

Intelligent Search Lifecycle Architecture for Mass Media Using SOA

Naser El-Bathy^{1,*}, Clay Gloster¹, Ghassan Azar², Mohammed El-Bathy²

¹Department of Electronics, Computer, and Information Technology, North Carolina A&T State University, Greensboro, NC, USA

²Department of Computer Science, Lawrence Technological University, Southfield, MI, USA
cgloster@ncat.edu, gazar@ltu.edu, melbathy@ltu.edu

Abstract Mass communication in general and mass media specifically confront rising challenges to collect information faster at lower cost. These challenges must be coordinated with organization's strategic, tactical, and operational needs via the alignment of Information Technology (IT) with the business strategy. A successful IT strategic plan is based on four main priorities. These are IT-Mass media alignment escalation, IT governance enhancement, IT significance improvement, and IT investments grade. In this study, the IT strategic plan follows a timetable via a roadmap and is exposed to alteration based on the business process changes. The development of IT strategic plan consists of five phases. These phases include the purpose of the plan, the business requirement, and the business requirements' validation. The phases also include the procedures needed for eliminating the gaps between the state of "to be" and the IT processes state, and the plan approval. The ultimate goal is a Modern Mass Media Industry. This paper provides an innovative solution to accomplish this goal by accelerating mass media processes time. The process' time is accelerated via an Intelligent Search Lifecycle Architecture (ISLA); which integrates Search Engine, Information Extraction, Information Retrieval, Data Mining and Data Warehouse using Service-Oriented Architecture (SOA). The design methodology of the study of this solution is a post-positivist approach to empirical research. A prototype is created and examined in order to validate the concepts.

Keywords SOA, ISLA, Data Mining, Modern Mass Media Industry, Data Warehousing

1. Introduction

In the past few years, Mass Media Industry has made major progress at several levels. However, the reality is that the industry's organizations still face severe obstacles mainly in gathering information. Such problem has been emerged due to imperfect and poor media processes. This paper applies Service-Oriented Architecture (SOA) principles by providing services that have concrete meaning on the business level to improve the capability of an enterprise. These services enable Information Technology (IT) architecture for Intelligent Search Lifecycle Architecture (ISLA) as a requirement for the modern mass media industry. This addresses new business requirements in the short term by reusing existing business logic and data models. It results in reducing cost, resources, time, and overheads, while minimizing risks, especially when compared to rewriting entire application systems. Implementing SOA provides acceptable benefits in terms of agility and integrity since it provides a long-term strategy to increase flexibility of an IT infrastructure[1].

The rest of this paper is structured as follow: Section two identifies the problem. Section three presents the outcomes of our research. Section 4 defines research contribution to knowledge. Sections 5 and 6 outline research analysis and design, and finally the conclusion is given.

2. Problem Identification

2.1. Research Problem

The origin of this research problem is lack of applications that if defined will allow the mass media including publishing and news paper organizations to meet its strategic, tactical and operational needs by accelerating processing time. These organizations' processes are subjected to major negative influences. Such negative influences mainly include inefficient access to information, information contradiction, irrelevant information, delays, redundancy, poor interface, and lack of data integration[2]. The reasons for the research problem are absence of the alignment between IT goals and business strategy and absence of IT planned processes. The research problem focuses on the development of intelligent information retrieval and intelligent web (web mining) technique[3, 4].

2.2. Research Question

* Corresponding author:

nielbath@ncat.edu (Naser El-Bathy)

Published online at <http://journal.sapub.org/arch>

Copyright © 2012 Scientific & Academic Publishing. All Rights Reserved

One specific research question which arises is: How does the integration of data mining and data warehousing using SOA accelerate search lifecycle processing time in the mass media organizations to meet its strategic, tactical and operational needs? This question focuses on the integration of data mining and data warehousing using SOA. "Figure 1" describes the Intelligent Search Lifecycle Architecture (ISLA) that forms the major components needed for answering the research question.

