A PROPOSED FRAMEWORK FOR KNOWLEDGE DISCOVERY IN ENTERPRISE ARCHITECTURE

M. El Kourdi, H. Shah, and A. Atkins
M.Elkourdi@staffs.ac.uk, H.Shah@staffs.ac.uk, A.S.Atkins@staffs.ac.uk
Faculty of Computing, Engineering and Technology, Staffordshire university, Octagon, Beaconside, Stafford, ST16 9DG, United Kingdom

Abstract

The aim of this research is to develop an information agent framework for knowledge discovery in enterprise architecture (EA). This framework is based on specific purpose ontology and knowledge discovery techniques. Such framework would facilitate strategic decision making for EA stakeholders by enabling them to analyze and monitor the portfolio of processes, data, applications, and organizational units in terms of their correlation and impact in the overall organization. This framework is very useful for affording key stakeholders with the appropriate view that is reliant on an accurate and concise picture of systems, applications, technologies and other infrastructure elements in the business and how these integrate to serve the enterprise. The paper discusses the concepts and components of this framework. Potential benefits of this framework over existing approaches are also discussed.

Keywords: Enterprise architecture (EA), ontology, EA stakeholders, intelligent agents, knowledge discovery/data mining, Business process mining (BPM).

1 INTRODUCTION

The value of EA has been recognized as a methodology to manage the organization’s complexity by stakeholders, which are concerned with supporting knowledge that will enable them to address issues such as Business/IT alignment, complexity reduction, EA portfolio management (Infosys, 2005, Luftman and McLean, 2004). Indeed, the supporting of knowledge is crucial to the fulfilment of their roles in order to enable them to be engaged efficiently and to mitigate the risks of IS projects. Business/IT alignment (Wegmann et al., 2005, Cumps et al., 2006), which is the integration of IT resources and business resources, is considered to be one of the major concerns of stakeholders to keep sustainable competitive advantages. There have been several surveys that have been conducted in order to study the major key concerns of EA stakeholders. In 2004, the society for information management SIM (Luftman and McLean, 2004) has reported in a formal survey that 70% from 182 responders of its members are concerned with IT-Business alignment. Infosys (Infosys, 2005) also carried out an electronic survey to investigate the key concerns, approaches, focus area, and key success factors among 45 IT decision makers and enterprise architects in large companies. The survey revealed that the major key concerns are IT cost reduction, technology and application portfolio simplification, enablement of business change, and business IT alignment.

Sustainable competitive advantages are based on several criteria, and it has been suggested that maintaining competitive advantages rely on Business/IT alignment(Wegmann et al., 2005, Cumps et al., 2006), an organization’s ability to leverage IT resources to organizational units and processes (Powell and Dent-Micalef, 1997, Ross et al., 1996), and strategic information system planning(Segars and Grover, 1999). However, because of the fluctuating nature of IT/Business alignment, its evolutionary aspect over time (Sabherwal et al., March-April 2001), and its relationship with other criteria such as resource planning impose a further focus on how organizations can obtain alignment.
Consequently, business/IT alignment is considered to be one of the major concerns of stakeholders for decision making to enable organizations to maintain competitive advantages. One of the primary objectives of Business/IT alignment is to make explicit how EA elements such as resource, processes, data, and organizational units are inter-related and organized to achieve strategic business objectives. To achieve this, it would be beneficial to develop a framework and a mechanism that allows stakeholders to visualize the correlation between different EA elements and architectural models. The knowledge of degree of correlation between different architectural elements will allow capturing the causal effects of removing or altering a given element (impact analysis). For instance, the substitution of an existing information system by an enterprise resource planning (ERP) software system necessitates an *a priori* knowledge of the inter-related processes, data, and organizational units for better decision making. Evaluating the correlation between different architectural models, on the other hand, is a means to predict the common practices in different EA business lines.

However, the level of details in the EA underlying repository databases is overwhelming, which hinder significantly the involvement of stakeholders in the development process of information systems in EA. In that sense, checking the correlation of EA elements is a difficult task because of the high dimensionality of the architectural models in the underlying repositories. Despite the existence of repositories capturing a wealth of enterprise architectural information it can be difficult to get an accurate picture of this information when needed and to exploit it in enterprise architectural terms for business decision making. As such, knowledge discovery in enterprise architecture repositories is very useful for affording EA stakeholders with the appropriate view that is reliant on an accurate picture of systems, applications, technologies and other infrastructure components in the business and how these integrate to serve the enterprise. Indeed, making the inter-relationship between EA elements explicit in terms of their correlation is problematic and impractical because of the massive divergent information stored in the underlying EA repositories. To cope with such complexity, an information agent framework for knowledge discovery in EA that makes use of specific purpose ontology as a knowledge base is proposed.

