

# Knowledge Based Consolidation of UML Diagrams for Creation of Virtual Enterprise

Debasis Chanda<sup>1</sup>, Dwijesh Dutta Majumder<sup>2,3</sup>, Swapan Bhattacharya<sup>4,5</sup>

<sup>1</sup>Jadavpur University, Kolkata, India

<sup>2</sup>Electronics & Communications Sciences Unit, Indian Statistical Institute, Kolkata, India

<sup>3</sup>Institute of Cybernetics Systems & Information Technology, Kolkata, India

<sup>4</sup>National Institute of Technology, Durgapur, West Bengal, India

<sup>5</sup>Department of Computer Science & Engineering, Jadavpur University, Kolkata, India

E-mail: [cdebasis04@yahoo.co.in](mailto:cdebasis04@yahoo.co.in), [ddmdr@hotmail.com](mailto:ddmdr@hotmail.com), [director@nitdgp.ac.in](mailto:director@nitdgp.ac.in)

Received December 25, 2009; revised January 27, 2010; accepted February 20, 2010

## Abstract

In this paper we address the problem related to determination of the most suitable candidates for an M&A (Merger & Acquisition) scenario of Banks/Financial Institutions. During the pre-merger period of an M&A, a number of candidates may be available to undergo the Merger/Acquisition, but all of them may not be suitable. The normal practice is to carry out a due diligence exercise to identify the candidates that should lead to optimum increase in shareholder value and customer satisfaction, post-merger. The due diligence ought to be able to determine those candidates that are unsuitable for merger, those candidates that are relatively suitable, and those that are most suitable. Towards achieving the above objective, we propose a Fuzzy Data Mining Framework wherein Fuzzy Cluster Analysis concept is used for advisability of merger of two banks and other Financial Institutions.

Subsequently, we propose orchestration/composition of business processes of two banks into consolidated business process during Merger & Acquisition (M&A) scenario. Our paper discusses modeling of individual business process with UML, and the consolidation of the individual business process models by means of our proposed Knowledge Based approach.

**Keywords:** Knowledge Base, Predicate Calculus, Service Oriented Architecture, UML, Fuzzy Data Mining, Cluster Analysis

## 1. Introduction

The virtual enterprise (VE), which is in general the collaborative partnership between business partners in value chains, has become a key factor for survival in the competitive business environment. The VE is a set of collaborative business processes to produce services and products. A VE is a temporary organization which is created according to a business opportunity and is dissolved when the business opportunity no longer exists. The VE is designed to increase competitiveness, to optimize resource utilization, to increase scale of the business, and to take advantages of the complementary capabilities of the business partners [1].

The VE is made up of the higher level value chains and each value chain is made up of the collaborative business processes for business execution in a loosely-coupled

way. Each business process is regarded as the core competitive functionality of the lower-level VE and the lower level VE repeatedly is composed of the lower-level value chains. This composition manner among the VEs, the value chains, and the collaborative business processes is repeated recursively until these cannot be divided any more. Representative modeling languages are the Unified Modeling Language (UML) and the Integrated Definition methods (IDEF) [1].

In the world of business the manifestation of VEs is through M&As (Mergers & Acquisitions)/JVs (Joint Ventures).

Service-oriented computing is becoming the prominent paradigm for leveraging inter-enterprise information systems to complete higher-order business transactions at the heart of the modus operandi of the virtual enterprise (VE). Business process is a set of linked activities that

create value by transforming an input into a more valuable output. SOA (Service Oriented Architecture) is a design framework for construction of information systems through combination of services. A service is a program unit which can be called by standardized procedures, and which can independently execute assigned functions [2].

There is an increasing acceptance of Service-Oriented Architectures (SOA) as a paradigm for integrating software applications within and across organizational boundaries [3]. While the technology for developing basic services and interconnecting them on a point-to-point basis has attained a certain level of maturity, there remain open challenges when it comes to engineering services that engage in complex interactions with multiple other services. A number of approaches have been proposed to address these challenges. One such approach, known as (process oriented) service composition, has its roots in workflow and business process management.

The purpose of Workflow Management Systems [4] is to execute Workflow Processes. Workflow Processes represent the sequences of activities which have to be executed within an organization to treat specific cases and to reach a well defined goal. Of all notations used in the Software Industry, UML [5] is one of the best accepted.

Several approaches (e.g., standards, reference models, architectures, frameworks, industry-neutral and industry-specific initiatives) have been developed mainly on technical aspects related to inter-organizational communication. Relevant European initiatives, frameworks and roadmaps to develop interoperability of enterprise applications and software are ATHENA ([www.athena-ip.org](http://www.athena-ip.org)), IDEAS ([www.ideas-roadmap.net](http://www.ideas-roadmap.net)), INTEROP NOE (<http://interop-noe.org>).

Fuzzy Data Mining scores over classical data mining techniques since it involves knowledge discovery from data expressed in real world terms. Existing work on Fuzzy Data Mining in the Banking domain is directed more towards managing customer relationships. Use of fuzzy mathematical data mining techniques for merger/consolidation of banks are a new area of work in the fuzzy mathematical data mining space.

The existing work in fuzzy data mining covers many functional areas like Customer Relationship Management (CRM), Web navigation, targeted customer campaign in retail banking, electrical load flow analysis and forecasting, transportation systems, medical systems, large database systems with multi-dimensional data, etc.

Most of the above mentioned works focus on classification of data, association rules and if-then analysis, clustering & pattern analysis, use of linguistic real world variables & terms, and other combinations of data mining & fuzzy mathematical techniques.

The goal of this paper is to present a modeling framework for the VE, which is focused on process composi-

tion by consolidating individual UML process models into a consolidated process model. This framework uses Predicate Calculus Knowledge Bases. We also propose a Data Mining model, using a fuzzy mathematical approach, which aims to discover knowledge in banking databases. This paper explores a fuzzy mathematical model of data mining to determine those candidate organizations that are suitable for merger.

The rest of this paper is organized as follows: Section 2 reviews the previous approaches which are related to our modeling framework. In Sections 3 & 4 we introduce our modeling framework and an example scenario to facilitate understanding. Finally, Section 5 provides some conclusions.

## 2. Related Work

Originally coming from the business world, service-oriented architecture (SOA) paradigm is expanding its range of application into several different environments. Industrial automation is increasingly interested on adopting it as a unifying approach with several advantages over traditional automation. In particular, the paradigm is well indicated to support agile and reconfigurable supply chains due to its dynamic nature. In this domain, the main goals are short time-to-market, fast application (re)configurability, more intelligent devices with lifecycle support, technology openness, seamless IT integration, etc. The current research challenges associated to the application of SOA into reconfigurable supply chains are enumerated and detailed with the aim of providing a roadmap into a major adoption of SOA to support agile reconfigurable supply chains [6]. In our Paper, we adopt the SOA paradigm to represent Banking Services.

Although the Web was conceived for human use, this infrastructure has evolved to allow computer programs to become major players of the Web. Recently, the trend in software development has been converging towards reusing and composing loosely coupled functionality accessible by the Web, commonly known as services. Service oriented computing methodology replaces the development of specific software components with a combination of service discovery, selection and engagement. A typical service-oriented architecture (SOA) has three main parts: a provider, a consumer and a registry. A registry provides the foundations for service discovery and selection. Up until now, the software industry has broadly adopted SOA by using Web service technologies. A Web service is Web accessible software that can be published, located and invoked by using the standard Web infrastructure [7]. Our approach proposes a Knowledge Base which is a repository of processes/services; discovery is affected by pattern search & unification/substitution.

The research ‘An XML-based schema definition for model sharing and reuse in a distributed environment’ [8] leverages the inherent synergy between structured modeling and the extensible Markup Language (XML) to facilitate model sharing and reuse in a distributed environment. This is accomplished by providing an XML-based schema definition and two alternative supporting architectures. The XML schema defines a new markup language referred to as the Structured Modeling Markup Language (SMML) for representing models. The schema is based on the structured modeling paradigm as formalism for conceiving, representing and manipulating a wide variety of models. Overall, SMML and supporting architectures allow different types of models, developed in a variety of modeling platforms to be represented in a standardized format and shared in a distributed environment. In our paper, we create a Knowledge Base that facilitates sharing and reuse in a distributed environment. Collaborative learning serves as an important part of e-learning, increasing interactivity and accessibility to various learning resources, either synchronously or asynchronously among users. Distributed interactivity through Web services forms the focus of this paper [2]. The paper reviews related work on service-oriented architecture (SOA), distributed infrastructure, business process management (BPM) and highlights the need to integrate SOA technologies for meaningful and interactive collaborative learning processes. The significance of the study is a SOA approach to enhance the interoperability, flexibility and reusability of e-learning content in a collaborative environment. In our Paper we focus on business services in the Banking domain, and consider how these services are consolidated. This is relevant from the BPM context.

The paper [1] suggests a framework for designing agile and interoperable VEs. Existing approaches, such as enterprise architecture, model driven architecture, domain specific methodology, meta-modeling, framework-based development, etc., are reviewed, revealing that none offers a complete solution that supports all the aspects of the VE modeling in an elegant manner. This work develops a systematic modeling framework by harmonizing these approaches and combining the individual advantages of each approach to produce integrated synergy effects. This modeling framework can be used for business managers or business domain experts to build an agile and interoperable VE quickly and systematically with insights. It also supports a coherent enterprise modeling in which various stakeholders having their own aspects and methodology, such as an IT manager and a business manager, can communicate effectively. Our Paper proposes a framework/architecture for consolidation of business services that may be adopted on enterprise-wide basis.

The vast diversity of implementation and support platforms for service-oriented architectures (such as Web,

Grid or even CORBA) increases the complexity of the development process of service-based systems. With the aim of facilitating the development of service oriented solutions, Marcos Lopez-Sanz *et al.* [9] propose the specification of an architecture centric model driven development method. To achieve this, they study the architectural properties of the SOA paradigm and follow a development approach based on the MDA proposal. MDA proposes a separation of the development process in abstraction levels. This makes MDA suitable to tackle the development of service-oriented systems. This paper describes a UML profile for the PIM-level service-oriented architectural modeling, as well as its corresponding meta model. PIM (Platform Independent Model) level is chosen because it does not reflect constraints about any specific platform or implementation technology. To exemplify and validate the profile, a case study is presented in which the proposed profile is used. They propose UML profile to be used in the definition of the architecture model for SOA at PIM-level. Our Paper proposes UML for knowledge capture, and adopts a Knowledge Based approach for arriving at a PIM level service oriented model for consolidation of services.

