Decision Support Systems 55 (2013) 270–283

Contents lists available at SciVerse ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

Formal workflow design analytics using data flow modeling

Sherry X. Sun *, J. Leon Zhao

Department of Information Systems, College of Business, City University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Received 24 August 2011
Received in revised form 25 September 2012
Accepted 21 January 2013
Available online 16 February 2013

Keywords:
Workflow design
Data dependency
Activity dependency
Activity relations
Business process automation

ABSTRACT

Workflow design has become a critical function in enterprise information management. However, only scant research attention has been paid to formal workflow design methodologies. As a result, existing design methods in business process management remain a manual and experiential effort and result in inefficiency in design tasks and potential errors in workflow models. Considering that there are hundreds and thousands of business processes in organizations worldwide, overcoming this deficiency will have an enormous technical and economic impact on enterprise information management. In this paper, we investigate the possibility of incorporating formal analytics into workflow design, thus alleviating the intellectual challenge faced by business analysts when creating workflow models. The workflow design analytics we propose helps construct a workflow model based on information about the relevant activities and their associated data. In addition, our workflow design approach also helps determine whether the given information is sufficient for generating a workflow model and ensures the avoidance of certain workflow anomalies. The significance of our study is to enable the transformation of workflow design from a manual and experiential effort into a more systematic and rigorous approach.

1. Introduction

Over the last decade during the drive towards e-business, many organizations have realized the importance of Internet-based business process automation as they face the challenge of obtaining competitiveness in the global marketplace. To achieve high levels of business process efficiency, many organizations resort to workflow management systems to integrate business activities that span multiple functional units, such as human resources, marketing, and manufacturing [31,33]. As a prerequisite for effective workflow management, a workflow model is required to specify the execution sequence of activities needed to support certain business functions, such as producing a product or offering a service. Thus, it is critical to design correct workflow models efficiently according to specific business requirements.

The existing workflow design approaches, such as the participative approach widely used in practice, mainly focus on collecting business requirements rather than providing a rigorous procedure for generating workflow models. The lack of formal design approaches often causes design problems. Recent empirical studies have shown an error rate of 10%–20% from over 2000 process models used in different industry practice, which include the widely used SAP reference model, the models used in a process reengineering project in the service sector in Germany, the models documented in the financial industry in Austria, the models used by consulting firms, and the models from IBM projects [20–22]. With the empirical evidence of the design problems, the importance of formal design methods starts to be recognized.

The process of designing workflow models can be partitioned into three stages. The first stage is *business analysis* consisting of design activities to understand business objectives, system environment, domain knowledge, and process management rules [37,38]. The second stage is *process analysis* that deals with the issues of activity identification and dataflow specification [14,32]. Activities can be identified by workers and managers in various business units since they generally know what they do on the daily basis. Input and output data for each task, i.e. dataflow, can be identified from documents, forms, and databases used in the business process [15,27,35].

The third stage, namely *model construction* to derive the correct activity sequencing, is an intellectually challenging step. First, no clear guidelines exist to help determine whether a given set of activities and related data are sufficient for designing a workflow. Moreover, due to the human cognitive limit, great effort is required to figure out all interrelations of different steps in a complex process with no guarantee for the resulting model to be error-free [21,22]. These challenges are compounded by the fact that a workflow can involve multiple business units and different organizations located in different geographic locations.

The common wisdom for dealing with potential errors in the stage of model construction has been to confirm the workflow models through simulation and verification before deploying them in practice [2,7,17]. However, it is a fundamental principle in software engineering that design errors should be prevented as early as possible [21,23]. The later the errors are identified, the more cost and effort are needed to fix the errors. This principle also holds for designing and automating business processes [5]. Consequently, we believe that once a set of activities and their

* Corresponding author.
E-mail addresses: sherry.sun@cityu.edu.hk (S.X. Sun), jlzhao@cityu.edu.hk (J.L. Zhao).

0167-9236/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.dss.2013.01.028

input and output data have been identified in the stage of *process analysis*, it is possible to derive a workflow model through analyzing the data dependencies among activities, namely dataflow analysis, so as to avoid certain potential design errors in the first place. Of course, this assumes the knowledge of all the activities and their input and output data can be obtained in the stage of *process analysis* given that the necessary information can be identified from interviews, discussion sessions, and documents by using the existing design methods [13,15,27,35]. This simplification makes it manageable to develop a first-cut workflow design method.

The scope of our research is that starting from a dataflow to derive a corresponding acyclic workflow consisting of five basic constructs [1,2,28]. A dataflow is specified with a set of business activities, the related routing constraints defined on the given business activities, and input and output data items for each business activity. In this paper, we focus on determining relations between business activities from the dataflow and then developing a control flow consistent with the dataflow. The basic idea is to decide how activities should be sequenced in a workflow through examining the dataflow in a process, i.e., the transformation from input data to output data via a sequence of activities. First, we define certain dataflow properties, including well-connectedness, completeness, and conciseness, to verify if a dataflow specification provides correct and sufficient information for constructing a workflow model. Moreover, we develop design principles for deriving potential activity execution steps, referred to as “activity relations”, from data dependency analysis. Third, we develop an analytical method to enable the creation of complete workflow models from activity relations. Fourth, we simplify workflow design by applying the concept of inline block and decoupling the issue of model correctness from the issue of workflow efficiency. Comparing with the existing workflow design methods that involve largely manual and experiential efforts, our method converts the design of complex workflow models from a task that depends on the knowledge and experience of workflow developers to a task that is supported with a systematic and mathematically rigorous approach. Therefore, we refer to such an analytical approach as workflow design analytics.