This paper is organized into the following. Section 2 presents an overview to enterprise architecture. Section 3 describes the state of the art by investigating similar approaches. Section 4 summarizes the main components of the proposed framework, which are ontology and knowledge discovery. Section 5 presents the proposed framework in terms of its building blocks. Section 6 and 7 present respectively the benefits of the proposed approach, and conclusions and future work.

### 1.1 Enterprise architecture

EA is described as organizing logic for business processes and IT infrastructure, reflecting the integration and standardization requirements of the company’s operating model in order to achieve business agility and profitable growth (Ross et al., 2006). EA is a means to afford the fundamental technological and business infrastructure for an IT strategy, and align business strategies with the implementation. EA is also considered to be as a formal way to manage the complexity of organizations in order to maintain competitive advantages. EA frameworks, on the other hands, identify the scope of EA and decompose various elements of the architecture onto structured levels and elements. More formally, EA frameworks describe a method for designing IS in terms of a set of building blocks and how these blocks fit together. Several EA frameworks such as ARIS (Scheer, 1999) and DODAF (DOD, 2006) have been adopted for operational use in many organizations.

### 2 RELATED WORK

To our knowledge, no work has been reported in the knowledge discovery or the mining of EA repositories, and the bulk of the work has been directed to business process mining (BPM) in order to capture and improve business processes from event–based logs, also referred to as audit trails (VanderAalst and Song, 2004, VanderAalst et al., 2003). More recently, business process mining
(BPM) has emerged as a discipline for reengineering business process. BPM provides techniques and methods for capturing process behaviour from existing data, stored in audit trails. BPM is considered to be as a potential approach to optimize business processes by deriving new business process logic. Normally, Business process creation is user vision dependent and might be unstructured. As business process consists of the creation, provisioning, and the enactment phase (O’Brien and Wiegand, 1998), BPM can be formally defined as a mechanism to derive a process model based on the data gathered in the enactment phase. However, the information stored in these audit trails is usually insufficient to advocate the inter-relationship and predict the association between different elements in EA. This is analogically to software engineering which emphasizes that tracking errors at the design phase will potentially reduce the maintenance cost of the system. Consequently, mining the EA repository that consist normally of design elements will have the potential to reduce this cost overhead and to mitigate errors raised at the earlier stages of the IS development process in EA. Figure 1 compare BPM and EA repository mining in terms of the potential benefits. As noticed in Figure 1, one of the drawbacks of BPM in terms of outcomes is its restriction to solely make corrective actions for business process models in EA repositories. EA repository mining, on the other hand, absorbs the benefits of making corrective actions with regards to all the architectural models in EA.

![Figure 1. EA Repository Mining Versus Business Process Mining.](image)

3 THE MAIN BUILDING BLOCKS OF A FRAMEWORK FOR KNOWLEDGE DISCOVERY IN EA

This section highlights the characteristics of the major building blocks of the proposed framework, which are ontology and knowledge discovery.

**Ontology**

Ontology, defines the concepts and terms and their interrelationship in a domain of discourse and is considered as a specification of conceptualization (Gruber, 1993). In regard to EA, several explicit and implicit ontologies have also been proposed (Fox et al., 1995, Wegmann and Naumenko, 2001, OMG, 1995-1996). Some good examples of explicit ontologies are SEAM (Wegmann and Naumenko, 2001) and TOVE (Fox et al., 1995). TOVE proposes organization ontology for enterprise modelling by deriving preliminary concepts for linking the structural and the behavioural aspects of the organization through the use of a set of competency questions. SEAM ontology defines an OO modelling concept of hierarchical systems using a UML like CAD tool. The SEAM terms of the ontology were
formalized by integrating part 2 and part 3 of for reference model for object distributed processing (RM-ODP) (OMG, 1995-1996). Implicit ontologies, on the other hand, are embodied in the concepts and the terms that are used in commercial EA tools such as ARIS toolset (Scheer, 1999) to serve as entities for organizational modelling.

