A Novel Design of Smart Evaluation on Job Vacancy Application System (SEJVAS) Using UML

Noraziah Binti Ahmad
University Malaysia Pahang
<noraziah@ump.edu.my

ABSTRACT: *Nowadays, job application is more preferred online than paper resumes. Smart Evaluation on Job Vacancy Application System (SEJVAS) has been developed by using Rapid Application Development (RAD). In this paper, new designs of the job vacancy application module are presented by using UML diagrams. In addition, this system applies the Rule-based technique in order to generate a shortlist for the job applications file that fulfills the company requirement. SEJVAS has been developed by using php programming language, php myadmin for the database and css for the interface. It can screen a range of applications and generate a short list of applicants to call for an interview. The simulation result shows the right candidate for the right profession can be selected without the worry of cost and time consuming.*

Keywords: SEJVAS, UML, RAD, Rule based techniques

Received: 19 August 2009, Revised 29 September 2009, Accepted 23 October 2009

1. Introduction

Today's hiring procedure involves a lot of people, paper work and time. The applications received for a vacant job are often countless. The company also has to hold time-consuming meetings to go through all the applicants that fulfill the minimum qualifications from the unsuitable. The process of evaluation and hiring decision making often takes weeks. Hence, a system that can cut the time and cost in this field is necessary. To assist in the process, the utilization of information technology, automated software and database can provide efficiency and effective solutions to the problems of mass data and information handling [1, 2].

Expert knowledge is often represented in the form of rules or as data within the computer. A rule- based system consists of IF-THEN rules, facts, and an inference engine controlling the application of the rules based on the facts [3]. Rapid Application Development (RAD) was developed initially by James Martin in the 1980s [4]. The RAD approach involves conciliation in usability, features, and execution speed. In the design phase through RAD, information gathered from the requirement and analysis are defined in a visual representation of the system. The system design includes developing action diagrams that define the interactions between processes and data and workflow of the intended system. The workflow and procedures of the job vacancy application module by deploying four types of UML diagrams include Use case, flowchart, sequence diagram and state diagram.

The Rule-based technique, also called Knowledge-based technique is a conventional rule-based expert systems using human expert knowledge to solve real world problems that normally would require human intelligence[5]. Online Job Application (OJA) system also used Java Server Pages (JSP) a server-side technology used to make the HTML more functional, and used in dynamic database queries. A Decision Support System (DSS) is a group of computer-based information systems including knowledge based systems that support decision making activities [6]. There are many approaches in DSS; one of them is Artificial Intelligence (AI).

The Smart Evaluation on Job Vacancy Application System (SEJVAS) based on Rule-based expert system technique. This technique is applied to check the applications for applicants that fulfill the minimum qualification for the job and eliminate the not. The rule-based technique is also able to sort the suitable candidates according to the degree to which he/she is more suitable for the position using merit value. Hence, produces a more reliable decision-support results from which the admin can select the candidate to call for an interview. The Smart Evaluation for Job Vacancy Applicationwas developed for Bina Integrated Technology Sdn. Bhd.

2. Methodology

The objective of Smart Evaluation for Job Vacancy Application is to develop a successively running web application for the back-end user only [7]. Knowledge base contains knowledge concerning the problem-solving area of domain. Knowledge of the domain is presented in IF (antecedent) THEN (consequent) form. The rules in the knowledge base are retrieved from the Position requirement table. The Database contains the database whereas to match against the rules stored in the Knowledge base. The data in the Applicant's table is weighed against the data in the knowledge base. When the antecedent part of a rule is satisfied by a fact, the consequent is said to be fired. For each antecedent that is fulfilled, a consequent of merit or weight is assigned. The inference engine carries out the reasoning. The applicant's database is screened and applications that do not fulfill the minimum qualifications are dismissed from the workflow. While applications that fulfill the minimum requirements of the position are then sorted according to the degree to which they are qualified. The system then displays the decision-supported results of the process. A system admin then selects which applicants are most suitable candidates for the job. The applicants who are not suitable are discharged from the workflow.

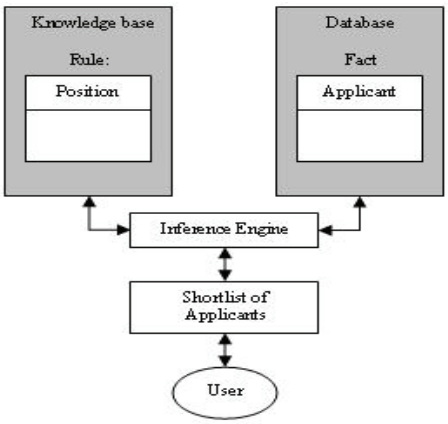

Figure 1. Block diagram of the Smart Evaluation for Job Vacancy Application system

2.1 Rapid Application Development

Rapid Application Development thus enables quality products to be developed faster, saving valuable resources. The RAD process model consists of five (5) phases, which are: Requirement and analysis; Design; Implementation; Testing; and Deployment RAD is a methodology for compressing the analysis, design, build, and test phases into a series of short, iterative development cycles. Iteration allows for effectiveness and self-correction. Studies have shown that human beings almost never perform a complex task correctly the first time. However, people are good at making an adequate beginning and then making many small refinements and improvements. This has a number of distinct advantages over the traditional sequential development model which has a rigid order of stages.

Unified Modeling Language (UML) is a standardized specification language for object modeling. A UML model is a graphical notation used to create an abstract model of a system. The workflow and procedures of the job vacancy application module is defined using UML diagrams;

i. Use case
ii. Sequence diagram
iii. State diagram
iv. Flowchart

2.1.1 Use Case

Figure 2 shows the interaction between the Job Vacancy Application module use case and the actors. The actors consist of two (2) actors which are; Applicant which is the class of all applicants who apply for the job and Admin which refers to the system admin.

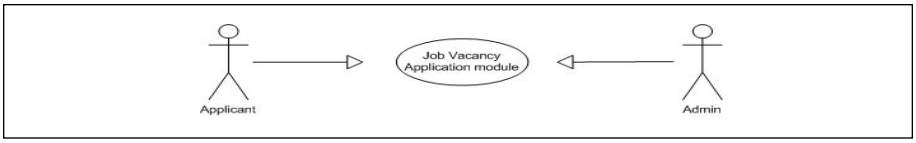

Figure 2. Use case diagram for Job application process

The Job vacancy application use case is initiated by the applicant where the applicant submits his/her application to the system. The Job vacancy application module screens the applications and rules out applications that do not qualify with the minimum qualifications. The Job vacancy application module then sorts the possible candidates according to the criteria required for the job. The Job vacancy application module then delivers the decision-supported results to the system admin where the admin selects which applicant is most suitable for the job. The job vacancy application starts when an applicant submits an application. The job vacancy application module automatically screens all the applications for a suitable applicant that fulfill the minimum requirement for the job such as; minimum CGPA of 3.00, minimum work experience more than 1 year, etc. Applications that do not fulfill the minimum qualifications are dismissed from the workflow. While applications that fulfill the minimum requirements of the position are then sorted according to the degree to which they are qualified. The system then displays the decision-supported results of the process. A system admin then selects which applicants are most suitable candidates for the job. The applicants who are not suitable are discharged from the workflow. The flow chart ends with the suitable candidates are sent an email to call for an interview.

2.1.2 Sequence diagram

One of the types of UML diagram is the sequence diagram. A sequence diagram shows different processes or objects that run simultaneously as parallel vertical lines and the messages exchanged between them in the order in which they occur [2]. Figure 3 displays the sequence diagram for the job vacancy application module.

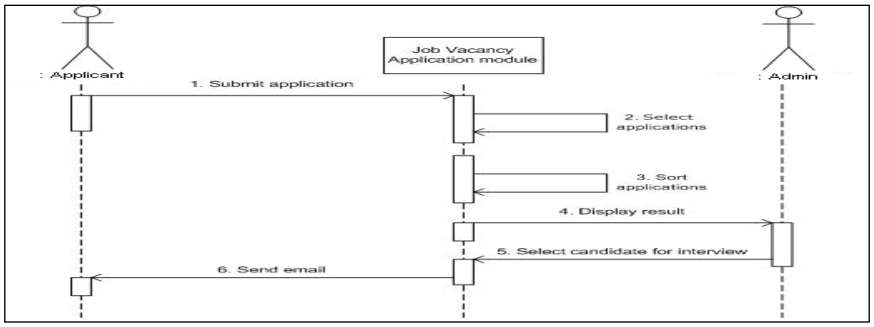

Figure 3. Sequence diagram for Job vacancy application module

This is the scenario for the job vacancy application process. The list of sequence from the diagram is:

a. The Job vacancy application module screens the applications submitted and dismiss applications that do not fulfill the minimum qualifications.

b. The Job vacancy application module sorts the competent applications.

c. The Job vacancy application module displays the results to the administrator.

d. The Administrator selects the most suitable applicant for an interview.

e. The Job vacancy application module sends an email to the interview candidate to call for an interview.

2.1.3 State Diagram

The state diagram, also known as the activity diagram in UML 1.0, is categorized under the Behavior diagrams in the UML diagram which emphasize what must happen in the system being modeled. A state diagram describes all of the possible states of an object as events occur. Figure 4 represents the state diagram for the job vacancy application module.

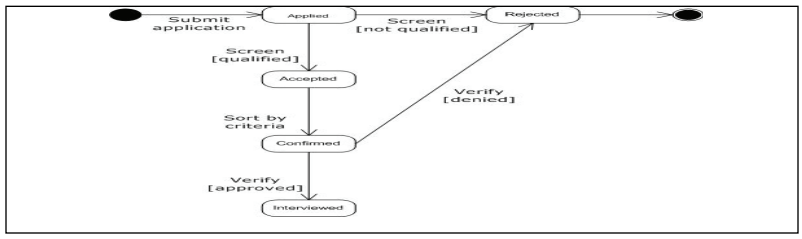

Figure 4. State diagram for Job vacancy application module

The arrows in Figure 4 represent transitions, from one state progressing to another one state. Transitions are the result of the invocation of a method that causes an important change in state. The filled circle refers to the initial state of an application. However, when an application in the Applied state is screened, it can either change into the Rejected state or the Accepted state. An application is in the Rejected state when it is not qualified in the screening process using the Rule-based technique. The application in the Accepted state change to the Confirmed state when it engages with the sort by criteria method where the Confirmed state is verified to change into the Interviewed state if approved or Rejected state if denied. The hollow circle containing a smaller filled circle indicates the final state of the application.

2.2 Rule-based Expert System Technique

The Rule-Based Expert system technique is the shell of knowledge-based systems. By associating the initiated rules with the database objects it is possible to build models and to apply these models to real world problems. Expert knowledge is often represented in the form of rules or as data within the computer. A rule-based system consists of IF-THEN rules, facts, and an inference engine controlling the application of the rules based on the facts.

3. Proposed Method

SEJVAS is developed for the use of the back-end user only, which is the administrator. The administrator can initiate requirements in accordance with the weighting of requirements and the system will examine all applications for the ones that are most suitable for the job [8]. The Rule-based expert system technique was implemented into the system using PHP language. The facts to be matched up against the rules were the applicant's data, whilst the rules were retrieved from the position's *requirements data*. The rules in the knowledge base were partitioned into six (6) categories, which are:

3.1 Prior rules

The prior rules are the rules that if not satisfied, the respected application will be eliminated from the workflow. The IF antecedent include; nationality, minimum and maximum age range and availability date. For an example, if an applicant does not fulfill the Malaysian nationality rule, the applicant will be removed from the process.

3.2 Highest qualification rules

In this category, and the categories after, the applicants have already fulfilled the prior rules requirement. The qualification rules are the rules concerning the education level of the applicant. This category includes three (3) entities, which are qualification field, qualification level and grade/CPA.

3.3 Required skills rules

In the required skills rules category, the IF domain includes the skills required for the position and its level of proficiency. For example, the required skill for an enterprise application developer position is MS Sharepoint, and the proficiency level is intermediate.

3.4 Preferred languages rules

In the preferred language rules category, the antecedent consists of the languages necessary for the position and its level of proficiency of 0 to 10. For example, the preferred language for a mandarin columnist position is Mandarin with a spoken Mandarin proficiency of 10 and a written Mandarin proficiency of 10.

3.5 Work Experience rules

Work experience category comprises of three (3) entities, which are; work experience, company industry and position level.

3.6 Salary rules

The minimum salary rule, the applicant's requested salary is measured up to the minimum salary for the position.
A flowchart diagram is a graphical notation of a system's workflow. Figure 5 illustrates the workflow of the job vacancy application use case.

The antecedent and consequent in each category in the knowledge base is defined by an array, and each array carries a merit/weight that is assigned according to the degree to which an applicant conforms to the position's requirements. At the end of each category, the total of the merit is calculated. For an example, in the highest qualification rules category;

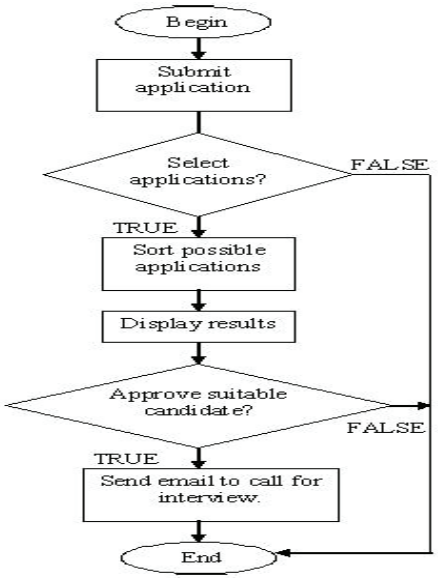

Figure 5. Flowchart for job vacancy application module

```
if ( $applicant_qualification_field = = $position_qualification_field ){ $point['field']=10; }
if ( $q_field != $row['qualification_field'] ){ $point['field']=5; }
if ( $applicant_qualification_level >= $position_qualification_level ){ $point['level']=10; }
if ( $q_level < $row['qualification_level'] ){ $point['level']=5; }
if ($applicant_qualification_grade >= $position_qualification_mingrade ){$point['grade']=10; }
if ( $applicant_qualification_grade < $position_qualification_mingrade){ $point['grade']=5; }
$qualification = (( $point['field'] + $point['level'] + $point['grade'] )/30)*20;
```

Figure 6. Antecedent and consequent

At the end of the workflow, each applicant's degree of suitability to the position is calculated in percentage. Figure 7 shows an example of the calculation. After the percentage of each applicant is calculated, the suitable candidates are sorted according to the degree to which they fit the position.

```
$percentage = $qualification + $work + $skills + $languages + $salary;
```

Figure 7. Total of the merit is calculated

4. Result

This system was developed using several applications and languages including; HTML, CSS, JS, and PHP using Editplus, MySQL for the database, Adobe Photoshop CS2 for graphical design, MS Visio for planning and MS Word for documentation. The hardware specification needed in the development of the Smart evaluation for job vacancy application system, Pentium M 1.73MHz and above, 1GB RAM, Minimum 4 GB free hard disk space, Network Adapter / Modem, DVD Writer 52X. There are three (3) tables created for this system; Admin, Position and Applicant. Figure 8-11 show SQL queries executed in creating these three (3) tables. Figure 5 refers to the new position page. In this page, the Bina web administrator is able to add new positions to offer at Bina Intergrated Technology.

