the internal consistency of a dependent variable that consists of other measurable. [29], this index is considered acceptable from 0.7 and has been used in numerous studies [30]. Discriminant validity is measured by the average variance extracted (AVE), the minimum cutoff value is 0.5 [31].

2.4 Structural Equation Model

The structural equation modeling (SEM) is a well-established statistical technique that has become popular in social science research [32], It is why the WarpPls 5.0® software for modeling of this technique is used. The versatility of SEM is that it can handle estimation problems involving latent variables and measurement errors [32].
Within the structural equation modeling three effects, the direct, indirect and total effects are analyzed.

3 Results

From the surveys and after purging the database, a total of 341 valid questionnaires were obtained, Figure 2 shows the percentage of registered industrial sector surveys show. In the graph we can see that the people who answered the questionnaire were the automotive sector, followed by the electronics industry and can see that there was a percentage of 1% unanswered.

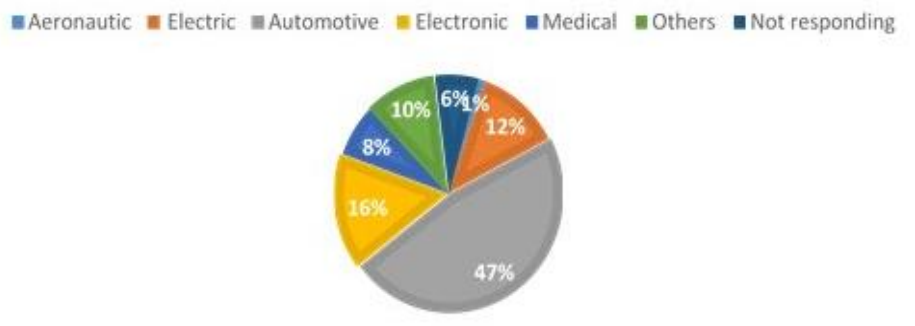


Fig. 2. Industry sector.

Table 3 shows the experience of people in his position shown and analyzed against gender, of the 317, where 24.3% are women and 75.7% men, also it notes that there are 159 people who have more than 2 years of experience which represents 50% of respondents, and finally notes that a total of 24 undeclared.

Table 3. Experience vs. Gender.

Years	Gender		Total
	Female	Male	
0-1	28	62	90
1-2	12	56	68
2-5	23	50	73
5-10	11	48	59
More than 10	3	24	27
Total	77	240	317

3.1 Validation

Table 4 shows the levels of quality and validity for each of the variables analyzed in the model. According to the R-Squared can conclude that the variables have predictive validity from a parametric view, this because the value is greater than 0.2, which is the minimum cutoff. Based on the positive values of Q-Squared and close to R-Squared, one can conclude that the model has predictive validity from a nonparametric view. The Cronbach's alpha indices show that it has adequate internal validity because the values are greater than 0.7. The birds above 0.5 show that all variables have convergent validity.

Table 4. Latent variables coefficients.

	Productivity Benefits	Equipment Suppliers	Work Culture
R-Squared	0.327	0.329	
Aadj. R-Squared	0.323	0.327	
Composite Reliab.	0.957	0.906	0.854
Cronbach´s alpha	0.948	0.876	0.784
Avg. Var. Extrac.	0.734	0.617	0.541
Full Collin. VIF	1.424	1.615	1.63
Q-Squared	0.329	0.33	

The Figure 3 shows the structural equation model with the results for each of the relationships among variables according to the methodology described above. Similarly, the figure shows for all parameters or latent variables their estimated values, the P-values that determined their statistical significance, and the R-squared values that indicated their percentage of variance.

3.2 Model Fit and Quality Indices

The following indices are used:

- Average path coefficient (APC) = 0.410, P<0.001,
- Average R-squared (ARS) = 0.328, P<0.001,
- Average adjusted R-squared (AARS) = 0.325, P<0.001,
- Average block VIF (AVIF) = 1.378, acceptable if <= 5, ideally <= 3.3,
- Average full collinearity VIF (AFVIF) = 1.557, acceptable if <= 5, ideally <= 3.3,
- Tenenhaus GoF (GoF) = 0.455, small >= 0.1, medium >= 0.25, large >= 0.36.

According to APC, ARS, AARS indexes, it can be concluded that the model variables are statistically significant, as they have a value of P <0.05 therefore valid. The index goodness of fit Tenenhaus with a score of 0.445 which is higher than recommended 0.36, indicating that the model is efficient.

3.3 Direct Effects

Based on the values of the betas that you have in Figure 2, hypotheses are concluded, which have been formulated based on the direct effects:

H_1 There is enough statistical evidence to say that latent variable equipment suppliers have a direct and positive impact on latent variable productivity benefits because when the first latent increases its variable standard deviation by one unit, the second does in 0.31 units.
H_2 There is enough statistical evidence to say that latent variable work culture has a direct and positive impact on the latent variable equipment suppliers because when the first latent variable increases its standard deviation by one unit, the second does it in 0.57 units
H_3 There is enough statistical evidence to say that latent variable work culture has a direct and positive impact on latent variable productivity benefits because when the first variable increases its standard deviation by one unit, the second does it in 0.35 units.

3.4 Indirect Effects

For this model you have only an indirect effect that occurs between the latent variables *work culture* towards *productivity benefits* through the mediating variable latent *equipment suppliers* with a size of 0.176 and P <0.001 value.

3.5 Total Effects

Table 5 shows the total effects, which are obtained from the sum of the direct and indirect effects that the variables have each other. Suppliers have a direct effect on productivity benefits, also, the work culture has a direct effect on the productivity benefits and finally the work culture has an indirect effect on the productivity benefits through equipment suppliers.

Table 5. Total effects.

	Productivity Benefits	Equipment Suppliers	Work Culture
Productivity Benefits		0.307	0.524
Equipment Suppliers			0.574
Work Culture			

4 Conclusions

According to the analysis of the results, it can conclude that equipment suppliers have a direct and positive about the productivity benefits impact, which is why it is very important to have a good relationship with them as an important part of the maintenance process and consequently the productivity benefits. In addition, it can be concluded that it is important to have a good working culture within the company and to equipment suppliers since this may depend we have a good relationship with them because the latent variable work culture has a direct and positive impact on the latent variable equipment suppliers. Finally, it is concluded that the work culture has a direct and positive impact on the productivity benefits because with good work practices can get many benefits.

References

1. Fullerton, R.R., Kennedy, F.A., Widener, S.K.: Lean manufacturing and firm performance: The incremental contribution of lean management accounting practices. Journal of Operations Management, 32(7-8), pp. 414–428 (2014)
2. Moore, R.: 7 - Lean manufacturing, in Selecting the Right Manufacturing Improvement Tools, Moore, R. (ed.) Butterworth-Heinemann: Burlington. pp. 135–158 (2007)
3. Santos, Z.G.D., Vieira, L., Balbinotti, G.: Lean Manufacturing and Ergonomic Working Conditions in the Automotive Industry. Procedia Manufacturing, Vol. 3, pp. 5947–5954 (2015)
4. Bortolotti, T., Boscari, S., Danese, P.: Successful lean implementation: Organizational culture and soft lean practices. International Journal of Production Economics, Vol. 160, pp. 182–201 (2015)
5. Sundar, R., Balaji, A.N., Kumar, R.M.S.: A Review on Lean Manufacturing Implementation Techniques. Procedia Engineering, Vol. 97, pp. 1875–1885 (2014)
6. Arslankaya, S., Atay, H.: Maintenance Management and Lean Manufacturing Practices in a Firm Which Produces Dairy Products. Procedia - Social and Behavioral Sciences, Vol. 207, pp. 214–224 (2015)
7. Rahman, C.M.L.: Assessment of total productive maintenance implementation in a semiautomated manufacturing company through downtime and mean downtime analysis. in Industrial Engineering and Operations Management (IEOM), International Conference (2015)
8. Chan, F.T.S.: Implementation of total productive maintenance: A case study. International Journal of Production Economics, Vol. 95, Vol. 1, pp. 71–94 (2005)
9. Cooke, F.L.: Implementing TPM in plant maintenance: some organisational barriers. International Journal of Quality & Reliability Management, Vol. 17, No. 9, pp. 1003–1016 (2000)

10. Ali, N.M.: Influence of Leadership Styles in Creating Quality Work Culture. Procedia Economics and Finance, Vol. 31, pp. 161–169 (2015)
11. Chlebus, E.: A new approach on implementing TPM in a mine – A case study. Archives of Civil and Mechanical Engineering, Vol. 15, No. 4, pp. 873–884 (2015)
12. Kamath, H.N., Rodrigues, L.L.R.: Simultaneous consideration of TQM and TPM influence on Production Performance: A case study on multicolor offset machine using SD methodology. Perspectives in Science.
13. Yun Zhang, H., Zhao Han, S., Xiao Ling, W.: Research of Buyer-Supplier Relationship Based on Agent Approach. In Wireless Communications, Networking and Mobile Computing (WiCOM), 4th International Conference on (2008)
14. Mirakhorli, A., Farahani, M.H., Ramtin, F.: New approach in supplier selection using linear physical programming. In Service Operations, Logistics and Informatics (SOLI) IEEE/INFORMS International Conference on (2009)
15. Yuqi, W.: An Application of the AHP in Supplier Selection of Maintenance and Repair Parts. in Information Science and Engineering (ICISE), 1st International Conference on. (2009)
16. Li, H., Pei, X.-M.: An Evaluation Method on Service Capability of Product Maintenance Service Provider. In System Science, Engineering Design and Manufacturing Informatization (ICSEM), International Conference on. (2010)
17. Lin, G.-P., Du, Y.-F. Shi-ming, L.: Study on Pricing of a Sort of Maintenance-Service Contract Based on Adjustment of Quantity and Cost. in Service Systems and Service Management, International Conference on (2006)
18. Eti, M.C., Ogaji, S.O.T., Probert, S.D.: Implementing total productive maintenance in Nigerian manufacturing industries. Applied Energy, Vol. 79, No. 4, pp. 385–401 (2004)
19. McKone, K.E., Schroeder, R.G., Cua, K.O.: The impact of total productive maintenance practices on manufacturing performance. Journal of Operations Management, Vol. 19, No. 1, pp. 39–58 (2001)
20. Singh, R.: Total Productive Maintenance (TPM) Implementation in a Machine Shop: A Case Study. Procedia Engineering, Vol. 51, pp. 592–599 (2013)
21. Willmott, P., McCarthy, D.: 2 - Assessing the true costs and benefits of TPM, in Total Productivity Maintenance (2 Ed.). Butterworth-Heinemann: Oxford. pp. 17–22 (2001)
22. Wang, Z.: Composite sustainable manufacturing practice and performance framework: Chinese auto-parts suppliers′ perspective. International Journal of Production Economics, Vol. 170, Part A, pp. 219–233 (2015)
23. Konecny, P.A., Thun, J.-H.: Do it separately or simultaneously - An empirical analysis of a conjoint implementation of TQM and TPM on plant performance. International Journal of Production Economics, Vol. 133, No. 2, pp. 496–507 (2011)
24. Cua, K.O., McKone, K.E., Schroeder, R.G.: Relationships between implementation of TQM, JIT, and TPM and manufacturing performance. Journal of Operations Management, Vol. 19, No. 6, pp. 675–694 (2001)
25. Wagner, S.M., Bode, C.: Supplier relationship-specific investments and the role of safeguards for supplier innovation sharing. Journal of Operations Management, Vol. 32, No. 3, pp. 65–78 (2014)
26. Grondys, K.: Economic and Technical Conditions of Selection of Spare Parts Suppliers of Technical Equipment. Procedia Economics and Finance, Vol. 27, pp. 85–92 (2015)
27. He, Y.: The impact of supplier integration on customer integration and new product performance: The mediating role of manufacturing flexibility under trust theory. International Journal of Production Economics, Vol. 147, Part B, pp. 260–270 (2014)
28. Hair, J.F.: Multivariate Data Analysis (7 Ed.). Prentice Hall (2010)
29. Cronbach, L.J.: Coefficient alpha and the internal structure of tests. Psychometrika, Vol. 16, No. 3, pp. 297–334.

30. Alcaraz, J.L.G.: A systematic review/survey for JIT implementation: Mexican maquiladoras as case study. Computers in Industry, Vol. 65, No. 4, pp. 761–773 (2014)
31. Kock, N.: Using WarpPLS in e-collaboration studies: What if I have only one group and one condition. International Journal of e-Collaboration, Vol. 9, No. 3 (2013)
32. Kreiberg, D., Söderström, T., Yang-Wallentin, F.: Errors-in-variables system identification using structural equation modeling. Automatica, Vol. 66, pp. 218–230 (2016)

Multicriteria 3PL Selection with Risk Considerations

J.C. Osorio[1], D.F. Manotas[1], J.L. García[2]

[1] Universidad del Valle, Escuela de Ingeniería Industrial, Cali, Colombia

[2] Universidad Autónoma de Ciudad Juárez, Departamento de Ingeniería Industrial y Manufactura, Mexico

{juan.osorio, diego.manotas}@correounivalle.edu.co, jorge.garcia@uacj.mx

Abstract. Supplier selection is a critical activity for organizations, because it directly affects their performance and affects the productivity and competitiveness of the firm. It has been widely accepted in academic literature that this problem has a clear strategic nature multi definition. Third Party Logistics Providers (3PL) have been widely promoted by the phenomenon of outsourcing, on which companies increasingly rely. On the other hand, risk management in supply chains (SCRM, Supply Chain Risk Management) is one of the issues that are gaining more strength today. We propose a multicriteria approach using fuzzy analytic Hierarchy Process with TOPSIS for the 3PL supplier selection with risk considerations. The model was validated for a real case in Colombia.

Keywords: Multicriteria approach, supplier selection, third party logistics, supply chain risk management, fuzzy AHP, topsis.

1 Introduction

Supplier selection (SS) has received considerable attention for its significant effect toward successful Logistic and supply chain management (LSCM) [2]. It is a strategic and long-term decision. A long term relationship with suppliers is greatly beneficial to the buyer, and offers advantages such as stability in future plans and strategies, reliability of product quality and service time, better understanding of the supply chain, potential discounts, and superior inventory utilization [7].

Choosing suppliers should not solely depend on their eligibility at the time of the decision-making. It is essential to predict and analyze the past changes and future potentials of suppliers, as it lowers the need to periodically assess and replace existing contracts. The overall market condition has distinctive effects and consequences on different suppliers, and so it is important to assess the relationships of suppliers and their relative changes throughout time, to assure an optimized supply chain [7].

Supply relationship management in supply chains seeks the participation of good suppliers providing low cost and high quality. A recent trend in 21st Century business is outsourcing product manufacturing. With an increase in outsourcing, offshore-sourcing, and electronic business, supply management decisions are becoming

evermore complex in a global market. Supply management strategies such as off-shore-sourcing can emphasize manufacturers at low cost locations such as China, India, or Vietnam, assemblers at hightech operations in Taiwan and Korea, and distributors where customers reside all over the globe. There are increased risks expected from differences in product quality, as well as differences in the probabilities of late delivery [28].

The supplier selection decision in a supply chain does not depend solely on cost or quality measures, but also on various risk and socio-economic factors, usually incorporated as constraints or filters. Many supply chain risks have been identified. Various risks can be the major factors that influence the supplier selection [28].

Vendor selection is of prime importance in the enterprise supply chain management. Modern trend of purchasing management is to reduce the number of vendors, and establish the long-term and stable partnership of mutual trust, benefit and support. These trends aggravate the risk of vendor selection. So enterprises in selecting vendors should measure all aspects of the performance of candidate vendors, not only the product quality, price, vendor performance and services, but also the risk factors of vendor selection [19].

Every supply chain is composed of a network of businesses where each depends on upstream business for supplies and services. A reliability chain exists along supply chains. The reliability of one company in a supply chain depends on the reliability of its predecessor. The reliability of that predecessor depends on the reliability of its own predecessor along the supply chain and so on [16].

The increasing need of the companies to focus on the core business object has generated a trend oriented to outsource different activities. In this context, supply chain activities have evolved from a first stage where we have companies that are responsible for their logistics processes up to the current trend with companies who have delegated all their logistics activities to specialized agents. (Osorio et al, 2015)

Table 1 presents the principals and recent works and technics used by the authors in the multicriteria supplier selection problem. The authors have been used combinations and hybrids tools.

We can mention these papers about multi criteria 3PL selection processes: [18] using fuzzy Delphi, fuzzy inference and fuzzy linear programming. They don´t present the risk like explicit criteria but consider 26 criteria and risk is implicit in many of them. [17] using fuzzy sets and they don´t include risk like criterion. [8] present a model that integrates case-based reasoning, rule-based reasoning and compromise programming techniques in fuzzy environment. Although risks are not explicit in the model, they are considered in the decision. Finally, [6] present an integrated fuzzy QFD and fuzzy AHP. In this case, risk is an explicit criterion.

2 Methodology

We proposed a FAHP integrated with TOPSIS methodology. We obtained the weight of criteria with the FAHP and then apply TOPSIS methodology to 3PL supplier selection according with the criteria. The needs are represented on the weights that Organization experts assigned to each criterion by applying the methodology FAHP.

For the weighting of the criteria It was used a classical Saaty scale but with triangular fuzzy numbers (Table 2). And with the results of the analysis will continue with the implementation of the TOPSIS methodology.