While a general EA ontology tends to gather all the information about the organization domain, specific ontologies depend mainly on the problem to be solved. Typically, General ontology tends to contain all the object types that are necessary for modelling organizations. Within an EA domain, the information agent needs specific purpose ontology as a knowledge representation scheme in order to analyze EA elements and monitor their correlation from a given stakeholder’s viewpoint. Accordingly, the agent has a set of goals that specify explicitly the terms of the problem to be solved. Such terms provoke the derivation of the relevant portion of the general EA domain ontology that is encapsulated implicitly in EA repository. In that sense, the specific purpose ontology is formed from the terms of the problem to be solved, and the relevant portion of the EA general ontology that is representative for problem solving tasks. As far as the correlation of EA elements is concerned, existing concepts from EA general ontology will be reused and supplemented with the needed concepts for impact analysis and similarity analysis in order to avoid re-inventing the wheels. More specifically, competency questions, which have been proposed as a way to engineer enterprises (Gruninger and Fox, 1994), are formed to express the goals of the agent. Competency questions serve as a basis to identify and capture the necessary concepts and their inter-relationships for the problem to be solved. Clearly, the needed concepts from reusable ontologies should be also identified by means of competency questions and combined with the new concepts that are needed by the agent to satisfy its goals. Some examples of these competency questions are as follow:

- What organization units use a given technical resource?
- Which technical resources support a given business process?
- Which data is consumed or generated by a given business process?
- Who are the process owners?

Figure 2 illustrates the building blocks of the specific purpose ontology. To this end, the development of specific purpose EA ontology as a knowledge representation scheme for the proposed agent is useful since it is the basis for semantic based processing in EA repository. The specific purpose ontology serves as a high level abstraction for structuring the EA elements in conceptual terms. The benefits of this specific ontology are two-fold. Firstly, it provides a mechanism for representing the domain knowledge of the agent by making explicit the vocabulary for annotating EA elements. Secondly, it operates as a mediator between the stakeholders’ analysis requests and the underlying EA repository.

![Figure 2. The specific purpose ontology in terms of its building blocks](image-url)
3.1 Knowledge Discovery

Due to the problems of massive amount of divergent information on enterprise databases, relevant information extraction has become a crucial task for strategic decision making. Knowledge discovery or data mining has emerged as a way to cope with such a complexity. One of the primary objectives of knowledge discovery is the derivation of useful rules and hidden patterns in large data sets in order to enhance the decision making strategy. In this research, association rules mining and concept based similarity learners, which are unsupervised knowledge discovery mechanisms, will be used to learn from the EA repositories to derive useful rules and concepts for impact analysis.

Mining association rules have been widely used in large database systems (Agrawal et al., May 1993, Cheung et al., Feb. 1996, May 1993, Fu and Han, Dec. 1995), large data sets (Klemettinen et al., 1994), and data warehouse (Chaudhuri and Dayal, 1997). Such applications include market basket analysis (Brin et al., May 1997), and the discovery of spatial association rules in geographic information database (Koperski and Han, Aug. 1995). Association rules are implications of the form (X \(\Rightarrow\) Y), which states that the transactions that contain X are most likely to contain Y. The relevance of the rule is based on two criteria. Firstly, the support which measures the percentage of transactions that contain both X and Y. And secondly, the confidence which measures the percentage of transactions that contain only X versus the items that contain X and Y (the probability that Y occurs when X occurs). Typically, association rules use database transaction for analysis purpose. On the other hands, EA repository databases hold the architectural elements, which are encapsulated in different architectural models. Similarly to database transactions, architectural models can be logically viewed as bag of concepts (EA elements) instances in EA repository database. As such, the application of association rules mining on the underlying EA repository databases that comprise these architectural models would be useful in predicting the combination and the correlation of EA elements. The intuition behind the use of association rules mining techniques stem from their simplicity and their power to derive useful information and facts, and hidden patterns from EA repositories regarding the correlation of different EA elements.

The structure of EA repository database is represented as concept taxonomy from high level concepts to primitive levels, as illustrated in Figure 4; necessitates the use of a technique that automatically estimate the semantic distances between values of categorical attributes. The aim of concept similarity estimation is to gain some insight as to whether different groups and models tend to group together. Such a correlation between different groups and architectural models in EA is useful since it will enable stakeholders to track different activities and common practices of an enterprise in different domains. Concepts similarities mining have been used to answer imprecise queries in information retrieval (Nambiar and Kambhampati, November 2003.), and to find interesting patterns in textual database (Feldman and Dagan, August 1995, Haveliwala et al., May 2002.). This research intends to employ concept similarities mining to cluster and derive the affinity between different groups and architectural models concepts in an EA repository database.