The impact of workflow design analytics is significant. First, workflow design analytics builds the foundation for a formal approach to deal with the complexity of workflow model construction, thus alleviating the intellectual challenge faced by business analysts. Second, analytical workflow design can help verify the sufficiency of the given process information, including activities and associated data, for workflow model construction. Moreover, our approach can help eliminate certain potential workflow flaws caused by violation of data dependencies [30] and inappropriate model structures [7]. This feature of our approach will reduce the costs of generating error-free workflow models. As our research goal is to enable the transformation of workflow design from a manual and experiential effort into a more systematic and rigorous approach, the contribution of our research goes well beyond the computational formalism presented in this paper. Successful application of our approach will have considerable economic implications for companies with complex business processes.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature. Section 3 presents the data dependency concepts needed for workflow design. Section 4 defines the principles for deriving basic workflow structures from analyzing data dependencies. Section 5 demonstrates how to generate complex workflow models step by step. Finally, Section 6 concludes the paper by outlining our contributions and future directions.

2. Literature review

Most work in the area of workflow management focuses on representing activity sequences through various graph-oriented modeling languages. For example, both Event-driven Process Chains (EPC), used to describe the SAP reference model [9], and Business Process Modeling Notation (BPMN), developed by the Business Process Management Initiative (BPMI), are notations intended not only for technical developers but also for business analysts. As efforts to make the two graphic notations formal, different formalisms have been added to both EPC and BPMN [11,20]. Another popular graphic tool, UML activity diagram, plays an important role in workflow modeling [12]. Due to the lack of formal foundation [19] in UML activity diagrams, propositional logic has been applied to verifying UML activity diagrams [7].

As a formal modeling language, Petri nets [1,2] can help determine if an activity sequence contains errors, such as deadlock, activities without termination or activation, and infinite cycles, due to its abundance of analysis techniques. Based on Petri nets, a new workflow modeling language, Yet Another Workflow Language (YAWL), has recently been developed with some extended expressive power for specifying different workflow structures [4]. Metagraphs, a graphic structure extended from directed graphs and hypergraphs, has also been used as a formal workflow modeling approach to supports analysis of the data transformation through a sequence of activities [6].

While a significant amount of research in the workflow area has focused on modeling, verification, and analysis issues [2,7,17], formal methods to generate workflow models, namely the workflow design approaches, have not received enough attention in the literature [31]. The most well-known workflow design approach is the participative approach [13], which adopts the principle of joint application design (JAD) consisting of a variety of methods for conducting workshops where users and technical developers work together to define requirements of information systems [10]. This approach is useful for collecting information needed for workflow design, such as business documents, business rules, relevant business activities, and data elements critical to execute the activities. However, the participative approach does not offer formalism for generating workflow models in a rigorous manner.

The method of Product-Based Workflow Design uses the relationship among data elements derived from product specifications as a starting point for workflow design [27,35]. Moreover, cost and flow time are considered as criteria for selection of workflow models. The policy-driven workflow mining method [15] focuses on deriving a workflow model from documented business policies. However, formal procedures are still needed on how to determine activity sequences and how to use parallelism and conditional routing on the basis of analyzing relationships among data.

As an important step in workflow management, dataflow analysis can help discover dataflow errors hidden in a workflow model after the model has been created [25,29,32,34]. However, it should be more efficient to prevent dataflow errors by taking dataflow into consideration at the time the workflow model is created. Therefore, in this paper, we propose to start workflow design with analyzing data dependency, a type of flow dependency that arises whenever one activity uses a resource produced by another activity [18]. Moreover, we propose a new concept called “activity relations” as a foundation for a formal workflow design method. The concept of activity relations is inspired by the concept of “log-based ordering relations” used in [3]. However, the concept of “activity relations” is different from “log-based ordering relations” in that the former describes the structures of a workflow model and the latter describes the event sequences found in an event log.

The detection and verification of errors have also been investigated in the area of rule-based systems [24,26,30,37,38]. Validation of rule-based systems requires the identification of anomalies such as redundancy, ambivalence, circularity, and deficiency at a large-scale since a rule-based system can have thousands of rules. However, workflow design faces a different challenge given that each step in a business process typically produces a data output while a rule in a rule-based system may or may not produce any data. Further, consideration of the two intertwined aspects, i.e., dataflow and control flow, is not explicit in rule-based systems but critical for preventing structural errors in workflow systems.

As shown above, the number of publications related to workflow design is still quite small, particularly in comparison with the area of database design. However, workflow design is becoming increasingly