The first step is to define the decision criteria. There are in the literature a lot of criteria, qualitative and quantitative that depends heavily on the organization and its supplier's policies. This methodology was validated in a Colombian company (Urbano y Muñoz, 2013).

Table 1. Principal papers about multi criteria supplier selection with risk considerations.

Approach	Articles
FMCDM (Fuzzy Multicriteria Decision Making)	[10] [9]
AHP and Entropy	[1]
Grey Relational Analysis	[26] [23]
AHP	[13] [16] [15] [14]
Fuzzy TOPSIS	[21] [24]
FAHP	[14] [3] [12]
Multiobjective programming, DEA	[27]
Multiobjective programming and ANP	[20]
Neuronal network y DEA	[22]
Fuzzy multiobjective	[28]
Fuzy TOPSIS and MCGP	[4]
Multiobjective model and AHP	[7]
QFD and linear programming	[5]
QFD and DEA	[11]

Table 2. Fuzzy AHP scale [25].

Linguistic scale for relative importance	Triangular fuzzy scale	Reciprocal of triangular fuzzy scale
Exactly the same	(1,1,1)	(1,1,1)
Same importance	(1/2,1,3/2)	(2/3,1,2)
Slightly important	(1,3/2,2)	(1/2,2/3,1)
Serious importance	(3/2,2,5/2)	(2/5,1/2,2/3)
More serious importance	(2,5/2,3)	(1/3,2/5,1/2)
Absolutely importance	(5/2,3,7/2)	(2/7,1/3,2/5)

This company is one of the largest and the most recognized in the health and beauty business.

It has customers in 15 countries: USA - El Salvador - Panamá - Guatemala - Perú - Ecuador - Costa Rica - México - Venezuela - Chile -Guinea - Aruba - Curazao - Bolivia - Honduras - El Salvador – Nicaragua - República Dominicana.

Its suppliers are locals and internationals. The principal transport mode is terrestrial and they use containers. The problem is to select a 3PL supplier for the container transport.

The volume of containers handled per year in this operation is 93 including both imports and exports.

There are 3 potential suppliers and the criteria are:

- **Structural Alignment.** It covers aspects related to cultural and technological compatibility of the companies involved. Also considers the interest and commitment to build and maintain an alliance. It includes cultural and technological alignment, experience and confidence.
- **Management Aligning.** Includes aspects of the affinity between the management styles of the companies involved, including the quality management system and corporate social responsibility programs and risk management (particularly operational risks). In this criterion we consider the quality, social responsibility and strategies alignment.
- **Financial skills.** Evaluate the aspects of financial position and bargaining power of suppliers. Also considers the evaluation of the costs of the logistics operation.
- **Operational skills.** Includes the activities and skills of logistics operation. It includes aspects such as delivery, service, flexibility and quality in the distribution. In this feature are implicit elements of risk because of the good performance of these duties, the success of the operation load is derived; therefore, poor performance or failures occurring increase the operational risk for the Company.

Of the four criteria presented, you can see that two of them present the risk.

Table 3 presents the criterion qualification for the potential suppliers. We used a 1 to 5 scale where 1 is associated with poor capacity and 5 is the better value.

Table 3. Supplier ratings for the criterion.

3PL suppliers	C1	C2	C3	C4
S 1	4	4.5	4	3.7
S 2	5	4.6	4	4.2
S 3	3	4.7	4	4

3 Results

Applying the proposed methodology, the following weights to the criteria (Table 4) were obtained. These weights are obtained from the FAHP methodology and judgments made by the experts included in the study.

Using these values and applying TOPSIS has the following proposal for the 3PL supplier selection. The table 5 presents the wigthed normalized matrix and the positive-ideal and negative-ideal solution. And finally in the table 6 the choice selection is showed.

With this approach we have incorporated the subjective elements of the experts and the operational risk is incorporated in the decision too.

Table 4. Criterion weighting (FAHP).

C1	C2	C3	C4
20,14%	19.27%	22.79%	37.80%

Table 5. The weighted normalized decision matrix and positive-ideal and negative-ideal solutions.

3PL suppliers	C1	C2	C3	C4
S 1	0.11393	0.10882	0.13158	0.20329
S 2	0.14241	0.11124	0.13158	0.23077
S 3	0.08545	0.11366	0.13158	0.21978
Positive-ideal solution	0.14241	0.11366	0.13158	0.23077
Negative-ideal solution	0.08545	0.10882	0.13158	0.20329

Table 6. Results.

	Ri
S 2	0.9683195
S 1	0.4086512
S 3	0.2659112

4 Conclusions

Although there are still many operational risk associated with supply chain elements, this work presents a first approach to joining one of the most important decisions of today's organizations, the selection of their 3PL supplier and the operational risk management. It is important in these decisions involve judgments of experts, because knowledge acquired over his career he is an invaluable element in decision making.

The 3PL supplier selection was made considering all relevant aspects for the company and according with the significance criteria defined by the experts involved in the decision-making process. Operational risk was included in both explicit like supply chain risk management program and implicit into the company's organizational competencies.

References

1. Ahuja, B.B.: Vendor Selection in Supply Chain Using Relative. Methodology (2010)
2. Chai, J.: Expert Systems with Applications Application of decision-making techniques in supplier selection : A systematic review of literature. Expert Syst. Appl. Vol. 40, No. 10, pp. 3872–3885 (2013)
3. Chan, F.T.S., Kumar, N.: Global supplier development considering risk factors using fuzzy extended AHP-based approach. Omega, Vol. 35, pp. 417–431 (2007)
4. Daneshvar, B.: Supplier selection using integrated fuzzy TOPSIS and MCGP : a case study. Procedia - Soc. Behav. Sci. Vol. 116, pp. 3957–3970 (2014)
5. Hassanzadeh, S., Razmi, J.: Expert Systems with Applications an integrated fuzzy model for supplier management : A case study of ISP selection and evaluation. Expert Syst. Appl., Vol. 36, No. 4, pp. 8639–8648 (2009)
6. Ho, W.: Strategic logistics outsourcing: An integrated QFD and fuzzy AHP approach. Expert Syst. Appl., Vol. 39, No. 12, pp. 10841–10850 (2012)
7. Hosseininasab, A., Ahmadi, A.: Selecting a supplier portfolio with value, development, and risk consideration. Eur. J. Oper. Res. Vol. 245, No. 1, pp. 146–156 (2015)
8. Işıklar, G.: Application of a hybrid intelligent decision support model in logistics outsourcing. Comput. Oper. Res. Vol. 34, pp. 3701–3714 (2007)
9. Kar, M.B.: A Network-TOPSIS Based Fuzzy Decision Support System for Supplier Selection in Risky Supply Chain. 2014 Seventh Int. Jt. Conf. Comput. Sci. Optim., pp. 288–293 (2014)
10. Karsak, E.E., Dursun, M.: Computers & Industrial Engineering An integrated fuzzy MCDM approach for supplier evaluation and selection q. Comput. Ind. Eng., Vol. 82, pp. 82–93 (2015)
11. Karsak, E.E., Dursun, M.: Expert Systems with Applications an integrated supplier selection methodology incorporating QFD and DEA with imprecise data. Expert Syst. Appl., Vol. 41, No. 16, pp. 6995–7004 (2014)
12. Kilincci, O., Onal, S.A.: Expert Systems with Applications Fuzzy AHP approach for supplier selection in a washing machine company. Expert Syst. Appl., Vol. 38, No. 8, pp. 9656–9664 (2011)
13. Kull, T.J., Talluri, S.: A Supply Risk Reduction Model Using Integrated Multicriteria Decision Making. IEEE Trans. Eng. Manag., Vol. 55, No. 3, pp. 409–419 (2008)
14. Lee, A.H.I.: A fuzzy supplier selection model with the consideration of benefits, opportunities, costs and risks. Expert Syst. Appl., Vol. 36, No. 2, pp. 2879–2893 (2009)
15. Levary, R.R.: Ranking foreign suppliers based on supply risk. Supply Chain Manag. An Int. J., Vol. 12, pp. 392–394 (2007)
16. Levary, R.R.: Using the analytic hierarchy process to rank foreign suppliers based on supply risks. Comput. Ind. Eng., Vol. 55, 535–542 (2008)
17. Li, F.: A 3PL supplier selection model based on fuzzy sets. Comput. Oper. Res., Vol. 39, No. 8, pp. 1879–1884 (2012)
18. Liu, H.T., Wang, W.K.: An integrated fuzzy approach for provider evaluation and selection in third-party logistics. Expert Syst. Appl., Vol. 36, No. 3, pp. 4387–4398 (2009)
19. Liu, L.: A conceptual framework for vendor selection based on supply chain risk management from a literature. Vol. 1, No. 3, pp. 1–8 (2011)
20. Malmir, R.: A Multi Stage Decision Making Model to Evaluate Suppliers by Using MOLP and ANP in a Strategic Approach. Vol. 2, Vol. 3, pp. 563–577 (2013)
21. Mehralian, G.: Developing a suitable model for supplier selection based on supply chain risks: An empirical study from Iranian pharmaceutical companies. Iran. J. Pharm. Res., Vol. 11, pp. 209–219 (2012)

22. Nourbakhsh, V.: Considering supply risk for supplier selection using an integrated framework of data envelopment analysis and neural networks. Int. J. Ind. Eng. Comput., Vol. 4, pp. 273–284 (2013)
23. Rajesh, R., Ravi, V.: Supplier selection in resilient supply chains : a grey relational analysis approach. J. Clean. Prod., Vol. 86, pp. 343–359 (2015)
24. Swain, D.P.: Supplier selection in risk consideration: a fuzzy based TOPSIS. Vol. 110, pp. 0–25
25. Tian, J., Yan, Z.F.: Fuzzy Analytic Hierarchy Process for Risk Assessment to General-assembling of Satellite. J. Appl. Res. Technol., 11, August, pp. 568–577 (2013)
26. Viswanadham, N., Samvedi, A.: Multi tier supplier selection for a sustainable global supply chain. IEEE Int. Conf. Autom. Sci. Eng., pp. 492–497 (2013)
27. Wu, D., Olson, D.L.: Supply chain risk, simulation, and vendor selection. Int. J. Prod. Econ., Vol. 114, pp. 646–655 (2008)
28. Wu, D.D.: Fuzzy multi-objective programming for supplier selection and risk modeling: A possibility approach. Eur. J. Oper. Res., Vol. 200, No. 3, pp. 774–787 (2010)

SCHASKOS: A SKOS-Based Tool for Supporting the Decision Making Process in a Supply Chain

Daniel Vázquez-Ramírez, Cristian Aaron Rodríguez-Enríquez,
Giner Alor-Hermández, Cuauhtémoc Sanchez-Ramírez,
María Antonieta Abud-Figueroa, Luis Ángel Reyes-Hernández

Instituto Tecnológico de Orizaba,
División de estudios de posgrado e investigación, Veracruz,
Mexico

danielvazquezrmz@outlook.com, {crodriguezen, l.a.reyes.h}@gmail.com,
{galor, csanchez, mabud}@itorizaba.edu.mx

Abstract. Semantic Web technologies provide innovative means for organizing, describing, and managing digital resources in a wide range of formats. From this perspective, Linked Data has recently gained some attention in the supply chain management field as an effective way to provide and share data across all the value chain. Semantic Web technologies can support the seamless integration of data, on-the-fly composition and interoperation of Web services, as well as more intuitive search engines. This search engines allows the supply chain data exploitation in order to get valuable knowledge. However, a successful implementation and use of Semantic Web technologies requires both information professionals and system developers. In this paper, we propose a Web-based tool called SCHASKOS that covers these needs by working with the W3C's Simple Knowledge Organization Systems (SKOS) by using Linked Data. The SCHASKOS tool is a Linked Data-based tool for non-experience information professionals that allows the SKOS and Linked Data adoption for supply chain knowledge management in organizations. Additionally, a case study of how the SCHASKOS tool works and their benefits are presented.

Keywords: Linked data, semantic web, SKOS, supply chain.

1 Introduction

The knowledge management in organizations has gained momentum during the last decade as many academic; executives and practitioners have recognized that applying knowledge will become the key competitive factor and the most valuable and strategic resource in their companies. Knowledge is the result of a process where an individual makes contact with data found in various kinds of "data sources" such as reports, letters, bills, memos, e-mails, notes and software to mention but a few [1]. Then this individual adds value to these "data sources" to convert data into information, and finally the information translates into knowledge when it is used for improving a process. In other context, supply chain management is an important area of research. Under the context of coordinating the supply chain, several Knowledge Management approaches have

been developed to extent supply chains, rather than single companies, are now seen as interoperable units of a value chain. The present and the future of the Knowledge Management in organizations is the exploitation of computational systems that are able to store, process and transmit knowledge from one unit to another in the value chain, in order to daily facilitate operations within an organization, and especially timely detect problems, and possible solutions. From this perspective, SKOS is a family of formal languages designed to represent thesauri, taxonomies, subject-heading systems, or any other type of structures of controlled vocabulary for the Semantic Web. These formal languages allow improving the Knowledge Management in a powerful semantic way. Taking into consideration that knowledge is an effective way to improve processes in organizations; in this paper we explore the use of SKOS for knowledge management of a supply chain.

This paper is organized as follows: Section 2 introduces a literature-review of tools for SKOS visualization, SKOS vocabularies transformations, and other related works with knowledge management in supply chains. The SCHASKOS architecture is presented in Section 3. In Section 4, a case study is presented. Finally, conclusions and future work are presented in Section 5.

2 State of the Art

Nowadays, they are different knowledge domains has been used SKOS as a mechanism for knowledge management in several research fields such as geography [2-3] and medical sciences [5], to mention but a few. However, in supply chain management field, the use of SKOS for knowledge management has not been reported before. In this section, we describe several related works to SKOS, these works were selected according to the following classifications: 1) visualization tools and 2) vocabularies transformations. Also, a review of related works with the knowledge management in supply chain is presented.

There are several SKOS visualization tools reported in the literature, for instance GeoNetwork [2] and ThManager [3], these tools have been used in the geography domain. In the GeoNetwork and ThManager research works, a multi-platform desktop and open source application were presented. GeoNetwork allows importing a thesauri in SKOS-based format and locate geospatial data through an interactive map. GeoNetwork support ISO standards for thesaurus creation such as: ISO 19115, ISO 19119 and ISO 19110 by using the ISO 19139, FGDC and Dublin Core.

In [4], ThManager was used to transform the Thesaurus for French Local Archives to SKOS-based format. Moreover, with ThManager an application for using the Thesaurus for French Local Archives was developed.

In [5] the authors presented a Web-based tool for the medical domain that allows the graphical visualization of the concepts contained in a SKOS vocabulary, and the different relationships among labels of each concept.

Vocbench [6] is a Web-based tool that supports an aquatic science vocabulary and a fishing thesaurus. Vocbench used the SPARQL query language and it was used to manage the AGROVOC thesauri.

SISSVoc [7] is a Linked Data API that allows the access to published vocabularies, which use the SKOS-based format. SISSVoc is used in a large number of projects, which mainly belongs to environmental sciences.

SKOS Play [8] is an open source web-based tool that allows the visualization of thesaurus, taxonomies, and controlled vocabularies expressed in SKOS. Another interesting tool for vocabulary management is Pool Party, a commercial Web-based tool that allows publishing thesaurus as Linked Open Data. The latest Pool Party version allows performing a vocabulary quality report through "The PoolParty SKOS Quality Checker". The quality checker is based on qSKOS (quality simple Knowledge Organization System).

TopBraid Enterprise Vocabulary Net [9] is a commercial Web-based tool that allows managing business vocabularies, which is mainly developed to business domain. In addition, TopBraid Enterprise Vocabulary Net allows to organizations the possibility of creating and editing ontologies or SKOS-based business vocabularies.

With regard to the knowledge management in supply chains, a literature review to supply chain is presented in [10], the objective of this review was to provide a taxonomy developed about supply chain management field used to assist in the search ranking in the supply chain field in order to provide a framework for categorizing related literature in supply chain management according to two criteria: content-oriented (level of analysis and element of the exchange within the studied network) and methodology-oriented (theoretical, empirical, prescriptive and descriptive).

In [11] the authors conducted a study of knowledge management in supply chain networks. Additionally, a knowledge management in supply chain conceptual model was proposed as an extension of Nonaka and Takeuchi model.

In [12] a literature review of supply chain was presented to assess the relationship between knowledge management and supply chain management. The literature review showed that the knowledge management is considered a tool for the supply chain integration. Additionally, the lack of ITs (Information Technologies) integration to improve knowledge sharing among supply chain was discussed.

In [13] a scientific framework to solutions adoption of knowledge management in a supply chain was presented. This framework used a hybrid multi criteria technique, which combined fuzzy Analytical Hierarchy Process (AHP) and fuzzy Technique for Order Performance by Similarity to Ideal Solution (TOPSIS).

In [14] a platform for knowledge sharing in a supply chain by using semantic Web technologies was proposed. The platform involves a semi-structured model to represent knowledge, an agent-based annotation process to resolve issues associated with the heterogeneity of knowledge documents. Additionally, an articulation mechanism to improve the effectiveness of interoperability by using two ontologies was presented.