Figure 9, Bina's web administrator is able to view all available positions posted. On this page, the admin can edit the position requirements, edit status of the position and delete the position. The admin can edit the position by clicking on the position title link where he/she will be redirected to the edit position page. The admin can also edit the status of the position (Vacant, In process, Closed). By default, the status of the position is 'Vacant'. Finally, the admin can delete the position by clicking on the 'Delete' icon.

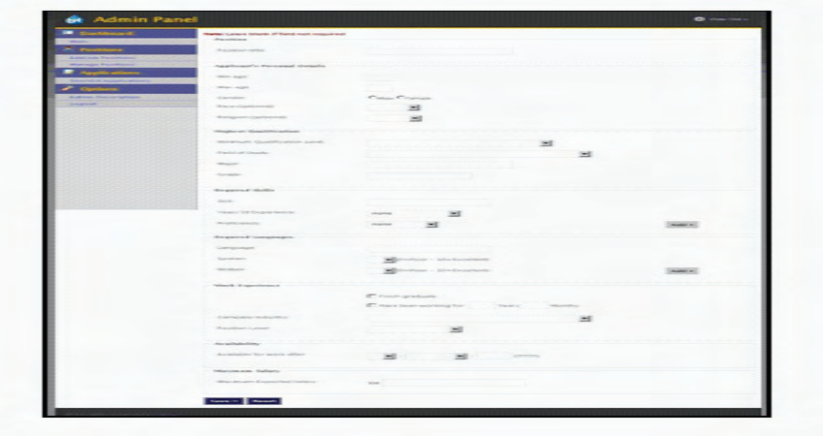

Figure 8. Add new position

Figure 9. Manage positions

Figure 10, the admin selects a position title from the dropdown menu and clicks the "Go" button. Then, the system will display all the applicants that applied for the respective position. There are many functions in this page. The admin can view the text resumes of the applicants by clicking on the applicant's name link. The admin can also skip the shortlist process and mark the applicants he/she want to call for an interview and click the Email button where an automatic email will be sent to the applicants. The admin can also rate the applicants where a certain leverage is given to the applicant when the short listing process. Finally, the admin can also mark and remove the applicants.

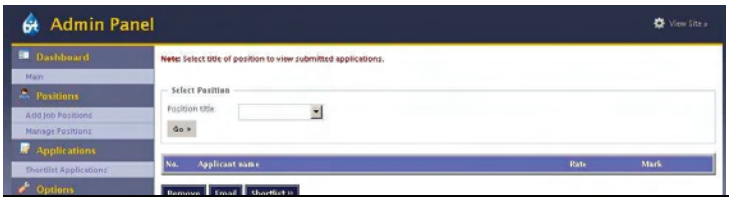

Figure 10. Shortlist applications – select position

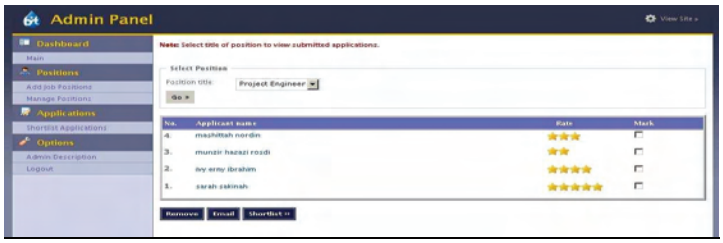

Figure 11. Shortlist list of applications for a position

5. Conclusion

SEJVAS is a web application that applies rule-based decision making to serve the purpose of cutting time and cost of a job application evaluation procedure. Current manual employee evaluation mostly relies on file-based method which is inefficient and disorganized. The process of screening these applications for the most suitable for the interview is tedious and takes a lot of time. It is particularly time-consuming and tiresome for jobs that receive a wide range of applications. Having to go through these applications one by one requires a lot of patience and professionalism. That is why most employee hiring decisions are often biased by personal opinions and emotions. Consequently, hiring the wrong person for the job will result in a waste of resources or even lose a lot of money.

References

[1] Moody, Daniel, L. (1998). Metrics for evaluating the quality of entity relationship models". *Lecture Notes Comput. Sci*., p. 211-225. DOI: 10.1007/b68220

[2] Noraziah**,** A., Norhayati, R ., Abdalla, Ahmed N., Roslina, A.H. Noorlin, M.A. Affendy, O.M. (2008). A Novel Database System Model Design for Tender Management System, *Journal of Computer Science*, Science Publications, USA, p. 463-466.

[3] Negnevotsky, M. (2005). *Artificial Intelligence: A Guide to Intelligent Systems.* Second Edition, Addison Wesley, England.

[4] Sommerville, I. (2005). *Software Engineering*. Seventh Edition, Addison Wesley, United States of America.

[5] Lisnyak, T. (2002). *An Agent-Based System for Intelligent Internet Shopping*, Master Thesis, University of Cincinnati.

[6] Turban, E., Aronson, J. E.(2001). Decision Support Systems and Intelligent Systems, Prentice Hall.

[7] Wikipedia. (2007). Software testing, http://en.wikipedia.org/wiki/Software_testing, accessed on 14 July 2007.

[8] Kingsbury, N. G. (1998). The dual-tree complex wavelet transform: A new technique for shift invariance and directional filters, Proceedings of the Eighth IEEE DSP Workshop, Utah.

Knowledge Based Flexible and Integrated PLM System at Ford

M.B. Raza[1], T. Kirkham[2], R. Harrison, Q. Reul[3].
[1]Loughborough University.
Loughborough, LE11 3TU, UK

[2]The University of Nottingham.
Nottingham, NG7 2RD, UK
Thomas.Kirkham@nottingham.ac.uk

[3]Vrije Universiteit Brussel
B.R.Muhammad@lboro.ac.uk

ABSTRACT: *This paper reviews the perennial problem of knowledge management / information integration in large scale, complex and knowledge intensive organisations such as automotive industries. Automotive industry is under increased pressure to produce low cost customised products using innovative agile manufacturing techniques. Presently this innovation has focused on the improved process development between different stages of Product Lifecycle Management (PLM). However in terms of implementation the application data management techniques have lagged behind leaving these processes disjointed and lacking in automation. Assembly line design and configuration consist of highly creative and complex tasks that involve extensive communication and information exchange among distributed teams. At Ford the assembly line design or reconfiguration process rely on PLM system to provide necessary information. This paper proposes an improved model based on innovation in the PLM to quickly adapt to the new feasible assembly line configuration that satisfies the ever changing user requirements. Building on existing work in the use of ontologies for knowledge management, the paper applies these techniques to PLM system. The implementation has been first applied to a prototype rig and then around a Ford production line in UK to efficiently exploit PLM systems using a state of the art web service infrastructure based upon ontology.*

Keywords: Product Lifecycle Management, Knowledge Management, Ontology, Automative Industry

Received: 21 September 2009, Revised 12 November 2009, Accepted 28 November 2009

1. Introduction

Agile manufacturing in vehicle assembly operations requires rapid configuration and/or re-configuration of assembly lines to support high levels of customisation in product design and manufacture. The adoption of this type of specialised production is vital for the future of manufacturing in developed economies [1]. At Ford UK, Loughborough University team has been studying relationships between the product design and the production line design phases. It is concluded that the information of product design needs to be quickly adapted to machine and line creation. This can be achieved by greater integration of production and enterprise knowledge into the manufacturing processes. However in complex, large scale production environments legacy systems and vendor specific technologies exist and persist. These systems slow down and break up the integration process, making it hard to achieve enterprise level agile processes to support manufacturing/ assembly processes.

The Business Driven Automation (BDA) project [19] at Loughborough University in partnership with Ford Motor Company UK focuses on addressing these challenges. The research explores the use of ontologies and recent advances in the semantic web technologies in factory automation systems over different lifecycle phases of products. This contribution proposes a method by which knowledge can be better managed in automated production systems using the Powertrain automated manufacturing environment as an example. Ontology based semantic technologies facilitate the suitable integration of disparate knowledge so that it is reusable by legacy applications.

2. Related work

2.1 Product Lifecycle Management (PLM)

Although Product Data Management (PDM) and PLM systems have significantly improved manufacturing efficiencies still significant limitations exist in terms of integration and right information retrieval tasks. PLM has its roots in the initial integration of applications with the manufacturing design process. Data created from CAD software has facilitated designers in the electronic creation, reuse and manipulation of product models [2]. The integration of design data in PDM systems has emerged to present a means by which distributed access to design data from design teams can be achieved [3]. However this combination of electronic design data and localized software management has proven inadequate for demands of increasingly streamlined business processes. Pressure soon formed on PDM systems to integrate with other elements of the enterprise including non-engineering areas such as sales, marketing and supply chain management [4].

Data used in PDM has therefore moved from a focus on product design to a need to present this data in order to enhance the wider manufacturing processes [5]. Thus PDM can now be seen as a legacy element of a wider PLM process that encompasses all elements of the product manufacturing process including processes and resources.

To deal with such complexity, PLM has largely developed around vendor specific or project / product specific environments [6]. PLM distinguishes itself from other enterprise application systems such as ERP, SCM and CRM by presenting ways to enable effective collaboration among networked participants specifically in the product design process [7, 8]. The enablement of PLM integration within and cross organisations in recent years has seen it move into the domain of service orientated computing [9]. Using web services, the concept of creating a loosely coupled PLM environment is seen as vital to support increasingly globalised manufacturing processes [10].

However, the shift to more flexible PLM implementations has been a challenge to both data integration and management. Current PLM systems though more flexible and promising to PDM, turned out to be document oriented, vendor specific and data management systems rather than knowledge management. Even using PLM, flawed coordination among teams, systems and data incompatibility and complex approval processes are common [11,12].

2.2 Knowledge Management and Engineering

Ontologies are often viewed as allowing more complete and precise domain models [20]. An ontology is commonly defined as: "a formal, explicit specification of a shared conceptualization" [22]. More specifically, an ontology explicitly defines a set of entities (e.g. classes, properties, relations and individuals) imposing a structure on the domain that is readable by both humans and machines. As a result, the domain knowledge represented in ontologies assists greater information sharing and re-use. Ontologies are developed and used because they enable among others [24]:

- to share knowledge - by sharing the understanding of the structure of information shared among software agents and people
- to reuse knowledge - ontology can be reused for other systems operating in a similar domain
- to make assumption about a domain explicit - e.g. for easier communication

Within multi-faceted complex production environments the use of ontologies has great potential to aid knowledge management. Ontologies are not only useful for achieving semantic interoperability on the web but also to coordinate a range of disparate expertise for large organisations. More specifically, it will enable different communities to infer the same meaning when information/ knowledge is exchanged across systems. For example, 'Rolls-Royce Company' now needs to coordinate information collection from various parts of the organisation since it has shifted its focus from selling products to providing services i.e. selling engine power instead of engine. In the 'Integrated Products and Services' (IPAS) project [21], ontologies of products and processes were developed allowing the information representation and sharing during servicing of Rolls-Royce engines.

Within Ford the same benefits can be realised through the use of ontologies particularly for an assembly line design / re-configuration. Vendor specific approaches to PLM and integration issues necessarily increase with the size of an organisation and ad-hoc integration often leads integration issues to re-emerge [13]. The PLM approach of working with existing systems and data formats may lead to the corralling of data in centralized repositories [14].

The rapid development of web services based computing has not only influenced the adoption of common standards to enhance business processes across enterprises, but has presented ways of integrating cross-organisational processes. In order to aid this integration, the concept of the 'semantic web' has been developed to better aid the integration of varying forms of distributed data using ontologies [15]. The purpose of re-configuration is to allow a manufacturing system to change rapidly and cost-effectively from its current to a new configuration without being taken off-line, maintaining system effectiveness

when product or production changes or breakdowns occur [23]. The integration of different sources of information in PLM application at Ford is achieved by means of a common vocabulary defined by an ontology. Linking the ontologies and semantic web services into PLM system will allow greater access to organisational data structures and improve processes and productivity at Ford.

3. Problem

3.1 Background

The PLM system at Ford is a complex aggregation of several domains working collaboratively to manufacture and assemble different variants of vehicles. PLM at Ford is managed using 'Teamcenter' that links product data from various CAD / CAM repositories. The machines on the line are designed collaboratively with external machine builders. The design format between the parties has to be agreed as part of this process and the final design is incorporated into Teamcenter. The third party companies don't have access to the Teamcenter repository and mostly rely on email based exchange of design data. By applying a more automated approach using SWSs, it is envisaged that both time and cost can be reduced.

3.2 Problem: Powertrain assembly systems

The Powertrain assembly plant and its relationship with PLM has been the core of the research. A typical Powertrain assembly process involves hundreds of individual parts and the impact of change in one part may cause a rippling effect in the whole assembly processes. A key role of PLM system should be to detect and manage this change and its effect which was found missing.

In addition, the current reconfiguration approach is largely based on the skill and knowledge of engineers rather than the actual process involved. Whenever there is any change in the product it is then essentially engineer's responsibility to examine the needs of the reconfigured system to support the new product.