Model-driven development [10] is a software development framework that emphasizes model-based abstraction and automated code generation. Service-based software architectures benefit in particular from semantic, ontology-based modeling. Claus Pahl’s paper [10] presents ontology-based transformation and reasoning techniques for layered semantic service architecture modeling. Integrated ontological layers support abstract domain modeling, architectural design, and interoperability aspects. Ontologies are beneficial due to their potential to formally define models, to allow reasoning about semantic models, and to automate transformations at all layers. Ontologies are suitable in particular for the Web Services platform due to their ubiquity within the Semantic Web and their application to support semantic Web services.

Model-driven development combines layered modeling techniques based on notations such as the Unified Modeling Language (UML) with automated transformations and code generation. Recently, ontology-based modeling has been investigated as a semantic modeling framework that enhances the semantic richness of the classical UML-based approaches. While formal modeling and reasoning is, to some extent, available in the UML context in form of the Object Constraint Language OCL, ontologies as logic-based formalisms offer full reasoning support. A second benefit of ontologies as sharable knowledge representations is the potential to easily reuse and share models. Providers need to provide an accurate description or model for a service that can be inspected by potential clients. In particular the attention that Web services have received recently emphasizes the importance of service-orientation as the architectural paradigm. Service-oriented architecture is becoming an

important software engineering paradigm. The focus [10] is the Web services platform based on techniques such as the service description language WSDL and the service invocation protocol SOAP, but also extensions like the service-based business process execution language WS-BPEL. This specific area is particularly suitable to demonstrate the benefits of semantic ontology-based modeling due to the component-orientation and distributed nature of service-based software development with its emphasis on provision and discovery of descriptions and on sharing and reusing of models and services. In addition to the modeling capabilities, ontologies also provide a formal framework that enables reasoning and transformation and, thus, supports the automation of development activities. Process-orientation and interaction and composition in distributed architectures are central for service-based software systems. Our paper adopts modeling technique based on UML based notation. We also propose a Knowledge Based framework for consolidation of business processes, instead of widely used IDE (Integrated Development Environment) based techniques for modeling business processes that use BPEL execution.

Choreography is an essential element of business integration that allows the modeling of the external behavior of services for a specific interchange or collaboration. Existing service-oriented technologies attempt to model such external visible behavior. However, they lack the decoupling and consistent semantic support required in heterogeneous B2B environments.

Arroyo *et al.* [11] describe the practical application of a semantic web service-based choreography framework, to generic Request for Quote (RFQ) and quote processes based on the OAGIS canonical model. In doing so, the paper has the objective of showing how existing limitations are overcome, by means of the intermediate structures that conciliate the heterogeneities between services from the semantic descriptions of the message exchange patterns (MEPs) they follow. We propose a consolidation/choreography framework that adopts Knowledge Based approach for Banking System.

David Chen *et al.* [12] define and clarify basic concepts of enterprise architectures. An overview on architectures for enterprise integration developed since the middle of the 1980s is presented. We adopt a Knowledge Based framework/architecture for process/service consolidation, which can be adopted on organization-wide basis.

The advances in information and communication technologies and economic factors impelled organizations to engage in new forms of collaboration, such as Collaborative Networks (CNs) [13]. They require adequate frameworks, architectures, tools and platforms to support interoperability among heterogeneous and geographically distributed organizations. As a consequence, businesses have migrated from traditional practices to

e-business. The CNs consist in heterogeneous and geographically distributed organizations with different competences, but symbiotic interests, which join and efficiently combine for a determined period of time their most suitable skills and resources to achieve a common objective. They require adequate technologies and ICT platforms to assure seamless interoperability among their member organizations. We propose a comprehensive Knowledge Base that supports interoperability.

Process interconnection mechanisms are necessary to coordinate geographically distributed business processes in order to strength awareness inside virtual enterprises, to facilitate multinational e-transactions, etc. Existing business process modeling and enactment systems (workflow systems, project management tools, shared agendas, to do lists, etc.) have been mainly developed to suit enterprise internal needs. Thus, most of these systems are not adapted to inter-enterprise cooperation. Karim Bana *et al.* [14] aim, through their paper, to present a model supporting dynamic heterogeneous workflow process interconnection. We propose consolidation of business processes of couple of banks during M&As (Mergers & Acquisitions), and the proposed framework is adapted to inter-enterprise cooperation.

Business Process Management Systems (BPMSs) are software platforms that support the definition, execution, and tracking of business processes. BPMSs have the ability of logging information about the business processes they support. Proper analysis of BPMS execution logs can yield important knowledge and help organizations improve the quality of their business processes and services to their business partners. Daniela Grigori *et al.* [15] present a set of integrated tools that supports business and IT users in managing process execution quality by providing several features, such as analysis, prediction, monitoring, control, and optimization.

The global market is willing to improve their competitiveness through collaborative work and partnerships, motivating the companies to look for enhanced interoperability between computer systems and applications. However, the large number of system's heterogeneity and the company's lack of resources and know-hows have been preventing organizations to move ahead in that direction. Today, the OMG's model-driven architecture (MDA) makes available an open approach to write specifications and develop applications, separating the application and business functionality from the platform technology. As well, the service-oriented architecture (SOA) establishes a software architectural concept that defines the use of services to support the requirements of software users, making them available as independent services accessible in a standardized way. Together, these two architectures seem to provide a suitable framework to improve company's competitiveness through the adoption of a standard-based extended environment, challenging and enhancing the interoperability

between computer systems and applications in industry. The paper by Ricardo Jardim-Goncalves *et al.* [16], after illustrating the general motivations the industrial SMEs have to adopt open architectures to achieve interoperability for extended and collaborative enterprise practices, presents the emerging model-driven and service-oriented architectures. We adopt the service oriented paradigm, and propose the consolidation of business processes for organizational interoperability during M&As (Mergers & Acquisitions).

Existing web services specifications lack an appropriate semantic framework for the creation and operation of emerging manufacturing systems such as the supply chains and extended/virtual enterprises (EE/VE) in order to execute business processes over internet. Semantic web services (SWS) technology aims to add sufficient semantics to the specifications and implementations of web services to make possible the (automatic) integration of distributed autonomous systems, with independently designed data and behaviour models. Defining data, behaviour and system components in a machine understandable way using ontologies provides the basis for reducing the need for humans to be in the loop for routine system integration processes. In their paper [17], Jagdev *et al.* present the design and implementation of a bid auction application for procurement automation within supply chains embedded in extended and virtual enterprises. They show how emerging semantic web services technologies such as the web services execution environment (WSMX) facilitate the creation of such applications, and discuss the advantages of such applications. We propose an approach that supports syntactically and semantically correct WFF (Well Formed Formulae).

Recently, interests in the notion of process-oriented knowledge management (PKM) from academia and industry have been significantly increased. Comprehensive research and development requirements along with a cogent framework, however, have not been proposed for integrating knowledge management (KM) and business process management (BPM), which was proposed independently. Based on a comprehensive framework that reflects lifecycle requirements of both KM and BPM, Jung *et al.* [18] in their paper proposes an architecture for integrating knowledge management systems (KMSs) and business process management systems (BPMSs) to combine the advantages of the two paradigms. The paper first defines the concept of process knowledge and classifies it into three types. Then, it suggests how the functionalities of existing KMSs and BPMSs must be extended to support the three types of process knowledge while satisfying the lifecycle requirements of both knowledge and business processes. We propose a Knowledge Based framework that supports the capture of business process knowledge.

Service-oriented computing is becoming the prominent paradigm for leveraging inter-enterprise information systems to complete higher-order business transactions at the heart of the modus operandi of the virtual enterprise (VE). The paper by Rezgui [19] describes research aimed at supporting the formation and operations of virtual enterprises through the setting-up of service-oriented workspace environments. The paper argues that a role based authorization approach to service invocation is necessary in order to enhance and guarantee the integrity of the transactions that take place in the business environment of a VE. We propose formation of virtual enterprise through virtual consolidation of banks. Our framework takes into account authentication, and can be extended for role based authorization.

The book by Luger [20] captures the essence of artificial intelligence -- solving the complex problems that arise wherever computer technology is applied. Luger demonstrates techniques and strategies for addressing the many challenges facing computer scientists today. Diverse topics on this exciting and ever-evolving field range from perception and adaptation using neural networks and genetic algorithms, intelligent agents with ontologies, automated reasoning, natural language analysis, and stochastic approaches to machine learning. We apply artificial intelligence concepts for developing knowledge bases towards modeling business processes in our work.

Over the last decade there has been a shift from “data-aware” information systems to “process-aware” information systems. To support business processes an enterprise information system needs to be aware of these processes and their organizational context. Business Process Management (BPM) includes methods, techniques, and tools to support the design, enactment, management, and analysis of such operational business processes. BPM can be considered as an extension of classical Workflow Management (WFM) systems and approaches. The tutorial by Wil M.P. van der Aalst [3] introduces models, systems, and standards for the design, analysis, and enactment of workflow processes. Petri nets are used for the modeling and analysis of workflows. Using Petri nets as a formal basis, contemporary systems, languages, and standards for BPM and WFM are discussed. Although it is clear that Petri nets can serve as a solid foundation for BPM/WFM technology, in reality systems, languages, and standards are developed in an ad-hoc fashion. To illustrate this XPD, the “Lingua Franca” proposed by the Workflow Management Coalition (WfMC), is analyzed using a set of 20 basic workflow patterns. This analysis exposes some of the typical semantic problems restricting the application of BPM/WFM technology. Our work also focuses on Business Process Modeling, and Petri Nets is one of the modeling techniques discussed in our paper.

The Workflow Patterns Initiative was established with the aim of delineating the fundamental requirements that arise during business process modeling on a recurring basis and describe them in an imperative way. The first deliverable of this research project was a set of twenty patterns describing the control-flow perspective of workflow systems. Since their release, these patterns have been widely used by practitioners, vendors and academics alike in the selection, design and development of workflow systems. The paper by Russell *et al.* [21] presents a systematic review of the original twenty control-flow patterns and provides a formal description of each of them in the form of a Coloured Petri-Net (CPN) model. It also identifies twenty three new patterns relevant to the control-flow perspective. Detailed context conditions and evaluation criteria are presented for each pattern and their implementation is assessed in fourteen commercial offerings including workflow and case handling systems, business process modeling formalisms and business process execution languages, which is relevant to our paper on Business Process Modeling.