In [15] a classification of knowledge in a supply chain was presented, identifying various kinds of knowledge and their definition. The classification was developed in five categories depending on the stages in a supply chain: (1) planning, (2) production, (3) storage, (4) delivery, and (5) transportation. As part of the research a case study of a food supply chain was defined. In Table 1, a summary about the state of art is presented. We have considered if the work is a SKOS-based tool (SKOS), Linked Data (LD) is used, and finally, if the research work was developed for Supply Chain Management (SCM).

Table 1. State of the art highlights.

Research Work	Feature		
	SKOS	Linked Data	Supply Chain Management
Open Source Geospatial Foundation [2]	Y	N	N
Lacasta Javier, et al [3]	Y	N	N
de Grimoüard-Sibille [4]	Y	N	N
Conway Mike, et al [5]	Y	N	N
Agricultural Information Management Standards [6]	Y	Y	N
JD Cox Simon [7]	Y	Y	N
Francart Thomas [8]	Y	N	N
Schandl Thomas, et al	Y	N	N
Morshed Ahsan, et al [9]	Y	N	N
Croom Simon, et al [10]	N	N/A	Y
Samuel Karine Evrard, et al [11]	N	N	Y
Marra Marianna, et al [12]	N	N	Y
Patil Sachin K, et al [13]	N	N	Y
Huang Chun-Che, et al [14]	N	N	Y
Almuiet Mohammad Zayed, et al [15]	N	N	Y

3 SCHASKOS Architecture

Our Knowledge Organization System called SCHASKOS has a layered design in order to organize its components. This layered design allows scalability, easy maintenance and multi-device presentation capabilities, because its task and responsibilities are distributed.

3.1 Architecture Description

The general architecture is shown in Figure 1. Each component of SCHASKOS has a function explained as follows:

Data layer: This layer stores concepts (SKOS), social network data; data source URIs for Linked Data and non-structured data (files and raw data) for knowledge management. This layer additionally contains all the configuration tables allowing the operation of the modules and services offered by SCHASKOS.

Data Access layer: This layer communicates with the *Data Access layer* and it contains the representation on flat objects of each relationship of the database management system (for relational data), a XML-based transformation schema for both non-structured data and social network data; and finally a SKOS data manager.

Core layer: This layer is responsible for interpreting data from the *Data Access layer* through the Linked Data engine and SKOS engine, and it builds a XML-based representation. This layer additionally allows the data knowledge management and its representation.
Service layer: This layer provides a set of services (modules) offered by SCHASKOS. These services represent different functions such as visualization, analysis, search, storage and data editing. This layer communicates with the data layer to obtain the required content at any time. In this layer, a module is located responsible for the representation of diagrams with different concepts in SKOS-based format obtained.
Integration layer: This layer allows the creation of informative diagrams, request management, and it determines the way to get the information and build the diagram requested by the user in the presentation layer.
Presentation layer: In this layer, SCHASKOS determines the diagram to be presented to the final user. This layer represents the view that the user has of the system by using HTML5 and CSS3. The presentation layer unknown events taken by SCHASKOS, it only gets the information and required diagrams to display to the end user in a graphical user interface.

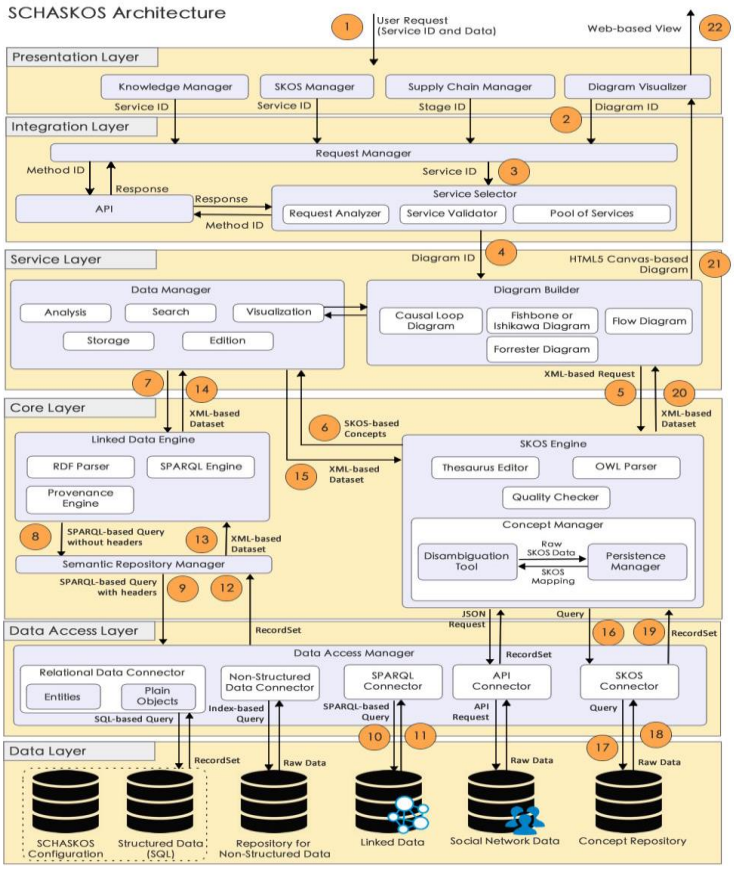


Fig. 1. SCHASKOS architecture.

3.2 Component Description

In SCHASKOS architecture, each component has a specific function that is explained in Table 2.

Table 2. Component Description.

Component	Functionality
Linked Data	This component represents the Linked Data cloud, which contains information available on various domain fields.
Concept Repository	This component is an extensive vocabulary related to the supply chain management.
SPARQL Connector	This component allows performing SPARQL-based queries over the Linked Data cloud by using a SPARQL endpoint.
SKOS Connector	The SKOS connector allows the access to the concept repository, which is requested by the SKOS Engine module for related concepts.
Linked Data Engine	This module contains a component to analyze RDF data because SKOS is RDF-based. Additionally, a SPARQL Engine is used to analyze the data provenance.
SKOS Engine	This component provides four sub-components, these components includes: 1) The concept manager module, 2) The disambiguation tool that identifies the most appropriate meaning among the concepts, 3) The persistence manager that is responsible of storing the knowledge in SKOS-based format according the information obtained in a XML-based file, and 4) The SKOS Mapping provides tags used to indicate the relation degree among the builder concepts and previously stored concepts by the persistence manager.
Data Manager	This module contains different operations that the system executes with the information provided by the user, such as edit, search, analyze, store and display.
Diagram Builder	This module allows building diagrams to analyze related information in a supply chain, in order to find problems and the expert takes a decision. SCHASKOS allows presenting the following diagrams to the final user: 1) Causal Loop Diagram, shows the structure and relationships between different variables in a system. This kind of diagram allows knowing the reasons why a variable can affect another, 2) Fishbone or Ishikawa Diagram, proposed by dr. Kaorou Ishikawa, allows graphically showing the causes of a problem, allowing a thorough analysis, and 3) Forrester Diagram, created by Jay Forrester, is used to observe the evolution of the variables through various equations that allows validate the model obtained in the causal loop diagram.
Request Manager	This module manages the requests made by the final user.
Request Analyzer	This component identifies the entire user's request.
Service Validator	This component, validate that the request service is valid and available for the user.
Pool of services	This component includes all services of the SCHASKOS tool.

Component	Functionality
API	This component allows the access to the methods available in SCHASKOS through a public interface.
Diagram visualizer	This module allows the user to view the built diagram by SCHASKOS.

3.3 Workflow Description

The workflow describes the functionality of SCHASKOS's architecture. The workflow is explained as follows:

(1) the user makes a request to SCHASKOS using a Web browser by indicating a URL,

(2) the *Presentation Layer* identifies the user's request and redirects the HHTP-based request to the *Integration Layer* where the HTTP-based request is analyzed by the *Request Manager Component* to determine what type of service is being requested,

(3) the Integration Layer uses the *Service Selector Component* to determine whether the HTTP request is valid. If a request for a valid service has been requested, the *Service Selector Component* is prepared to invoke the appropriate service from the *Service Layer*,

(4) if the service is valid and available, a query is carried out to recover the service identifier in the *Pool of Services Component* provided by SCHASKOS. In the event that the service is not available or is invalid, SCHASKOS returns an error message with the failure reasons and a list of available services,

(5) if the requested service is the generation of a Diagram, the request is redirected to the *Diagram Builder Component* in the Service Layer where it performs an XML-based request to the *SKOS Engine Component* through the *Core Layer*,

(6) the *SKOS Engine* transform the XML-based request in SKOS format, and redirects the request to the *Data Manager Component* in the Service layer in order to store the data,

(7) the concepts are directed to *Linked Data Engine component* in the Core Layer where its origin is identified by using the *Engine Provenance Component*,

(8) the Core Layer uses the *SPARQL Engine component* available in the *Linked Data Engine module* to build and run a query without using headers to find information in the semantic repository, without indicating a source of information,

(9) the *semantic Repository Manager module* builds a SPARQL-based query by using headers, it is sent to the *Data Access Manage*r through the Data Access Layer,

(10) the Data Access Layer uses the *SPARQL Connector component* to execute the SPARQL-based query built on the Core Layer on the Linked Data Cloud to locate information from a particular source, (10) if the query returns values, they are sent back to the *SPARQL Connector component* available from the Data Access Layer,

(11) the Data Access Layer stores the results in a Record Set to be sent to the *Semantic Repository Manager component*,

(12) the *Semantic Repository Manager* transforms the result stored in the Recordset to XML-based format to be sent to the *Linked Data Engine module*,

(13) the *Service Layer* uses the *Data Manager* to store the result set in XML-based format sent by the *Linked Data Engine*,

(14) the XML-based data set is sent to *SKOS Engine module* to be transformed to SKOS-based format,
(15) a query is executed to be sent to *SKOS connector module* available in the Data Access layer for obtained information of the concept repository,
(16) the query is executed on the *repository concepts* to compare the information obtained with the concepts related to supply chain,
(17) the data obtained are sent to the *SKOS Connector module* to be stored in a Recordset,
(18) the Recordset with the information obtained is sent to *SKOS Engine module* to be stored in SKOS-based format,
(19) in order to build the diagram requested by the user, it is necessary to transform the information SKOS to XML-based format, the dataset in XML-based format is sent to the service layer to build the diagram,
(20) the *Diagram Builder module* is responsible for building the diagram applied to all information obtained using HTML5 Canvas, which is sent to the presentation layer for display to the user. The diagram constructed is displayed to the user using a Web browser, at this point; the user has the opportunity to make decisions.

4 Case Study: Generating a Ishikawa Diagram to Help the User in the Decision Making Process

SCHASKOS is a Linked Data-based tool that allows knowledge management in organizations and generating Ishikawa, causal and Forrester diagrams by adopting Linked Data and SKOS to facilitate information management. The main idea of SCHASKOS is to help users better manage their information using the labels provided by SKOS. SCHASKOS transforms this information into a diagram for identifying changes and possible problems occurring within the organization and so, allows the user to make decisions in order to improve the performance of a supply chain. To explain SCHASKOS functionality, the following case study is presented for generating a diagram (Ishikawa or Fishbone, causal and Forrester) in order to help in the decision-making process. Let us suppose that an organization dedicated to dairy milk production will keep management for their full supply chain. In this context, there is a constraint: if the administrator of the supply chain is a non-experienced computer user, and therefore does not have the required skills to develop an Ishikawa, causal or Forrester diagram. As a solution to this issue, SCHASKOS provides a set of graphic interfaces that allow any user capture and transform information into a diagram by using SCHASKOS capabilities. To do this, the user must be follow the next steps: (1) the user accesses SCHASKOS via an authentication mechanism by entering a login name and password, (2) once the user has been authenticated, the user selects the option to create a new project and entering a name for the project, (3) The next step is to capture the information of each link in the supply chain, since suppliers have supplied to the organization, changes in raw materials or the method for making products until the reasons for a product return, (4) Once the user has captured all the available information, the next step is to transform it into a diagram to help user in the data analysis The user select the option “Diagram builder” and choose the diagram type, (4) SCHASKOS transform the information obtained in the Linked Data Cloud and the

concept repository in a diagram. The diagram generated is available in the application interface (this is depicted in Figure 2), and (5) once SCHASKOS has finished the building diagram, the user can view details selecting a cause in the case of Ishikawa diagram. This process is depicted in Figures 2 and 3.

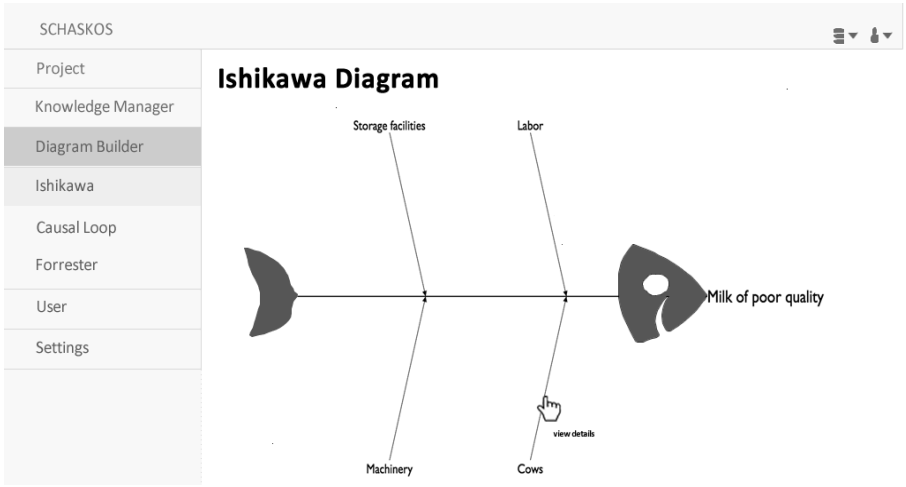


Fig. 2. Ishikawa Diagram generated for SCHASKOS.

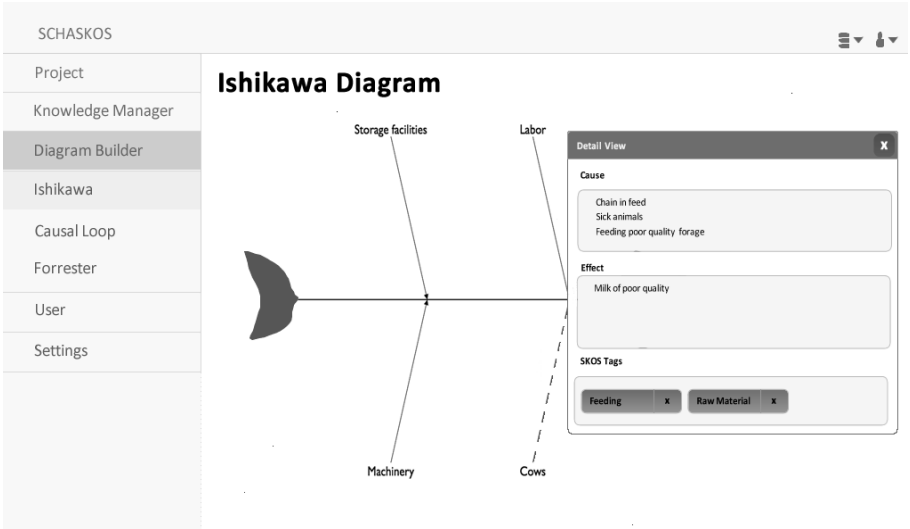


Fig. 3. Details of a cause in the Ishikawa diagram shown by SCHASKOS.

As shown in Figures 2 and 3, the Ishikawa diagram visualization tool provided by SCHASKOS presents a detail view modal window when clicking in a category of causes of the "Milk of poor quality" event. The Figure 3 shows in a dashed line, the selected category. In the modal window, different causes related with the "Cows" are shown. In this regard a side effect of feeding cows whit poor quality forage insides in Milk quality.

5 Conclusions and Future Work

The knowledge has become a strategic asset for organizations, due to this; organizations must invest in developing the best strategy for applying the knowledge assets in their organizations. From this perspective, a knowledge organization system allows to organizations to manage data, and transform these data to business knowledge. In this context, SCHASKOS transform the information for the generation and development of Ishikawa, Causal or Forrester diagram to end-users, in order to provide a graphic tool for helping on the decision-making process. Finally, with the use of SKOS, the decision-making process can be improved with knowledge sharing and management in supply chains. As future work, we have considered to analyze the backward effects of decision making across the supply chain. Additionally, we propose to include the ability for connecting SCHASKOS with external data sources to extract the information and build other kind of diagrams. At the same time, a supply chain ontology in the tool will be developed in order to infer new knowledge that can be useful on the decision-making process.