4. Case study - implementation

The focus of this paper is rapid reconfiguration of assembly lines in Powertrain assemblage through the assimilation of PLM data. This has been achieved by integration of services into the PLM tool. To automate (fairly) this task, product (engine) and resource (line) link points i.e. dependency relationships need to be defined at early stages of design and made available to be searched, analysed and implemented on 'when and where required' basis. This level of integration will link the PLM system improvements to both the machine data integration and also to the enterprise computing applications at Ford.

4.1 Environment

The testing environment is constructed using Festo automation rig components supplied by Ford which use the same interfaces as the machines on the Powertrain line at Ford plant in Dagenham, UK. Control of the interfaces was linked to a web service enabled control application developed for the SOCRADES EU Framework 6 project [16]. Live data of execution from the line is available through web service interfaces on the control application. The Festo Rig can be seen in Figure 1.

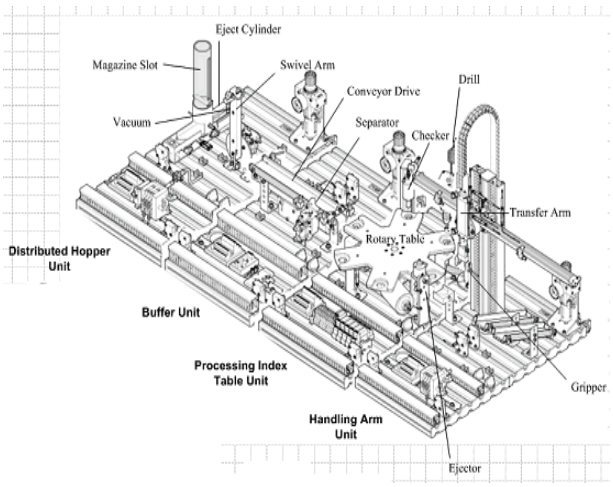

Figure 1. Festo Rig layout

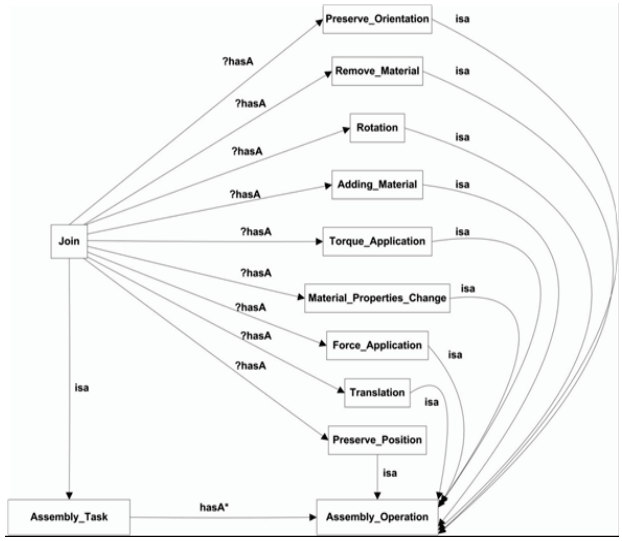

Figure 2. Assembly Process Captured as an Ontology

4.2 Ontology

A general layout of the line is captured as ontology and an instance of it is defined as 'Festo Rig' with the current layout of the prototype line where each component (independent work unit) has had its CAD data translated into ontology. This allowed the line components and layouts to be interrogated by SWSs through ontologies. The translation of product and resource data in Teamcenter (stored in multiple CAD formats) into the OWL format [18] is achieved via the use of ontology design tool Protégé [17]. A visualization of a simple ontology to represent a typical assembly task is illustrated in Figure 2.

The ontologies facilitate improvement compared to the previous method of human based re/configuration of the line. For example, if the user wants to query about all sensor elements being used on a specific workstation or characteristics of a workpiece required to carry out a successful assembly operation, can be answered with the help of ontologies and services.

Figure 3. Ontology of the Festo Rig

4.3 Service Interaction

In order to demonstrate the use of SWSs and the enhanced use of data in PLM, the rig was implemented with several supporting services. These supporting services provide the function to both support the line design process in PLM and also add live analysis of line execution using the data from the orchestrator linked to the rig. The main simplified elements of the system can be seen in Figure 4.

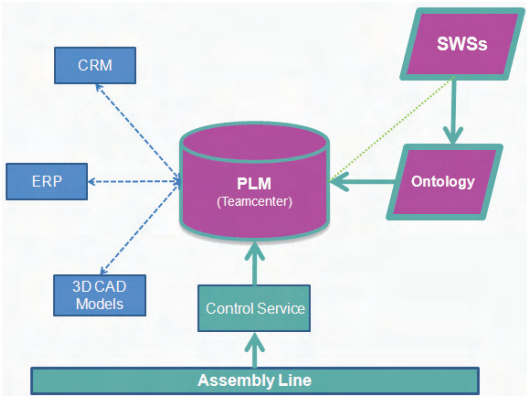

Figure 4. Main elements of implemented system.

Various collaborative designers, while using PLM, can access the requisite heterogeneous data by using a semantic query which would target PLM system through ontology as shown above.

4.4 Constraint Checking

Central to the use of the ontologies is the ability for rules to be conducted on them. For example if a new product is added to the line with an increased weight it is imperative that the components in the line can support that weight. Therefore a key aim of the services is to quickly check these new constraints and areas where redesign is needed. The core constraints defined in the ontology are the product vs. resource dependency relationships. These assembly constraints are represented explicitly using OWL triples and SWRL rules. For example, assembling crankshaft with block, this assembly operation is stored in PLM and can be readily accessed by querying through service and checked against defined rules and constraints for changed product as shown in Figure 5.

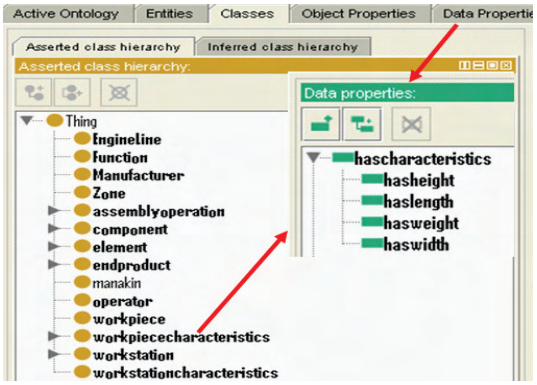

Figure 5. Rules and dependency relations in ontology

5. Use cases

The framework implemented had two main use cases demonstrating the use of semantic web services for knowledge capture and it use with the production process.

5.1 Live System

The live system uses PLM services to notify events from the line to shop floor engineers / interested parties such as if there is a jam on a conveyor or the current load level of the indexing table exceeds etc. This information can then be used by the PLM user to aid in the diagnosis of errors in the assembly line. If there are errors the PLM service uses the line ontology to find a remedy to treat the error. This process automates some of the response using the knowledge from Teamcenter (via the ontology) in a standardised way which previously was provided by a production engineer.

Using the ontology the PLM services can instruct the control mechanism to notify dependant stations that an error has occurred on the line and even request a halt in the production. In the case of a multi routed system the PLM could use the SWRL rules

defined in ontology of the line to diagnose a possible alternative route for the workpiece. Notifications to engineers can be enhanced with the PLM services of the line from the ontology and live line data.

Apart from the direct benefits on the assembly line, the use of the ontology outside of the line will enable the notification of other appropriate services in the supply chain. For example in live system the ontology may be used to order a replacement part for the line by interrogating the components affected by the error. A supporting knowledge base of previous faults linked to probable causes could aid in this process and potentially enhance production output.

5.2 Production Reconfiguration

Engine assembly line is a highly sophisticated and complex combination of sequential operations and activities which are mostly automatic. One of the aims of this research study was to develop methodologies to facilitate Ford company to visualise, model and re-configure new/changed assembly line fairly automatically for building new/changed engines. The emergence of SWSs opens new prospects for integrating a wide array of manufacturing resources in a cost effective manner which is implemented to reconfigure the line to support a new product. Here the line ontology in its current live configuration is used alongside ontologies of the product and equipment. An illustration is shown in Figure 6 where screwing equipment and its ontology are captured. This concept was used to capture knowledge of all the equipment in a particular zone of the assembly line.

Based upon the concepts shown in Figure 6, an ontology of a complete zone of Powertrain assembly line with several work-stations was captured and tested for changed product scenarios, a snapshot of the ontology in graphical form in Protégé is shown in Fig. 7.

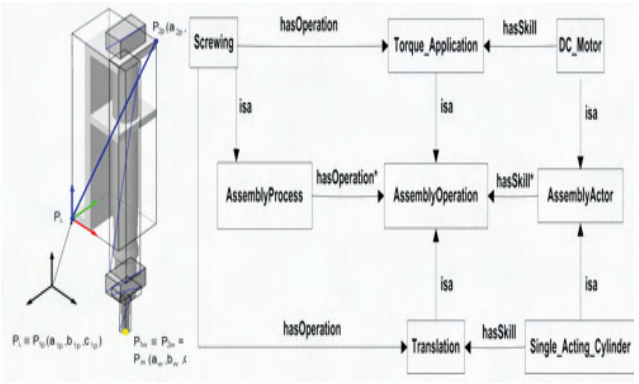

Figure 6. Screwing equipment vs screwing process ontology

Figure 7. Ontology of the engine assembly line

Key concepts introduced into the ontology of the line are work piece, work piece characteristics, workstation, workstation characteristics, assembly operation, OEM, operator and end product. Based upon these concepts, relations among products, processes and resources are established into the ontology. For example, a certain workstation performs particular assembly tasks on specific products to achieve a definite objective. With the help of this knowledge in ontology, a quick evaluation of many potential configurations is possible as well as the best suited one for a changed product.

5.3 Testing

An initial version of the PLM service was demonstrated on the Ford rig at 'EU IST 2009' event in Lyon, France. Here the rig was installed and the PLM services were linked to the control application (Field Transfer Block- FTB) of the rig. A user interface was created to allow the compatibility of new products when applied to the line to be assessed. The report illustrated the points on the line that would need to be reconfigured to support the new product using the ontology of the line. A separate ontology was created to manage the knowledge associated with line errors in order to improve both response time and error detection during commissioning. Again this was demonstrated at the event using a user interface that communicated with the PLM service.

An error case was set up on the rig that brought both of the processes together where a fault on the line was detected and an appropriate response was selected by the PLM service. This response involved the selection of a new line for the product and the matching that was needed along with the appropriate notifications to ERP, engineers etc. In terms of functionality this demo was rather basic but does demonstrate the core concept that both configuration and error responses to events within a Ford production line can be enhanced by the use of knowledge captured using ontologies.

5.4 Evaluation

Currently the complex process of designing a line for a new product is both a manual and unpredictable process. The commissioning phase takes a large amount of time as the new configuration is slowly tested and errors are ironed out. The current commissioning process can be seen in Figure 8 with caption "AS-IS".

The main problems in the process can be seen as the large amount of time taken between line specification and launch along with the unexpected costs of the process. A key aim of PLM at Ford has been to improve the speed and reduce costs by improving the efficiency of the line commissioning process. However the structure of the process can be seen to remain the same.

Using services to the process can be enhanced using knowledge from the PLM service linked to ontologies. Using ontologies the new product specification can be compared to the line layout in a more automated fashion. With the help of ontology, this process is becoming smoothened and helping Ford engineers to perform parametrical relationship analysis between engine and workstation with relevant assembly processes through ontology.

For example this approach will automatically pick out any issues with the dimensions of the product and the dimensions of certain elements in the line. These issues can be sent to the product designer and changes can be made before the commissioning process starts. As illustrated in Figure 8, caption "BDA" this reduces process and product development time and costs, it also influences the smoothing of the unexpected costs in the reconfiguration process.

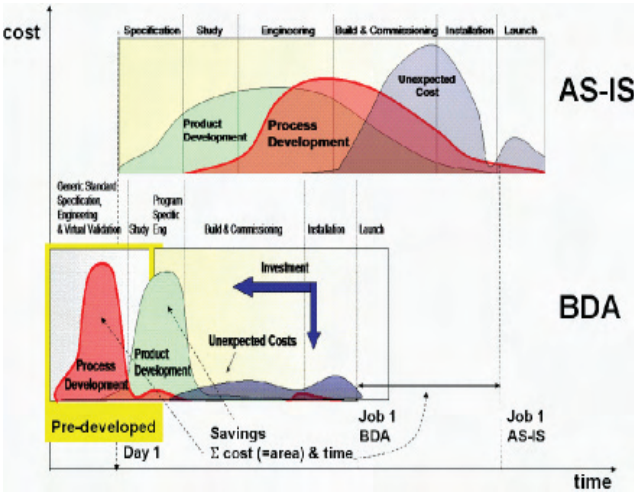

Figure 8. Process improvements with knowledge-based PLM

It is envisaged that adding knowledge to the PLM system the unexpected errors in the line commissioning process could be predicted in a more stable way. Thus rapidly increasing the time to develop and commission the product and line. The improvements illustrated in Figure 8 will be achieved using services to analyze ontologies of the line and the components within it against new product designs in an automated manner.

6. Future work

The paper advocates the migration from labour intensive integration techniques to more sophisticated ontology based services in PLM systems to result intelligent information integration and knowledge reuse. The development and the use of ontologies to aid the PLM system in Ford are currently focused on a use case at a specific Ford plant.

By taking this approach the use of ontologies to express knowledge in the system and aid integration in the PLM process will be tested in far more robust way than illustrated in this paper. As a result of these investigations it is the aim of the research in the final part of the BDA project at Loughborough University to produce a PLM approach using ontologies to apply to other areas of Ford and possibly through multiple ontologies based upon hybrid approach of data integration. In the next phase of the project, we will work on updating the ontology automatically as the line or product changes. New concepts, properties or values of properties will be extracted from the legacy systems such as PLM systems and added to the ontology automatically so the ontology will be dynamically updated. Also, we are planning to study whether ontology and its graphical components can be integrated into PLM systems to get the benefits of both.

Ontologies for real world geographically distributed applications could be quite complex and need to be modularized as the perspective increases. This modularity would help in easy maintenance and updating of ontologies.

To apply the results of the project in other enterprises or to existing areas of Ford production a retrospective approach at ontology capture needs to be defined. This is a key area of research as the PLM system is a vital source of data for automotive industry which is being under-used due to difficulties at integrating the information stored within it.