The books by Rambaugh *et al.* [22] and Priestley [23] articulate Object Oriented Modeling and Design concepts including Unified Modeling Language (UML), which is one of the approaches discussed in our work.

Dirk Beyer, Andreas Noack and Claus Lewerentz have discussed querying and manipulating of relations based on predicate calculus in their paper Efficient Relational Calculation for Software analysis [5]. We adopt a predicate calculus knowledge base for consolidating business processes.

The paper by Chanda *et al.* [24] proposes a new paradigm for combined operations of Distributed Architecture and Distributed Data Base Management System (DDBMS), in a banking system, representing two separate banks. In the proposed technology, assuming that all the branches are connected in a network of adequate bandwidth, Service Oriented Architecture allows the operation of business processes and services within well-defined Web Services assigned to branches, without destroying the existing mechanized system operation of the bank / branch. In this system, all the branches of collaborating banks act as technologically autonomous business units, so that the advantages of large scale banking system and small scale banking operations are preserved without destroying each other. The emerging developments of network and bandwidth technologies promise that such architecture will be a viable and cost effective alternative to a large scale, centralized server based architecture.

The paper by HJ Koehler, U Nickel, J Niere, A Zuen-dorf [25] proposes to use SDL block diagrams, UML class diagrams, and UML behavior diagrams like collaboration diagrams, activity diagrams, and statecharts as a visual programming language. They describe a modeling approach for flexible, autonomous production agents,

which are used for the decentralization of production control systems. In order to generate a (Java) implementation of a production control system from its specification, they define a precise semantics for the diagrams and how different (kinds of) diagrams are combined to a complete executable specification. Our paper focuses on modeling business processes rather than implementation.

Nadhan [26] and Channabasavaiah *et al.* [27] discuss various approaches to solving the key challenges organizations face when implementing a service-oriented architecture, and is relevant to our work which focuses on services based business processes.

Ram Mohan [4] and Nitsure [28] provide economic perspectives to bank consolidation, which is also the subject of our work.

Wil van der Aalst *et al.* [29] discuss business-management context within which workflow management systems operate, as well as model and analyze processes; our work also models and analyzes processes.

Simha *et al.* [30] discuss customer loyalty analysis. One of the major conceptual analyses in analytical CRM is customer loyalty. Current methods use user defined and statistical methods for loyalty analysis. One of the major limitations of these methods is that they assign every customer into exactly one segment, which may be counter intuitive. In this paper an approach using fuzzy logic has been discussed. The proposed method uses fuzzy context model to extract the customer membership to a segment of interest. Our paper uses fuzzy linguistic variables which are linguistic expressions for set memberships.

With the traveling records of browsers, one can analyze the preference of pages, further understand the demands of consumers, and promote the advertising and marketing. In their study, Chen *et al.* [31] use maximum forward reference (MFR) algorithm to find the travel pattern of browsers from web logs. They employ fuzzy data mining technique that combines Apriori algorithm with fuzzy weights to determine the association rules. From the yielded association rules, one can be aware of the information consumers need and which Webs they prefer. This is important to governmental institutions and enterprises. Enterprises can find the commercial opportunities and improve the design of Webs by means of this study. Governmental institutions can realize the needs of people from the obtained association rules, make the promotion of policy more efficiently, and provide better service quality. We propose a fuzzy data mining approach that determines association rules in banking databases.

In real world applications, databases are constantly added with a large number of transactions and hence maintaining latest sequential patterns valid on the updated database is crucial. Existing data mining algorithms can incrementally mine the sequential patterns from databases with binary values. Temporal transac-

tions with quantitative values are commonly seen in real world applications. In addition, several methods have been proposed for representing uncertain data in a database. In their paper, Subramanyam *et al.* [32] propose a fuzzy data mining algorithm for incremental mining of sequential patterns from quantitative databases. Proposed algorithm uses the fuzzy grid notion to generate fuzzy sequential patterns validated on the updated database containing the transactions in the original database and in the incremental database. It uses the information about sequential patterns that are already mined from original database and avoids start-from-scratch process. Also, it minimizes the number of candidates to check as well as number of scans to original database by identifying the potential sequences in incremental database.

Han [33] discusses data mining concepts and techniques in his book, which is relevant to our work.

Mining maximal frequent item sets is one of the most fundamental problems in data mining. In this paper Yang [34] studies the complexity-theoretic aspects of maximal frequent item set mining, from the perspective of counting the number of solutions. He presents formal proof that the problem of counting the number of distinct maximal frequent item sets in a database of transactions, given an arbitrary support threshold, is #P-complete, thereby providing strong theoretical evidence that the problem of mining maximal frequent item sets is NP-hard. This result is of particular interest since the associated decision problem of checking the existence of a maximal frequent item set is in P. He also extends the complexity analysis to other similar data mining problems dealing with complex data structures, such as sequences, trees, and graphs. In our work, we also look at mining data from transactional databases.

With the explosive increase of information, data mining techniques are frequently employed to identify trends in the warehouse that may not be readily apparent. In their paper Jin *et al.* [35] apply fuzzy data mining techniques to security system and build a fuzzy data mining based intrusion detection model. Through normalizing the data set and building fuzzy similar matrix of the network connections in the data set, network connections are clustered into different classes. We propose a fuzzy clustering algorithm for data mining of banking databases.

In a competitive environment, providing suitable information and products to meet customer requirements and improve customer satisfaction is one key factor to measure a company's competitiveness. In their paper, Tai *et al.* [36] propose a preference perception system by combining fuzzy set with data mining technology to detect the information preference of each user on a web-based environment. Our paper proposes a data mining approach in the context of competitive scenarios that trigger M&As (Mergers & Acquisitions).

Data mining is a domain difficult to cope with for various reasons. Most of the databases are complex, large, and contain heterogeneous, imprecise, vague, uncertain, incomplete data. Furthermore, the queries may be imprecise or subjective in the case of information retrieval, the mining results must be easily understandable by a user in the case of data mining or knowledge discovery. Fuzzy logic provides an interesting tool for such tasks, mainly because of its capability to represent imperfect information, for instance by means of imprecise categories, measures of resemblance or aggregation methods. Bernadette Bouchon-Meunier [37] focuses his study on the use of similarity measures which are key concepts for many steps of the process, such as clustering, construction of prototypes, utilization of expert or association rules, fuzzy querying, for instance. The paper considers a general framework for measures of comparison, compatible with Tversky's contrast model, providing tools to identify similar or dissimilar descriptions of objects, for instance in a case-based reasoning or a classification approach. The paper presents some real-world problems where these paradigms have been exploited among others to manage various types of data such as image retrieval or risk analysis. Our Paper proposes data mining of banking databases using real-world terms as well as similarity measures.

Data mining techniques can be used to discover useful information by exploring and analyzing data. The aim of the article by Hu [38] is to propose a fuzzy-data mining method to find a compact set consisting of fuzzy if-then classification rules with high classification capability using the genetic algorithm. Furthermore, for not reducing the usefulness of the proposed method for classification problems with high dimensional feature space, the curse dimensionality resulting from the grid partition is overcome in the proposed method by employing the principal component analysis to reduce the dimensions. Through computer simulations, it can be seen that the proposed method is comparable to the other fuzzy classification methods on the well-known iris data, the appendicitis data, and the cancer data. This is relevant to our work which considers banking data that is high dimensional in nature.

The study by Huang *et al.* [39] proposes a knowledge discovery model that integrates the modification of the fuzzy transaction data-mining algorithm (MFTDA) and the Adaptive-Network-Based Fuzzy Inference Systems (ANFIS) for discovering implicit knowledge in the fuzzy database more efficiently and presenting it more concisely. Our paper proposes a fuzzy data mining knowledge discovery model.

Medical data is often very high dimensional. Depending upon the use, some data dimensions might be more relevant than others. In processing medical data, choos-

ing the optimal subset of features is such important, not only to reduce the processing cost but also to improve the usefulness of the model built from the selected data. The paper by Ghazavi *et al.* [40] presents a data mining study of medical data with fuzzy modeling methods that use feature subsets selected by some indices/methods. The paper shows that feature selection is important to mining medical data for reducing processing time and for increasing classification accuracy. However, not all combinations of feature selection and modeling methods are equally effective and the best combination is often data-dependent, as supported by the breast cancer and diabetes data analyzed in this paper. Our paper proposes data mining of banking databases which are also high dimensional.

In his paper, Angryk [41] introduces a new method which enables utilization of uncertain data for precise decision rules learning. The paper focuses investigation on a proximity-based fuzzy relational database as it provides convenient mechanisms for the storage and interpretation of uncertain information. In proximity-based fuzzy databases the lack of certainty about obtained information can be represented via insertion of multiple (*i.e.* non-atomic) attribute values. In addition the database extends classical equivalence relations with fuzzy proximity relations, which provide users with extraordinary analytical capabilities. This paper takes advantage of both of these properties when developing the approach to induction of decision trees from imperfect information. Our paper proposes fuzzy equivalence relations to discover knowledge in fuzzy relational databases.

In their paper, Chaing *et al.* [42] are interested in mining the data with natural ordering according to some attributes, and the time series data is one of this kind of data. The problem of mining the time series data is that the quantity at different time may be very close or even equal to each other. To solve this problem, they propose a fuzzy linguistic summary as one of the data mining functions in our KDD (Knowledge Discovery in Databases) system to discover useful knowledge from the database. Our Paper adopts an approach which uses fuzzy linguistic variables.

Dutta Majumder *et al.* [43] & [44] propose data mining models, using a fuzzy mathematical approach, that aims to discover knowledge in agricultural databases, and is relevant to our work which similarly proposes to discover knowledge in banking databases.

Books by Dutta Majumder *et al.* [45], Klir *et al.* [46], Adriaans *et al.* [47], Han *et al.* [48] and Silberschatz *et al.* [49] provide concepts on fuzzy sets, databases and data mining, which are the subjects of our work.

The Annual Report [50] comprises various banking reports including Balance Sheet and Profit & Loss Account which we refer to in our work.

### 3. Fuzzy Data Mining Framework for Creation of Virtual Enterprise

#### 3.1. Proposed Approach

The proposed Data Mining model, using a fuzzy mathematical approach, adopts fuzzy cluster analysis concept for determining the suitability of merger of two banks/financial institutions.

The paper comprises of the following sections: Mathematical Preliminaries as given in Subsection 3.2, Banking System Reports are analyzed in Subsection 3.3 in a tabular form, Cluster Based Fuzzy Data Mining Model is in Subsection 3.4, Fuzzy/Data Mining Model for Relations Database is in Subsection 3.5, Benefits of the Proposed Method of Discovery of Hidden Knowledge is in Subsection 3.6.