References

1. Hafeez, K., Rodiguez-Falcon, E., Abdelmeguid, H., Malak, N., Training, P.: Knowledge Management in Supply Chains. In: ICSTM. (2000)
2. https://climate.iarc.uaf.edu/geonetwork/docs/eng/users/admin/index.html.
3. Lacasta, J., Nogueras-Iso, J., López-Pellicer, F.J., Muro-Medrano, P.R., Zarazaga-Soria, F.J.: ThManager: An open source tool for creating and visualizing SKOS. Information Technology and Libraries, Vol. 26, pp. 39–51 (2013)
4. Sibille-de Grimoüard, C.: The Thesaurus for French Local Archives and the Semantic Web. Procedia-Social and Behavioral Sciences, 147, pp. 206–212 (2014)
5. http://www.zora.uzh.ch/98982/22/Developing%20a%20Web%E2%80%90based%20SKOS%20Editor.pdf.
6. http://aims.fao.org/vest-registry/tools/vocbench-2.
7. Cox, S., Yu, J., Rankine, T.: SISSVoc: A Linked Data API for access to SKOS vocabularies. Semant, Web (2014)
8. http://labs.sparna.fr/skos-play/about.
9. Morshed, A., Dutta, R.: Machine Learning Based Vocabulary Management Tool Assessment for the Linked Open Data. International Journal of Computer Applications, Vol. 60, (2012)
10. Croom, S., Romano, P., Giannakis, M.: Supply chain management: an analytical framework for critical literature review. European journal of purchasing & supply management, Vol. 6, pp. 67–83 (2000)
11. Samuel, K.E., Goury, M.-L., Gunasekaran, A., Spalanzani, A.: Knowledge management in supply chain: An empirical study from France. The Journal of Strategic Information Systems, Vol. 20, pp. 283–306 (2011)
12. Marra, M., Ho, W., Edwards, J.S.: Supply chain knowledge management: A literature review. Expert Systems with Applications, Vol. 39, pp. 6103–6110 (2012)
13. Patil, S.K., Kant, R.: A fuzzy AHP-TOPSIS framework for ranking the solutions of Knowledge Management adoption in Supply Chain to overcome its barriers. Expert Systems with Applications, Vol. 41, pp. 679–693 (2014)
14. Huang, C.-C., Lin, S.-H.: Sharing knowledge in a supply chain using the semantic web. Expert Systems with Applications, Vol. 37, pp. 3145–3161 (2010)
15. Almuiet, M.Z., Salim, J.: Knowledge flow in supply chain manufacturing: Case study in food manufacturing firm. Procedia Technology, Vol. 11, pp. 463–470 (2013)

Functional Tester Runner Tool

Leticia Dávila-Nicanor[1], Aldo Benhumea-Peña[1],
Carlos Omar González-Morán[1], Pedro Mejía-Álvarez[2]

[1] Centro Universitario UAEM Valle de México, CDMX, Mexico

[2] Centro de Investigación y de Estudios Avanzados del IPN, CDMX, Mexico

{ldavilan, cogonzalezm} @uaemex.mx, pmalvarez@cs.cinvestav.mx

Abstract. The sheer complexity of Web applications and inadequate approaches to analyze, design or update such projects generate incorrect environments, which are the source of potential risks. Poor applications reflect low availability, maintenance difficult and questionable efficiency in the services they provide. At present the quality assurances schemes still performed manually. A group of evaluators are based on a test matrix to run tests, and thus according to the degree of experience is that the results are interpreted by each evaluator. This process requires much time and effort, since the goal is to test the system thoroughly to identify as many potential errors in the system. This proposal focuses on the development of an assessment tool with the ability to emulate a group of virtual testers in real contexts of tests for applications on the Internet. This system generates controlled testing using virtual evaluators that emulate the activity of people atmosphere. The results have been satisfactory, currently we worked on the proposal to create a factory abstract of test cases to extend the functionality tool.

Keywords: Testing tool, web applications, reliability.

1 Introduction

Software systems are exposed to significant changes during its development, maintenance and evolution. These affect system functionality and quality you must possess a productive area. The central problem is to have functionality changes that impact directly on the software architecture and design. Functional and non-functional requirements set out in the statement of the purpose of developing a software system. Regarding functional requirements in architecture and software components of the system it is designed. When a system is undergoing maintenance and even evolving relationship has clear requirements with system design is central because it's the basis to extend or upgrade the functions of a system. However, it is up to the test phase when the functionality of the system is evaluated.

This situation is much more noticeable on Internet system, in this case the number of users to whom it may concern, emerging technologies, processes and services that cater as in the case of banking portals and business sites, now that provides a special

relevance. The complexity of applications and inadequate approaches to analyze, design or update such projects generate incorrect environments which are the source of potential risks. Poor architectures reflect low availability, maintenance difficult and questionable efficiency in the services they provide.

Companies that develop software development processes have been used systematically highlight the need for and select tools that benefit the operating costs in software testing. Thus, they have useful and efficient tools, according to their contexts of operation is a necessity that the software project leaders and administrators argue from the last decade of the twentieth century [1]. The interest of companies and organizations that develop and test software systems are focused on reducing production costs for software testing. These costs are assessed on the duration of the evaluation process; the effort hours/man is implicit in the test coverage and computing resources consumed. Thus, assessment tools have important expectations such as tangible improvements in test coverage and implementation of a greater number of test cases at lower cost. Productivity and quality thresholds are dictated by the policies of the company and the project managers, some companies require very high thresholds. Today a lot of organizations perform evaluations of their systems manually, with a group of people called (QA) that are based an array of tests to run tests manually and functionality according to the degree of experience is that the results are interpreted by each evaluator. This process is very expensive because the goal is to test the system thoroughly, to locate as many potential errors in the system.

According to studies Miranda and Jelinski the trend's still exponentially [2]. So, to test a system it is requiring thousands of evaluations. In studies that have been conducted [3] and [4], 100 tests functionality of an application 1200 lines of code carried by a single evaluator lasts about 30 days, in the embodiment of 5000 evaluations it would take approximately 1500 days, namely 4.10 years, this time lapse exceeds any economic development, human resources and the projected time in any case. If we combine this fact that having all these assessments don't guarantee that a software system operate without failure, the problem is even more critical.

This proposal is based on the development of efficient software with the ability to emulate a group of virtual testers in real contexts of reliability testing for Internet applications. In developing the tool they have combined several techniques to make assessment process reliability for Internet system. In this case turned on a group of virtual evaluators is emulated; for this process elements and statistical simulation were taken for analysis of the activity of each of the evaluators elements of analysis used by the scheme compilers high-level languages.

The paper is organized as follows; in section 2 of the theoretical framework and related work is discussed. Section 3 operation executing functional test is described. Section 4 architecture and the main executor of test algorithm is presented and finally in Section 5 conclusions and future work are described.

2 Related Work

Web applications possess unique characteristics that make web testing and quality assurance different from the corresponding traditional techniques. Web applications can be characterized by the following aspects [5]. Massive Access of users, this

simultaneous access of the users in these applications is part of the essence of systems. Web applications provide cross-platform universal access to web resources for the massive user population. For the users it should be transparent that these web applications provide this service to millions of other users. Difficulty of establishing causes of the errors. Since web applications may be accessed by millions of users, errors have a big impact. Finding the origin of errors in web Applications may be difficult and its recovery time may not be immediate, given the great number of software elements that intervene.

The integration of diverse software elements for an application on the Internet, Web users employ different hardware equipment's, network connections, operating systems, middleware and web server support. In a web application, two main components are always required: the backend and the frontend. The backend is the software required for an application on the Internet to operate. Among the most important software found in the backend are: the database servers (MySQL, Oracle, Informix, DB2, among those most important), Web Servers (Apache, Netscape Enterprise Server, Netra of Sun, etc.), and the interface programming languages (HTML, XML, PHP, Servlets-Java, Live - Wire, etc.). The frontend is the software required on the part of the client to allow the systems to access the Web. Among the most important software found in the frontend are: Navigators (Explorer, Netscape), which contain plug-in software such as presentations of Macromedia, and languages like JavaScript.

Diversity of frameworks are develop to operate and to maintain a Web site. The development of a Web site requires of a great team of people with different profiles and backgrounds. These teams include programmers, graphic designers, and usability engineers, specialists in information integration, network experts and database administrators. This diversity of personnel profiles makes reliable web applications development difficult and sometimes unpredictable. Because of the above characteristics web-based systems tend to evolve rapidly and undergo frequent modifications, due to new technological and commercial opportunities, as well as feedback from the users. In consequence web-based systems are very sensitive to errors. Most work on web applications has involved making them more powerful, but relatively little has been done to ensure its quality. The most important quality attributes demanded by web-based systems are reliability, usability and security. Additional important quality attributes are availability, scalability, maintainability and time-to-market [6].

2.1 Automated Assessment Tools for Web Systems

In the article by Robert M. Poston and Michael P. Sexton [1] an entire study on the needs and requirements that must cover the evaluation tools from the point of view of managers and project leaders who develop software business addresses. Until then there wasn't clarity on the needs assessment tools that were developed, misinformation predominated and tools hitherto existing lacked related to the sector which evaluates software goals, this work focuses on developing a summary precisely these needs. Table 1 shows a product form requirements evaluators sectors in companies that develop software.

The companies generally want to see greater development productivity and software quality as a result of the incorporation of new assessment tools. The concern is the

potential for improving productivity and quality of the evidence, whether a tool can have tangible and substantial improvements. Productivity and quality thresholds are dictated by the policies of the company and the project managers. Some companies require very high thresholds.

Table 1. Form needs related to the testing process [1].

Test-productivity (cost) data	Present data from recent Project	Predicted data if new tools are not acquired data
Predicted cost of testing in staff months		
Predicted cost for all testing		
Predicted cost per testing activity		
Planning testing		
Defining test objectives		
Designing tests		
Constructing test environments		
Executing tests		
Evaluating tests and software		
Test-quality data		
Test coverage		
Requirements coverage		
Input coverage (valid, invalid, etc.)		
Output coverage		
Structure coverage (DU path, branch, etc.)		
Predicted data if new tools are not acquired		

Since 2000 year, companies such as IBM. Hewlett Packard and Rational Software, has been concerned about the need to develop tools that can be used in the evaluation process of software development, the latter being an international economic activity, it is very important that the software products operate with high standards [7]. However the quality of an Internet system according to Jeff Tian studies [6] has six aspects: Reliability; Security; Usability; Availability; Maintainability and Scalability. Instruments or tools that are automate the evaluation of this process with different approaches. Below is a comparison chart is presented, the most important tools available today.

Table 2 shows some of the most important characteristics of the instruments most comprehensive evaluation exist in this case most licensing are summarized; such is the case of Web Link [8], Mercury LoadRunner [9], Rational Functional Tester [10] and WAPT [11]. There are other less robust instruments that do not take licenses for its implementation [12], [13], such as: Apache JMeter (evaluates efficiency) Curl-loader

(evaluates efficiency), Selenium (simulates multiple access). Cloud-based tools: Blitz (evaluates performance), Testize (evaluates usability).

Table 2. Comparative table of tools that automate the testing process.

Name	Description	Quality attribute	Advantages	Disadvantages
Web Link (Rel software)	It is a software which is responsible for checking links and web addresses.	Availability.	– Check links – Has spell check – Make content validation – Send emails reports	Costs: 3000 bonds $145 to $1,195 (limited links).
Mercury LoadRunner (Hewlett Packard)	Testing tool for software application performance.	Efficiency.	– Emulates a lot of users interacting with a specific application at the same time. – It can measure response times of processes.	Costs
Rational Functional Tester(IBM)	Tool for functional testing and automated regression.	Functionality.	– Automated testing. – Tests based on data. – Test Script (manual sequences).	$7,017.20
WAPT (SoftLogica, 2014)	It is a tool that evaluates the performance of web applications and interfaces related by generating virtual evaluators.	Efficiency.	– Supports load specified users at all times.	Users have a Static activity.

3 Functional Tester Runner Tool Architecture

The Functional Tester Runner (FRT) tool has been based on the development of a software system with the ability to emulate a group of virtual testers in real contexts of reliability testing for Internet applications. The property to be evaluated is *Reliability*, which is defined as the *probability that a free operating system failures within a specified framework of time and under specific environmental conditions* [14]. This group of virtual testers perform tests and obtain metrics on a defined Web application in a defined time frame. The group´s activity is based on a black box testing technique,

this is combined with statistical simulation [15] and analysis of the activity of each of the evaluators used the analysis approach language compilers high level.

When designing our system architecture (see Figure 1), it can be seen that control passes from the *Arrival test evaluator* to the *Process test state*, control then forks to two concurrent flows, the enclosing object will be in the *Web application* state and the *Testing activity state*. Furthermore, while in the *Testing activity state*, the enclosing object will be in the *Test path* and *URL_conexion* state. The *analyzer* reads the Web server file answer and calls to the parser program to make a new question related with the test case. The *analyzer* instance reads the response from the Web server and calls the *parser* instance to make a new request that is related to the test case. This focus is important in the information upgrade, where the select and update activity are involved more than one form. Finally, a report suite based on programing shell, reads the log files and make the final report. This report included the metric result. The tool nowadays to make a defect density report. The defect density is a software metric.

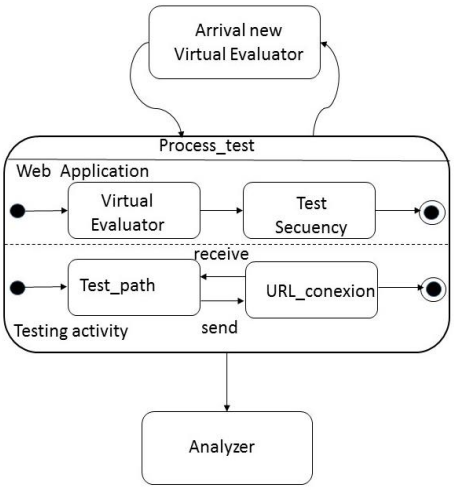


Fig. 1. System architecture to run functional tests.

Fig. 2 shows the sequence diagram for the module arrivals. In the instance initialization *(Init)*, data structures, statistical counters and the first arrival is scheduled initialized. The *new_test_application* instance aims to estimate the time of arrival of virtual evaluators check whether waiting in line at the pool of servers and if the system has capacity to serve more requests. If the architecture has the required capacity, an event access to the virtual evaluator with prior knowledge to assess the application is activated. The *process_test* instance contains the main process where a call is made to *Departure_pro()* method in turn causes a flame to its corresponding instance. In the instance Departure user activity it is established, taking into account its assessment coverage. In Statistics, the results of the evaluation process previously conducted by virtual testers and statistics are generated are stored. The simulation conditions are as follows:

– The time of arrival for evaluators to the system is determined by an exponential distribution $\mu = 5$.

- Server queues *M/M/1* type in the pool of Web servers were used.
- To determine whether the user expects the server's attention in the queue, a probability of *4/(n+1)*, where n represents the current queue size (number of users on the system) was used.
- For reasons of reliability in the operation of the system can work properly even with evaluators *k* (where *k* = 3400).
- Depending on the application and coverage of evaluation, evaluators can perform only *x* types of transactions. The operating time depends on the evaluators related transactions allocated to each evaluator coverage.

The activity of the evaluation process that makes each evaluator as shown in Fig. 3 by a sequence diagram component test process. When they arrive the virtual evaluators them is assigned a test case according to the coverage assessment by *Secuency_path* instance, the script is formed into a loop to form a list of nodes associated with the application components to evaluate this process done with the interplay between *Secuency_path* and *Component* instances. When the virtual evaluator executes a sequence of each node evaluation by *test_path* instance, the *send_receive* method sends requests to the server where the application and evaluated the responses are analyzed by the Analyzer component is housed.

Analyzer component in the activity analysis of the results obtained in the evaluation process by the sequence diagram is developed (see Fig. 4). In this case the results obtained by the virtual evaluators analyzed through *test_case_analysis* instance and its *verify_answer_test* method sends the requested response. The *scanner_html* instance examines the contents of the response and using the *parser_test_case* instance *make_next_test* method and the following request is built

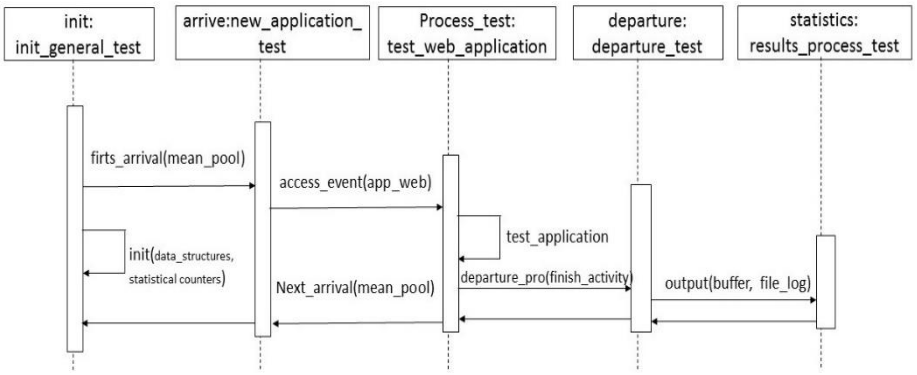


Fig. 2. Diagram for component sequenced arrivals.

This tool was used to evaluate the reliability attribute under real conditions. Evaluation was conducted by our testing tool FTR using concurrent test threads (system testers). Each test thread is responsible for executing a specific functionality test specified on its corresponding test case file. Test cases are generated randomly and test data is prepared to perform functionality tests. For each test case, a specific test profile indicates a specific path of navigation (type of test) and the view that the tester will test. The test thread has access to test cases (which contain the test data and the test profile) and the activity log files. The activity log files are files that contain the activities

performed by the test thread. The analyzer reads the activity log files and produces an error log file which contains the specific faults detected and the defect density computed.