7. Conclusion

This paper describes how existing PLM systems can be used as a Knowledge Management (KM) tool to solve the semantic interoperability problem of heterogeneous data. The main objective of PLM as a KM tool is to improve the capabilities of technology intensive organisations to monitor and respond to technological and product changes. Using knowledge based services a new layer of manufacturing management can be envisaged that will aid the entire production lifecycle. Large amounts of product and machine component data exists at Ford in under-utilised databases due to the inability of existing integration approaches to systematize and relate the available knowledge.

The development of a series of ontologies to both represent and capture this data will rapidly improve the production process in large scale manufacturing/assembly processes. This approach allows services such as in the case of PLM to better understand product and line design allowing this data to feed automated processes to aid agile manufacturing. While some basic concepts are proved successfully, room for improvement is acknowledged. The initial aim of our work is to prove the concept by introducing and exploiting domain ontologies which we believe have been demonstrated successfully at Ford. Further experimentation, larger KBs, and other legacy databases / applications are needed to test and improve the service based PLM concept.

8. Acknowledgements

The authors gratefully acknowledge the support of 'UK EPSRC' and Loughborough University's 'IMCRC' through Business Driven Automation (BDA) project. We would like to thank all the project participants and engineers who have contributed in this research from Loughborough University and participant companies especially Ford Motor Company, UK.

References

[1] Maskell, B. (2001). *The Age Of Agile Manufacturing, Supply Chain Management: An International Journal*. 6.

[2] Bryant, R., K.-T. Cheng, A. Kahng, K. Keutzer, W. Maly, R. Newton, L. Pileggi, J. Rabaey, A. Sangiovanni-Vincentelli. 2001*Limitations and challenges of computer-aided design technology for CMOS VLSI. Proc DODEN IEEE PAD* 89(3) 341–365.

[3] Philpotts, M. (1996). An Introduction to the Concepts, Benefits and Terminology of Product Data Management, *Industrial Management & Data Systems*, 4. 11-17

[4] Pikosz, P., Malmqvist, J. (1996). Possibilities and Limitations when Introducing PDM Systems to Support the Product Development Process, *In:* Proceedings NordDesign'96, p 165-175.
[5] Crnkovic I., Asklund U., Persson-Dahlqvist A., (2003). Implementing and Integrating Product Data Management and Software Configuration Management. Artech House,
[6] Ameri, F., Dutta, D. (2005). Product lifecycle management: closing the knowledge loops, *Computer-Aided Design & Applications*, 2 (5) 577-590.
[7] Morris, H., Lee, S., Shan, E., Zeng, S. (2004). Information integration framework for product life-cycle management of diverse data, Journal of Computing and Information Science in Engineering, Transaction of the ASME **4**. 352–358.
[8] Westkaemper, E., Alting, L., Arndt, G. (2000). Life cycle management and assessment: approaches and visions towards sustainable manufacturing. Ann. CIRPÐMfgTechnol. 49 (2) 501-522.
[9] Kimura, F., Kato, S. (2003). Life Cycle Management for Improving Product Service Quality, University of Tokyo, Tokyo, Japan.
[10] Yaoguang, Hu., Rao, Wang. (2008). Research on collaborative design software integration based on SO.A" *Journal of Advanced Manufacturing Systems*. 7 (1) 91-94.
[11] Qui, R. G. (2007). A Service-oriented Integration Frame-work for Semiconductor Manufacturing Systems, *International Journal of Manufacturing Technology and Management.* 10(2/3) 177-191.
[12] Ming, X. G., Yan, J. Q., Wang, X. H., Li, S. N., Lu, W. F., Peng, Q. J., Ma, Y. S. (2008). Collaborative process planning and manufacturing in product lifecycle management, *Comput. Ind*. 59 (2-3) 154-166.
[13] Kopácsi, S., Kovács, G., Anufriev, A., Michelini, R. (2007). Ambient intelligence as enabling technology for modern business paradigms. *Robot. Comput.-Integr. Manuf.* 23 (2) 242-256.
[14] Lund, J. G. (2006). The storage of parametric data in product lifecycle management systems. Masters Thesis, Brigham Young University.
[15] Ramdas, K. (2003). Managing Product Variety: An Integrative Review and Research Directions, Production & Operations Management.
[16] Berners-Lee, T (1998). Semantic Web Road Map. *http://www.w3.org/DesignIssues/Semantic.html , last accessed 01/07/09*
[17] Socrades homepage: www.socrades.eu *last accessed 01/07/09*
[18] Knublauch, H. Fergerson, R.W. Natalya. Noy, F. Musen, M.A.(2004). The Protege OWL Plugin: An open development environment for semantic web applications. *In*: 3rd International Semantic Web Conference (ISWC 2004), Hiroshima, Japan.
[19] Bechhofer, F. Harmelen, J. Hendler, I. Horrocks, D.McGuinness, P. Patel-Schneider, et al. (2003). "*OWL Web Ontology Language Reference*." W3C Proposed Recommndation, from http://www.w3.org/TR/owl-ref/
[20] Business Driven Automation Project Home page: http://www.lboro.ac.uk/departments/mm/research/manufacturing-systems/dsg/index.htm *last accessed 01/07/09*
[21] Huhns, M., Singh, M. Ontologies for agents, *Internet Computing* 1 (6) 81-83.
[22] Fowler, D. W. Reul, Q., Sleeman. D (2008). IPAS ontology development, *In*: Proceedings of the 3rd International Workshop on Formal Ontology Meet Industry Workshop (FOMI 2008), p. 120-131, Torino, Italy, June.
[23] Gruber, T. R. (1993). Towards Principles for the Design of Ontologies Used for Knowledge Sharing. *In: Formal Ontology in Conceptual Analysis and Knowledge Representation*, 907-928
[24] Kordic V., Lazinica, A., Merdan, M. (2005). Future of Manufacturing: Concepts of Autonomy and Self organisation, *International Journal of Advanced Robotic Systems*, 2 (1) 12.
[25] Obitko, Marek ., Marik, Vladimir (2002). Ontologies for Multi- Agent Systems in Manufacturing Domain, *In*: Proceedings of the 13th International Workshop on Database and Expert Systems Applications (DEXA'02), IEEE.

IPOMS: an Internet Public Opinion Monitoring System

Jie Ding, Jungang Xu
School of Information Science and Engineering
Graduate University of Chinese Academy of Sciences
Caixa Postal Beijing 2707
China
{dingjie.gucas, xujungang}@gmail.com

ABSTRACT: *In this paper, an IPOMS (Internet Public Opinion Monitoring System) is proposed. This system can collect web pages with some certain key words from Internet news, topics on forum and BBS, and then cluster these web pages according to different 'event' groups. Furthermore, this system provides the function of automatically tracking the progress of one event. With this system, supervisors can know what is exactly happening and what has happened from different views, which can improve their work efficiency a lot. This system is composed of web crawler, html parser and topic detection and tracking tool. Because of the existence of numerous data in web pages, in order to improve efficiency of Internet public opinion analysis, the technologies of web page cleansing and k-d tree algorithm in topic tracking are adopted.*

Keywords: Public opinion, Clustering, Topic tracking, Web page cleansing, k-d tree

Received: 10 September 2009, Revised 30 October 2009, Accepted 9 November 2009

1. Introduction

Public opinion refers to the society and politics attitude toward the social administrator in certain social space [1]. Public opinion online is called Internet public opinion.

There are three characteristics of Internet public opinion. First, Internet public opinion emerges rapidly, with great influence on society. Second, there are large amount of comments to the relevant news and hot topics. Third, different people may have different opinions on the same event due to their different position, personal quality and breakthrough point.

Owing to the above three points, there are three kinds of requirement of monitoring Internet public opinion. First, New public opinion should be found quickly. Second, the progress and change of public opinions should be tracked. Moreover, history and present public opinion situation should be displayed in various formats, so that the supervisor can analyze, research and judge them.

There are several challenges in monitoring Internet public opinion. Technically, it is difficult to obtain a large number of relevant web pages rapidly at first. It is also difficult to judge the relevant degree of two text sections. Moreover, it is difficult to process the public opinion change quickly, because the web pages are enormous in quantity and processing speed is one bottleneck.

Traditional way that public security department processes Internet public opinion is manual, which results in waste of manpower, limited speed of information processing, relatively narrow goal range to research and judge, single form, slow response speed, hard to find relations among information.

In this paper, one new way of monitoring Internet public opinion is proposed, and the corresponding system is designed and realized.

The remainder of this paper is organized as follows. Section 2 discusses the source and concepts of topic detection and tracking. Section 3 describes the architecture of the system. Section 4 proposes the features of the system. Section 5 discusses system evaluation. Section 6 summarizes current work and outlines future work.

2. Topic detection and tracking

Topic Detection and Tracking (TDT), originally introduced by Defense Advanced Research Projects Agency (DARPA), is a research program concerned with organizing a stream of broadcast and print news stories by the events that they discuss [2]. TDT encompasses several tasks, but one of them requires that a system gather arriving news stories into clusters that correspond to real-world events. That task is known in the community as either 'cluster detection' or just 'detection'.

Topic detection and tracking has been widely studied for years [2][3][7][8]. CMU (Carnegie Mellon University) and Umass (University of Massachusetts) have carried on similar research, and have obtained the positive appraisal [2] [3] [4].

2.1 Event detection

Event detection can be defined as 'discovering new or not found event in continuous bunch of news' [4], divided into retrospective detection and online detection.

2.2 Event tracking

The purpose of event tracking is to sort out following text in previous event [2]. It is a kind of application with categorized files. CMU adopted kNN classification (k-Nearest Neighbor Classification) to do this, and changed kNN in general M-way into 2- way kNN [3].

2.3 Web page cleansing

The purpose of web page cleansing is to improve the efficiency of the entire system by means of removing html tags.

3. The architecture of IPOMS

In this research, we improve the previous work on topic detection and tracking, and represent the result in Internet browser. This system mainly includes spider, web page parser and detection and tracking tool. Figure 1 illustrates the system architecture of the IPOMS.

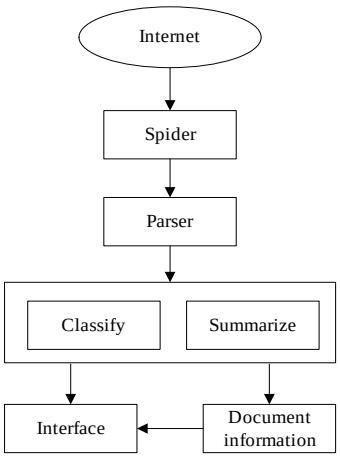

Figure 1. The architecture of IPOMS

3.1 Spider

We adopt the network spider to crawl web pages. Most TDT systems are designed for analyzing news stories, which are very "clean". This system can collect web pages from several news sites and big forums.

3.2 Web page parser

The collected web pages are in many formats, such as html, shtml, php, which influence the quality of outcome and the efficiency of the entire system, the collected web pages should be cleaned immediately. The main function of the web page parser is to remove 'noise' in web page, leaving web page link, head, title, time, text and first-level title.

We adopt the method of DOM (Document Object Model) tree [4] to get link, title, time, first-level title and text, combining with the method of Embley [5]. We adopt web page parser to construct a DOM tree, set the sub-tree number with maximum covered area number, search the tree with depth first search method, and write the detected text in document. If the leaf node is null, no content is written, else if the leaf node is '\ n', blank is written, which means it has relation with the previous text. When the text under a text node is searched, a tag is written in document, which means the partitioned tag with another section. The first-level title is written in document, under the head, with a "*" as a tag. The tracking and sectioning of text area are shown in Figure 2.

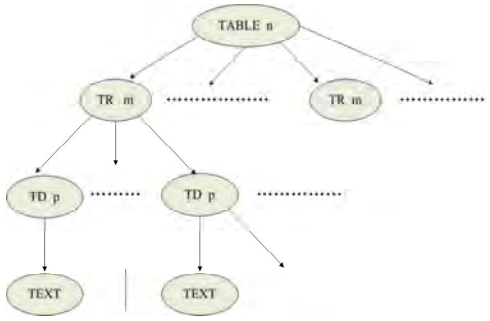

Figure 2. Tracking and sectioning of the text area

3.3 Topic detection and tracking tool

3.3.1 Sensitive word list and degree

According to the sensitive words that public opinion experts provide, we build one sensitive word list, and then set the degree for each sensitive word (according to its importance, sensitive degree), and some of them are set as keywords.

3.3.2 Document vector model

With TF-IDF (Term Frequency-Inverse Document Frequency) method, we transfer text into vectors. Calculate formula is shown in formula 1:

$$w_{ij} = tf_{ij} \times idf_i \tag{1}$$

w_{ij} stands for the weight of word i in file j. tf_{ij} stands for the frequency of word i in file j. idf_i stands for the reciprocal of the file frequency of word i.

To strengthen the importance of one keyword in critical location, we enhance the weight of keywords that appear in head, title, first-level title of that web page.

3.3.3 Cluster

We adopt the method CMU proposed to cluster the vectors [2]. First we calculate the similarity degree between every two vectors. We adopt cosine formula to calculate similarity degree:

$$sim(x,c) = \frac{\sum_{j=1}^{M} w_{jx} \times w_{jc}}{\sqrt{(\sum_{j=1}^{M} w_{jx}^2) \times (\sum_{j=1}^{M} w_{jc}^2)}} \tag{2}$$

$sim(x,c)$ stands for the similarity degree between the vector x which comes from a new text and a cluster c, w_{jx} stands for the weight of word j in cluster c, M stands for total amount of words in the vector space.

We adopt the k-means method to cluster the vectors [5] [6].

3.3.4 Topic detection

Each cluster is viewed as an event, the average weight of which is then calculated. Firstly, we calculate the similarity degree between the vector coming from a new webpage and the existing average vector.

Considering the importance of an event diminishing with the elapsing time, and the same key word may have a completely different meaning, we adopt the time span calculation method. The calculation formula is shown as below:

$$score(x) = 1 - \max_{c_{i \in window}}\{(1 - \frac{k}{m}) \times sim(\vec{x}, \vec{c_i})\} \tag{3}$$

In this formula, x stands for a new document vector, c_i stands for the number i cluster in the time area, i stands for the total amount of vectors in the vector space, k stands for the number of added document vectors coming between x and the latest one. If the outcome score is greater than the default value, the new file is viewed as one new topic.