![Figure 4. Taxonomy of an EA Repository Database.](image-url)
4 INFORMATION AGENT FRAMEWORK FOR KNOWLEDGE DISCOVERY IN EA

Intelligent software agents are an emerging technology that promises to provide systems with a form of autonomy, responsiveness, and adaptability to structure and environment. Recently, agents are classified into several categories such as reactive agents, cognitive agents, and information agents (Woolbridge, 2002). Such a classification is based on the features and the functionalities afforded by agents in the environment in which they operate. Currently, there is no agreed formal and shared definition of an agent among research communities. In this paper, the IBM definition of agents that describes intelligent agents, also referred to as information agents, as software entities that carry out some set of operations on behalf of a user or software with some degree of autonomy, and in doing so, employ some knowledge or representation of the user's goals or desires will be complied to (Franklin and Graesser, 1996).

In this paper, it is proposed to the development of an information agent framework in order to analyze and monitor different elements and distinct groups from an EA repository databases. Differently from multi-agents systems, the information agent depicts an intelligent software that presumes the ontology as a knowledge base in order to carry out some reasoning tasks. This framework makes use of a specific purpose embedded ontology representing user’s goals, and a knowledge discovery inference engine that will be used to retrieve relevant rules and facts that are reliant to impact and similarity analysis in the EA repository databases. As depicted in figure 3, the framework consist a set of building blocks which are ontological formulation, knowledge extraction, knowledge discovery, and results presentation.

Ontological formulation deals with defining the sufficient knowledge items that an agent needs to satisfy the goal of an EA stakeholder. These goals are expressed by the analysis requests initiated by EA stakeholders, as illustrated in figure 3. In this context, Ontological formulation specifies which objects need to be talked about and which one might be ignored to let the inference procedures operate on the developed taxonomies comprised in the specific purpose ontology; in order to derive the necessary knowledge items to realize the goal. Yet, Ontology formulation feed forward the EA artefacts or the knowledge items that are needed to be extracted from the EA repository database. EA artefacts denote the EA elements and the architectural models that are encapsulated in the EA repository. Knowledge extraction intends to apply information retrieval techniques in order reduce and index the architectural models in the underlying EA repository databases to tabular structures. These structures are means to organize and store the relevant ontological artefacts that represent the knowledge items from a specific architectural viewpoint. As a pruning strategy, knowledge extraction serves as a basis for applying knowledge discovery techniques that we have presented so far as one of the main components of the proposed framework. Association rules and concepts similarity mining (clustering) are two knowledge discovery techniques that have been presented in terms of their potential benefits to oversee the overall organizational impact in EA, as well as the similarities between the dispersed architectural models and groups. The discovered patterns will be presented and interpreted to provide a feedback to EA stakeholders.
5 THE BENEFITS OF THE PROPOSED FRAMEWORK

In the context of EA, the proposed framework, along with its knowledge discovery scheme is thought to be useful in many ways and are summarized into the following:

- Impact analysis: EA knowledge discovery is very useful in terms of predicting the elements that will be affected if an enterprise plans to change or remove a component.
- Resource oversight: the overlapping and the redundant resources will be identified by a means of their correlation and dependency to the other EA elements.
- Traceability: As such, one of the primary objectives of traceability is to advocate the coherence of distinct elements in EA. The EA repository mining scheme is very useful in terms of revealing the correlation between different elements.
- Social networks manipulation: The cross-organizational relationships can be identified through clustering different groups in terms of the used models and elements. Yet, the common activities among different organizational units can be visualized.

Consequently, stakeholders can track EA elements in terms of their evolutionary aspect by enabling them to refine and take new initiatives with regards to planning different EA roadmaps. Moreover, this framework will provide stakeholders with feedback from the organization concerning the consistency and the correlation of the EA elements. This feedback will enable stakeholders to mitigate the risks and to make corrective actions in the earlier stages of IS project lifecycle.

6 CONCLUSIONS AND FUTURE WORK

In this paper, an information agent framework, along its components have been outlined in order to satisfy the key concerns of stakeholders embodied in Business/IT alignment. These components include a specific purpose ontology, ontological formulation, knowledge extraction, knowledge discovery, and Results interpretation and presentation. The paper also highlights and discusses association rules and concept similarity mining as knowledge discovery techniques to serve as a learning scheme of the proposed framework. Future work will focus on the development of specific purpose EA ontology for impact analysis and similarity analysis in order to serve as a basis for implementing the proposed framework.

References