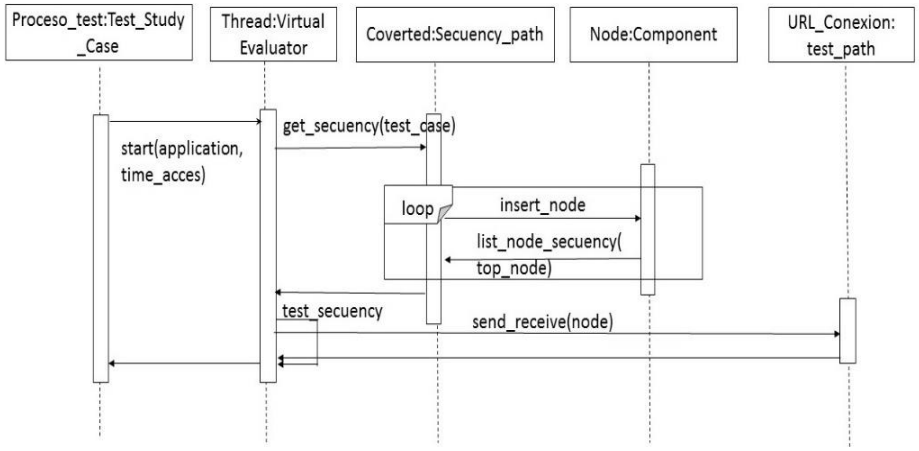


Fig. 3. Diagram for evaluation process.

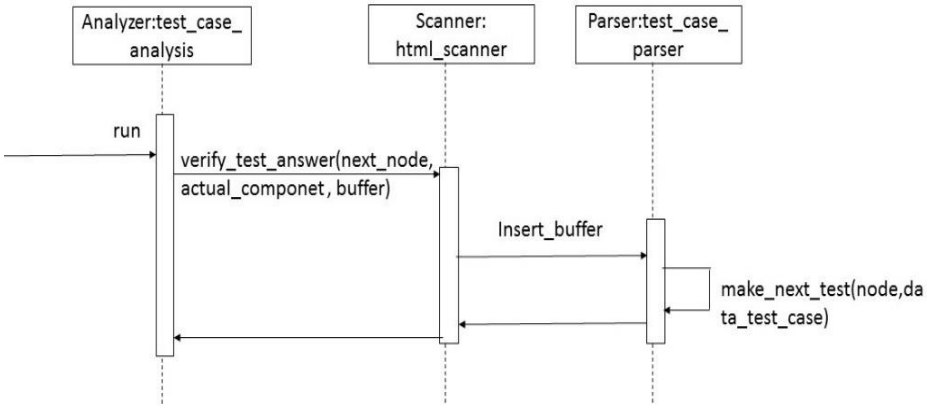


Fig. 4. Flow chart of activity Analyzer.

4 FTR Tool Implementation Results

The functional tester runner is implemented in Java programming language and has been used as an instrument of evaluation and analysis of the reliability of the systems on the Internet [3]. This tool was used to evaluate the reliability attribute on SOGU system on the publications Reliability improvement with PSP of Web-based software application [4], with the ability to emulate a group of virtual testers in real contexts of reliability testing for Internet applications. The process evaluation was conducted by our testing tool FTR using concurrent test threads (system testers). Each test thread is responsible for executing a specific functionality test specified on its corresponding test

case file. Test cases are generated randomly and test data is prepared to perform functionality tests. On each test case a specific test profile indicates a specific path of navigation (type of test) and the view that the tester will test. The test thread have access to test cases (which contain the test data and the test profile) and the activity log files. The activity log files are files that contain the activities performed by the test thread.

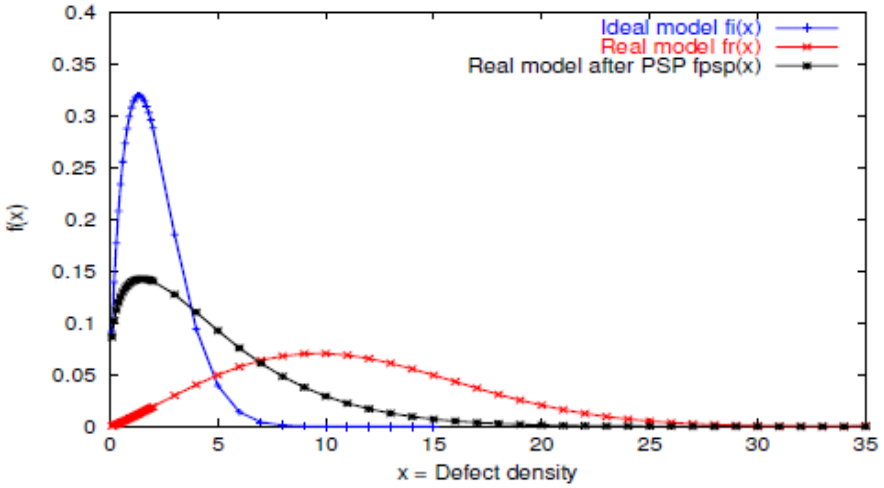


Fig. 5. Graph of SOGU system assessment before and after applying PSP - SEI CMM.

The graph in Figure 5 reports the results of the research [4] where the SOGU system was evaluated before and after applying the Personal Software Process SEI-CMM [16] and [17]. The evaluation was performed by the tool described here. In this case the model $f_r(x)$ describes the evaluation of the system before applying PSP and $f_{psp}(x)$ describes the evaluation system after applying PSP.

5 Conclusions

In conclusion we can say that today software development is an economic activity, where quality assurance is an imperative; however, the existing assessment processes are still expensive. That's why, it's appropriate to develop an instrument to support and contribute to reducing operating costs in software systems in organizations.

In this way, our research shows the development and implementation of a tool to assess the reliability of Web applications. This tool simulates a virtual group of evaluators who perform the testing process according to a scheduled coverage of specific cases. The combination of statistical simulation techniques to the process of assessing the reliability of the software is an effective scheme respect of the testing process serving virtual testers, control of variables such as weather and operating conditions can be handled efficiently in the context of concurrent processes.

For purposes of our research was relevant implement an automated assessment tool, because the approximate time 5000 for manual testing required approximately 4.16 years ; accordingly , such an approach far exceeds available resources in any organization. To evaluate system SOGU's reliability, were performed 100 tests, with a

single virtual evaluator, we need 30 days with the manual process. With the FTR tool operating on CISC architecture in a Red Hat Linux 7.0 platform only they took 5 hours to do 500 tests. In environment with a RISC processor and Solaris 2.7 platform with only they took 5 hours to do 5000 tests [3, 4].

In general, it is projected that the FTR can compete with tools developed by private companies such as Rational Functional Tester or Mercury Load Runner, but without restriction involving licensing; since the FTR is provided with mechanisms that improve its efficiency and scope in relation to the type of system to be evaluated.

6 Future Work

In the Functional Tester Runner tool, developing test cases is according to the analysis of traceability. Every test case is designed and coded. In the test execution, coverage is limited by the cases already scheduled. Currently we worked on the proposal of creates a factory of abstract test cases. With the current approach, the generation of test cases has new scope, because they are generated dynamically according to the information traceability test matrix. A very important advantage is that the test matrix has the ability to upgrade and expand. This approach improves the use of the tool from any perspective. In the evolution of software requirements, the functionality may vary according to the current context of system operation to generate new test cases you will only need to update the traceability matrix testing. In this case, the use of design patterns Abstract Factory and Builder [18] allow you to build test cases to run time.

References

1. Poston, R.M., Sexton, M.P.: Evaluating and Selecting Testing Tool. Software, IEEE, Vol. 9, Issue 3, pp. 33–42 (1992)
2. Horgan, A. H. Krauser, E.W.: Incremental Regression Testing. IEEE Proceedings of the Conference on Software Maintenance, pp. 348–357 (1993)
3. Dávila-Nicanor, L., Mejía-Alvarez, P.: Reliability improvement with PSP of Web-based software application. Computer Science & Engineering: An International Journal, Vol. 2, No. 4, pp. 106–112 (2012)
4. Dávila-Nicanor, L., Mejía-Alvarez, P.: Reliability improvement with PSP of Web-based software application. Computer Science & Engineering: An International Journal, Vol. 2, No. 4, pp. 106–112 (2012)
5. Kallepalli, C., Tian, J.: Measuring and Modeling Usage and Reliability for Statistical Web Testing. IEEE Transactions on Software Engineering, Vol. 27, No. 11 (2001)
6. Offutt, J.: Quality Attributes of Web Software Applications. IEEE Software, Vol. 19, Issue 2, pp. 25–32 (2002)
7. Software Engineering Technical Subcommittee of the IEEE Computer Society: IEEE Guide for the Use of IEEE Standard Dictionary of Measures to Produce Reliable Software, IEEE Std. 982.2-1988 (1988)
8. Li, W., Moore, A.W.: Classifying HTTP Traffic in the New Age. SIGCOMM'08: proceedings of the 2008 SIGCOMM Conference and Co-Located, pp. 17–22 (2008)
9. Stantchev, V., Schröpfer, C.: Techniques for service level enforcement in webservices based systems. 4th International Conference, GPC 2009, Geneva, Switzerland, May 4-8. Proceedings, Series Title Lecture Notes in Computer Science, Springer-Verlag, pp. 25–35 (2009)

10. Davis, C., Chirillo, D., Gouveia, D., Saracevic, F., Bocarsley, J.B., Quesada, L., Thomas, L.B., van Lint, M.: Software Test Engineering with IBM Rational Functional Tester: The Definitive Resource. Goebel, A.: Pearson plc (2010)
11. Carvalho, R.E.: Ames Res. Center, NASA, USA, Williams, J., Sturken, I., Keller, R.: Investigation Organizer: the development and testing of a Web-based tool to support mishap investigations. Aerospace Conference, IEEE, pp. 89–98 (2005)
12. Vithani, T.: BITS Pilani, Dubai, United Arab Emirates; Kumar, A. Presentation 5. A comprehensive mobile application development and testing lifecycle. IT Professional Conference (IT Pro), pp. 1–3 (2014)
13. Thanawala, P., Pareek, J., Shah, M.: OntoB Aeval: Ontology Based Automatic Evaluation of Free-Text ResponseTechnology for Education (T4E), IEEE Sixth International Conference on, pp. 189–190 (2014)
14. Mussa, J.D.: Software Reliability Engineering. McGraw Hill, (2000)
15. Law, A.M., Kelton, W.D.: Simulation Modelling and Analysis, McGraw-Hill Series in Industrial Engineering and Management Science (2000)
16. Humphrey, W.S.: Introduction to the Personal Software Process. Addison Wesley (1997)
17. Humphrey, W.S.: Using a Defined and Measured Personal Software Process. IEEE Software, Vol. 13, No. 3, pp 77–88 (1996)
18. Gamma, E., Helm, R., Johnson, R., Vlissides, J.: Design Patterns. Elements of Reusable Object-Oriented Software. Addison Wesley, Pearson Education (2005)

Use of Transportation Methodology to Maximize Profits of a Private Transporter

Oziely Armenta[1], Aide Maldonado-Macías[1], Liliana Avelar Sosa[1], Guillermo Cortés Robles[2], Jorge Limón[3]

[1] Universidad Autónoma de Ciudad Juarez,
Department of Industrial Engineering, Ciudad Juarez,
Mexico

[2] Instituto Tecnológico de Orizaba,
Division of Research and Postgraduate Studies, Orizaba,
Mexico

[3] Universidad Autónoma de Baja California,
Faculty of Engineering, Architecture and Design, Mexicali,
Mexico

danyelita_3@hotmail.com,{amaldona,liliana.avelar}@uacj.mx, gcrobles@itorizaba.edu.mx,
jorge.limon@uabc.edu.mx

Abstract. This study addresses the transport of fresh fruit and vegetables from an independent transporter in Ciudad Juarez, México. Even though the main issue relies on the transport itself, an alternate objective function was proposed to maximize profits of the transporter by considering certain constraints. The main objective of this research is to determine the optimal amount of fruit and vegetables that should be purchased to be transported later to the different destinations to generate the maximum profitability that is possible. The problem is tackled by means of classical methods of problem solving. Moreover, the simplex method employed enables to identify the constraints and determine the optimal solution to the problem that satisfies transport requirements and generates the highest profitability. Finally, results demonstrate that the maximum profit that can be gained is $1,840.00 Mexican pesos during a single trip, satisfying demand of three different businesses.

Keywords: Optimal solution, transport, simplex method.

1 Introduction

This study describes the problem faced by a private transporter in the transport of fresh fruit and vegetables in Ciudad Juarez, Mexico. An alternate objective function was proposed to maximize profits of the transporter by considering restrictions of product capacity, costs, and sale prices.

The independent transporter has experienced monetary losses, since he does not rely on a plan determining the amount of product transported that is profitable to his business. Therefore, the main goal of this study is to find the optimal configuration of

products that must be purchased to be later transported, and which could generate the maximum profitability. In order to achieve this, the research relies on the simplex and transportation methods. Demand is considered weekly, for every two weeks, and monthly.

2 Literature Review

Literature has addressed several methods to solve maximization and transport problems. However, this research views them as two separate elements, since the transportation method is usually used for minimization. Therefore, both this one and the simplex method are here combined to maximize resources. Similarly, since they have been largely considered as classical methods for the solution of this type of problems, the two remain applicable to and efficient for real and contemporary situations.

This second section of the paper is divided in several subsections that introduce basic concepts from the literature review. Experts in the field of operations research are [1, 2, 3]. As experts, they recognized advantages in these methods; as example, the Simplex method implies a linear-programming algorithm that can solve problems having more than two decision variables and generating a series of solutions in tabular form much faster than other algorithms. The efficiency of these methods is also equally important for computer programming as the need of having lower processing power when they are applied. Additionally, an advantage of the transportation method is that the solution process involves only the main variables; artificial variables are not required. In the other hand, the simplex method presents some limitations to solve programming problems, it can only be used in certain linear programming problems and is difficult to adapt it to solve transportation problems; also, the simplex method involves time-consuming computations in the solution of problems that involves a large number of variables and constraints.

Figure 1 represents the methods used to solve this problem and explains the elements considered for the given solution.

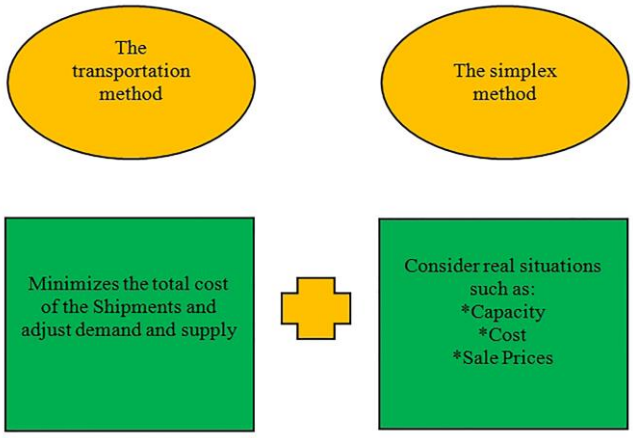


Fig. 1. The Transporting and Simplex Methods.

2.1 The Transportation Method

This model seeks to define a supply network, the amount of product to be produced, the product destination, and the production centers by considering both constraints from resources and costs derived from materials storage and transport [4]. The main goal of this method is to determine the route for materials/product transport that implies the total minimum cost.

Data of this model must include:

1. Supply levels from every source and the amount of demand requested by every destination
2. Unit cost per product for every destination

Since there is only one kind of product, a same destination may receive its demand from one or several sources. The objective of this model is therefore to determine the amount of product that would be sent to every source in order to minimize the total cost of transportation [5]. The basic assumption of the model is that transport costs from one route are directly proportional to the number of units transported. The definition of transport unit depends on the kind of product that is handled.

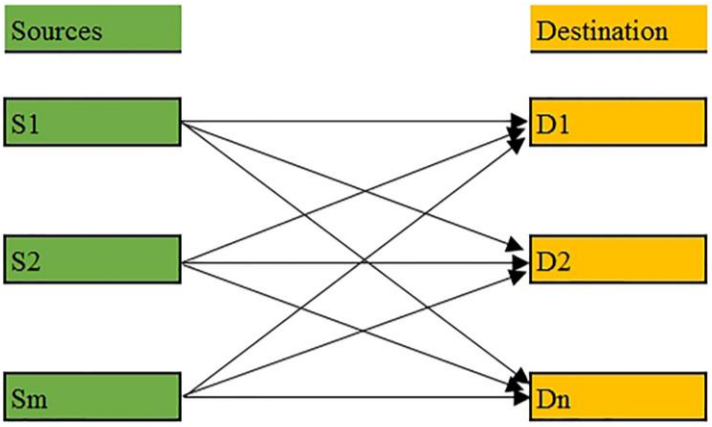


Fig. 2. Sources and destinations.

Every problem explained by this table can be considered as a transportation problem, even without considering its real physical context. For this reason, the transportation method has become one of the most important techniques in linear regression [6].

The following diagram represents the transportation model as a network with m sources and n destinations. Sources and destinations are depicted by nodes, and the arch connecting a source with its destination stands for the route used to transport the product. The amount of supply in source i appears as a_i, while demand from destination j is represented by b_j. Similarly, C_{ij} refers to the unit transport cost between source i and destination j.