#### 3.2. Mathematical Preliminaries [46]

1) A crisp relation  $R_c(X,X)$  is reflexive iff  $\langle x,x \rangle \in R_c$  for each  $x \in X$ , that is, if every element of  $X$  is related to itself. A fuzzy relation  $R_f(X,X)$  is reflexive iff  $R_f(x,x) = 1$

2) A crisp relation  $R_c(X,X)$  is symmetric iff  $\langle x,y \rangle \in R_c$ , it is also the case that  $\langle y,x \rangle \in R_c$ , where  $x,y \in X$ . Thus, whenever an element  $x$  is related to an element  $y$  through a symmetric relation,  $y$  is also related to  $x$ . A fuzzy relation is symmetric iff  $R_f(x,y) = R_f(y,x)$

3) A crisp relation  $R_c(X,X)$  is transitive iff  $\langle x,z \rangle \in R_c$ , whenever both  $\langle x,y \rangle \in R_c$ ,  $\langle y,z \rangle \in R_c$ , for at least one  $y \in X$ . In other words, the relation of  $x$  to  $y$  and of  $y$  to  $z$  implies the relation of  $x$  to  $z$  in a transitive relation. A fuzzy relation  $R_f(X,X)$  is transitive if  $R_f(x,z) \geq \max_{y \in Y} \min [R_f(x,y), R_f(y,z)]$  is satisfied for each pair  $\langle x,z \rangle \in X^2$ .

4) The transitive closure of a crisp relation  $R_c(X,X)$  is defined as the relation that is transitive, contains  $R_c(X,X)$ , and has the fewest possible members. For fuzzy relations, this last requirement is generalized such that the elements of the transitive closure have the smallest possible membership grades that still allow the first two requirements to be met.

Given a relation  $R(X,X)$ , its transitive closure  $R_T(X,X)$  can be determined by a simple algorithms that consists of the following three steps :

$$R' = R \cup (R \circ R)$$

If  $R' \neq R$ , make  $R = R'$  and go to step 1.

Stop:  $R' = R_T$



5) A fuzzy binary relation that is reflexive, symmetric, and transitive is known as a fuzzy equivalence relation or similarity relation.

6) A fuzzy binary relation that is reflexive and symmetric is called a fuzzy compatibility relation.

7)  $\alpha$ -cut of a fuzzy set A is the crisp set  $\alpha A$  that contains all the elements of the Universal set X whose membership grades in A are greater than or equal to the specified value of  $\alpha$ .

$$\alpha A = \{x \mid A(x) \geq \alpha\}$$

### 3.3. Banking System: Reports

To facilitate understanding of our proposed framework we consider the Banking System as example. The Banking System (like any other organization) is characterized by its reports viz. Balance Sheet and the Profit & Loss Account. As shown above in **Table 1**. These reports manifest the performance of the Bank. The above table shows the main characteristics of Balance Sheet and Profit & Loss Account, as well as the associated attributes.

In addition to the main reports (*i.e.* Balance Sheet and Profit & Loss Accounts) there are various Schedules to the Balance Sheet that provide details on each of the Balance Sheet attributes e.g. Capital, Reserves & Surplus, Deposits, Borrowings, etc.

### 3.4. Cluster based Fuzzy Data Mining Model

Clustering is the process of grouping data into classes or

clusters, so that objects within a cluster have high similarity in comparison to one another, but very dissimilar to objects in other clusters. Dissimilarities are assessed based on the attribute values describing the objects.

Cluster Analysis focuses mainly on distance based cluster analysis, objective being to reduce the intra-cluster distance and increase the inter-cluster distance.

#### 3.4.1. Fuzzy Cluster Analysis in the Context of Banking & Financial Services

The Banking System may be expressed in terms of the relations referred to in Subsection 3.3. The attributes of each of the relations may be expressed by the Fuzzy Linguistic Variables viz. Very High, High, Medium, Low & Very Low, which may, in turn, be denoted by a set of integers such as {5, 4, 3, 2, 1}. Thus, a point in a multi-dimensional space may represent each record. The groups or clusters, thus formed, reflect in a natural way the structure of given data.

Every fuzzy equivalence relation (a relation that is reflexive, symmetric and max-min transitive) induces a crisp partition in each of its  $\alpha$ -cuts. The fuzzy clustering problem can thus be viewed as the problem of identifying an appropriate fuzzy equivalence relation on given data. Although this cannot be done directly, we can readily determine a fuzzy compatibility relation (reflexive and symmetric) in terms of an appropriate distance function applied to given data.

Let us define a fuzzy compatibility relation  $R$ , on our data set ( $n$   $p$ -tuples of  $R^p$ ), in terms of appropriate distance

**Table 1. Banking system reports.**

Report	Relations	Attributes
Balance Sheet	Capital And Liabilities	<ul style="list-style-type: none"> <li>○ Capital</li> <li>○ Reserves and Surplus</li> <li>○ Deposits</li> <li>○ Borrowings</li> <li>○ Other Liabilities and Provisions</li> </ul>
	Assets	<ul style="list-style-type: none"> <li>○ Cash &amp; Balances with the Central Bank</li> <li>○ Balances with Banks &amp; Money at Call and Short Notice</li> <li>○ Investments</li> <li>○ Advances</li> <li>○ Fixed Assets</li> <li>○ Other Assets</li> </ul>
	Income	<ul style="list-style-type: none"> <li>○ Interest earned</li> <li>○ Other income</li> </ul>
	Expenditure	<ul style="list-style-type: none"> <li>○ Interest Expended</li> <li>○ Operating Expenses</li> <li>○ Provisions &amp; Contingencies</li> </ul>
Profit & Loss Account	Net Profit	<ul style="list-style-type: none"> <li>○ Appropriations</li> <li>○ Transfers to                             <ul style="list-style-type: none"> <li>▪ Statutory Reserve</li> <li>▪ Capital Reserve</li> <li>▪ Investment Fluctuation Reserve</li> <li>▪ Revenue Reserves</li> <li>▪ Interim Dividend</li> <li>▪ Proposed Dividend</li> <li>▪ Dividend Tax</li> </ul> </li> </ul>

function of the Minkowski class by the formula given below, for all pairs  $\langle x_i, x_k \rangle \in X$  where  $q \in R^+$  (the set of all non-negative real numbers) and  $\delta$  is a constant that ensures that  $R(x_i, x_k) \in [0,1]$ . Clearly  $\delta$  is the inverse value of the largest distance in  $X$ .

$$R(x_i, x_k) = 1 - \delta \left( \sum_{j=1}^p |x_{ij} - x_{kj}|^q \right)^{1/q}$$

To illustrate the clustering method based on fuzzy equivalence relations, let us use a small data set consisting of the following five points in  $R^2$ , as furnished in **Table 2**.

Where say  $X_{k1}$  represents Investments, and  $X_{k2}$  represents Advances.

As the first step, we perform the analysis for  $q=2$ , which corresponds to the Euclidean distance. First, we need to determine the value of  $\delta$  in the above equation. Since the largest Euclidean distance between any pair of given data points is 4 (between  $x_1$  and  $x_5$ ), we have  $\delta = 1/4 = 0.25$ . Now, we can calculate membership grade of  $R$  by the above equation. For example,

$$R(x_1, x_3) = 1 - 0.25 \{ (3-1)^2 + (4-1)^2 \}^{0.5} = 0.1$$

When determined relation  $R$  may conveniently be represented by the matrix

$$R = \begin{bmatrix} 1 & .65 & .1 & .21 & 0 \\ .65 & 1 & .44 & .5 & .21 \\ .1 & .44 & 1 & .44 & .1 \\ .21 & .5 & .44 & 1 & .65 \\ 0 & .21 & .1 & .65 & 1 \end{bmatrix}$$

The relation is not max-min transitive; its transitive closure is

$$R_T = \begin{bmatrix} 1 & .65 & .44 & .5 & .5 \\ .65 & 1 & .44 & .5 & .5 \\ .44 & .44 & 1 & .44 & .44 \\ .5 & .5 & .44 & 1 & .65 \\ .5 & .5 & .44 & .65 & 1 \end{bmatrix}$$

This relation induces four distinct partitions of its  $\alpha$ -cuts:

$$\alpha \in [0, .44] : \{ \{ x_1, x_2, x_3, x_4, x_5 \} \}$$

**Table 2. Representative data set.**

K	1	2	3	4	5
$X_{k1}$	1	2	3	4	5
$X_{k2}$	1	2	4	2	1

$$\alpha \in (.44, .5] : \{ \{ x_1, x_2, x_4, x_5 \}, \{ x_3 \} \}$$

$$\alpha \in (.5, .65] : \{ \{ x_1, x_2 \}, \{ x_3 \}, \{ x_4, x_5 \} \}$$

$$\alpha \in (.65, 1] : \{ \{ x_1 \}, \{ x_2 \}, \{ x_3 \}, \{ x_4 \}, \{ x_5 \} \}$$

The cluster tree for Euclidean distance is shown below

Thus we have multidimensional classification based on cluster analysis. Multidimensional classifications imply similarity of concerned records in a cluster, or in other words, close association. We may observe from the above that there is close association between points  $(x_1, x_2)$  and between points  $(x_4, x_5)$ .

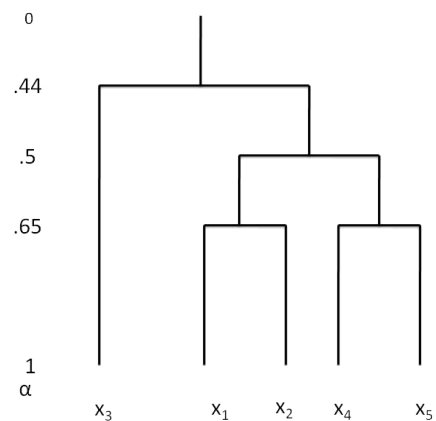
The above exercise considers two attributes. The clustering may be carried out with any number of desired attributes. For example, it may be desired to carry out the analysis with the following four attributes: Investment, Advances, Fixed Assets & Borrowings. Then if two candidate banks belong to different clusters as a result of cluster analysis (based on  $\alpha$ -cut partitions), we may conclude that these banks complement each other, and would therefore, be RIGHT CANDIDATES for CONSOLIDATION / MERGER. Whereas, if two candidate banks belong to the same cluster, they share the same features, and may not be the right candidates for merger.