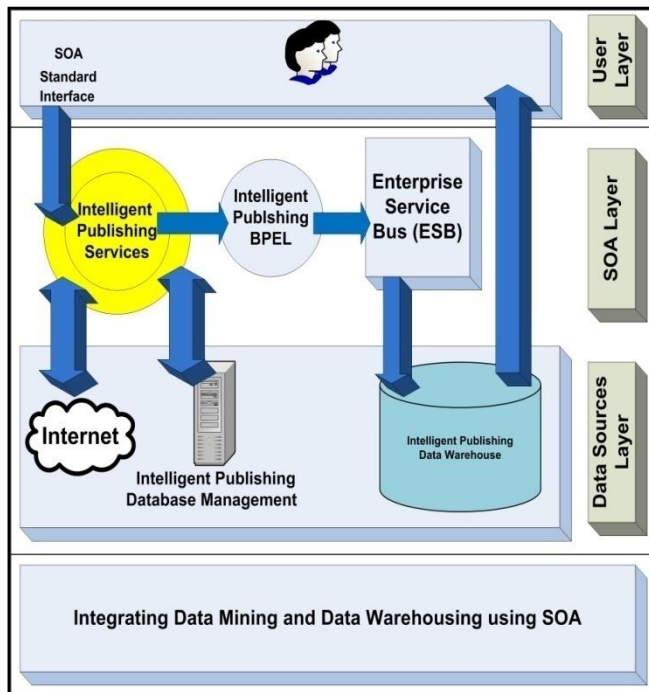


Figure 1. An intelligent integration of data mining and data warehousing with SOA

2.3. Significance of Research

A study of a mass media organization and its domain, defined here as technology employed, product scope, customer orientation, and markets served is important for two reasons. First, the study of this research is a new concept. It defines new mass media search lifecycle concept as a requirement for a modern mass media industry. Second, researchers have

often studied general technical types of software which cannot entirely solve mass media search lifecycle processes time at lower cost.

3. Research Objective

The objective of this research is providing a solution to a very specific problem instance in the area of data mining, data warehousing, and service-oriented architecture in mass media industry. The solution is efficient and computerized Intelligent Search Lifecycle Architecture. It must be developed within the enterprise's framework to establish business requirements analysis details about information that describes management seniors' strategic values, management middle level tactical value, and low level operational value[3]. The levels of strategic, tactical, and operational information compose a hierarchy of information as shown in Figure 2.

The ultimate goal is to modernize mass media industry. However, determining the requirements of such modernization is the main part for successfully achieving the following:

3.1. Mass Media Organizations Objectives

Publishing and newspaper organizations are major elements of mass media. These organizations are known as knowledge-intensive firms. Their general goal is investing its capabilities in producing, distributing, and delivering products that are of distinct, high quality at low cost and on time while meeting the customers' needs. This goal can be achieved through the capabilities of the organizations' main business processes and their devoted social actors. Those are the editors, journalists and image artists who utilize their empirical knowledge and skills in the processes of designing, building, and operating the specified activities [5].

However, the talents and abilities of IT professionals have more important responsibilities comparing to the editors, journalists and image artists. They outline the quality and the distinct capabilities of these organizations[5].



Figure 2. Information architecture

3.2. Architectural Capability

1. SOA is a software architecture form in which software systems can be developed via services composition[5, 6].
2. SOA is a design practices and principles.
3. SOA is a rising successful method that concentrates on loosely coupled, standard-based, and protocol[8-10]. Oracle SOA suite enables business architecture by supporting the development of integrated software systems as reusable web services[12].

3.3. IT Strategic Plan

A successful IT strategic plan is based on IT-business alignment escalation, IT governance enhancement, IT significance improvement, and IT investments priority[13].

The IT Strategic Plan Phases are the purpose of the plan, the business requirement, and the business requirements' validation. The phases also include the procedures needed for eliminating the gaps between the state of "to be" and the IT processes state, and the plan approval[13].