Therefore, if X_{ij} stands for the amount of product transported from source i to destination j, the linear programming (LP) model that represents the transportation model can be described as it follows:

$$Minimize\ Z = \ S_i = 1\ m\ S_j\ = 1\ n\ C_{ij}\ X_{ij}\ , \tag{1}$$

subject to:

$$S_j = 1\ n\ X_{ij}\ \leq\ a_i\ ,\quad i = 1,2,\dots,m, \tag{2}$$

$$S_i = 1\ m\ X_{ij}\ >=\ b_j\ ,\quad j = 1,2,\dots,n, \tag{3}$$

$$X_{ij}\ >= 0 \quad \text{for every } i \text{ and } j.$$

The first set of constraints states that the sum of deliveries from one source cannot be higher that this source's supply capacity. Likewise, in the second set, it is required that the sum of deliveries to a destination satisfies this destination's demand.

The model thus implies that the total supply $Si = 1\ m\ ai$ must at least equal the total demand $Sj = 1\ n\ bj$. In this case, the resulting formulation is known as a balanced transportation model. Unlike in the regular transportation model, in this new one all constraints are equations.

That is:

$$SX_{ij}\ =\ a_i,\quad i = 1,2,\dots,m,$$
$$SX_{ij}\ =\ b_j,\quad j = 1,2,\dots,n.$$

However, in real situations, supply is not necessarily equal to or higher than demand, but a transportation model can always be balanced. In addition to its functionality when modeling certain practical situations, balance is important to develop a solving method that fully exploits the special structure of the transportation model. The following two examples depict the idea of balance as well as its practical implications.

2.2 The simplex method

This method was first published by George Dantzig in 1947. It involves an iterative algorithm that, through iterations, sequentially approaches the optimal problem of linear programming if there is any [1, 2, 3]. Also, it was first computationally implemented in 1952 for a problem including 71 variables and 48 equations. The solution to such problem lasted 18 hours. Eventually, in 1956, IBM implemented the RSLP1 code with 4Kb in RAM, which could handle LPs with up to 255 constraints [2]. The simplex method is based on the premise that the optimal solution to a LP problem lies on a vertex or on the boundary of the domain of feasible points (this only in special cases). Therefore, the sequential search of the algorithm is based on the progressive evaluation of such vertexes until the optimal one is reached. However, in order to apply the simplex method to a linear model, this model must be presented in a standard form, which will be described below.

2.3 Standard form of a linear programming model

Consider the standard form of a LP model as it follows:

Min $$c1x1 + c2x2 + \ldots + cnxn,$$
$$a11x1 + a12x2 + \ldots + a1nxn = b1,$$
$$a21x1 + a22x2 + \ldots + a2nxn = b2,$$
$$am1x1 + am2x2 + \ldots + amnxn = bm,$$
$$x1, \ldots, xn \geq 0.$$

Expressed in matrix form as:

$$Min\ cTxAx = b,$$
$$x >= 0.$$

It is assumed without loss of generality that the LP model is in its standard form.

3 Methodology

This section introduces the solution to the problem of fruit and vegetables transportation to the independent transporter in Ciudad Juarez. The different stages that lead to the solution are described in detail.

First, the Vogel approximation method is considered to solve the issue. Second, the three product destinations and their demand requested in the actual period are identified. These destinations focus on grocery sales in different parts of the city. Afterwards, sources are also defined, although, in this case, they are replaced by products offered.

When adjusting the supply and demand to solve the transportation problem, there is one essential element to consider: the fact that the amount of demand is provided by every destination and it does not change within the supply period. However, supply may be freely adjusted in order to meet the total demand of fruit and vegetables. Product units for both supply and demand are expressed in kilograms.

For the optimization of profits generated, data is presented by using a LP model as it is shown below:

$$\sum_{i=1}^{m}\sum_{j=1}^{n} u_{ij}x_{ij} \tag{4}$$

$$\begin{aligned} Max\ z = {} & 5x_{11} + 5x_{12} + 5x_{13} + 6x_{21} + 6x_{22} + 6x_{23} + 8x_{31} + 8x_{32} + 8x_{33} - 1.8x_{11} - 1.8x_{12} \\ & - 1.8x_{13} - 2.3x_{21} - 2.3x_{22} - 2.3x_{23} - 3.2x_{31} - 3.2x_{32} - 3.2x_{33} \\ & - 2{,}100\ , \end{aligned} \tag{5}$$

subject to:

$$\text{Fixed Cost: } x_{11} + x_{12} + x_{13} + x_{21} + x_{22} + x_{23} + x_{31} + x_{32} + x_{33} = \$2{,}100, \tag{6}$$

$$\text{Capacity in Kilograms: } x_{11} + x_{12} + x_{13} + x_{21} + x_{22} + x_{23} + x_{31} + x_{32} + x_{33} \leq 1{,}100, \tag{7}$$

$$\text{Chili Pepper: } x_{11} + x_{12} + x_{13} = 335, \tag{8}$$

$$\text{Watermelon: } x_{21} + x_{22} + x_{23} = 425, \tag{9}$$

$$\text{Mango: } x_{31} + x_{32} + x_{33} = 270, \tag{10}$$

$$\text{Location 1: } x11 + x21 + x31 = 460, \quad (11)$$

$$\text{Location 2: } x12 + x22 + x32 = 220, \quad (12)$$

$$\text{Location 3: } x13 + x23 + x33 = 350, \quad (13)$$

$$x_{ij} \geq 0.$$

where
u_{ij} are pofits from transporting product units i to the destination j,
x_{ij} is flow of product units i to the destination j.

4 Results

The problem was introduced in the Excel Solver tool. Table 1 and Table 2 are data tables, and afterwards, the solution found is presented [7].

Table 1. Costs, sale prices, and fixed cost (in Mexican pesos).

	Cost	**Sale Price**	**Fixed cost**
Chili pepper	$1.80	$5.00	$2,100
Watermelon	$2.30	$6.00	
Mango	$3.20	$8.00	

Table 2. Results from the matrix.

	Location 1	**Location 2**	**Location 3**	**Supply**
Chili pepper	182	57	96	335
Watermelon	157	110	158	425
Mango	121	53	96	270
Demand	460	220	350	

The result obtained is:

$x11$=182, $x12$=57, $x13$=96, $x21$=157, $x22$=110, $x23$=158, $x31$=121, $x32$=53, $x33$=96

Profitability found in this solution is = $1,840.50 (Mexican pesos).

5 Conclusions

Using the simplex and the transportation methods together allows for the consideration of particular constraints to find the optimal solution to the problem that also satisfies transport requirements and generates the maximum profit. Constraints ensure 1) that the amount of product transported by every source equals the amount available, and that 2) the amount of product transported equals the amount required by every destination [7].

Therefore, based on the objectives initially proposed, it is concluded that considerable maximization of resources was actually achieved, since transport routes

selected for the delivery of fruit and vegetables were the optimal ones. In the end, the independent transporter to whom support was provided felt satisfied with results obtained.

Nevertheless, the most effective solution to a transportation problem is not always found, since it is not an easy task to identify and differentiate the variables involved in every problem. Moreover, solutions to transportation problems may be complex. Therefore, it is crucial to correctly identify variables and expectations from every result obtained in order to correctly define the model [8-10].

Finally, advantages and disadvantages of both the simplex and transportation methods must be fully visualized to determine whether they may be applicable in a given situation and in which ways [11].

References

1. Taha, H.: Investigación de operaciones. Novena edición Pearson Educación, México (2012)
2. Taha, H.: Investigación de Operaciones. Pearson Educación, (2004)
3. Cely: Hybrid Meta-Heuristic Algorithms for Minimizing Entropy in Heat Transfer Problems that May Be Present in Electronic Circuits2. Print version ISSN 0123-2126 (2011)
4. García: Modelos para la planeación de transporte terrestre eficiente para el sector de hidrocarburos. Bogotá (2014)
5. García, A.: Investigación de operaciones. Available at: garciaescalona1.blogspot.com /2012/10/origen-de-la-investigacion-de.html (2012)
6. Garcia, J., Maheut, J.: Modelos y Métodos de Investigación de Operaciones. Procedimientos para Pensar. Available at personales.upv.es/jpgarcia/LinkedDocuments/ modeladomatematico.pdf (2012)
7. Ayil, E., Cauich, R., Moo, A., Urcid, A.: Investigación De OperacionesII. Available at: www.itescam.edu.mx/principal/sylabus/fpdb/recursos/r66942.PDF (2010).
8. Hillier, F., Lieberman, G.: Introducción a la Investigación de Operaciones. Available at: www.dropbox.com/s/xtlnulxtayqh2si/Introduccion%20a%20la%20investigacion%20de% 20opera%20-%20Frederick%20S.%20Hillier.pdf (2010)
9. Nahmias, S.: En S. Nahmias, Análisis de la producción y las operaciones. México: McGraw-Hill, pp. 303–305 (2007)
10. Prawda: Métodos y Modelos de Investigación de Operaciones I: Modelos Determinísticos, Limusa, México (2004)
11. Cerda, I.: Optimization using the gradient and simplex methods. Talanta, Vol. 148, pp. 641–648 (2016)

SoLDES : Service-oriented Lexical Database Exploitation System

Mehdi Ben Abderrahmen[1], Bilel Gargouri[2], and Mohamed Jmaiel[1,3]

[1] ReDCAD Laboratory, University of Sfax
ENIS, B.P 1173, 3038 Sfax, Tunisia
mehdi.benabderrahmen@redcad.org
[2] MIRACL Laboratory, University of Sfax
FSEGS, BP. 1088, 3018 Sfax, Tunisia
bilel.gargouri@fsegs.rnu.tn
[3] Digital Research Center of Sfax,
B.P. 275, Sakiet Ezzit, 3021 Sfax, Tunisia
mohamed.jmaiel@enis.rnu.tn

Abstract In this work, we focuses on the assisted exploitation of lexical databases designed according to the LMF standard (Lexical Markup Framework) ISO-24613. The proposed system is a service-oriented solution which relies on a requirement-based lexical web service generation approach that expedites the task of engineers when developing NLP (Natural Language Processing) systems. Using this approach, the developer will neither deal with the database content or its structure nor use any language query. Furthermore, this approach will promote a large-scale reuse of LMF lexical databases by generating lexical web services for all languages. For evaluating this approach we have tested it on the Arabic language.

Keywords: LMF, Lexical Database, Exploitation, Requirement, Interrogation, Web Service, Automatic Generation.

1 Introduction

The majority of NLP systems require lexical resources which make lexical component one of the most important in the field of NLP. Lexical resources are the key element for NLP systems. On the basis of the different needs, several studies dealt with modeling and implementing these resources in different forms (i.e. simple lexicons, relational lexical databases, XML lexical databases). They tried to cover most of the linguistic levels [6]. However, these studies have some drawbacks that can be divided into two classes: some problems related to the lexical product contents (i.e. linguistic coverage) and others related to their interrogation (i.e. integration with the NLP systems). For the first class, it can be noted that lexical products remain very dependent on target NLP systems in the choice of lexical entry structures and the implementation technology. The linguistic coverage is consequently limited. For the second class, the diversity

of model and platform implementation limits the possibilities of interoperability and reuse, in particular, when integrating the same lexical product into different NLP systems. This diversity leads to the heterogeneity of the interrogation mode, update and result presentation.

In order to standardize lexical database design, the internal description format and the multilingual representation, the LMF (Lexical Markup Framework) [7], which is a novel standard under the reference ISO-24613, was proposed. This project proposes an approach for the design of XML databases starting from a single meta-model in UML for all the programming languages and enables to represent multilingual lexical data in connection with all the linguistic levels. Some illustrations of LMF were already proposed (i.e. El-Madar Dictionary for the Arabic language [11] [12] and Morphalou [14] database for the French language). However, the LMF project is interesting only in lexical data representation. It has not yet covered the interrogation and exploitation of the lexical database for the possible needs of the NLP systems.

In this work, we are interested in the use of LMF lexical databases in order to satisfy the whole requirements (i.e. NLP systems requirements, user requests, import and export of external resources). In particular, we focus on the NLP system requirements. In this context, we propose an oriented service based system for the exploitation of the lexical services of LMF standard lexical databases. This system considers two main phases in the deployment and exploitation of LMF services:

The first phase corresponds to the set up of a service oriented architecture for LMF services allowing the interrogation of LMF database. It covers three main steps starting from i) the specification of the main NLP requirements, then ii) their formulation to concrete queries and finally iii) the implementation and the deployment of web services that execute these queries to interrogate LMF database. In this phase, all the steps are realized off-line. An NLP application could then find out a specific web service among the offered ones to interrogate the LMF database. This proposal made the task of engineers easier when developing NLP applications by discharging them from mastering the database structure, a query language or even from formulating queries.

The second phase corresponds to the proposal of an automatic lexical web service generation approach based on the NLP requirements. In fact, in the case where an NLP application developer couldn't find a web service corresponding to his specific requirement, the approach enables to handle this specific requirement online and generate then automatically the corresponding query and web service. The resulted service is then added to the lexical web service library to enrich the exploratory capacity of the given LMF database.

The remainder of this paper is organized as follows: in the next section, we will present our service-oriented LMF database exploitation system that we called SoLDES. We will present the different steps leading to the realization of such system. Then in the third section we will present a requirement-based approach for the generation of lexical web services. We will give details about the lexical web service generation tool developed in this context. After that, we will give

some illustrations of the use of our system and examples of the generation of some lexical services starting from the expression of specific requirements. Before the conclusion, we will show some related works that tried to offer solutions in order to have access to lexical resources in general.

2 Service-Oriented LMF Database Exploitation System: SoLDES

The exploitation and the reuse of lexical databases in general, go through the provision of a well designed interface to query the stored lexical data. The development of NLP (Natural Language Processing) applications becomes then easier and more structured. Such interface could be seen as a specific querying system dedicated to Standard Lexical Database like LMF databases. In this direction, we propose our service-oriented exploitation system called SoLDES to interrogate LMF databases (fig.1). This section will give an overview of the SoLDES system then detail the three main steps leading to the realization of such system.

2.1 SoLDES Overview

SoLDES, as any service-oriented system, provides an extensible web service library for NLP application developers. These web services are lexical services allowing the interrogation of LMF databases. They are developed based on NLP system requirements depicted as queries. In fact, the NLP requirements in term of lexical resources were firstly identified. Then we have matched to each requirement a corresponding query (cf. 2.2). Similar queries were gathered to provide parameterized queries (cf. 2.3). Finally, we have designed a service model for each query pattern (cf. 2.4). Hence, the proposed services are developed regarding the depicted service model corresponding to the NLP requirements. Once the web services are developed and hosted in the service library, their corresponding WSDL file (contains the web service functionality description) is generated and published into a specific UDDI representing a lexical service catalog. This service directory enables the full deployment of lexical web services. It will be consulted by NLP application developers looking up for a specific lexical service fulfilling their requirements. If a service is found, the developer subscribes to this service to enable its invocation and achieve the NLP application development. It should be noted that sometimes, a specific requirement could not be satisfied by only one lexical service, but rather by several ones. For this reason, we consider a service composition layer in the client side which enables the web service composition to provide an added-value service that could meet complex lexical requirement of NLP applications. We will detail an example in (cf. 4). The SoLDES system presented could be seen within a business model involving three main actors:

– Lexical Service Provider LSP: this is the actor who develops and provides lexical web services. He is in direct contact with LMF database.

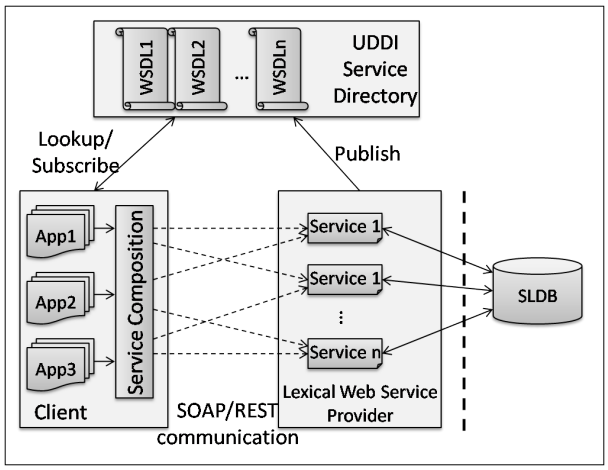


Fig. 1. Service-oriented LMF Database Exploitation System: SoLDES

– Lexical Service Client LSC: he is an actor who will consume lexical services, mainly represented by NLP applications;
– Lexical Service Directory LSD: this is the UDDI directory allowing the publication of lexical web services by LSP. It offers the LSC lookup tools to look for a specific lexical service and make subscription once found.

In this way, the only concern of a client using SoLDES is to consult the LSD and invoke the needed lexical service that will directly provide required lexical information from LMF database. Hence, the client is discharged from likely arduous tasks like:

– Knowing and mastering the structure of LMF database, its DTD or XML schema, used extensions, connection mechanisms;
– Learning a query language (like XQuery) and mastering the query formulation, even the complicated ones, usage flexibility, etc. The interaction with LMF database is used exclusively through web services;
– Ensuring integration of LMF database to NLP applications, no more effort will be needed since the web services became the most widespread integration tool.

The client, in general, only has to understand the web service functionality which is independent from both the implementation technologies and the LMF database location. Hereafter, we will detail the three main processes involved in the SoLDES system, namely, the requirements identification, the query formulation and the web service implementation.