Figure 3. shows topic detection flow.

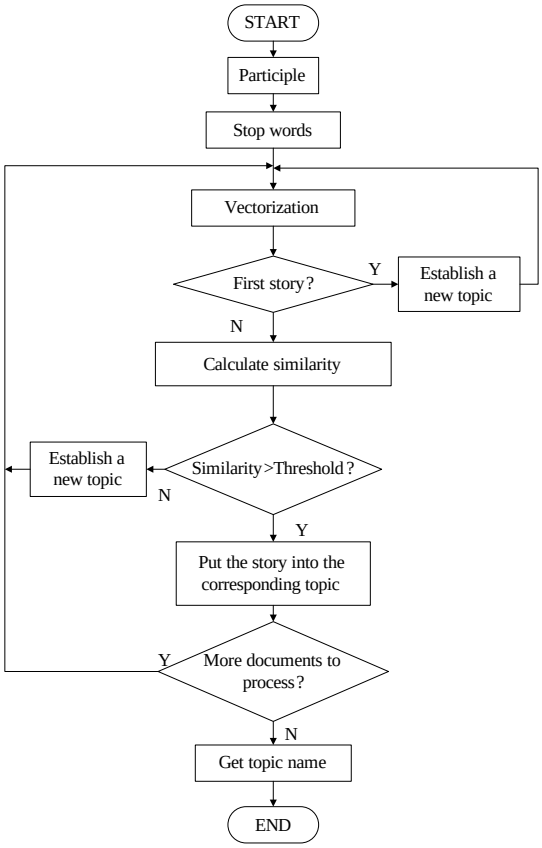

Figure 3. Topic detection flow

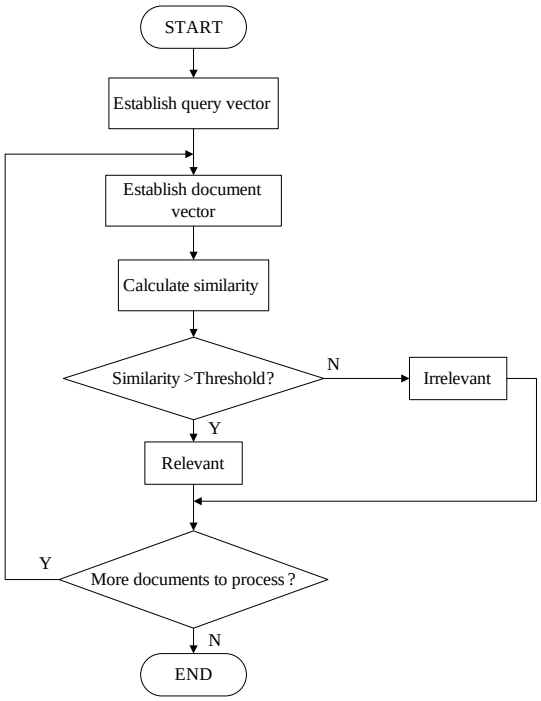

Figure 4. Topic tracking flow

3.3.5 Topic tracking

If the similarity degree between the vector coming from a new web page and one of the existing average vectors is lower than the default value, the new vector is viewed as the part of the existing topic.

This is a category process, and we adopt the k-d tree method to do this [7]. The time complexity is more competitive than kNN method [7].

Figure 4 shows topic tracking flow.

4. The features of IPOMS

IPOMS has four main features as follows.

1. It can grab web pages from web news sites and forums, which expands the range of monitoring.
2. In the step of parsing web page, it can transfer web pages into different formatted text files, and get web links, head, title, time and first-level title out of the text files. This step can improve the efficiency and accuracy of processing text.
3. It adopts the algorithm of k-d tree instead of kNN, which improves the efficiency of the system a lot.

5. System evaluation

5.1. Evaluation criterion

The $(C_{Det})_{Norm}$ evaluation metric which is widely used in TDT methods is used to evaluate the performance of our system. System performance, the miss probability and false alarm probability of the topic i (i =1, 2, …, tn, tn is the number of topics) are defined as follows:

$$\text{Miss}_i = \frac{\text{undetected stories about topic i}}{\text{total stories about topic i}} \tag{4}$$

$$\text{FA}_i = \frac{\text{false stories detected about topic i}}{\text{total false stories about topic i}} \tag{5}$$

The average miss probability, average false alarm probability and $(C_{Det})_{Norm}$ are shown as below:

$$P_{Miss} = \sum_i Miss_i / tn \tag{6}$$

$$P_{FA} = \sum_i FA_i / tn \tag{7}$$

$$(C_{Det})_{Norm} = \frac{C_{Miss} \cdot P_{Miss} \cdot P_{Target} + C_{FA} \cdot P_{FA} \cdot P_{-Target}}{\min(C_{Miss} \cdot P_{Target} + C_{FA} \cdot P_{-Target}} \tag{8}$$

C_{Miss} and C_{FA} are the cost of miss and false. P_{Target} is the prior probability of miss and false of target topic, $P_{-Target} = 1 - P_{Target}$. C_{Miss}, C_{FA} and P_{Target} are preseted, and the values vary in different methods. In this paper, these values are 1.0, 0.1 and 0.02.

5.2 Experiments

This research adopts the data from sogou lab, which includes 13,560 Chinese reports from October 1, 2007 to March 30, 2008. We consider the first 1,000 stories and corresponding topics as the training linguistic data, and consider the remaining 12,560 stories as test linguistic data, and these stories belong to 20 topics.

The miss probability is 0.3550, the false alarm probability is 0.0097, and $(C_{Det})_{Norm}$ is 0.4012.

(1) Spider

Figure 5 shows the output of spider.

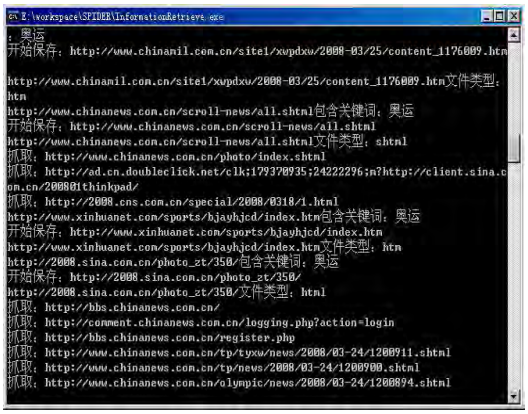

Figure 5. The output of spider

(2) Html parser

Figure 6 shows the output of html parser. The results include link, title, first-level title, and text.

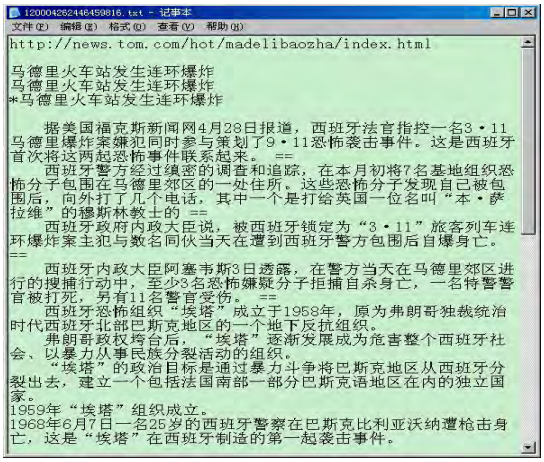

Figure 6. Parsed text

(3) Vectors

Figure 7 shows the vectors.

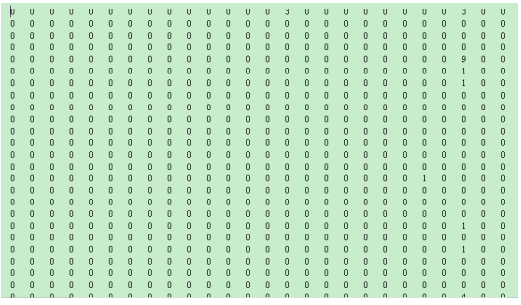

Figure 7. Vectors

(4) Clusters

Figure 8 shows the output of clusters.

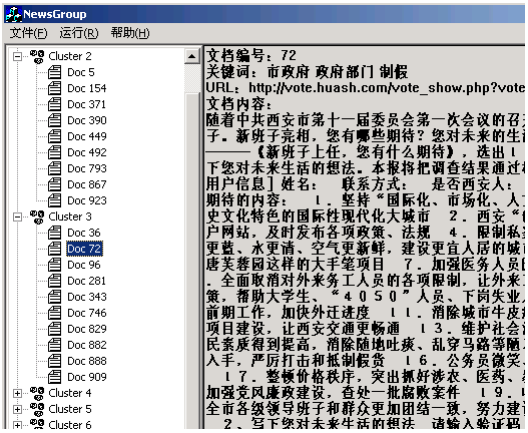

Figure 8. Clusters

6. Conclusions

In this paper, we adopt the method of topic detection and tracking, and propose a system which can efficiently monitor the public opinion. This system extends the range of monitoring web, improves the efficiency with web page parser and k-d tree algorithm. The results show that the system can help the supervisor track the progress of hot topics.

References

[1] Laihua, W., Yi, L (2005). An Overview of China 2004 Public Sentiment Research. Xinhua Digest, 2005, 18, pages 133-134.

[2] James, A., Papka, R., Lavrenko, V (1998). On-line New Event Detection and Tracking. In: Proc. of the 21st International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 37-45. Melbourne, Australia, 1998.

[3] Yiming, Y., Jaime, C., Ralf, D.B., Thomas, P., Brian, T.A., Xin, L (1999). Learning Approaches for Detecting and Tracking News Events. IEEE Intelligent System, 1999, 14(4), pages 32-43.

[4] James, A., Jaime, C., George, D., Jonathan, Y., Yiming, Y (1998). Topic Detection and Tracking Pilot Study: Final Report. In: Proc. of the DARPA Broadcast News Transcription and Understanding Workshop, pages 194-218. Virginia, USA, 1998.

[5] MacQueen, J (1967). Some Methods for Classification and Analysis of Multivariate Observations. In: Proc. of the 5th Berkeley Symposium on Mathematical Statistics and Probability, pages 281-297. Berkeley, California, USA, 1967.

[6] Huajun, Z., Qicai, H., Zheng, CH., Wenying, M., Jinwen, M (2004). Learning to Cluster Web Search Results. In: Proc. of the 27th International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 210-217. Sheffield, United Kingdom, 2004.

[7] Jon, B (1990). K-d Trees for Semi-dynamic Point Sets. In: Proc. of the 6th Annual Symposium on Computational Geometry, pages 187-197. Berkley, California, USA, 1990.

[8] Kuo, ZH., Juan, Z., Ligang, W (2007). New Event Detection Based on Indexing-tree and Named Entity. In: Proc. of the 30th International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 215-222. Amsterdam, Netherlands, 2007.

Authors Biographies

Jie Ding obtained his B.Sc. degree in Electronic Engineering from Beijing University of Aeronautics & Astronautics in Beijing (China) in 2005 and his M.Sc. degree in Computer Science from Graduate University of Chinese Academy of Sciences in Beijing (China) in 2009. His research interests include topic detection and tracking, information retrieval.

Jungang Xu obtained his Ph.D. degree in Computer Science from Graduate University of Chinese Academy of Sciences in Beijing (China) in 2003. He is an Associate Professor in School of Information Science and Engineering, Graduate University of Chinese Academy of Sciences since 2006. His research interests include topic detection and tracking, information retrieval, data management and data mining.

Search Engines and Arabic Language

Salah S. Al-Rawi[1], Belal Al-Khateeb[2]
[1]College of Computers, Al-Anbar University
Ramadi, Iraq
salah-s@computer-college.org

[2]College of Computers, Al-Anbar University
Ramadi, Iraq
belal@computer-college.org

ABSTRACT: *This paper presents an attempt to show the efficiency of some search engines in handling Arabic keywords. To achieve this, a comparison was made among the number of retrieved pages, retrieving time, and stability (in both the number of retrieved pages and the order for each retrieved page) for each of the 20 Arabic keywords selected (with its roots) after being simultaneously entered to the selected four search engines. Search engines tested in this experiment were Google, Yahoo, Al-hoodhood and Ayna. Google was the best search engine among the four selected search engines according to the results obtained over 10 weeks of experimenting.*

Keywords: Search engines, World Wide Web, Arabic keywords, Arabic roots, Information retrieval

Received: 17 October 2009, Revised 27 November 2009, Accepted 4 December 2009

1. Introduction

Arabic is a language that is being increasingly used by Internet users despite many significant problems. First time users face many difficulties when trying to read Arabic web sites. In major part, these difficulties arise from the way of representing Arabic in multiple character sets and the characteristics of the Arabic script itself [1].

With the extremely rapid growth rate of the Internet and the spread of textual information in a whole host of languages other than English on the web, retrieval of documents in these languages is becoming problematic more and more. Rules, theories, algorithms, and retrieval methods designed and developed for English and other morphologically similar languages may not necessarily apply in different linguistic environments. In the context of languages that differ fundamentally from English in morphology and word-formation rules, the problem tends to be even more exigent. Being the essence of written and spoken information queries, words are hugely the most important elements of expression and are the building blocks of meaningful information exchanges [2].

Initially, most of the available electronic databases were in English. Search and retrieval software, indexing methods, and user interfaces were designed specifically for this language. Since English is no longer the sole language used on the Internet, Information Retrieval (IR) systems have been developed for languages other than English. In tandem, search engines were being progressively modified to handle these languages [3][4][5]. Traditional IR environment and popular search engines face real challenges when Arabic is the language used due to the radical differences in morphology and words formation rules between Arabic and English. These rules are based on a root and pattern system that has been long thought to be a major factor in hindering IR operations. Finding all possible words that have a common Arabic root might not necessarily lead to better IR performance. Despite the use of advanced word stemming and root extraction algorithms in Arabic IR field, researchers still fail to answer many questions [6]. This paper investigates the handling of Arabic words in English and Arabic search engines. Retrieval environment is represented by Google, Yahoo, Al-hoodhood, and Ayna. Also, it presents specific approaches to assessing stemming and root-based retrieval methods to lodge the peculiarities of Arabic word formation rules within the skeleton of this environment. The following section will briefly present the information retrieval. Search Engines will be described in section 3 and the implementation that has been conducted will be dealt with in section 4. Experimental designs and their results are discussed in section 5, while section 6 gives the concluding notes of this work. Finally, some suggestions for future work will be highlighted in section 7.