3.4. Outcome Analysis

The purpose of outcome analysis is essentially for the proof of the theory that involves new principles. The outcomes of the research are obtained by designing and developing a unique prototype model. The outcomes are valid to support the research study. The solution can be applied to solve similar practical problems in various domains and therefore the study is a contribution. These outcomes are:

- Constructing ISLA maintains building of connected software systems by the use of SOA.
- The modernization of the mass media including publishing and newspaper organizations is accomplished by applying ISLA
- The implementation of SOA results in decreasing application development time and reducing integration costs. The results also encompass leveraging existing investment in IT assets and boarding decision making throughout the value chain with automated workflows[6].
- Aligning IT with the business strategy is a major outcome. This is achieved by applying an IT strategic plan and governance framework that enable successful execution of

the research solution according to the business goals.

4. Literature Review

During literature review, the area of interest is determined. It is based on professional experience gained from working in the fields of Information Technology, Journalism, Academic, and Academic degrees. Therefore, it is identified in integrating data mining and data warehousing using service-oriented architecture to accelerate mass media search lifecycle processes.

A fair amount of time is dedicated for searching for articles, studies, interviews, and other references to specify a major problem in this area. The search processes have also focused on the implementable techniques that can solve such problem.

This review is intended to summarize current knowledge about mass media mainly publishing and newspaper organizations. It is a crucial mechanism for the development of a new theory to accelerate processing time, and provide a new direction for future research and management of the alignment of IT with mass media organizations strategy.

The major elements of the literature are data sources, service oriented architecture, and data warehouse.

4.1. Data Sources

Databases take the most important part in the computers' business applications. The data modeling is considered crucial design thesis in developing a database. The data modeling is a theory or specification. It describes the structure and the usage of the database[14]. An object class stores the data. "Figure 3" shows an object database model for ISLA.

The data warehouse term is defined as an organization storage area for a homogeneous, coherent, clean and integrated data electronically produced by different operational systems. It also supports tools for extracting, transforming, and loading data into the storage area[15,16]. Data warehouses can be designed and deployed using Oracle Warehouse Builder (OWB) which entirely can be part of SOA approach.

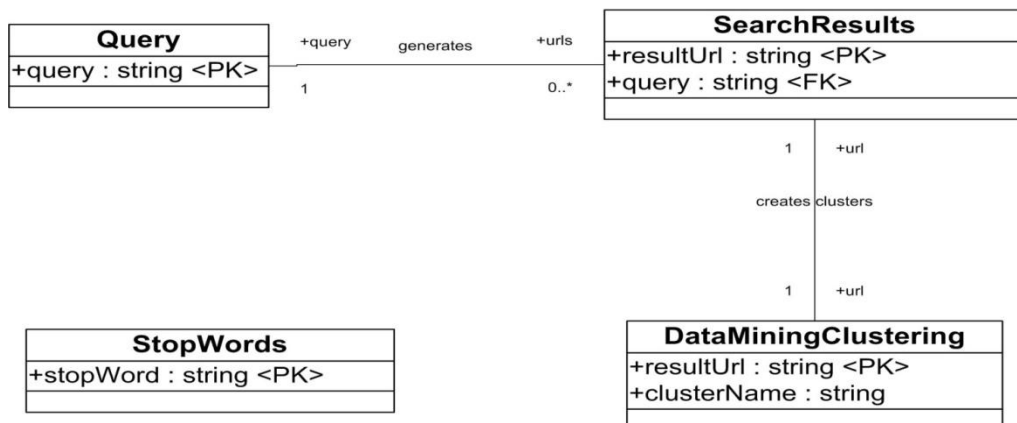


Figure 3. Object class data model for ISLA

Several objects of OWB can be published as Web Services. This enables manipulating the object's functionality using Simple Object Access Protocol (SOAP), and Web Service Descriptive Language (WSDL) standards to completely be included within SOA. In turn, the design of ETL can be orchestrated within Oracle BPEL Process Manager[17].

4.2. Service-Oriented Architecture

SOA allows IT infrastructure to be flexible and adaptable. The standards of web services encompass Web Services Description Language (WSDL), Extensible Mark-up Language (XML), and Simple Object Access Protocol (SOAP).

In this paper, Web services function is a two-step process. Publishing the service is the process that makes a service available. Composing, or orchestrating, the services into business flows is the process that organizes multiple services within end-to-end business process. The orchestration is supported by Business Process Execution Language (BPEL). The components of Oracle SOA Suite that are utilized in this paper include Integrated Service Environment (ISE) to develop services, Oracle BPEL Process Manager, Oracle Enterprise Service Bus (ESB), and Oracle Application Server.