2.2 Requirement Identification

Starting with NLP leader applications such as “Segmenter”, “Lemmatizer” and “Morphological Analyzer”, the first step is to determine the set of lexical needs required by these applications according to the studied linguistic level (morphological, syntactic or semantic). As shown in Fig.2, a NLP application (i.e. App.1) can use one or more linguistic levels. Indeed, the “Morphological Analyzer”, for example, requires only a morphological analysis but the “Segmenter” needs, the syntax analysis, as well. Each level expresses a number of lexical needs to be met from lexical database (LDB). A lexical database that covers all the aspects of a natural language must provide lexical information about all the linguistic levels and thus will satisfy all types of needs that can express an NLP application. During the requirement acquisition step we tried to identify the most generic needs by a flattened way starting with basic needs. Thus, the most complex needs can be solved through a decomposition into basic needs. The requirement identification operation requires the involvement of linguistic experts (in our case, the target language is Arabic) and the NLP application developers.

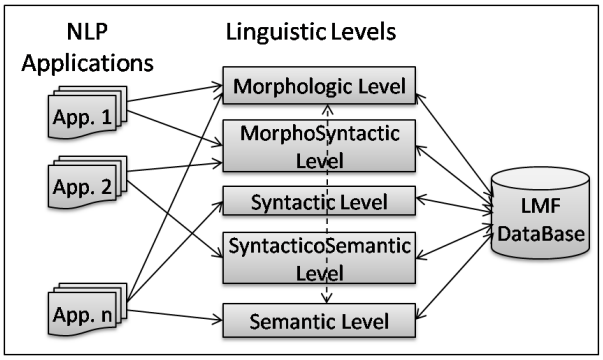


Fig. 2. Requirement identification through linguistic level

2.3 Query Formulation

The requirements identified in the previous step are expressed in a human-readable manner (abstract form). In order to make them understandable using a machine, they have to be transformed and formulated as queries (concrete form). In fact, for each requirement, we have matched a query expressed manually with XQuery language. This transformation process requires good skills and knowledge of both XQuery syntax and LMF database structure. This is what we have actually done to spare the complexity for lexical service clients. The result of the transformation is a set of various queries. Normally, the next step is to implement a web service to execute each query. However, we noticed the existence

of similarity between some queries having the same output type. Indeed, these similar queries correspond, in general, to the same requirement but with different input parameters. In this case, we gathered each set of similar queries into one parametrized query implemented by only one configurable service (service with several methods) instead of having a web service for each similar query. The main objective of this operation is organizational. For example, there could exist two similar queries: the first deals with the consistency between Prefixes and Suffixes and the second with the consistency between Prefixes, Infixes and Suffixes. The similarity here resides in the inputs "Prefix" and "Suffix" common in both queries. As a result, only one service will be devoted to both queries but with different capabilities.

2.4 Web Service Implementation

After the formulation of the query emerging from NLP requirement identification, the development of the web service will be performed. Indeed, a service is an application entity that handles several tasks, like:

- The connection to the lexical database. In our case, the connection and then the interrogation of the lexical database are made through DataDirectXQuery implementing the XQJ (API XQuery for Java);
- Catching NLP requirements as input parameters for the service. This information is read as external variables whose values will be injected in the XQuery query;
- Executing the query and giving back result to the NLP application.

A service could be composed of one or several methods according to the requirement type (elementary or complex). All the aforementioned tasks are achieved as processing within the service methods. Therefore, the granularity of a service is defined by the association of a service method with an elementary requirement. Hence, a service with several methods generally fulfills a complex requirement. When a service contains only one method, the corresponding requirement should be simple (elementary).

For example, a consistency service that asks whether the pair (affixes, root) is consistent or not. This service contains several methods as shown below.

```
1 boolean coherenceRootSuffix (String root, String suffix)
2 boolean coherencePrefixSuffix (String prefix, String suffix)
3 boolean coherencePrefSufInf (String prefix, String suffix, infix String)
```

The development of a lexical web service is followed by the generation of a standard WSDL file description. This file contains a service capability description which will help index and reference the service within a lexical service directory (UDDI). The service capability description precises for each method its inputs and output types and the service address. The deployment of a lexical web service is done by adding both the WSDL file to the Lexical service directory (UDDI) and the lexical web service to the service library. The lexical web service is hence

ready for use by NLP application. The latter has only to look for lexical web service corresponding to its requirement to exploit LMF database capabilities.

2.5 Discussion

The SoLDES system as proposed made the task of engineers easier when developing NLP systems by discharging them from mastering the database structure, a query language or even from formulating queries. The interrogation of an LMF database was transformed to a web service application allowing the fluency of the interoperability between NLP system and LMF databases. However, since the SoLDES system is based on a three-step process performed manually and offline, a consistent problem is raised; What happens when an NLP system lexical requirement doesn't match any lexical web service provided by the library? This will lead to the limitation of the exploitation of lexical databases. Although the requirement identification step has involved linguistic experts, all requirements couldn't be conceived and then expressed. There are always some NLP applications that have a specific requirement not already expressed before and should be considered and covered. For this reason, we proposed as a second phase of SoLDES system, a lexical web service generation approach enabling to overcome this problem. This approach enables to automate all the steps leading to the generation of a lexical web service. This is detailed in the next section.

3 Requirement-Based Lexical Web Service Generation Approach

3.1 Approach Overview

In order to cover the aforementioned problem, we proposed an approach for the generation of lexical web services. This approach is composed of three milestones with everyone of which is based on the use of a specific module. In the

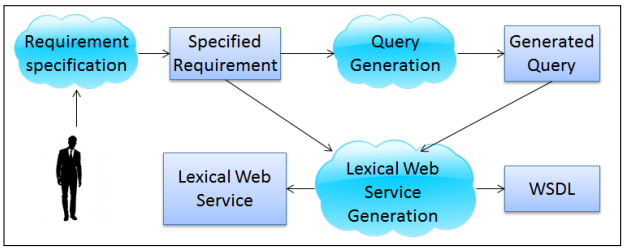


Fig. 3. Requirement-Based Lexical Web Services Generation Approach

first milestone, the NLP developer specifies his requirements in term of Inputs

and Outputs (Input and Output here represent the values taken from the Data Categories Registry DCR which is a standard under the number ISO-12620 [10]) through a user-friendly interface belonging to the first module called Requirement Specification. The Input part is generally conditioned by an LMF-QL grammar compliant formula. At the end of this step, the Requirement Specification module generates an abstract form of the user's need.

In the second milestone, the Query Generation module transforms the abstract form of the requirement to a concrete form written in XQuery.

Finally, in the last milestone, the two forms (abstract and concrete) go to the third module which will generate both the lexical web service and its description in WSDL language. After the generation of the query based on the requirement provided by the user itself through the user-friendly interface, the tool also helps the generation of a web service corresponding to the generated query regardless of its complexity. The result of the web service generator tool is two outputs: the web service code and the WSDL file.

3.2 Requirement Specification

The requirement specification module helps NLP developer through a user-friendly interface to easily specify his needs in terms of lexical data from LMF database. This requirement is defined by a pair (Input, Output) which represents its abstract form. The Output represents the required result, but the Inputs represent a filter that can restrict the scope of the possible required result. For the definitions of this pair, we used data categories taken from the DCR (Data Categories Registry). To define Inputs, we removed all types of abstract symbols, particularly, those used in other languages such as XQBE [4], in order to facilitate the requirement specification. Then, we replaced these symbols by a selection of attributes used in the LMF database, which represent data categories (DC). For the definition of the query Output, we proceeded in the same manner, with the difference that for the Input we were obliged to work only with attributes (leaves), and for the Output we could work either with attributes or elements (Internal Nodes). The query Input must be precise and expressed in term of attribute-value; consequently an attribute, which is in an internal node (element which contains many entries), cannot belong to the Input of the query. More information about this step could be found here [2].

3.3 Query Generation

This module handles the abstract form of the user's requirement. The Output of the generation corresponds to an XQuery based query meeting the expressed requirements. The facility introduced by this tool lies in the fact that the Input elements make the database structure abstract and do not refer to any particular level of the database. These levels which are generally present in queries, will be determined by data categories selected in the requirements. With this intention, we proposed a high-level language called LMF-QL [2]. This language allows specifying the logical relation between different data categories selected

in Input part. The LMF-QL grammar is composed of a set of terminals Σ, a set of non terminals V, an axiom S and a set of transformation rules P. Among the terminals, we used logical operators of two types: unary operator 'not' that can be applied on one data category or on an expression and which represents disjunction, and binary operators 'or'/'and' that can be applied to two data categories.

```
1  G= (Σ, V, S, P)
2  Σ= {or, and, not,(, ), E_i}
3  V = {S}
4  P = { S –> (S) | not(S) | (S or S) | (S and S) | E_i}
```

The query generation is done after the requirement specification (abstract form) by translating this form into XQuery syntax (concrete form). The generated queries may have different forms depending on the number of entries set in the input and the output of the query. The Table 1 shows the classification of queries according to the number of inputs and outputs. Each form represents a general

Table 1. Query class patterns

Number of Inputs	Number of Outputs	Result Type	Requirement Type	Number of Classes of patterns
1	0	Boolean	Existence	2
2	0	Boolean	Consistency	4
$i \geq 1$	$j \geq 1$	Different types	General Requirement	2^{i+j}

model for a class of requirements as shown in table I. The first class is to express the requirement of the existence of a category of data. It is divided into two patterns because it has two different query models. The second class is used to test the consistency between the two categories or the other data. For example, we can test the consistency between the prefix and the suffix. The third class represents the general case. The number of requests generated depends on both the number of entries and that of outputs. In general, the number of classes verifies equation (1):

$$Numberofclasses = 2^{(NumberofInputs+NumberofOutputs)} \quad (1)$$

3.4 Lexical Web Service Generation Tool

Like the automation of the query generation, the service generation represents an efficient step to ease the interrogation of LMF databases. Besides, having a

standard interface, like web services, increases the interoperability aspect of the lexical resources. In this context, we developed a specific tool based mostly on the query generated by the Query Generator Module (see fig.4). The query is firstly

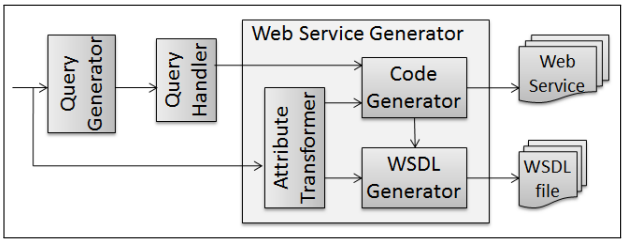


Fig. 4. Lexical Web Service Generation Tool

sent to the Query Handler Module in order to check its consistency. It's a kind of query compiler that verifies the syntax of the query before execution, which can help avoid exceptions due to syntax when calling the web service. The query can then be transferred to the Code Generator Module to start manufacturing the web service. The last cited module also needs some inputs that will be gathered from Attribute Transformer module. These inputs give information about the user's needs required in the service code generation. The Attribute Transformer is responsible for gathering information and transforming it to the suitable forms in order to make it available for both modules "Code Generator" and "WSDL Generator".

Attribute Transformer As already mentioned, the user's requirements are involved in all the steps of lexical service generation process. The main role of the Attribute Transformer Module is to gather the user's requirement information in its abstract form in order to transform it to suitable variables that will be used later by both modules "Code Generator" and "WSDL Generator". The information about the user's requirements is the inputs and output parameters like their Data Categories, variables names, entry options, etc. This information could also be a simple text that the user uses to describe query capabilities. The Attribute Transformer Module catches this information and structures it by creating dedicated variables with suitable types. The goal is to make it ready for use by the next modules.

LWS Code Generation This module will build the web service around a service skeleton that contains 4 main parts:

- Connection to the LMF database.
- Query parameter binding operation
- Execution of the query

– Preparation of the result of the query

a) The connection to the LMF database: This part of the service code allows the connection to the lexical database. Actually, there is only one database we are working with. Besides, the database network address and the connection and configuration parameters are hardcoded. However, this information could be easily provided by a user as input parameters.

b) Query parameters binding operation: The code generated in this part serves the preparation of the query before being executed. In fact, each query contains input and output parameters. The values of these parameters have to be prepared before being injected in the query for execution. Two cases are supported (both cases can be combined):

– The first case where the input value is already expressed by the user while providing the requirements (handwritten or chosen among predefined values list of the corresponding data category). In this case, the input value is hardcoded in the query. No input value injection will be done.
– This the case where the user doesn't provide values to his inputs but rather chooses to provide them as external values (as variables). Only the data category of each input is specified. In this case, the inputs are considered as arguments in the web service method. These arguments have to be bound respectively to the variables already declared in the query. Thereby, the inputs values will be correctly transmitted to the query when invoking the web service. This case enables the reusability of a configurable lexical query.

In this last case, the Code Generator Module requires the user's requirements, precisely information about the inputs and output parameter (data category, variable name, value entry option, etc), to generate the binding operation code. This information is caught and structured by the Attribute Transformer Module to make it available for other modules (like WSDL Generator explained later).

c) Execution of the query: Once prepared, the query has to be executed. This part of the service skeleton enables the generation of the service code responsible for the execution of the query. The execution is based on the DataDirectXQuery implementing the XQJ (API XQuery for Java).

d) Preparation of the result of the query: The execution of the query generates an output corresponding to the user's requirement. This part of the service code retrieve the result of the query execution and transforms it to the suitable format specified by the user when generating the query. The possible formats can be a simple string, XML section, Array, Object, etc. The generated service code is obtained as Java class source code. We used Apache Axis native JWS (Java Web Service) to expose and deploy our code as a web service. The generated bundle contains a deployable web service with all required resources ready for use in web application. However, our web service generator tool helps modify the WSDL file automatically generated by Apache Axis Framework. Indeed, we propose to enrich the basic WSDL file with lexical information that helps index and reference lexical web service in the lexical service directory UDDI.

WSDL File generation The last remaining step is to generate the WSDL file corresponding to the generated web service. Once again, the user's requirements are involved. Indeed, the module WSDL Generator enriches the basic Axis generated the WSDL file with specific annotations that help NLP application developers to find out the most appropriate web services. The content of these annotations is provided by the Attribute Transformer Module. It contains a human readable description of the generated query provided by the NLP developer when specifying his requirement. Besides, it also contains the definition of the query input and output parameters representing the lexical data categories. There are two kinds of annotations enriching the WSDL file:

a) Annotation describing the query capability: The content of this annotation is a human readable text provided by the user when specifying his requirements. It describes the functionality of the generated query. In fact, there is only one annotation of this type per generated WSDL file. This annotation is added as a predefined XML element called <wsdl:documentation> under the WSDL element <wsdl:operation> under the WSDL element <wsdl:portType> representing the method of the service. This is an example:

```
1  <wsdl:portType name="Affixes_Service">
2    <wsdl:operation name="Prefixes_List">
3    <wsdl:documentation>This method gives the list of prefixes</wsdl:documentation>
```

b) Annotation describing the input and output parameters: This annotation define each input and output of the query. Actually, there are annotations as many as inputs and outputs. This kind of annotation is added using the predefined XML element <wsdl:documentation> under the element <wsdl:part> under the WSDL element <wsdl:message>. The text added in the XML tag <wsdl:documentation> is the definition of the data category of the corresponding input or output.

```
1  <wsdl:message name="SyntacticFunctionRequest">
2    <wsdl:part name="syntacticHead" type="xsd:string">
3      <wsdl:documentation>syntacticHead: central element of a subcategorization
           frame</wsdl:documentation>
4    </wsdl:part>
5    <wsdl:part name="VoiceProperty" type="xsd:string">
6      <wsdl:documentation>VoiceProperty: is the class of properties that concern the
           grammatical encoding of the relationship between the verb and the nominals in
           a subject-predicate configuration.</wsdl:documentation>
7    </wsdl:part>
8  </wsdl:message>
```

3.5 Synthesis

The automatic generation of lexical web services based on the NLP requirements helps overcome the lack of new specific web service in an initial service library of SoLDES system. Hence, SoLDES became a complete service-oriented architecture covering even new specific lexical requirements. NLP developers no longer

fear the complexity of LMF database query with such a tool. In order to assess and to show the usability of SoLDES system, we will give, in the next section, some illustrations of the use of our system and examples of the generation of some lexical services starting from the expression of specific requirements.

4 Illustration: Lexical Services for Arabic Language

In this section, we present a set of lexical service requirements identified in the context of Arabic language although most of these requirements could be the same for other languages. Then we present an example of implementation of some services. We will show later our service library. We finish this section by giving two case studies that give an example of how to integrate these services in the context of an NLP application.

4.1 Characteristics of the Arabic Language.

The recent years have seen a considerable progress in the field of NLP. The Arabic language does not make exception but it has been much less studied from the data-processing point of view than English or French. This may be due to difficulties related to this language. Indeed, by its morphological and syntactic properties, the Arabic language is considered a difficult language to master in the NLP field [3]. Among these properties we can quote:

- The letters change form of presentation according to their position in the word which is written from the right to the left.
- An Arabic word is written with consonants and vowels. The vowels are added above or below letters (بْ, بِ, بُ, بَ). They facilitate the reading and the correct understanding of a text to differentiate words having the same representation.
- In Arabic, a word can mean a whole sentence thanks to its mixed structure which is an agglutination of elements of grammar.
- Arabic is an inflected language. Indeed, the terminations make it possible to distinguish the mode from the verbs and the function of the names.