### 3.5. Fuzzy Data Mining Model Using Relational Databases

Let us consider the following fuzzy relations where Fuzzy Linguistic Variables viz. {Very High, High, Medium, Low and Very Low} are denoted by a set of integers such as {5, 4, 3, 2, 1}

**Table 3** is characterized by the attributes bank\_id & Bank name. This table provides the ids and the names of the banks under consideration. For the purpose of codification, we have considered simple numerical codification for the bank\_ids. The bank\_id uniquely identifies each bank, and is the primary key for this master data table.

The relation Balance Sheet captures the Balance Sheet attributes furnished in **Table 1**. In **Table 4**, the primary



key is constituted by a combination of the bank\_id & the year. This combination identifies each record uniquely. The attribute values furnished in the table above are representative in nature.

It may be noted that the relation geography\_penetration (Table 5) has been defined keeping the geographic penetration of the concerned bank in mind, which will vary from region to region. Thus region-wise penetration values have been shown as geog\_1, geog\_2, geog\_3, etc. This relation, therefore, gives us in insight into the geographic penetration character of the bank. For example, geog\_1 may be Northern Region, geog\_2 may be Western Region & geog\_3 may be Southern Region. A yardstick of geographic penetration may be the number of branches the bank has in that region. In this table a combination of the bank\_id & year uniquely identifies each record. The attribute values furnished in the table above are representative in nature.

We may thus have the following fuzzy compatibility relation defined for domain universal set {1, 2, 3, 4, 5}

Table 6 defines the fuzzy compatibility relation. For example, in case of geographies the fuzzy compatibility relation may be defined on the nearness of geographies to the base geography.

Thus, it would be possible to query the above database with the following question: "Which similar banks have similar penetration in geography 1 for year 2007?"

Table 3. Relation bank\_name.

bank_id	Bank name
1	Bank_1
2	Bank_2
3	
4	

Table 4. Relation balance sheet.

bank_id	year	attribute_1	attribute_2	attribute_3	..
1	2007	5	2	3	
2	2007	4	1	5	
3	2007	4	4	4	
4	2007	2	3	5	

Table 5. Relation geography\_penetration.

bank_id	year	geog_1	geog_2	geog_3
1	2007	5	1	4
2	2007	2	3	2
3	2007	1	5	2
4	2007	4	2	1

The answer is obtained by accomplishing a relational join on the relations "bank\_name", "balance\_sheet" and "geography\_penetration" on the attribute bank\_id, selecting the tuples where geog\_id=1 and year=2007, then projecting the bank\_ids and balance\_sheet attributes as sets, where threshold attribute\_1 (e.g. assets) ≥ .6 and threshold penetration for geog\_1 ≥ .8.

On performing the relational join, we have the following relation (relation\_1), as illustrated in Table 7.

Now, we observe that attribute\_1 (assets) for bank 1 and 2 are similar in the sense that both are > the threshold value of 0.6 (since the compatibility of 5 with 4 from the fuzzy compatibility relation is .8; compatibility of 5 with 2 is .4) However, only for bank 1 the threshold geography penetration for geog\_1 ≥ .8 (since the compatibility of 5 with 5 from the fuzzy compatibility relation is 1; compatibility of 5 with 2 is .4). Thus, for the criteria defined by us, only bank\_id 1 qualifies.

The result is shown below in Table 8.

It may, therefore, be immediately observed that fuzzy relational databases provide us with insights (knowledge), which would otherwise not be possible with crisp relational databases, where we would not be able to define similar banks, similar attributes (assets) or threshold values. Moreover, search time is faster, as we are dealing with a few integers only when considering the attributes (e.g., Balance Sheet attributes like capital, reserves & surplus, deposits, etc.) Moreover, the above methodology also provides us with mechanisms of arriving at strong or

Table 6. Fuzzy compatibility relation.

	5	4	3	2	1
5	1	.8	.6	.4	.2
4	.8	1	.8	.6	.4
3	.6	.8	1	.8	.6
2	.4	.6	.8	1	.8
1	.2	.4	.6	.8	1

Table 7. Fuzzy relational join.

bank_id	attribute_1	geog_1
1	5	5
2	4	2
3	4	2
4	2	4

Table 8. Fuzzy relational join.

bank_id	asset
1	5

weak association rules, based on the threshold values. Thus, it may be possible to ascertain which banks would be most suitable CANDIDATES for the CONSOLIDATION/merger process, based on the defined criteria.

As observed, the above methodology leads to knowledge discovery, and is, therefore, an effective data mining process that may be applied on relational databases.

Moreover, multi-dimensional data mining on relational databases using the above approach is also possible by sequential relational joins with relations containing the relevant dimensions (attributes). For example, we may accomplish a relational join between the relation\_1 above and the relation schedule\_1 (a schedule to the Balance Sheet), and may obtain further insights (knowledge), into the database under consideration.

### 3.6. Benefits of the Proposed Method of Discovery of Hidden Knowledge

#### 1) *Extraction of realistic hidden patterns:*

The proposed method is closer to real world expressions (since linguistic variables are used) as compared to existing classical methods, e.g. it would be more realistic to convey that there is strong association between (high Investments, high Fixed Assets and low Borrowings) instead of say (Investment Rupees 400 Crores, Fixed Assets Rupees 700 Crores and Borrowings Rupees 300 Crores).

#### 2) *Outlier analysis:*

There may exist data objects, which do not comply with the general behaviour or model of the data. Such data objects, which are grossly different from or inconsistent with the remaining set of data, are called outliers (or exceptions), and may be of particular interest.

#### 3) *Fuzzy relational databases:*

Fuzzy relational databases provide us with insights (knowledge), which would otherwise not be possible with crisp relational databases.

#### 4) *More Generalized Method:*

Fuzzy Mathematics is more generalized than Crisp Mathematics. All methodologies developed with crisp approach can always be addressed with fuzzy approach.

## 4. AI Based Consolidation of UML Diagrams

The global economic scenario (past, current & present) has been witness to continuing M&As (Mergers & Acquisitions), especially in the Banking & Financial Institutions sphere. The ongoing economic downturn has further led to M&As being an ever increasing strategic phenomenon for Banks & Financial Institutions. In order to avoid legal, financial and other complex problems related to M&As of two banks/financial institutions, we have presented a proposal of virtual consolidation for all practical purposes. For the purpose of Virtual Consolida-

tion, there would be a Memorandum of Agreement between 2 banks for implementation of services to the public, so that each bank can use the other's resources at the required time and place, as if the two banks are one bank.

From the Systems point of view, Virtual Consolidation of two banks during mergers and acquisitions requires the composition/consolidation of the business processes of the two banks. Virtual consolidation implies that the collaborating banks act as technologically autonomous business units, so that the advantages of large scale banking system and small scale banking operations are preserved without disturbing each other [24].

The orchestration/composition of the business processes is modeled by means of our proposed framework based on an AI [20] approach, which leads to consolidation of individual processes modeled by UML Diagrams.

### 4.1. Proposed Approach

Our Paper proposes an AI based approach for realization of services and business processes (composition of services realized through their orchestration).

In our paper we discuss relevant schemes for modeling business process viz. UML & Predicate Calculus (AI). We carry out a comparative study reflecting how our proposed Knowledge Based approach is an improvement on the UML as well as Petri Net based modeling and design approaches for business process composition.

There remain open challenges when it comes to engineering services that engage in complex interactions with multiple other services. A number of approaches have been proposed to address these challenges. One such approach, known as (process oriented) service composition, has its roots in workflow and business process management. The idea of service composition is to capture the business logic and behavioral interfaces of services in terms of process models.

In the area of workflow, it has been shown that Petri nets, is a valuable technique for the simulation of business processes. The purpose of Workflow Management Systems is to execute Workflow Processes. Workflow Processes represent the sequences of activities which have to be executed within an organization to treat specific cases and to reach a well defined goal. Of all notations used in the Software Industry, UML is one of the best accepted. In particular, the activity diagrams of the UML notation seem to be very suitable for proposing approaches to represent Workflow Processes as these diagrams represent basic routings encountered in Workflow Processes which are the sequential routing (the sequential execution of activities), the parallel routing (two or more activities executed simultaneously) and the selective routing (when a choice must be made between two or more activities). However, UML notations have their limitations when they are used for specifying the real time characteristics of Workflow Management Systems.

Our analysis of our practical experience on modeling business processes (within the SOA paradigm) with UML shows that UML is useful for static relationships, but falls short for dynamic properties, even when using sequence diagrams, and activity diagrams, because of the absence of concept of tokens and places that characterize Petri Net notation. Also, UML notation is still semiformal in nature, and lacks well defined semantics. The Class Diagrams, Sequence Diagrams as well as Activity Diagrams are three distinctly separate representations in UML without check for integrity. On the other hand, Petri Nets, though being formalized, are not useful for modeling of static relations, since they are not designed for that purpose. To overcome these challenges we propose a modeling framework that would provide a single and comprehensive knowledge repository for static relationships as well as dynamic properties.

For our purpose, the representative domain of discourse considered is the deposit function of two banks viz. Bank1 and Bank2. The realization of the business process ‘display balance’ for the consolidated bank (following the merger of the two representative banks) requires the orchestration/composition of the two business processes for each of the two banks into one consolidated business process. The orchestration/composition of the business process ‘display balance’ is achieved by means of our proposed AI approach based on Predicate Calculus and search techniques based on pattern search [6].

**4.2. UML Based Modeling of Business Processes**

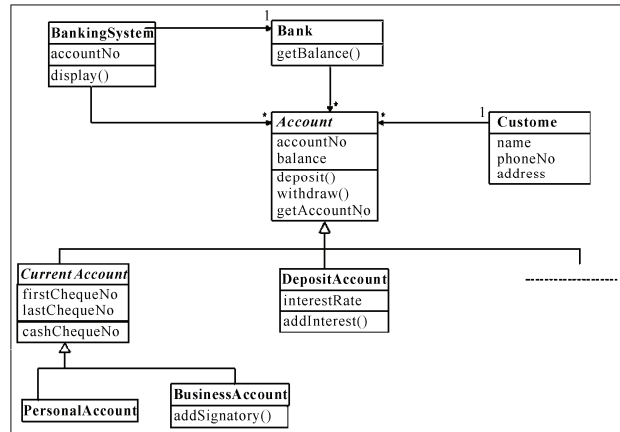
For our purpose, the representative domain of discourse considered is the deposit function of two banks viz. Bank1 and Bank2. For each of these banks, we consider the knowledge represented by the static structure (class) diagrams and interaction diagrams (sequence diagrams) in an object oriented approach. The sequence diagrams show the realization of the services ‘display balance’ for the two banks.