4.3. Data Mining

Data is a pool of simple symbols. Data mining is an element of a larger procedure known as knowledge discovery. The knowledge discovery is obtaining beneficial knowledge from a titanic data pool[18,1]. This research implements theoretical clustering data mining technique in the provided solution. It applies k-means clustering algorithm as a data mining technique.

An intelligent Web is a web mining that its essential benefit is satisfying the users' needs, ranking the resources based on the user's concerns, relating resources to a user search request, and enabling data-and text-mining methodologies in search lifecycle architecture[16].

5. Research Analysis

The research is planned to review recent knowledge about the mass media including publishing and newspaper industries' processes, discover knowledge gaps, and identify future research target. The goal of the research is determining required information for accelerating search processes based on the alignment between IT and the business strategy.

6. Research Design and Procedures

Based on the research question, the study of this research follows an exploratory approach in which a mixed method is established for examining and evaluating the prototype. A mixed method encompasses the quantitative and qualitative approaches as subjective and objective procedures[19]. In a mixed method, the study of this research tends to build the knowledge allegation on practical arguments such as out-

come-directed, or problem-centered[20]. In a quantitative approach, the knowledge is developed via the implementation of postpositivist claims. Examples include lessening to particular questions and testing the theory. Postpositivist viewpoint is the thinking after positivism, disputing the conventional certainty of the knowledge. In a qualitative approach, knowledge allegations are constructed by the knowledge requester based on either the viewpoints of constructivist or advocacy/participatory viewpoint or both[20].

The research design applies a postpositivist approach to empirical research in terms of:

- Literature review: the area of interest is identified
- Assessment of the established theory: a theoretical framework has been developed (Unique Intelligent Search Lifecycle Architecture) to derive workable and testable question
- Theoretical conjecture identifies all the involved variables and describes the possible relationship between them
- Empirical generalizations: a series of clear statements are produced which is testable against further evidence in different domains
- Measuring instrument: prototype model
- Confirmation and refined theory: confirming or refining the theoretical conjecture is conducted based on testing and analysing the evidence in order to bring the original theoretical conjecture closer into the evidence of the findings of the empirical research

A transformative strategy procedure is conducted based on mixed approach to introduce the structure of the problem's topic.

The following are the procedures of the research:

6.1. Developmental Research

A road map is outlined and used as guidance for developing the study of this research[3]. It is based on the following processes:

- Selection of a general area and topic: The topic is original and provides an answer to research question
- Stream of research: The chosen topic is very specific
- A page of definition of terms
- A method of problem solving:
 1. Understanding the problem
 2. Creating a plan
 3. Executing the plan
 4. Looking back to review and discuss the solution

6.2. Research Prototype

The prototype follows the Architected Rapid Application Development (ARAD) model. The prototype requires several iterations before evolving into the final product. During the prototype process model, the study of this research specifies four stages of the prototype's activities. These are planning, analysis, development, and transition. The prototype architecture is based on three-tiered web-enabled prototype architecture.

The prototype intelligent processes are Information Re-

trieval (IR) and Web Mining (WM).

This study developed Java classes to build an easy intelligent information retrieval system. These classes are published and consumed as web services. IR is involved in locating and levelling documents that satisfy the users' information search criteria. Therefore, the study develops an automatic information extraction that extracts the user's related information in a brief structure[14]. "Figure 4" demonstrates the activities of IR processes.