4.2 Requirement Identification

The lingware development complexity is due to the multitude of needs that may present a linguistic application in terms of information, in particular, of lexical nature. In our experimentation, we studied a set of applications dealing with the morphology of Arabic in order to identify their lexical needs. Hereafter, we give the results of this study for some of these applications namely: Arabic Text Tagger, Morphological analyzer, spellchecker, education through interactive software.

Arabic Text Tagging This application can be defined as the set of operations that can switch between plain text, free of linguistic information, and a sequence of elementary lexical units (lemmas) accompanied by morphosyntactic labels. This definition implies successively the choice of the basic unit of segmentation, the process of segmentation itself, the lemmatization of the units and the association of linguistic information to the lemmas. The phase of lemmatisation requires the exploitation of the Lexical DataBase in order to join to a lemma all morphological features that correspond to it. These features can change according to the lemma [17].

Table 2. Text Tagging

Requirement	Input	Output
Part-of-Speech Tagging	Verb	Conjugation Mode , GramPerson, Scheme
	Deverbal	Type, Chained Verb, GramGender, GramNumber
	Noun	human characteristics Human/Non-Human, Proper Noun/Common Noun, GramGender, GramNumber
	Function Word	GramGender, GramNumber

Morphological Analysis and Synthesis This is a program that can recognize a word in the various forms it can take in sentences. For each found form, the elements must be isolated and morphological features deduced out of the context associated with them. The morphological processing of Arabic must cover both generation processes (or synthesis) and analytical (or recognition). The synthesis process must allow the generation of a word starting from a root, a scheme and a set of morphological specifications. Based on [9], we were able to identify all the needs for these two processes. Table 3 shows these needs.

SpellChecker This kind of application deals with the errors related to the lexical level. Error handling does not take into account the context of the word to check. Lexical errors are those related to membership of the words to the language. The need for the LDB within the framework of this system can be summarized in the Table 4.

Interactive Teaching The teaching of Arabic through interactive software must be able to intervene both in recognition (which corresponds to certain aspects of the difficulties encountered when reading or looking for a word in a

Table 3. Morphological Analysis and Synthesis

Requirement	Input	Output
Morphological Analysis	Base	Boolean (Exist or not)
	P, S et I	Boolean(Exist or not)
	P, S et I	Boolean(Consistent or not)
	Word	Root, chained word
Morphological Synthesis	Root	Boolean(Exist or Not)
	P, S, I et Root	Boolean (Consistent or not)
	Root	List of Associated Base
	Root	Exist or Not

Table 4. SpellChecker

Requirement	Input	Output
Orthography	Word	Exist or Not
	Selection	List of Prefixes
	Selection	List of Suffixes
	Selection	List of Infixes
	Selection	List of Roots
	Prefix, Suffix	Boolean (Consistent or Not)
	Prefix, Suffix et Infix	Boolean (Consistent or Not)

dictionary or BDL) and in producing (as regards the construction problems of Arabic words at the time of expression). Both approaches will need two processes of analysis and synthesis which have been mentioned in Table 6. Thus, we could address in the Table 5 the interactive teaching needs.

Table 5. Interactive Teaching

Requirement	Input	Output
interactive teaching	Verb	Grammatical Features
	Non Diacritical Word	List of all Diacritical Forms
	Diacritical Word	PartOfSpeech, (Prefix, Infix, Suffix) list of chained roots
	Root	All associated forms

Synthesis All the previously cited lexical needs were translated into XQuery queries and its corresponding lexical web services that were generated using

Table 6. conjugation synthesis and analysis

Requirement	Input	Output
synthesis Conjugation	Verb	Conjugated forms in all modes
	Verb + conjugation mode	Conjugated forms in this modes
	Verb + conjugation mode + Person	Conjugated Form
Analysis Conjugation	Conjugated Form	Person, Gender and Number
	Conjugated Form	Related forms
	Accomplished Form	Unaccomplished Form

our SoLDES system. All these generated services could be then used in the development of NLP applications.

4.3 Arabic Text Spell Checker (ATSC): First Case Study

To experience our approach, we proceeded to the redevelopment of the system [8] "Arabic Text Spell Checker". The choice of this system is justified by the richness of its lexical resource requirements and also the availability of its code. ATSC uses a lexicon that is formed by the following files: a file for roots, a file for the prefixes, a file for the suffixes, a file for the infix, a file containing the consistency matrix between prefixes and suffixes, and a file containing the consistency matrix between prefixes, suffixes and infixes. During the redevelopment of ATSC, we focused on the file access. Indeed, we proceeded in two stages:

- locating the access levels to files in the code of the application and determine the needs that involve from this access.
- Searching from our lexical web service library services that can meet these needs. A new service has been developed for every unmet need.

Second, we tried to see if some ATSC features could be provided by one or more of our services. Thus, we have located the "decomposition" service which allows for a correct word to give its components in terms of root, infix, prefix and suffix. As experimental results, we note that after using our SoLDES system, we have reduced in the ATSC new version the number of used java classes from 22 classes in the original version to 7 classes in the new service-oriented version. Furthermore, the use of SoLDES system for generating lexical web services has led to the reduction of the number of lines of code. This is due to the replacement of the methods of existence check of the suffixes, prefixes, infixes and root by the simple service invocation from our Lexical Service Directory (LSD). The methods of verification in the original version are very complex with the instructions of

opening and reading files, the iterative loops and multiple conditional blocks. All this is replaced by the simple method binding() for the service invocation which considerably reduced the effort and the time of the development.

4.4 Automatic Summarizer for Arabic: Second Case Study

For the evaluation of our work, we will present another case study that integrates some of the lexical services from our library during the process of summarizing an Arabic text that we first present. The production process of an automatic summary is based on five modules namely the segmentation module which is the first step that enables to segment the source text into sentences[1]. The result is sent to another module called the morphological analysis module to give the morphological features of each word of the text. The syntactic analysis module receives the output of the morphological analysis module and shows the syntactic structure of each sentence in the summary. The extraction module for the relevant sentence extracts the most important phrases. Finally, the revision module helps obtain a refined extract by eliminating the redundant phrases. In this process, we will focus on modules that can offer low-level linguistic information, namely morphological and syntactic analysis.

In this context, among the requirements, a developer of an automatic summarizer for Arabic concretely requires two specific low-level lexical services that we call "Requirement 1" and "Requirement 2".

Requirement 1 Here, the developer needs to look for the morphological features of a given lexical entry. He uses our tool to precise the different parameters of his requirement. In this case, the input is a lexical entry having a "Written Form" as Data Category. The developer has to choose to provide this input as an external value. In fact, no value will be entered at this level. The value of the lexical entry will be provided later while using the generated service. The output which the developer wait for is a "WordForm" element supporting several kind of Data Categories like "Grammatical Number", "Grammatical Person" and "VerbFormAspect". The developer chooses the XML section as output format and integrate it in the whole summarizer system. The generated query is shown here:

```
1 declare variable $x as document−node(element(*,xdt:untyped)) external;
2 declare variable $u1 as xs:string external;
3 declare variable $u2 as xs:string external;
4 for $a0 in
5 $x//LexicalResource/Lexicon/LexicalEntry/WordForm where $a0/DC/@att="writtenForm"
      and $a0/DC/@val="$u1"
6 return for $a1 in $a0 where $a1/DC/@att="voice" and $a1/DC/@val="$u2"
7 return <result>{$a0}</result>
```

Once the query is generated, the developer generates the corresponding Web Service to be integrated in the whole system. We present here the result of the invocation of the generated service with the values of data categories (the variable $u1 will take the value "كَانَ" and $u2 will take the value "activeVoice".

```
<?xml version="1.0" encoding="UTF-8"?>
<result>
   <WordForm>
        <DC att="writtenForm" val="كَانَ"/>
        <DC att="grammaticalNumber" val="singular"/>
        <DC att="grammaticalGender" val="masculine"/>
        <DC att="verbFormAspect" val="accomplished"/>
        <DC att="voice" val="activeVoice"/>
        <DC att="person" val="thirdPerson"/>
    </WordForm>
</result>
```

Requirement 2 In this case, the developer requires to know the syntactic function of a lexical entry having the voice property value equal to "Active Voice" and having a given syntactic behavior. Actually two inputs are given here:

- Input 1: represents the syntactic behavior having "Type" as data category and will be provided as external value.
- Input 2: having the "Voice Property" as data category and its value is "activeVoice". It will be a hardcoded value.
- Output: having the data category "SyntacticFunction".

Hereafter, we present the query corresponding to this requirement.

```
declare variable $x as document-node(element(*,xdt:untyped)) external;
declare variable $u1 as xs:string external;
for $a0 in
$x//LexicalResource/Lexicon/SubcategorizationFrame where $a0/DC/@att="type" and $a0/
      DC/@val="$u1"
return for $a1 in $a0/LexemeProperty where $a1/DC/@att="voice" and $a1/DC/@val="
      passiveVoice"
return
for $a2 in $a1/../SyntacticArgument/DC where $a2/@att="syntacticFunction"
 return <syntacticFunction>{$a2/@val}</syntacticFunction>
```

The result of this query will be as follows after invoking the service with the value of $u1="فِعْلٌ مُتَعَدِي لِمَفْعُولَيْنِ أَحَدُهُمَا بِأَدَاةٍ":

```
<?xml version="1.0" encoding="UTF-8"?>
<syntacticFunction val="نَائِبُ الفَاعِلِ"/>
<syntacticFunction val="مَفْعُولٌ بِهِ"/>
```

5 Related Work and Discussion

The main objective of our current research is to promote the use of lexical resources through enhancing the interoperability between involved actors, namely,

end users applications, NLP applications and lexical databases. The first solution is the use of normalization (standard) of principles and methods related to the lexical resources in the context of multilingual communication and cultural diversity. This is done through the adoption of LMF standard for lexical databases. This can bring NLP application developers to less dependency on a specific lexical database structure, but this is not enough. In fact, a service oriented approach effortlessly allowing the access to lexical standard resources represents a second step solution we are adopting. This solution was also adopted by several scientific communities having looked to facilitate the extraction of lexical resources according to the needs of users or NLP applications. The proposed work in [5] is based on the use of web services for the use of NLP features in a multicultural context. However, the proposed services don't tackle the access to the lexical resource .Besides, the proposed work does not address the Arabic language and its specificities since it doesn't adopt LMF standard. Finally, the generated web service is specific to their platform while our work generates a web service according to the need of the user and can be used by any NLP application. In the same context, researchers in [16] present an architecture to connect customers to NLP frameworks through the use of web services. However, NLP subsystems do not support standard database and the proposed approach does not display solution for the needs of the customer's alignment with the standard structure. Other works propose the use of a web service but are limited to a specific usage. We mention the work [15] which deals with the development of a RESTful Web service to access WORDNET-type semantic lexicons. It generates LMF compliant XML data. We can also mention the web service tool for automatic extraction of MWE (Multi-Word Expression) lexicons [13]. The usage here is limited to the creation of Lexical Resources.

6 Conclusion and Future Work

The research carried out by the ISO around LMF project showed the interest granted by all the community of NLP researchers to the lexical component in this field. Starting from this, our work takes the continuation of the representation aspect to cover the LMF lexical database exploitation. In addition, on the basis of various needs in lexical resources and technical difficulties facing users (i.e., NLP applications developers), we proposed to use a Requirement-based approach for the generation of lexical web services to ensure easier exploitation of lexical resources and minimize the developers' efforts. Indeed, this approach discharges the developers from several tasks: knowledge of the database structure, master of a query language, etc. Moreover, this approach gives the possibility of serving different users in different corners of the world starting from a single database. On another side, the interrogation of the lexical database remains transparent for the users. Indeed, we worked with parametrized queries generating according to the listed needs. The query list remains extensible to cover new needs. The results of our work are being tested for many linguistic levels in the framework of Arabic spell checker application and during a summarizing

process. The lexical database (ElMadar Dictionary)[4] used in our work is LMF compliant. Currently, we are planning to test our approach on various lexical databases in conformity with LMF, possibly for other languages. In the near future, we will work on the refinement of the description generated with the lexical service (WSDL description file) in order to facilitate the discovery of our services. Subsequently, we will study the integration with applications and composition scenarios that could give more help to assist NLP developers in the development of new lingwares.

References

[1] Abdallah, M.B.: Réalisation d'une plateforme de résumé automatique de textes arabe. Master's thesis, Ecole Nationale d'Ingénieurs de Sfax (2008)

[2] Abderrahmen, M.B., Gargouri, B., Jmaiel, M.: LMF-QL: A graphical tool to query LMF databases for NLP and editorial use. In: Vetulani, Z., Uszkoreit, H. (eds.) Human Language Technology. Challenges of the Information Society, Third Language and Technology Conference, LTC 2007, Poznan, Poland, October 5-7, 2007, Revised Selected Papers. Lecture Notes in Computer Science, vol. 5603, pp. 279–290. Springer (2007), http://dx.doi.org/10.1007/978-3-642-04235-5_24

[3] Aljlayl, M., Frieder, O.: On arabic search: Improving the retrieval effectiveness via a light stemming approach. In: Proceedings of the Eleventh International Conference on Information and Knowledge Management. pp. 340–347. CIKM '02, ACM, New York, NY, USA (2002), http://doi.acm.org/10.1145/584792.584848

[4] Braga, D., Campi, A., Ceri, S.: Xqbe (xquery by example): A visual interface to the standard xml query language. ACM Trans. Database Syst. 30(2), 398–443 (Jun 2005), http://doi.acm.org/10.1145/1071610.1071613

[5] Bramantoro, A., Tanaka, M., Murakami, Y., Schafer, U., Ishida, T.: A hybrid integrated architecture for language service composition. In: Web Services, 2008. ICWS '08. IEEE International Conference on. pp. 345–352 (Sept 2008)

[6] Francopoulo, G.: Proposition de norme des Lexiques pour le traitement automatique du langage. http://pauillac.inria.fr/atoll/RNIL/TC37SC4-docs/N07.pdf (2003)

[7] Francopoulo, G., George, M.: Model Description, pp. 19–40. John Wiley & Sons, Inc. (2013), http://dx.doi.org/10.1002/9781118712696.ch2

[8] Hamadou, A.B.: Vérification et correction automatiques par analyse affixale des textes écrits en langage naturel : le cas de l'arabe non voyellé. Ph.D. thesis, Faculté des Sciences de Tunis (1993)

[9] Hassoun, M.: Conception d'un dictionnaire pour le traitement automatique de l'arabe dans différents contextes d'application. Ph.D. thesis, Université Claude Bernard-Lyon I (1987)

[4] elmadar.miracl-apps.com

[10] ISO 12620:2009: Terminology and other language and content resources – Specification of data categories and management of a Data Category Registry for language resources, vol. 2009. ISO, Geneva, Switzerland (Dec 2009)
[11] Khemakhem, A., Gargouri, B., Haddar, K., Ben Hamadou, A.: LMF for Arabic, pp. 83–98. John Wiley & Sons, Inc. (2013), http://dx.doi.org/10.1002/9781118712696.ch6
[12] Khemakhem, A., Gargouri, B., Hamadou, A.B.: LMF standardized dictionary for Arabic language. In: Proceedings of the 1st International Conference on Computing and Information Technology (ICCIT 2012). pp. 522–527 (2012)
[13] Quochi, V., Frontini, F., Rubino, F.: A MWE acquisition and lexicon builder web service. In: Kay, M., Boitet, C. (eds.) COLING 2012, 24th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, 8-15 December 2012, Mumbai, India. pp. 2291–2306. Indian Institute of Technology Bombay (2012), http://aclweb.org/anthology/C/C12/C12-1140.pdf
[14] Romary, L., Salmon-Alt, S., Francopoulo, G.: Standards going concrete: from LMF to Morphalou. In: The 20th International Conference on Computational Linguistics - COLING 2004. coling, Genève/Switzerland (2004), https://hal.inria.fr/inria-00121489
[15] Savas, B., Hayashi, Y., Monachini, M., Soria, C., Calzolari, N.: An lmf-based web service for accessing wordnet-type semantic lexicons. In: Chair), N.C.C., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., Tapias, D. (eds.) Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10). European Language Resources Association (ELRA), Valletta, Malta (may 2010)
[16] Witte, R., Gitzinger, T.: A General Architecture for Connecting NLP Frameworks and Desktop Clients using Web Services. In: Kapetanios, E., Sugumaran, V., Spiliopoulou, M. (eds.) 13th International Conference on Applications of Natural Language to Information Systems (NLDB 2008). LNCS, vol. 5039, pp. 317–322. Springer, London, UK (June 24–27 2008)
[17] Zaafrani, R.: Développement d'un environnement interactif d'apprentissage avec ordinateur de l'arabe langue étrangère. Ph.D. thesis, ENSSIB/Université Lumière Lyon 2 (2002)

Impreso en los Talleres Gráficos
de la Dirección de Publicaciones
del Instituto Politécnico Nacional
Tresguerras 27, Centro Histórico, México, D.F.
Abril de 2016
Printing 500 / Edición 500 ejemplares