For knowledge representation, we have considered the UML diagrams furnished in **Figures 1, 2, 3** and **4**. These diagrams represent some patterns of invariant features. The realization of the process ‘display balance’ for Bank1 is shown in the Sequence Diagram **Figure 3**, and that for Bank2 is shown in **Figure 4**. It may be noted that the process ‘display balance’ is a Service.

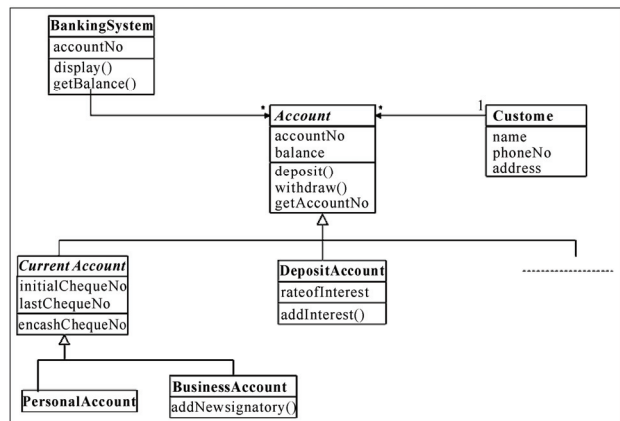
For Bank1, the BankingSystem object plays the role of controller (controlling the sequence of various messages). The Bank object is a feature of the Banking System and is responsible for maintaining information about the Account entity objects. However, there is a fundamental difference in design for Bank2. For Bank2, we consider that the Banking System object plays the role of controller and is also responsible for maintaining information about the Account entity objects. For both the above, an instance of the customer actor has been considered. The

sequence diagrams show how messages are passed between objects in an interaction.

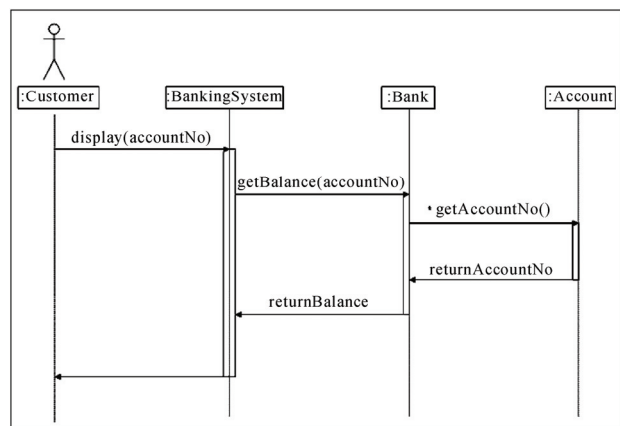
The messages in the sequence diagram reappear in the class diagrams (**Figures 1 & 2**) as operations belonging to various classes. **Figure 1** shows the banking system and bank classes, along with the customer and account



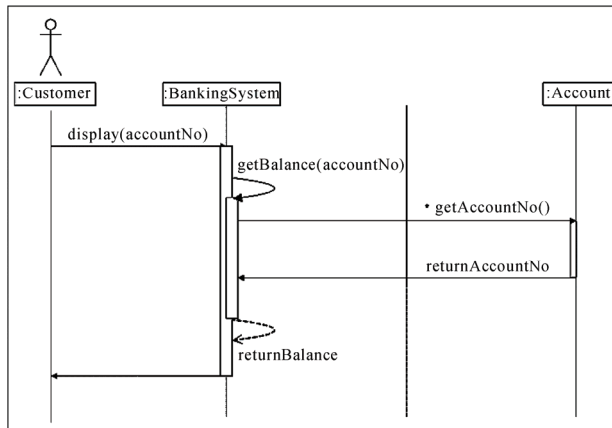
**Figure 1. Analysis class model for deposit function of Bank1.**



**Figure 2. Analysis class model for deposit function of Bank2.**



**Figure 3. Sequence diagram for displaybalance: Realization of the basic course of events for Bank1.**



**Figure 4. Sequence diagram for displaybalance: Realization of the basic course of events for Bank2.**

classes complete with attributes and operations. This is in keeping with the key business requirement, that in a banking system, customers hold accounts. A customer may hold many accounts. **Figure 2** shows the banking system, customer and account classes.

A perusal of **Figures 1** and **2** would also reveal differences in the attribute and operation terminologies.

### 4.3. Knowledge Representation Using Predicate Calculus

For the domains of discourse (Deposit Functions of Bank1 & Bank2), expressing the knowledge represented in **Figures 1, 2, 3** and **4**, we arrive at the following representative set of Predicate Calculus expressions:

- 1) class (bankingsystem)
- 2) bankingsystem (banksystem1)
- 3) banksystem1 (private (int (AccountNo)), display ( ))
- 4) class (bank)
- 5) bank (int(getBalance(AccountNo)))
- 6) multiplicity (bankingsystem , bank, one)
- 7) abstractclass (account)
- 8) account (protected (int (AccountNo)), protected (int (Balance)), deposit ( ), withdraw(), int (getAccountNo()))
- 9) abstractclass (currentaccount)
- 10) subclass(account,currentaccount)
- 11) depositaccount (protected (int(InterestRate)), addInterest ( ))
- 12) currentaccount (protected (int (FirstChequeNo, protected (int (LastChequeNo), cashCheque()))
- 13) subclass (account,depositaccount)
- 14) subclass (currentaccount,personalaccount)
- 15) subclass (currentaccount,businessaccount)
- 16) businessaccount (addSignatory())
- 17) has (customer,account)
- 18) multiplicity (bankingsystem,account, many)
- 19) multiplicity (customer, account, many)
- 20) navigation (banksystem1, bank, unidirectional)

- 21) navigation (bankingsystem, account, unidirectional)
- 22) class (customer)
- 23) customer (private (string (Name)), private (int (PhoneNo)), private (string (Address)))
- 24) navigation (account, customer, bi-directional)
- 25) bank (bank1)
- 26) bankingsystem (banksystem1)
- 27) bankingsystem (banksystem2)
- 28) customer (p\_kumar)
- 29) p\_kumar(prakash\_kumar,12345678,po\_box\_19\_AB\_Avenue\_Kolkata)
- 30) has (p\_kumar, da\_1001)
- 31) depositaccount (da\_1001)
- 32) da\_1001 (12345, 2500)
- 33) banksystem2 (private (int(AccountNo)), display(), getBalance())
- 34) equivalent (InterestRate,RateofInterest)
- 35) equivalent (addSignatory(),addNewsignatory())
- 36) equivalent (InitialChequeNo,FirstChequeNo)
- 37) bank (bank2)
- 38) equivalent (encashCheque(),cashCheque)
- 39) customer(p\_dev)
- 40) depositaccount (dep\_1005)
- 41) has (p\_dev, dep\_1005)
- 42) dep\_1005 (2468,5000)
- 43) p\_dev(pratik\_dev,98765432,street\_10\_number\_18\_C D\_Avenue\_Kolkata)

All static structure features of UML including inheritance, composition, aggregation, etc. can be properly represented in Predicate Calculus terminology. In Predicate Calculus, predicates, which represent relationships between zero or more objects, are expressed in lower case, variables start with upper case, constants in lower case, functions start with lower case. *The above Knowledge Base itself can be distributed or centralized in nature depending on the requirement scenario.*

#### 4.3.1. Process/Service Representation Using Predicate Calculus

The process / service 'display balance' for Bank1 is a goal in AI. The goal for Bank1 (**Figure 3**) is the expression: displaybalance (bank1). This is represented by the implication:

$$\{\text{customer (X)} \wedge \text{message (customer (X), bankingsystem (P), display (AccountNo))} \wedge \text{message (bankingsystem (P), bank (R), getBalance(AccountNo))} \wedge \text{message (bank (R), account(N), getAccountNo ( ))} \rightarrow \text{displaybalance (bank1)}$$

The process/service (goal) for Bank2 (**Figure 4**) is the expression: displaybalance (bank2). This is represented by the implication:

$$\{\text{customer (Y)} \wedge \text{message (customer (Y), bankingsystem (Q), display (AccountNo))} \wedge \text{message (bankingsystem (Q), getBalance(AccountNo))} \wedge \text{message (bankingsystem (Q), account (Z), getAccountNo ( ))} \rightarrow \text{displaybalance (bank2)}$$

playbalance (bank2)

The above processes/services (goals) are WFF (Well Formed Formulae) in AI, having proper syntax & semantics. Also, the inferences (services/goals) feed into the Knowledge Base & become a part of it (Inferences are goals or sub goals).

**4.3.2. Proposed Model: Solution Graphs for Realization & Implementation of the Processes/Services**

The graphical representation uses predicate calculus expressions in list syntax.

1) Solution Sub Graph for Bank 1

The Solution Sub graph for Bank 1 has been furnished in **Figure 5**.