2. Information Retrieval (IR)

Up until the 1990s, efforts of specialists in Arabic computing concentrated on presenting the language in a computer environment and finding solutions for display and coding problems. In the early 1990s, interest in Arabic IR became visible and research was conducted on the automation of Arabic online library catalogs and on IR issues [6]. IR involves many strategies each of which comes with its own features that can be used to retrieve information efficiently. *Boolean Search*, *Serial Search,* and *Cluster-Based Retrieval* are among these strategies [7].

Compared to English, redundancy in Arabic was assumed to be higher, because Arabic words are derived from roots according to certain patterns, depending on fixed rules, in addition to suffixes, prefixes and infixes [3]. Also by comparing the results with these from research on English, Arabic was found to have a greater redundancy, and the average word length for Arabic is greater than English, making Arabic potentially more compressible than English [6].

Root indexing was used to index Arabic documents because root indexing increases recall and circumvents composite problems created by Arabic morphology. A root index term would retrieve all variations of this root and abolish the need to use complex queries while searching [8].

3. Search Engine

Search engine technology has to advance hugely in order to keep pace with the web growth. Examples of web growth include the increased number of web pages, documents and web queries posted on the Internet [9] [10].

It is not easy to evaluate the information retrieval system for the World Wide Web (WWW) environment. The difficulty originates from the lack of standard test data and it can also be attributed to the highly subjective nature of the conception of relevancy of WWW pages retrieved in relation to the user's information needs [11].

Precision is always reported in formed information retrieval experiments. However, there are variations in the way it is calculated depending on how relevance judgments are made [12] Search engine stability problems were investigated in several studies performed by Bar-IIan, and several measures to evaluate search engine functionality over time were outlined in these studies [12]. Bar-IIans' measures are based on the technical relevance concept which is the document defined to be technically relevant if it fulfils all the conditions posed by the query [13]. For the purpose of updating search engines, a tool, generally called a spider, is used. Spiders clean hundreds of thousands of pages a day. To find information independently, many spiders also track the links on a page hence it is possible for a spider to index a web site even if that web site was not submitted to the search engine [14].

A search engine such as Google is designed to avoid disk seeks whenever possible, and this has a substantial effect on the design of data structures [15] [16]. In Google, several distributed crawlers do the web crawling (downloading of web pages). Web crawling is the backbone to the search engine. There is a URL (Uniform Resource Locator) server that sends lists of URLs to the crawlers to be fetched. Fetched web pages are then sent to the store server which then compresses and stores the web pages into a reservoir [17].

They may only use a small database from which to create a set of results to the users (Yahoo for example only indexes a very small proportion compare to a billion pages indexes by Google) or they may not be updated particularly quickly (All the web is updated every fortnight or so, while Google is updated monthly). These spider programs may not be very fast, which means that their currency might not be a real reflection of the state of play on the Internet [8] [18].

4. Implementation

In this research study, four search engines were selected in order to sustain a good comparison for Arabic keywords. Of the search engines selected, two are general search engines (Google [19][20][21][22] and Yahoo [23]) while the remaining are Arabic language search engines (Al-hoodhood [24] and Ayna [25]) that employ stemming and root indexing. These search engines were chosen because they are broadly used as general search engines. The test included using these search engines to search for a specified word, search for a specified word by its root, and then evaluating the stability of each search engine in terms of the number of retrieved pages and the order of each one. Search was designed to compare the performance of Google with Yahoo, Al-hoodhood and Ayna, and evaluate stemming as an alternative to root retrieval. Experiments were conducted using a computer with 1.7 GHz processor, 256 MB RAM, and windows XP operating system.

5. Results and Discussion

This study has been conducted in two phases. The first part was intended to determine the speed of loading results. In this phase, after selecting twenty different Arabic words (each with its root); each word was entered as an input in the four selected search engines simultaneously. A record was kept for the total number of pages resulted and the retrieval time. Table (1) shows the selected (20) words which were entered simultaneously to the four search engines and the number of results from each search engine with the relative time spent for searching and retrieving the results.
This process was repeated for the roots of the selected words (as shown in table 2). The purpose of

this phase was to maintain a good comparison between the selected search engines in the number of retrieved pages and time. To achieve this, the total number of retrieved pages (Total-Pages) was calculated by summing up the number of the retrieved pages for all the entered search keywords. Similarly, the total time of retrieving (Total-Time) was calculated by summing up the time required to retrieve each keyword. Then, Total-Pages were divided over Total-Time and the results obtained were collated in an ascending order for the four search engines to know which of the selected four search engines is faster in retrieving (the first one is faster than the second and so forth).

In the second phase of the work, five words were taken out of the selected 20 words with its roots. Each of the five words and its root were entered simultaneously into the selected four search engines and the process was repeated for ten weeks. A record was kept for the first twenty retrieved pages resulted for every week of the ten weeks period. The retrieving time was omitted at this part of the study as the aim of this phase was to compare the selected search engines from the results retrieval stability standpoint. Table (3) illustrates the stability of each search engine in terms of the number of the retrieved web pages for each word of the selected five words. Whereas tables (4 and 5) and figures (1 and 2) show the stability of each search engine in terms of the order of the retrieved web pages for each word of the selected five words. Figures in tables (4 and 5) were calculated by taking the first twenty pages resulted in the first week as a measure to assess how stable the search engine was in retrieving the same web pages in the coming weeks. For example, as it is clear in table 4, in the second week, Google retrieved eleven pages from the twenty that were retrieved in the first week; while Yahoo retrieved only four. Al-hoodhood retrieved 20 and Ayna retrieved 13 for the same week. These results underline two points; one is that Al-hoodhood and Ayna are more stable than Google and Yahoo. The other is that Google and Yahoo are more flexible in updating their databases (by adding new pages for the same subject).

Google		Yahoo		Al-hoodhood		Ayna	
Results	Time (sec.)	Results	Time (sec.)	Results	Time (sec.)	Results	Time (sec.)
2,630,000	0.55	1,480,000	0.22	41,904	2.000	3,182,550	0.991
531,000	0.37	338,000	0.14	13,007	1.000	641	0.7369
810,000	0.35	291,000	0.17	1,879	1.000	599,270	0.5383
625,000	0.38	342,000	0.11	8,058	1.000	715	0.2934
1,620,000	0.18	621,000	0.13	10,543	1.000	1,308,300	0.4291
378,000	0.17	300,000	0.17	8,715	1.000	5,366,480	0.4687
2,260,000	0.19	1,060,000	0.17	29,384	1.000	5,488,980	0.2081
109,000	0.21	59,400	0.10	11,480	1.000	326	0.4435
320,000	0.02	213,000	0.16	3,728	1.000	1,058,400	0.3454
227,000	0.08	348,000	0.12	4,298	1.000	301	0.2497
20,000	0.36	8,190	0.45	158	1.000	182	0.0489
835,000	0.90	456,000	0.12	12,210	1.000	15,224,790	0.396
8,400,000	0.35	7,290,000	0.15	281,677	1.000	26,972,050	1.154
3,240,000	0.55	1,850,000	0.10	19,879	1.000	11,157,300	0.2697
1,490,000	0.53	725,000	0.08	7,249	1.000	3,369,240	0.2071
11,700,000	0.07	4,770,000	0.12	124,130	2.000	68,771,010	0.2313
23,000,000	0.64	18,600,000	0.11	50,881	1.000	69,492,290	0.308
11,600,000	0.49	4,530,000	0.10	22,911	1.000	7,636,160	0.2994
3,600,000	0.57	5,240,000	0.11	28,853	1.000	8,698,480	0.2855
9,220,000	0.34	3,460,000	0.12	69,798	1.000	25,194,820	0.1831

Table 1. Loading speed of the selected search engines on Arabic Keywords

Google		Yahoo		Al-hoodhood		Ayna	
Results	**Time (sec.)**	**Results**	**Time (sec.)**	**Results**	**Time (sec.)**	**Results**	**Time (sec.)**
3,490,000	0.45	2,720,000	0.11	78,448	3.000	4,975,460	0.2354
4,080,000	0.44	2,920,000	0.23	87,941	4.000	5,375,300	0.2518
1,650,000	0.09	730,000	0.16	11,548	1.000	2,130,520	0.227
9,520,000	0.14	4,310,000	0.27	210,369	9.000	45,192,700	0.8705
624,000	0.26	266,000	0.22	5,301	1.000	676,690	0.5099
2,790,000	0.11	1,240,000	0.20	43,565	2.000	2,681,770	0.1814
1,530,000	0.43	947,000	0.11	26,478	2.000	2,329,950	0.4935
28,800	0.12	23,800	0.38	156	1.000	86	0.2234
331,000	0.14	240,000	0.15	3,661	1.000	415	0.708
5,250,000	0.24	3,440,000	0.27	149,919	7.000	12,096,630	1.7774
403,000	0.18	280,000	0.13	6,412	1.000	886,410	0.2275
6,800,000	0.04	3,850,000	0.27	89,188	1.000	15,803,970	0.6742
10,400,000	0.04	5,222,000	0.24	304,339	1.000	33,520,410	0.504
1,730,000	0.26	1,090,000	0.18	18,752	1.000	10,378,200	0.3234
1,180,000	0.28	755,000	0.21	26,647	1.000	941,290	0.2456
17,000,000	0.19	7,330,000	0.09	140,149	2.000	74,913,160	0.1938
20,200,000	0.51	8,630,000	0.15	39,221	1.000	70,907,410	0.4054
2,940,000	0.25	1,620,000	0.22	42,182	2.000	186	0.0449
19,500,000	0.37	12,900,000	0.03	291,731	15.000	91,490,840	0.5056
1,520,000	0.34	982,000	0.20	44,441	5.000	5,410,090	0.3271

Table 2. Loading speed of the selected search engines on Arabic roots

6. Conclusion

Analysis of tables 1 and 2 was performed by summing up the results of each search engine and dividing it by the sum of the retrieving time. (Bil, u can delete this green bit because there is no need to repeat the procedure in conclusion section) It is concluded that Google is the best search engine in dealing with Arabic keywords. Yahoo is the second, while Ayna comes third and Al-hoodhood is the last one. The results show that Google is faster and can retrieve a large number of results comparing with others; also they reflect that although there are search engines specialized in Arabic keyword, they still have limited abilities in comparison with the general purpose ones (Google and Yahoo).

Analysis of table 3 revealed that Google is the best search engine when it comes to dynamic update of web pages with stability in dealing with Arabic keywords. Yahoo falls behind Google in the second position to be followed by Ayna which comes third while Al-hoodhood is the last one (no update occurred in Al-hoodhood during the search time). These results clearly sho w that Google is capable of rapid dynamic updating to its database in a short time compared with other search engines. Similarly, it is easily concluded that Al-hoodhood is the slowest one in that update.

Conclusions drawn from analyzing tables 4 and 5 are compatible with the above and demonstrate that Google is the best search engine in maintaining the retrieval of the same results from week to week with dynamic update of web pages in dealing with Arabic keywords. Again, Yahoo follows Google as the second; while Ayna comes third leaving Al-hoodhood sitting at the bottom as the fourth (no update occurred in Al-hoodhood during the search time).

	Google	Yahoo	Al-hoodho	Ayna
Week1	403,000	280,000	6,412	886,410
Week2	445,000	329,000	6,412	1,046,640
Week3	493,000	238,000	6,412	677,670
Week4	482,000	252,000	6,412	1,335,250
Week5	667,000	338,000	6,412	1,335,250
Week6	510,000	240,000	6,412	1,335,250
Week7	402,000	220,000	6,412	1,214,710
Week8	475,000	304,000	6,412	1,089,760
Week9	462,000	335,000	6,412	861
week10	423,000	300,000	6,412	862

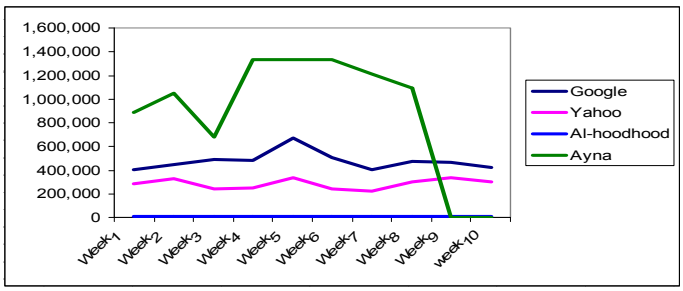

Table 3. The stability of the four search engines on the selected 5 words in terms of retrieved pages

	Google	Yahoo	Al-hoodhood	Ayna
Week1	20	20	20	20
Week2	11	4	20	13
Week3	11	7	20	11
Week4	0	15	20	5
Week5	0	16	20	14
Week6	14	13	20	20
Week7	11	6	20	18
Week8	0	2	20	6
Week9	8	14	20	20
Week10	15	18	20	20

Table 4. The stability of the four search engines on the selected 5 keywords in terms of the order of retrieved pages

	Google	Yahoo	Al-hoodhood	Ayna
Week1	20	20	20	20
Week2	11	6	20	7
Week3	8	11	20	8
Week4	0	15	20	5
Week5	0	11	20	8
Week6	10	9	20	20
Week7	10	11	20	20
Week8	0	5	20	17
Week9	15	19	20	4
Week10	18	19	20	16

Table 5. The stability of the four search engines on the selected 5 roots in terms of the order of retrieved pages

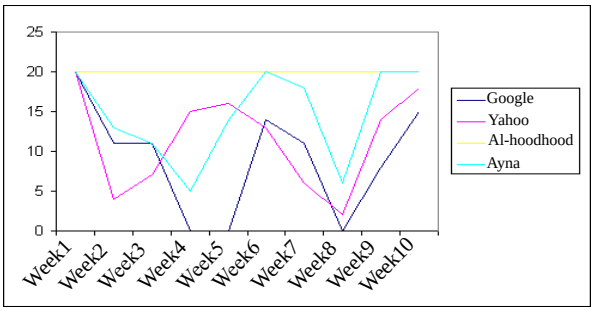

Figure 1. The stability of the four search engines on the selected 5 keywords in terms of the order of the retrieved pages

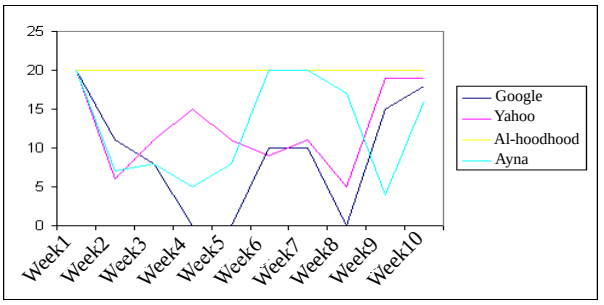

Figure 2. The stability of the four search engines on the selected 5 roots in terms of the order of the retrieved pages

7. Future Work

Research ideas are plenty in the web search engines' rich environment. There are many issues that need to be looked at when attempting to define new methods to search the web in a more meaningful way. Recommendations to addressing present and future issues in developing a web search are:

1 Design a smart algorithm to decide what old web pages should be re-crawled and what new ones should be crawled.

2 Developing a metasearch engine that improves the efficiency of web searches by downloading and analyzing each document and then displaying results that show the query terms in context. This helps users to more easily decide on the relevancy of the document without having to download each page.