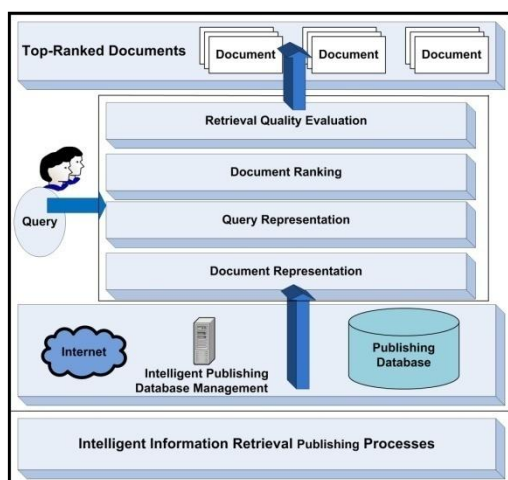


Figure 4. Information retrieval text-based processes

An intelligent Web is a web mining that its essential benefit is satisfying the users' needs, ranking the resources based on the user's concerns, relating resources to a user search request, and enabling data-and text-mining methodologies in search lifecycle architecture. Web mining is a data mining applied to the web[14]. ISLA forms natural groupings of pages whose characteristics are:

- Only relevant documents
- Only unique URLs
- Only unique contents
- Only top ranked documents are clustered

Prototype walkthrough methodology evaluates the architectures of extensive, complicated, and heterogeneous system. Walkthrough supports dynamic development such that the architectural demonstration can be recommended precisely during walkthroughs with the project's stakeholders[22].

The walkthrough participants of this study are the computer science and management professors, and researchers of North Carolina A&T State University, Wayne State University and Lawrence Technological University, and SOA engineers in North Carolina and Michigan States in the United States of America. They discovered the problems related to the phases of the design, development phase, testing, usability, and maintenance. They evaluated the prototype to ensure that it satisfies requirements of this research. They verified that the solution is conceptualized. They verified that the document findings answer the research question and solve the research problem.

6.3. Research Findings

The working prototyping technique is evidence that proves the validity of the new concept and answers the research question.

The design and the development of the prototype demonstrate the conceptual model of this study.

Modernization of mass media including publishing and newspaper organizations is a result of the ISLA new technique. It causes the acceleration of processing time due to the following results:

- Computerizing the new concept
- Integration of search engine, information extraction, information retrieval, data mining and data warehousing using SOA
- Using SOA assures and proves the following:
 1. A alignment of IT goals with business strategy
 2. Opportunities for dynamic governance alignment of business and IT activities
 3. Business changing demands and IT are met
 4. Greater performance of applications by the management of loose coupling
 5. Information Technology infrastructure can be changed and modified to easily meet the business' requirements
 6. Modifications don't require replacing or modifying the existing source code
 7. Developing software systems in a short period of time
 8. SOA integrative software systems are powerful in terms of:

- Flexibility
- Agility
- Reusability

7. Conclusions

Integration of search engine, information extraction, information retrieval, data mining and data warehousing mechanisms using service-oriented architecture is an efficient methodology that can be implemented to develop Intelligent Search Lifecycle Architecture (ISLA).

The implementation of ISLA leads to reducing cost, resources, time, and overheads, while minimizing risks. Implementing ISLA provides acceptable benefits in terms of agility and integrity since it provides a long-term strategy to increase flexibility of an IT infrastructure.

The benefits of using ISLA also encompass leveraging existing investment in IT assets and boarding decision making throughout the value chain with automated workflows.

Implementing ISLA achieves a successful alignment of Information Technology (IT) with the business strategy. The importance of such alignment leads to enhancing the IT abilities to fulfil the essential obligations required for pushing the organizations to be part of modern publishing and newspaper industry.

8. Recommendation

The following recommendations enhance an organization's ability to collect information faster at lower cost and to make accurate decisions:

- Organizing basic and advanced SOA training programs to improve the technical skills of the IT engineers and allowing the journalists to learn these skills as well
- Development of an effective IT strategic plan that should satisfy the stakeholders' requirements, business trends, and IT governance model to improve and support decision-making
- Development of Intelligent Enterprise Architecture (IEA) systems to make the alignment of technology strategy with business goals possible
- IT ability to deliver business value to the organization can be accomplished by supporting new products, processes, and opportunities

Dedication

The primary author of this paper, Dr. Naser El-Bathy, has dedicated this research to the Egyptian writer, Ibrahim El-Bathy, who passed away in 1979. His methods in writing and dedication to the publishing and newspaper industries are what drove him to choose this concentration for this research. Ibrahim El-Bathy achievements in the field of publishing and newspaper industries are reported in several Arabic newspapers and magazines during his life and after his death.