Here the set of unifications that satisfies each sub goal

is returned by the pattern\_search algorithm [6]. The following unifications/substitutions take place, using the unification algorithm [6]. *Security through authentication is inherent in the unification/substitution process:*

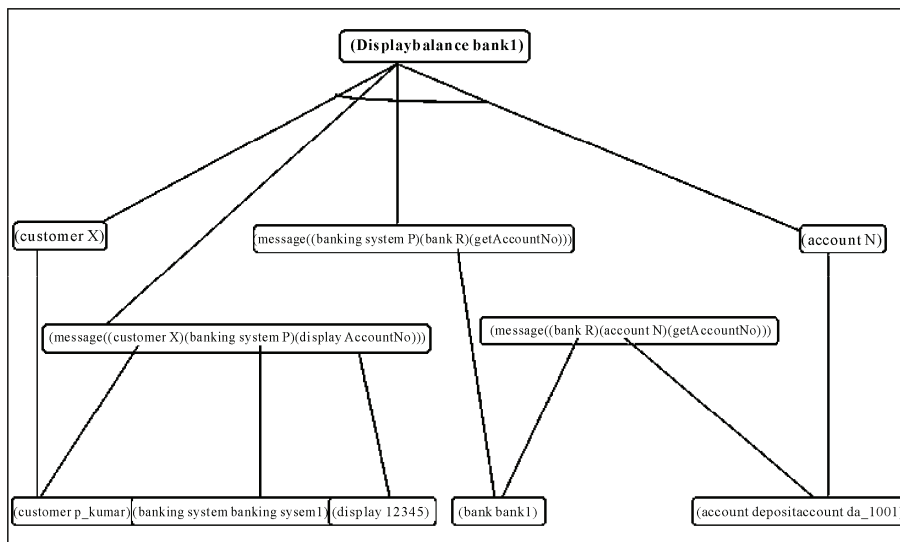
{p\_kumar/X}, {banksystem1/P}, {bank1/R}, {da\_1001/N}, {12345/AccountNo}

2) Solution Sub Graph for Bank2

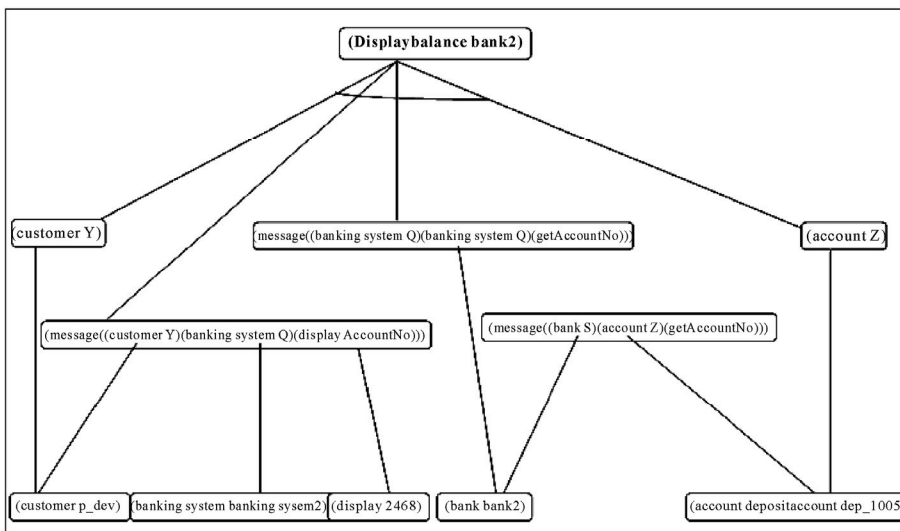
The Solution Sub graph for Bank2 has been furnished in **Figure 6**.

The set of unifications that satisfies each sub goal is returned by the pattern\_search algorithm [6]. The following unifications/substitutions take place using the unification algorithm [6]. *Security through authentication is inherent in the unification/substitution process:*

{p\_dev/Y}, {banksystem2/Q}, {bank2/S}, {dep\_1005/Z}, {2468/AccountNo}.



**Figure 5. Solution sub graph for Bank1.**



**Figure 6. Solution sub graph for Bank2.**

### 4.3.3. Consolidated Scenario: Process Composition

Let us look at the scenario when there is consolidation of the services, e.g. let us consider that the Consolidated Process/Service (Goal) is the expression:  $\text{displaybalance}(\text{bank\_newbank})$ . Let us assume that the services are realized based on the input provided by the customer. The goal for the consolidated scenario is represented by:  $\text{displaybalance}(\text{bank1}) \vee \text{displaybalance}(\text{bank2}) \rightarrow \text{displaybalance}(\text{bank\_newbank})$ . From the point of view of dynamic behavior of systems, this is synonymous to OR-Split/Or-Join. The goal gives a valid result if the customer of the new bank is either a customer of Bank 1 OR if he is a customer of bank 2, but he is simultaneously not a customer of both the banks.

Now if the customer is having an account in both the banks (Bank1 & Bank2), the above goal representation to evaluate to a business fault condition. However, if we have the following goal representation for the 'displaybalance' service, the fault condition would not occur. From the point of view of dynamic behavior of systems, this is synonymous to AND-Split/AND-Join.

$(\text{displaybalance}(\text{bank1}) \wedge \text{displaybalance}(\text{bank2})) \rightarrow \text{displaybalance}(\text{bank\_newbank})$

The new service  $\text{displaybalance}(\text{bank\_newbank})$  is an orchestration of the two services  $\text{displaybalance}(\text{bank1})$  &  $\text{displaybalance}(\text{bank2})$ . Atomic services as well as composite services (which have been arrived at by orchestrating atomic/composite services) can be well realized by means of Goals & Sub goals.

The above services (goals) are WFF (Well Formed Formulae) in AI, having proper syntax & semantics, thus assuring quality of the composed services. Also, the inferences (services/goals) feed into the Knowledge Base (KB) & become a part of it.

*Instead of using the  $\vee$  &  $\wedge$  grammar, we could have used predicate calculus terms for Fork, Split & Join, by representing them in the predicate calculus knowledge base. Predicates can be used for denoting various types of services in the KB, e.g. synchronous, asynchronous, manual, etc. Business Rules can also be expressed as predicates.*

## 5. Improvements Achieved through Proposed Knowledge Based Modeling Framework with Respect to Business Process Composition

We discuss features that are realizable through our proposed knowledge based architecture, but not achievable through UML / Petri Net based modeling.

### 5.1. Full Fledged Modeling

The proposed knowledge based modeling framework

provides a single and comprehensive knowledge repository of static relationships as well as dynamic behavior of a system, especially with respect to business process composition. In this framework the consistency of representation is automatically checked. Dynamic behavior modeling is visual in nature, by means of solution graphs for realization of the individual & composite processes / services.

### 5.2. Knowledge Based Inference & Implementation Framework

The proposed architecture provides a framework for implementation of the individual & composite processes / services as goals. These goals are WFF (Well Formed Formulae) in AI, having proper syntax & semantics, thus assuring quality of the composed services. Also, the inferences (services/goals) feed into the Knowledge Base (KB) & become a part of it.

### 5.3. Process/Service Registry Discovery

The Knowledge Base (KB) serves the purpose of Process/Service Registry since service information would be maintained in the KB itself on the form of predicates. Service Discovery is through pattern search & substitution/unification. Service binding is implemented through substitution/unification.

### 5.4. Service Routing & Data Transformation

Messages can be defined in the Knowledge Base (KB) as predicates. (However, message and transport protocols are out of scope of our discussion). The KB also contains the Routing & Transformation logic in the form of predicates.

When the routing/transformation logic is a constituent of the KB, it becomes inherently a constituent of the inference process for realization of goals/sub goals.

Moreover, various Triggers (human intervention, temporal activations, messages), can also be defined and stored as predicates in the KB.

### 5.5. Flexibility

Business Rules, Routing Logic, Transformation Logic, etc., are part of the KB.

Instead of using the  $\vee$  &  $\wedge$  grammar, we could have used predicate calculus terms for Fork, Split & Join, by representing them in the predicate calculus knowledge base. Predicates can be used for denoting various types of services in the KB, e.g. synchronous, asynchronous, manual, etc. Business Rules can also be expressed as predicates.



## 5.6. Pattern Based KB

Integration Patterns, Business Patterns, Workflow Patterns, other relevant patterns can be modeled / defined as predicate calculus expressions and maintained in the KB. The patterns can be searched and retrieved by means of Pattern Search & Unification / Substitution. This is a key strength of our proposed Knowledge based Architecture.

## 5.7. Business Intelligence

The Knowledge based Architecture provides Business intelligence by way of discovery of hidden knowledge through inferences.

## 5.8. Security

Security through authentication is inherent in the unification/substitution process. This has been discussed in Subsection 4.3.2.

## 5.9. Quality

The proposed Architecture uses syntactically & semantically correct Well Formed Formulae (WFF) for realizing & orchestrating services thus assuring quality of the orchestrated/composed processes.

## 5.10. Extensibility & Performance

The Predicate Calculus based Knowledge Base can be appropriately extended as per business requirement. For example, if another Bank is merged with our New Bank, the Knowledge Base can be appropriately augmented. It would not entail any change to the Pattern Search & Unification algorithms. Here we can observe again that the proposed Architecture is highly extensible as well as flexible.

## 5.11. Reusability

Our proposed Architecture scores high on the reusability front as the realized services (Goals & Sub-goals) become part of the Predicate Calculus Knowledge Base; these Goals & Sub-goals can be reused from the Knowledge Base. Also, the Inferences derived from the Production Rule based Expert System can likewise be incorporated to supplement the Knowledge Base, and reused.

## 5.12. Governance & Policy Management

The Governance of the Knowledge Base is achieved by means of Policies in the KB which can be revised as per the changing business requirements.

## 6. Discussion: A Comparative Study

Figure of Merit / relative Complexity Factor:

- Static structure Knowledge Representation with
  - Knowledge base: least Complexity
  - UML: medium complexity
  - Petri Nets: high complexity
- Dynamic behavior representation with
  - Petri Nets: least complexity & most user friendly
  - Knowledge base: medium complexity
  - UML: high degree of relative complexity

### 6.1. Benefits of Using Knowledge Base for Service Representation & Service/Process Composition

The following benefits are realized when a Knowledge based modeling approach is adopted vis-à-vis other approaches

- Comprehensive representation of static structure relationships as well as dynamic behavior in a system
- Knowledge base for inferring new knowledge
- Inferences become a part of Knowledge Base (Inferences can be goals & sub goals)
- Efficient search features
- Syntactically & semantically correct Well Formed Formulae (WFF) for realizing & orchestrating services thus assuring quality of the orchestrated/composed processes
- Implements Security through Authentication
- Capability for representing atomic services, as well as composite processes & services
- Flexibility to configure business rules
- Extensible Architectural framework
- Provides a high level implementation framework for realization of processes/services

### 6.2. Shortcomings of the Proposed Approach

- Deeper investigation of security in relation to the banking system operations is desired.
- Comparative study of UML modeling and BPMN (Business Process Modeling Notification) standard would be addressed in future.

## 7. Conclusions

In our Paper we have discussed how consolidation of UML Diagrams is possible by using a Knowledge Based approach. The modeling of individual Processes/Services, as well as composition of Processes into consolidated processes has been shown in the Paper. Our proposed modeling approach, which provides comprehensive and full fledged modeling of both static and dynamic behavior

**Table 9. Comparison of modeling approaches.**

UML	Petri Nets	Knowledge Base
Comprehensive & user friendly representation of static structure relationships in a system	Complex representation of static structure relationships	Represents static structure relationship of a system comprehensively by means of predicate calculus knowledge base
Unable to visually represent dynamic behavior of a system	Comprehensively represents dynamic behavior of a system; Established patterns & extended patterns for Business Process/Workflow modeling	Dynamic behavior of a system visually possible through goal representation; can represent all the Petri Net Patterns through Predicate Calculus Knowledge Base & Unification/Substitution (Basic Patterns can be represented through AND/OR Rules in Knowledge Base & Further Patterns can be represented as Inferences)
Semiformal notation	Based on Semantic Network architecture	Syntactically & Semantically complete Predicate Calculus based Architecture
Does not have search provisions	Does not have search provisions	Efficient search features

of a system, improves on prevalent modeling techniques based on UML (Unified Modeling Language) & Petri Nets.