3 For solving Arabic language problems, Unicode must be possible to be handled, which is just one out of several possible encoding sets.

4 Supporting query refining.

5 The addition of more search engines together with using additional samples in the experiments.

References

[1] Sanan, M., Rammal, M., Zreik, K. (2008). Internet Arabic Search Engines Studies, *In*: 3rd International Conference on Information and Communication Technologies: from Theory to Applications. ICTTA, Damascus, p. 1-8

[2] BCS British Computer Society (2005). The BCS Glossary of ICT and Computing Terms, Pearson Education, UK.

[3] Moukdad, H. (2006). Stemming and root-based approaches to the retrieval of Arabic documents on the Web, *Webology*, Article 22.

[4] Douglas, Comer (2008). Computer Networks and Internets with Internet applications, Prentice hall international, INC., USA.

[5] Christopher, D. (2008). Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press, UK.

[6] Al-Khadady, Saba Abdul Khaliq (2002). *Internet and Arabic Search Engines*, M.Sc. Thesis, Iraq,.

[7] Rijsbergen, Van (1979). Information Retrieval, Butterworth, London.

[8] Jassim, Khalid Shaker (2005). Comparison of Efficiency of some search Engines on the Internet, M.Sc Thesis, Iraq.

[9] Monge, Peter, R., Contractor, Noshir S. (2003). Theories of Communication Networks, Oxford University Press, INC., UK.

[10] Stott, D., Moran, D. (2000). Information and Communication, Springer, London.

[11] Marchiori, Masimo (2000). The Quest For correct information on the web: Hyper search Engines, Depantment of Pure Application Mathematics University of Padova, Italy.

[12] Bar-Ilan J., Evaluating the stability of the search tools Hotbot and Snap: a case study, Online Information Review, Emerald, Bradford, ROYAUME-UNI, INIST-CNRS, Cote INIST, 2000, p. 439-450.

[13] Mike Thelwall, The Responsiveness of Search Engine Indexes, *Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics* 2001.

[14] Levene, Mark (2006). An Introduction to Search Engines and Web Navigation, Pearson Education, UK.

[15] Danny Sullivan, Search Engine Features for Webmasters (2002). [online] available from <http://searchenginewatch.com/showPage.html?page=2167891> .

[16] Danny Sullivan, How Search Engines Work (2007). [online] available from <http://searchenginewatch.com/showPage.html?page=2168031>

[17] Brin, Sengey., Page, Lawrence (1994). The Anatomy of lange-scale Hypertextual web search Engine, Computer Science Department, Stanford University.

[18] Multi-search Engines - a comparison (2003). [online] available from <http://www.philb.com/msengine.htm> .

[19] Google [online] available from <http://en.wikipedia.org/wiki/Google>.

[20] All About Google [online] available from http://www.google.com/about.html.

[21] Google Help Central [online] available from http://www.google.com.au/help.

[22] Sullivan, Danny (2007). Major Search Engines and Directories [online] available from <http://searchenginewatch.com/showPage.html?page=2156221> .

[23] Barlow, Linda (2004). A Helpful Guide to Search Engines [Online], available from <http://www.monash.com/ spidap3.html> .

[24] http://www.alhoodhood.com/about.html.

[25] http://www.aynacorp.com/About/6.html.

An Empirical Investigation of the Information Technology Implementation in Saudi Arabia

Saleh Al-zharani
Department of Information Systems
Faculty of Computer and Information Sciences
Al Imam Muhammed Bin Saud University
Riyadh
Saudi Arabia
drsalehz@hotmail.com

ABSTRACT: *This research paper addresses the issues affecting information technology development and deployment extensively in Saudi Arabia. The issues specified in this paper have addressed the environment in IT implementation processes, especially with regard to the question of the perceptions and difficulties of the implementers at both government and private sectors. This study will provide an exploratory focus on the major problematic issues surrounding IT implementation in the country and how implementers perceive them. The study has identified 16 specific issues under 8 thematic divisions of information technology and studied the difficulties of selected implementers. The implementers were classified into proficient and novice and the significant differences of them are also addressed. Based on the identified problems, effective solutions were also presented for a long term strategic planning and implementation.*

Keywords: IT implementation, IT Governance, IT architecture

Received: 11 March 2009, Revised 18 April 2009, Accepted 22 April 2009

1. Introduction

Information technology has become an essential tool as well as the means for a large sector of users and facilitates the daily lives in every area. It is one of the most prominent resources in the modern era, which many people tend to depend.

The use of computers at the beginning of the evolution of information technology was limited to a few groups of specialists in research centers and scientific laboratories due to the cost and lack of the knowledge of its use. With the lower cost of computers and increased specialists in many sectors, the use of computer has expanded to include many applications in government and private agencies. The applications have potential to process voluminous data and the use becomes widespread and frequent. The availability and use of information systems and technologies has grown almost to the point of being commodity like in nature, becoming nearly as ubiquitous [1].

However, the use of computers in recent years has spread dramatically to become part of the lives of individuals as well. The reason behind the widespread of the computer in recent years is the global networking of computers, "the Internet". Although we all live in the information age, there is an unequal participation in the societies and countries due to skewed deployment of information technology. The people who use the technology now face a variety of obstacles and issues. Many organizations and countries suffer from several problems in dealing with technology; and, Saudi Arabia is not in isolation from these problems, and it is facing many of them.

2. Extracting efficiencies and synergies in Information Technology

Thus, to promote the application further and to offset many limitations, the difficulties are being now documented at many organizations and countries. As countries differ significantly in terms of the nature and magnitude of the problems, researches studied the difficulties independently.

Even the Information and communication technology (ICT) is flourishing in the Arab world at a rapid pace, the ICT implementation is not going smoothly or without major problems that hinder and slow down the progress in many countries in the area. [2]. With the emergence of an expanding interdependent global economy, information systems (IS) strategists need

to face the challenges of internationalization. The growth of multinational business has led many corporations to support significantly high level of IS operations and IS applications in Arab countries. [3].

The survey by *Tiamiyu* [4] (2009) revealed that in Nigeria due to either lack of computing technologies in most of the agencies or of their ineffective exploitation, the majority of the personnel were, still unaware of, or unimpressed by, the productivity potentials of using computers. There is a considerable pressure on most organizations to make their operational, tactical and strategic process more efficient and effective. [5].

The preceding discussions on information technology problems lead to understand us in two directions. First, is the existence of the problems across countries and second is the variance in the problems among them. This, it is imperative to discuss each country's problem in their own perspective based on empirical data.

3. Information Technology Architecture

Huber [6] suggests that Information Technology (IT) is a variable that can be used to enhance the quality and timeliness of organizational intelligence and decision making thus promotes organizational performance. However, Huber's analysis was offered at a time when IT was making its first major inroads into organizational life and subsequently the researchers have extended and updated Huber's research. It was applied to the examination of organizational functioning by describing the impact of IT on a broader array of organizational characteristics than was addressed in Huber's work. [7].

The information technology measurement framework was proposed by many researchers such as, *Markless* and *Streatfield,* [8], *Mendonca et al.,* [9] and *Kaplan* and *Norton* [10, 11, 12]. These initial frameworks need to be assessed and implemented. Keeping in view of the localization and its influence, we believe the significance of generating local framework to address the national characteristics. In order to offer a more encompassing view of IT and organizational functioning, we examine IT as a moderator of the relationship between organizational characteristics and several organizational outcomes, most importantly, efficiency and innovation (See Figure 1). The characteristics and capabilities of countries can be determined by the two levels of Technology extraction, viz., high and low. It is tangible to conclude that the high extraction leads to higher degree of efficiency, knowledge codification and innovation. On the other hand, failure of extraction will have a bearing on the execution of technology.

The framework proposed will enable to draw a plan of understanding the IT implementation in nations such as Saudi Arabia.

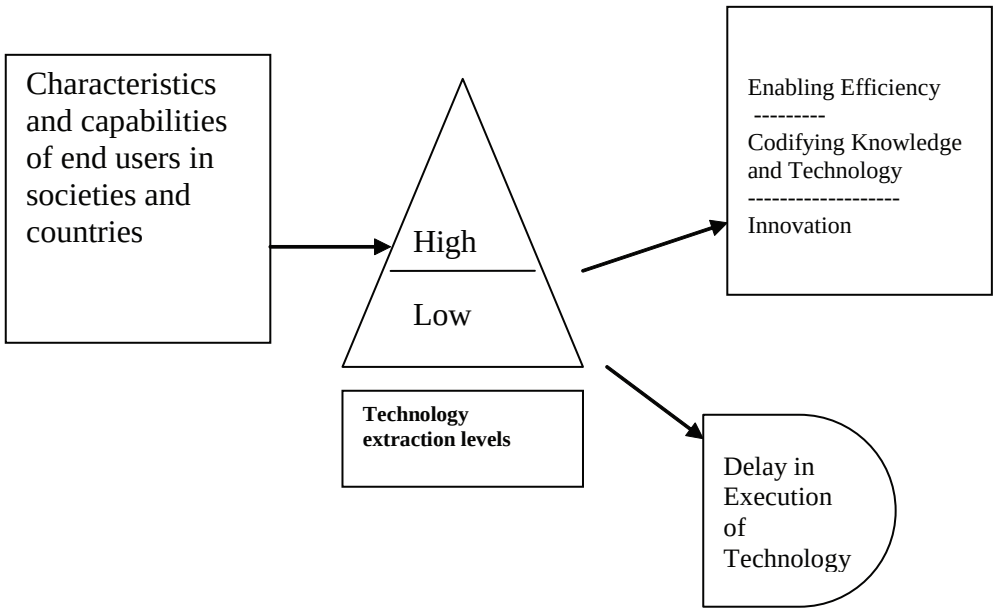

Figure 1. Role of Information Technology in the contribution to countries

4. Research Problem

It is clear from any questions that the information age is the reality and both public and private institutions are experiencing an increase in the use of a variety of information technologies (IT). Realistically, it has become nearly impossible for an individual or country to work without the use of one or more IT. From the period of the application of IT, it is viewed as the means for addressing the organizational challenges. However, it is the pertinent to pose a question – whether the application and exploitation is equal across countries. If not, what mechanisms need to be introduced to track the problems and how the local conditions can be monitored and recorded for proposing remedies?

The gap and the problems that exist in the societies and countries in terms of the effect and implementation failure were extensively treated by *Gottschalk,* [13, 14, 15], *Boar* [16] and *Lee* and *Bai* [17].

There are multiple, interrelated issues which impact effective IT implementation in Saudi Arabia. These issues must be identified and addressed before an IT implementation model for Saudi Arabia can be developed.

5. Research community and the data for the study

The research community focused in this study consists of computer system operating users located in the government as well as private sectors. We have limited our study to nine organizations that include the government sector and private sector.

A total of 120 questionnaires were distributed during the period April, 2009, and collected on various dates in May, 2009. These One hundred and twenty questionnaires were distributed in 9 organizations. In addition, several initiatives were taken to ensure a satisfactory response rate such as: the front page emphasized the assurance of respondent's confidentiality; the number of questions was limited to fit a four-page layout. Also, only necessary and relevant questions were asked to avoid redundancy and to maintain reliability.

Our study has two divisions as follows:

1 - The first part deals with the basic data of the members of the community study.

2 - The second part deal with – IT work and service that is divide into nine sections as follows:

- Consists of six questions of data management and data infrastructure.
- Consists of three question of personal computer and operating system
- Consists of five questions of collaboration tools.
- Consists of five questions of enterprise applications
- Consists of six questions of networking and communications
- These Section of security divide into two part:

-Consists of four questions about some application of security apply in its organization.

-Consists of eight questions about some problem in security facing Daly while they work.

-Consists of three questions about software and web development.

6. Data Analysis

The data extraction through the respondents was carried out in a longitudinal study. The user population has the following types and characteristics.

6.1 User population

The studied users have heterogeneous characteristics. In information access, use and exploitation, the research literature has identified two classes of end-users. These two classes of end-users such as 'proficient' and 'novice' have significant impact on the analysis of results. There is a wide variation between these two types with respect to their characteristics.

The Characteristics and Definition of Proficient and Novice: The proficient not includes the factor, 'experience' rather it denotes the extent of 'ability' the end users have. It happens that many experienced users do not possess skills and techniques in the implementation of information technology. In many organizations, the entry-level users score over others. The classification

for this study is thus the division of them in to the above two categories. To ascertain their levels, the division head is asked to specify the levels of them. The table 1 reflects their categories.

	Total	Novice	Proficient
Area- Hardware Type - Novice Proficient	19	11	8
Area - software novice Proficient	31	7	24
Area - networking novice Proficient	25	6	19
Area - services novice Proficient	2	1	1
Area - manager novice Proficient	5	1	4
Total	82	26	56

Table 1. Classification of IT implementers

The classification of the users is based on the decision of their respective division heads as they assess the IT implementers for a longer period. The table 1 indicates the number of implementers in five identified areas.