REFERENCES

- [1] A. Ronnie, "A new generation of Middleware solution for a Near-Real-Time data warehousing architecture," *Electro/Information Technology IEEE International Conference*, Chicago, IL, United States, pp. 192-197, May 2007.
- [2] W. Richard, *Data management: databases and organizations*. Wiley, 2006.
- [3] C. Perks, T. Beveridge. *Guide to Enterprise IT Architecture*. New York: Springer-Verlag, 2003.
- [4] D. Remenyi, B. Williams, A. Money, E. Swartz. *Doing Research in Business and Management – An Introduction to Process and Method*. London. *Thousand Oaks, New Delhi: SAGE Publications*, 2005.
- [5] B. Thomas and M. Ciaran, "Shaping information and communication technologies infrastructures in the newspaper industry: cases on the role of IT competencies," *International Conference on Information Systems, Proceedings of the 20th international conference on Information Systems*, Charlotte, North Carolina, United States, pp. 364 – 377, 1999.
- [6] G. Qing and L. Patricia, "A stakeholder-driven service life cycle model for SOA," *ACM*, New York, NY, USA, pp. 1-7, 2007.
- [7] D. Mark, "Surfing the net for software engineering notes," *ACM*, vol. 30, Number 6, 2005.
- [8] S. Derek, J.A. Hamilton, and M. Richard, "Supporting a service-oriented architecture," *Society for Computer Simulation International*, San Diego, CA, USA, pp. 325-334, 2008.
- [9] K. Dirk, B. Karl, and S. Dirk, "Enterprise SOA service-oriented architecture best practices," Prentice Hall, 2005
- [10] P. Mike and H. Willem-Jan, "Service oriented architectures: approaches, technologies and research issues," *The VLDB Journal*, Springer Berlin / Heidelberg, vol. 16, Number 3, pp. 389-415, 2007.
- [11] D. Vladimir, "Development of applications with service-oriented architecture for grid," *ACM New York, NY, USA*, Vol. 374, 2008.
- [12] S. Deborah, "Oracle SOA suite quick start guide 10g (10.1.3.1.0), Oracle, 2006.
- [13] A. Cullen, M. Cecere, C. Symons, B. Cameron, L. Cardin, L. Orlov, and B. Belanger, "The IT strategic plan step-by-step deliver an actionable plan in a reasonable timeframe," *Forrester Research, Inc.*, 2007.
- [14] A. Rajendra and L. Pawan, "Building an intelligent web – theory and practice," Jones and Bartlett Publishers, Sudbury, Massachusetts, 2008
- [15] W. Cuiru and L. Shuangxi, "SOA based electric power real-time data warehouse," *IEEE Workshop on Power Electronics and Intelligent Transportation System*, pp. 355 – 359, 2008.
- [16] K. Michael, B. Arthur, and L. Philip, "Database systems: an application-oriented approach," Pearson Education, Inc., 2005.
- [17] P. Padmaja and S. Vishwanath, "Oracle warehouse builder data modeling, ETL, and data quality guide 11g" Release 2 (11.2), 2009.
- [18] C. Marcello, M. Giuseppe, and T. Gianfranco, "A web text mining flexible architecture," *World Academy of Science, Engineering and Technology* 32, pp. 78-85, 2007.
- [19] M. Mahmud, "A mixed method for evaluating input devices with older persons," *ASSETS'06*, Portland, Oregon, USA. *ACM*, pp. 295 – 296, October 22-25, 2006.
- [20] J. Creswell, "Research design: qualitative, quantitative, and mixed methods approaches," *SAGE Publications International Educational and Professional Publisher*, Thousand Oaks, London, New Delhi, PP. 39-40, 2003.
- [21] P. Chang, "Research methodology – life cycle of a dissertation project in information systems and methods of problem solving," *Proceeding of IMSCI/EISTA*, vol. 3, pp. 40-46, Orlando, Florida, 2008.
- [22] S. Haynes, L. Skattebo, J. Singel, M. Cohen, and J. Hime-Iright, "Collaborative architecture design and evaluation," *University Park, Pennsylvania, USA, ACM*, pp. 219 – 228, 2006.