This approach would not only be valid for the Banking/Financial Services domain, but for other domains as well.

The Fuzzy Mathematical Model discussed in this paper leads to Management Intelligence. Discovery of hidden knowledge, patterns and new association rules in real world terms enable the Manager to get the best out of investments, implement effective control of operational costs and manage banking resources optimally, as well as gain insights into the feasibility of merger of two banks. Moreover, this Framework may be applied for M&As in other domains, as well.

## 8. References

- [1] T. Y. Kim, S. Lee, K. Kim, and C. H. Kim, "A modeling framework for agile and interoperable virtual enterprises," *Computers in Industry*, Vol. 57, pp. 204–217, 2006.
- [2] F. F. Chua and C. S. Lee, "Collaborative learning using service-oriented architecture: A framework design," *Knowledge-Based Systems*, Vol. 22, pp. 271–274, 2009.
- [3] W. M. P. van der Aalst, "Business process management demystified: A tutorial on models," *Systems and Standards for Workflow Management*.
- [4] T. T. Ram Mohan, "Bank consolidation: Issues and evidence," *Economic and Political Weekly*, pp. 1151–1161, 2005.
- [5] D. Beyer, A. Noack, and C. Lewerentz, "Efficient relational calculation for software analysis," *IEEE Transactions on Software Engineering*, Vol. 31, No. 2, pp. 137–149, February 2005.
- [6] G. Candido, J. Barata, A. Walter Colombo, and F. Jammes, "SOA in reconfigurable supply chains: A research roadmap," *Engineering Applications of Artificial Intelligence*, Vol. 22, pp. 939–949, 2009.
- [7] M. Crasso, A. Zunino, and M. Campo, "Easy web service discovery: A query-by-example approach," *Science of Computer Programming*, Vol. 71, pp. 144–164, 2008.
- [8] O. El-Gayar and K. Tandekar, "An XML-based schema definition for model sharing and reuse in a distributed environment," *Decision Support Systems*, Vol. 43, pp. 791–808, 2007.
- [9] M. L. Lopez-Sanz, C. J. Acuna, C. E. Cuesta, and E. Marcos, "Modelling of service-oriented architectures with UML," *Electronic Notes in Theoretical Computer Science*, Vol. 194, pp. 23–37, 2008.
- [10] C. Pahl, "Semantic model-driven architecting of service-based software systems," *Information and Software Technology*, Vol. 49, pp. 838–850, 2007.
- [11] S. Arroyo, M. Sicilia, and J. M. Doderio, "Choreography frameworks for business integration: Addressing heterogeneous semantics," *Computers in Industry*, Vol. 58, pp. 487–503, 2007.
- [12] D. Chen, G. Doumeingts, and F. Vernadat, "Architectures for enterprise integration and interoperability: Past, present and future," *Computers in Industry*, Vol. 59, pp. 647–659, 2008.
- [13] C. M. Chituc, A. Azevedo, and C. Toscano, "A framework proposal for seamless interoperability in a collaborative networked environment," *Computers in Industry*, Vol. 60, pp. 317–338, 2009.
- [14] K. Baina, K. Benali, and C. Godart, "DISCOBOLE: A service architecture for interconnecting workflow processes," *Computers in Industry*, Vol. 57, pp. 768–777, 2006.
- [15] D. Grigori, F. Casati, M. Castellanos, U. Dayal, M. Sayal, and M. C. Shan, "Business process intelligence," *Computers in Industry*, Vol. 53, pp. 321–343, 2004.
- [16] R. Jardim-Goncalves, A. Grilo, and A. Steiger-Garcia, "Challenging the interoperability between computers in industry with MDA and SOA," *Computers in Industry*, Vol. 57, pp. 679–689, 2006.
- [17] H. Jagdev, L. Vasiliu, J. Browne, and M. Zaremba, "A semantic web service environment for B2B and B2C auction applications within extended and virtual enterprises," *Computers in Industry*, Vol. 59, pp. 786–797, 2008.
- [18] J. Jung, I. Choi, and M. Song, "An integration architect-

- ture for knowledge management systems and business process management systems,” *Computers in Industry*, Vol. 58, pp. 21–34, 2007.
- [19] Y. Rezgui, “Role-based service-oriented implementation of a virtual enterprise: A case study in the construction sector,” *Computers in Industry*, Vol. 58, pp. 74–86, 2007.
- [20] G. F. Luger, “AI structures and strategies for complex problem solving,” Pearson Education, Fourth Edition, 2006.
- [21] N. Russell, A. H. M. ter Hofstede, W. M. P. van der Aalst, and N. Mulyar, “Workflow control-flow patterns: A revised view.
- [22] J. Rambaugh, M. Blaha, W. Premerlani, F. Eddy, and William Lorensen, “Object oriented modeling and design,” Pearson Education.
- [23] M. Priestley, “Practical object oriented design with UML,” Tata McGraw-Hill Publishing Company Ltd, 2nd ed., 2005.
- [24] D. Chanda, D. D. Majumder, and S. Bhattacharya, “Virtual consolidation: A new paradigm of service oriented distributed architecture for indian banking system,” *Proceedings of International Conference on Emerging Applications of Information Technology*, Elsevier, Kolkata, pp. 57–62, 2006.
- [25] H. J. Koehler, U. Nickel, J. Niere, and A. Zuendorf, “Integrating UML diagrams for production control systems,” *IEEE Computer Society, 22<sup>nd</sup> Annual Conference on Software Engineering*, 2000.
- [26] E. G. Nadhan, “Service-oriented architecture: Implementation challenges,” White Paper in [www.microsoft.com](http://www.microsoft.com), 2004.
- [27] K. Channabasavaiah and K. Holley, “IBM global services,” E. M. Tuggle, IBM Software Group, “Migrating to a service - Oriented architecture,” White Paper in [www.ibm.com](http://www.ibm.com), 2004.
- [28] R. R. Nitsure, “Basel II norms: Emerging market perspective with indian focus,” *Economic and Political Weekly*, pp. 1162–1166, 2005.
- [29] W. van der Aalst and K. van Hee “Workflow management: Models, methods, and systems,” MIT Press, 2002.
- [30] J. B. Simha and S. S. Iyengar, “Fuzzy data mining for customer loyalty analysis,” *9th International Conference on Information Technology*, Vol. 6, No. 18–21 pp. 245–246, December 2006.
- [31] Q. Z. Chen, J. H. Han, W. X. He, K. J. Mao, Y. G. Lai “Utilize fuzzy data mining to find the travel pattern of browsers,” *The Fifth International Conference on Computer and Information Technology*, No. 21–23, pp. 228–232, September 2005.
- [32] R. B. V. Subramanyam and A. Goswami “A fuzzy data mining algorithm for incremental mining of quantitative sequential patterns,” *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 13, No. 6, pp. 633–652, December 2005.
- [33] J. W. Han, “Data mining: Concepts and techniques,” Morgan Kaufmann Publishers Inc., San Francisco, CA, 2005.
- [34] G. Z. Yang, “The complexity of mining maximal frequent itemsets and maximal frequent patterns,” *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Seattle, August 22–25, 2004.
- [35] J. Hai, J. H. Sun, H. Chen, and Z. F. Han “A fuzzy data mining based intrusion detection model,” *10th IEEE International Workshop on Future Trends of Distributed Computing Systems (FTDCS’04)*, pp. 191–197, 2004
- [36] W. S. Tai, C. T. Chen “A web user preference perception system based on fuzzy data mining method,” *Information Retrieval Technology, Lecture Notes in Computer Science*, Vol. 4182, 2006.
- [37] B. Bouchon-Meunier “Similarity management for fuzzy data mining,” *2007 International Conference on Intelligent Systems and Knowledge Engineering, ISKE*, 2007.
- [38] Y. C. Hu, “A new fuzzy-data mining method for pattern classification by principal component analysis,” *Cybernetics and Systems*, Vol. 36, No. 5, pp. 527–547, July–August 2005.
- [39] M. J. Huang, Y. L. Tsoua, and S. C. Lee “Integrating fuzzy data mining and fuzzy artificial neural networks for discovering implicit knowledge,” Elsevier, 2006.
- [40] S. N. Ghazavi and T. W. Liao “Medical data mining by fuzzy modeling with selected features,” *Artificial Intelligence in Medicine*, Vol. 43, No. 3, pp. 195–206, July 2008.
- [41] R. A. Angryk, “Similarity-driven defuzzification of fuzzy tuples for entropy-based data classification purposes,” *2006 IEEE International Conference on Fuzzy Systems*, pp. 414–422, 2006.
- [42] D. A. Chiang, L. R. Chow, and Y. F. Wang “Mining time series data by a fuzzy linguistic summary system,” *Fuzzy Sets and Systems*, Vol. 112, No. 3, pp. 419–432, June 2000.
- [43] D. D. Majumder and D. Chanda “Datamining & knowledge discovery using a fuzzy mathematical approach for the indian agricultural system management,” *Fuzzy Logic and its Application to Technology and Management*, Narosa Publishing House, pp. 73–80, June 2006.
- [44] D. D. Majumder and D. Chanda, “Study on a framework for agricultural forecasting systems: An application of information technology & datamining techniques in the Indian scenario,” presented in an International Conference on “Recent trends & new directions of research in cybernetics & systems theory” at IASST, Guwahati, India, January 2004.
- [45] D. D. Majumder and S. K. Pal “Fuzzy mathematical approach to pattern recognition,” John Wiley & Sons, N.Y., 1986.
- [46] G. J. Klir and B. Yuan “Fuzzy sets and fuzzy logic theory and applications,” Prentice-Hall of India Private Limited, New Delhi, 2002.
- [47] P. Adriaans and D. Zantinge, “Datamining.”
- [48] J. W. Han and M. Kamber “Datamining concepts and techniques,” Morgan Kaufmann Publishers, San Francisco, 2001.
- [49] A. Silberschatz, H. K. Forth, and S. Sudarshan “Database system concepts,” McGraw Hill, International Edition, 2002.
- [50] Canara Bank Annual Report, 2007–2008.