Measure	**Factor Loadings**	**t-value**
A. Data Management and Data Infrastructure (n=82)		
Difficulty in back-up of data (valid =81; Missing = 1)	.59	12.15
Difficulty in Data Recovery (valid =79; Missing = 3)	.59	18.32
Difficulty in Data Security	.62	18.76
Supporting for business intelligence using data (valid =80; Missing = 2)	.91	22.43
Supporting from data centre for data analysis (valid =80; Missing = 2)	.61	17.59
B. Personal Computer and Operating Systems		
Convenient type of computer to use	.88	20.24
Prefer operating systems:		

Limitations in using the OS they have (valid =78; Missing = 4)	.63	17.65
Collaboration Tools		
Using E-Mail Client:	.79	19.54
Using Web Presentation Software (valid =78; Missing = 4)	.69	17.82
Using Spreadsheets	.68	16.82
Using Word Processing	.87	19.85
Enterprise Applications		
Using ERP application: (valid =79; Missing = 3)	.64	16.42
Using of Workflow Management tool (valid =80; Missing = 2)	.62	16.98
Document management tool (valid =80; Missing = 2)	.74	16.72
Using Groupware	.57	14.54
Networking and Communications		
Using LAN:	.66	18.02
Using WAN:	.62	17.76
Using Ethernet	.61	17.52
Using Network Interface Cards (valid =76; Missing = 6)	.60	16.85
Using VPNs (valid =74; Missing = 8)	.59	16.86
Security		
Using Encryption applications	.73	17.45
Using Digital Signatures application	.59	18.64
Using E-commerce Security applications	.61	19.49
Experienced Intrusion problem	.86	19.05
Experienced Tampering problem (valid =68; Missing = 14)	.73	16,88
Experienced Virus problem	.82	17.52
Experienced Spyware problem	.65	14.62
Experienced E-mail Fraud problem (valid =64; Missing = 18)	.77	18.68
Experienced Phishing problem	.66	17.56
Software and Web Development		

Using ASP, PHP, HTML,XML Applications / Languages	.58	16.96
Using Java, Middleware Applications / Languages	.52	15.76
Using Programming Languages	.51	15.46
Storage Management		
Prefer storage managements to store data	.86	18.96
Network Management		
Difficulty in network technology	.65	18.52

Fit indexes

Goodness-of-Fit Index (GFI) = .91

Adjusted Goodness-of-Fit Index (AGFI) = .89
t values are significant at p < .05

The t values are measured at where valid data is total sample.

Table 2 - Factor Analysis Data

6.2 Measurement Issues

The total factor analysis data is given in the table 2. All scale items of the data about implementers were measured with a 2-level scale, ranging from '*difficulty'* and '*no difficulty'*, largely on the basis of validated scales. The operation was largely based on an instrument from *Ad de Jong* and others' tolerance model. [18] (2003). The usefulness and ease-of-use scales were measured using a scale designed initially by *Davis* [19] (1989). Innovativeness and risk aversion were measured using items from a scale developed by *Grewal, Mehta, and Kardes* [20] (2000). However, the scale of *Ad de Jong* was modified by this study as *Jong* has employed a 7 point scale.

For the factor loadings, two systems were used to test the factors and item loadings of the scale constructs. We first used the coefficient and the factor structure (through principal component analysis) for all the scale items simultaneously. The factor structure was achieved with items loading on the assumed dimensions. In addition, we performed a confirmatory factor analysis (CFA) In addition the fit indexes of the proposed factor model, construct reliabilities of the scales, and confirmatory factor loadings with *t*-values for each item are represented in Table 2. The proposed factor model generated indexes as defined below that reads two measures called the Good fit (Goodness-of-Fit Index [GFI] = .91 and Adjusted Goodness-of-Fit Index [AGFI] = .89.

The factor loads are tested by different researchers earlier that used the coefficients. According to *Nunnally* and *Bernstein* [21], if the Coefficients of all measures were equal to or greater than .80, then it implies that reliability which is deemed acceptable (*Nunnally* and *Bernstein 1994*).

In addition, Chi-square difference tests with 1 *df* were used to test for unity between pairs of constructs. All tests were significant at the .05 significance level.

As the general over all aim of the work is to study the problems and limitations of the use of IT which was made into operation as a given users deployment and problem levels. Each respondent was asked to indicate how many times he or she actually used a specific IT application and what is the level of the problem in usage.

The supplementary testing was carried out to reinforce the inferences. We have used the confirmation of the division leader about the views and scores given by the participative users. Besides, a comprehensive list of 20 software applications was drafted, and individual team members were asked to indicate (a) which three IT applications they used most frequently and, consequently, (b) their usage rates for these applications. The group-level variable inter-team network concerns a dummy indicating whether a team participated in a network of multiple teams. Finally, the demographic variables computer training, work training, and age served as control variables when testing the data.

Variable	Mean 1	Mean 2	SD 1	SD 1
Data Management and Data Infrastructure	36.3	63.2	0.32	2.45
Personal Computer and Operating Systems	37.33	62.66	0.21	2.43
Collaboration Tools	73	27	5.65	0.75
Enterprise Applications	46.66	63.34	1.22	3.45
Networking and Communications	53	47	2.33	1.3 5
Security	37.8	62.2	1.33	2.67
Software and Web Development	72.3	37.7	6.54	0.65
Storage Management	70.7	29.3	7.54	0.23
Network Management	25.6	74.4	0.69	2.45

Mean 1 and Mean 2 denotes the Bi-variable either 'yes' or 'no'

Table 3 - Group level correlations

Table 3 indicates means, standard deviations, and individual-level as well as group-level correlations between levels of IT problems as identified by the users either as low or high. In the research on averages, the issue is that corrections for individual-level measurement error should be made first, before comparing individual and aggregate-level correlations [22] (*Ostroff 1993*). Therefore, we calculated individual- level correlations between the antecedents and IT adoption variables after increasing the reliability for those level constructs that had lower reliabilities. Overall, the results indicate some increase of the individual-level correlations but do not imply major changes in the magnitude differences between the two levels novice and proficient.

When analyzing the group level outcome measures, group user preferences and the ratings measured on the two point scale, ranging from difficult to no difficult *which was* gathered from the users expression. Table 3 represents the overall means, standard deviations, and the correlations between the adoption level of standardized and customized IT problems and difficulties and the outcome variables are presented.

	IT problems Standardized		
Variables	Coefficient	(SE)[a]	Δ Magnitude Coefficient[b]
Data Management and Data Infrastructure	–.130	(.124)	
Personal Computer and Operating Systems	.098	(.059)	.332
Collaboration Tools	–.172	(.095)	
Enterprise Applications	.049	(.068)	
Networking and Communications	–.046	(.166)	.225**
Security	2.122	(.133)	.345*
Software and Web Development	–.104	(.078)	
Storage Management	.112	(.124)	.184*
Network Management	–.054	(.096)	

a. Standard errors are in parentheses.

b. Differences in magnitude between individual-level and group-level coefficients were tested by means of raw-score analyses and reflected by the presented group-level coefficients.

*p < .05. **p < .01.

Table 4 Multilevel Analyses

The results of the multilevel analysis are presented in Table 4 for the nine identified information technology issues. To begin with, the coefficient values show that team members' perceptions of difficulties stand varied levels. The standardized errors encompass a considerable part of between-groups variance. Regarding the possible impact of one limitation on other, strong positive effects exist of the individual-level within a group of problems, which implies that homogeneity exists among the group of problems. Furthermore, it appears that individual- level perceived problem is not significantly related to team members' perception level, whereas it shows a significant independent relationship of the problems in IT. These findings imply that heterogeneity also exists besides the general consensus on certain individual problems. This is perhaps due to the two divisions of novice and proficient users. In addition, individual-level ease of use has a positive impact on users' perception of problems in IT, whereas no significant relationship emerges with respect to team members' views.

7. Discussions

The preceding discussions shed light on the fundamental results that we obtained out of the statistical analyses of the perceived data of the Saudi Arabian users. Therefore, the following discussion will present the comprehensive results.

In Saudi organizations, considerable interest for the implementation of information technology was shown, where evidences and the direct observation reflect. Some barriers were found which prevent the provision of adequate services. Further problems were seen such as - the lack of co-ordination within and between Saudi organizations, as well as issues associated with technical, behavioral as structural factors. Some of the results could be summarized as follows:

All Saudi organizations suffered from the absence of trained personnel with sufficient ICT knowledge, experience and skills to manage and use the information systems efficiently and too take the maximum advantage of the information technology.

The non-availability of information technology training programs and the absence of co-ordination among the organizations were the major problems in adopting the ICT in the country. In the government sector, most of employees in the IT are related to the network division (30.0%), and another one third (33.3%) belongs to IT hardware area whereas the in the private sector (mostly hospitals) the share is different as 40% belong to 'network' division and half (48.1%) are related to the software. This has some impact on perceived and documented implementation difficulties. In the private sector, the identified problems are lesser than government.

8. Conclusion

The principal focus of this research was to determine what issues perceived as being the most problematic; and what is the magnitude with regard to IT planning, procurement, and implementation in Saudi Arabia. Each stage in the development and deployment process was viewed individually in relation to the fundamental issues in order to better understand the impact of each one on the process.

The changes in technology have considerable influence when viewed in the context of strategic national planning for IT. Information technology, as a natural evolution is constantly changing. The issue of rapidly changing technology is also a major issue which we intend to address in the subsequent studies. It is evident that in the last decades the high rate of innovations in information technology replaces the old ones quite speedily replacing or enhancing previous innovations. The significant application of the information technologies is to make dynamic progress in the way we communicate and function. Quick innovations and strong systems can drive technology and the time available to understand and implement is less. The window for opportunity on the new and innovative is quite large. It is evidenced that the information technology is producing new hardware, software, network and systems. It is found that the fundamental breakthroughs in this arena occur at the astonishing rate of 18-24 month intervals [23].

The considerable scale and range of investments required to raise Middle Eastern technology and innovation performance suggests policymakers should now explore new financing methods to build infrastructure and innovation capabilities. [24-26]. Thus it is understandable that the technology certainly has serious ramifications for evolving strategic planning and decisions. It is the organizations who can take initiatives to build a strong and viable system for technology application and improvement which will enable the people to implement technologies with ease.

References

1. Dewett, Todd.,. Jones, Gareth R (2001). The role of information technology in the organization: a review, model, and assessment *Journal of Management* 27 (2001) 313–346.
2. Samir N. Hamade (2009). Information and Communication Technology in Arab Countries: Problems and Solutions, *In*: Information Technology: New Generations, Sixth International Conference on Information Technology: New Generations, p. 1498-1503.
3. Abdul-Gader, A. H (1997). Information systems strategies for multinational companies in Arab Gulf countries, *International Journal of Information Management*. 17. 3-12 .
4. Tiamiyu, M.A. (2000). Information technology in Nigerian federal agencies: problems, impact and strategies, *Journal of Information Science*, 26 (4) 227–237.
5. Ho, Chin-Fu (1993).Information technology implementation strategies for Manufacturing organizations: A strategic alignment approach, *In:* Pan pacific conference on information systems p. 219-226.
6. Huber, G. P (1990) A theory of the effects of advanced information technologies on organizational design, intelligence, and decision making. *Academy of Management Review, 15* (1). 47–71.
7. Dewett, Todd., R. Jones, Gareth The role of information technology in the organization: a review, model, and assessment, *Journal of Management* 27. 313–346.
8. Markless, S., Streatfield, D. (2001). Developing performance and impact indicators and targets in public and education libraries, *International Journal of Information Management*, 21 (2) 167-179.
9. Mendonca, M. G., Basili, V.R., Bhandari, I. S., Dawson, J. (1998). An approach to improving existing measurement frameworks. *IBM Systems Journal*, 37 (4) 484-501.
10. Kaplan, R. S., Norton, D.P. (1992). The Balanced Scorecard: measures that drive performance, *Harvard Business Review*, 70 (1).71-79.
11. Kaplan R. S., Norton, D. P. (1993). Putting the Balanced Scorecard to work, *Harvard Business Review***,** 71(Sept.), 134-149.
12. Kaplan R. S., Norton, D. P. (1996), Using the Balanced Scorecard as a Strategic Management System", *Harvard Business Review***,** 74 (1). 75- 85.
13. Gottschalk, P, Implementation of formal plans: the case of information technology strategy, *Long Range Planning*, 32 (3) (1999). pp.362-372.
14. Gottschalk, P. (1999). Implementation predictors of strategic information systems plan, *Information & Management,* 36. 77-91.
15. Gottschalk, P. (1999). Implementing predictors of formal information technology strategy, *Proceedings of the 32nd Hawaii International Conference on System Sciences*, 7, p 7035.
16. Boar, B (2001). *The art of strategic planning for information technology,* John Wiley & Sons, Inc., New York.
17. Lee, G., Bai, R. (2003). Organizational mechanisms for successful IS/IT strategic planning in the digital era, *Management Decision*, 41(1) 32-42.
18. Ad de Jong, Ko de Ruyter, Jos Lemmink, The Adoption of Information Technology by Self-Managing Service Teams, *Journal of Service Research*, 6 (2) November 2003 162-179.
19. Davies, Davis., Fred D, (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology, *MIS Quarterly*, 13 319-39.
20. Grewal, Rajdeep, Raj Mehta, Frank R. Kardes (2000). The Role of the Social-Identity Function of Attitudes in Consumer Innovativeness and Opinion Leadership, *Journal of Economic Psychology*, 21, 233- 52.
21. Nunnally, Jum C., Ira H. Bernstein, *Psychometric Theory*, 3rd ed. New York, McGraw-Hill, 1994.
22. Ostroff, Cheri (1993). Comparing Correlations Based on Individual-Level and Aggregated Data, *Journal of Applied Psychology*, 78 (4). 569-82
23. Braithwaite, Timothy, (1996). The Power of IT: Maximizing Your Technology Investments. Milwaukee, WI: ASQC Quality Press. p. 40.
24. Soutmitra, Dutta (2009). Promoting Technology and Innovation: Recommendations to Improve Arab ICT Competitiveness. P.81-96. Weforum.
25. Mrad, F. (2005). Meeting Arab Socio-economic Development through ICT." Presentation at the School of Computer Science, Carnegie Mellon. Available at www.cs.cmu.edu/~cfr/talks/2005-Feb-4.ppt.
26. ITU (International Telecommunication Union). 2007. World Telecommunication/ICT Development Report 2006: Measuring ICT for Social and Economic Development. Geneva : International Telecommunication Union.