Available online at www.sciencedirect.com
ScienceDirect
Advanced Engineering Informatics 22 (2008) 28–44
ADVANCED ENGINEERING INFORMATICS
www.elsevier.com/locate/aei

Challenges, approaches and architecture for distributed process integration in heterogeneous environments

William J. O'Brien [a,*], Joachim Hammer [b], Mohsin Siddiqui [a], Oguzhan Topsakal [b]

[a] *Department of Civil, Architectural and Environmental Engineering, University of Texas at Austin, 1 University Station C-1752, Austin, TX 78712-0273, USA*

[b] *Department of Computer & Information Science & Engineering, University of Florida, Box 116125, Gainesville, FL 32611-6125, USA*

Received 29 November 2006; received in revised form 14 August 2007; accepted 17 August 2007

Abstract

Integration and coordination of distributed processes remains a central challenge of construction information technology research. Extant technologies, while capable, are not yet scalable enough to enable rapid customization and instantiation for specific projects. Specifically, the heterogeneity of existing legacy sources together with firms' range of approaches to process management makes deployment of integrated information technologies impractical. This paper reports on an architecture for distributed process integration named process connectors that addresses heterogeneity in a scalable manner. The process connectors architecture incorporates two key approaches that address heterogeneity over varying time scales. The SEEK: Scalable Extraction of Enterprise Knowledge toolkit is reviewed as a mechanism to discover semantically heterogeneous source data. The SEEK approach complements existing data integration methods for persistent sharing of information. To make use of shared data on a per project basis, a schedule mapping approach is presented that integrates firms' diverse individual schedules in a unified representation. The schedule mapping approach allows integration of process views that have different levels of detail, while allowing participants to maintain their own view of the process. Collectively, SEEK and the schedule mapping approach facilitate a broad range of analyses to support coordination of distributed schedules. While this paper focuses primarily on schedule process integration, the process connectors architecture is viewed as providing a broad solution to discovery and integration of firms' process data.

Keywords: Process integration; Scheduling; Distributed schedule coordination; Construction management; Information discovery; Legacy integration

1. Introduction

This paper reports on on-going research by the authors to create new mechanisms for distributed process integration in the construction and project management domain. The construction industry poses significant challenges for integration given a large number of firms of varying sophistication and a corresponding variety of data formats including a range of proprietary legacy sources. Technical integration challenges are made more difficult given the business climate that includes short-term associations and differing business practices (especially practices that involve firms operating on different levels of detail, making integration and constraint propagation exceptionally challenging). These collective integration challenges exceed the ability of extant systems to create scalable, rapidly deployable information systems that address coordination needs. Our paper reports on a broad architecture for distributed process integration with specific emphasis on (1) rapid discovery of legacy data and (2) mapping and management of discovered data to support process coordination.

Discovery of firms' data remains a significant challenge to integration. Despite the advent of modern data

* Corresponding author. Fax: +1 512 471 3191.
E-mail addresses: wjob@mail.utexas.edu (W.J. O'Brien), jhammer@cise.ufl.edu (J. Hammer), mohsink@mail.utexas.edu (M. Siddiqui), otopsaka@cise.ufl.edu (O. Topsakal).

1474-0346/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aei.2007.08.008

standards, firms have a broad choice of how to implement these standards (particularly for process information). Further, many firms use a variety of applications for process management, and many of these are homegrown or dated applications. To semi-automatically discover legacy data whatever the source, we review the SEEK (scalable extraction of enterprise knowledge) toolkit. SEEK provides a generalized, structured approach using data reverse engineering and wrapper development technologies to overcome the challenges of integrating information resident in heterogeneous legacy sources. SEEK exploits meta-data (schema, application code, triggers, and persistent stored modules) to develop a semantically enhanced database schema for the data source. The data source owner is presented with a graphical user interface (GUI) displaying the correspondence matches and allowing the owner to approve or change the matches. This allows rapid generation of a wrapper that translates the legacy source data to a common schema.

Building from the wrapper generated by SEEK (or related toolkits), discovered data must be made available for process coordination. We present a mapping approach for process coordination building on integration of schedules. Schedules provide a skeleton for including additional resource information and related constraints, enabling rich process description and coordination support. Our particular contribution is a tree based schedule mapping approach to integrate schedules of varying levels of detail that have $m{:}n$ matches among sets of activities (i.e., corresponding schedules that do not fall neatly into a simple hierarchical arrangement). A tree based approach provides flexibility to add a variety of constraint information at each leaf node based on data availability at the firm, as well as reason at different levels of decomposition in the tree. The schedule mapping approach, by accommodating different views of schedules, also allows firms to maintain their own view of the schedule rather than have a view that is imposed by others.

Implementation of these approaches is within the modular and scalable process connector architecture. Process connector components include a stub at each firm (incorporating a SEEK generated wrapper to legacy sources) as well as a bridge component between stubs to enable process coordination. The stub component serves as an information gateway to a firm's legacy information and handles information gathering, consolidation, and transfer back to the firm. The bridge component connects stubs. It can make recommendations for process coordination among firms using simple algorithms or provide collective information to more complex decision support applications. Recommendations are received by the participating firms through their respective stubs and can be approved or rejected (thereby invoking a new coordination process at the bridge). Process connectors thus support distributed process coordination across firms.

This paper first reviews the challenges of distributed process integration from both a semantic and a process integration perspective, reviewing the current state of the art and defining challenges in Section 2. Section 3 presents an overview of the process connectors architecture in response to the challenges detailed in Section 2. Section 4 of the paper reviews the SEEK architecture and approach to extracting semantically heterogeneous data from diverse legacy sources. While SEEK provides an approach to discovery heterogeneous data, Section 5 of the paper presents a schedule mapping approach to integrating heterogeneous schedules. Section 6 brings together the SEEK and schedule mapping approach within the process connectors architecture, presenting a broad and scalable approach to process integration.

2. Distributed process integration: review and challenges

2.1. Examples

Distributed process integration is a widespread challenge, with associated acknowledgement of the potential value of such integration. Despite many on-going efforts, the challenges are deep and many problems posed by construction projects remain unsolved or even unaddressed. As noted by Turk, Construction Informatics is an evolving field [1]. Within his ontology, our work involves several core themes including managing information, communication, and processing. Before reviewing the literature, we illustrate the challenges in two motivating examples. The first describes the challenges associated with differing representations of schedules (i.e., the issue of firms maintaining a distinct local view of a given process apart from that of other firms). This first example also illustrates some of the semantic integration challenges that exacerbate challenges in information integration. The second example further describes the challenges of data integration with respect to schema nomenclature.

Our first example is a mapping between the schedule of a construction manager (Centex) and a subcontractor (Miller) and is depicted in Fig. 1 (further details of the example are found in [2]). This figure shows a set of activities for both Centex and Miller that represent the same physical activities on a project. Corresponding activities are mapped between the two schedules. Note that these mappings exist on several levels (e.g. 1st floor to 1st floor and activity to activity). Most important, note that the mappings between individual activities do not have a 1:1 or even $1{:}n$ mapping in many cases (i.e.., mappings 3, 6, 9 and 11 in Fig. 1). A 1:1 one mapping (such as mapping 2) makes it easy to coordinate these two schedules as the set of information for each activity is largely the same (i.e., start–finish dates, duration, etc.). Similarly, a $1{:}n$ mapping could be considered a hierarchical mapping where the construction manager's or general contractor's schedule summarizes a more detailed breakdown by the subcontractor. Such a hierarchical level makes it easier to integrate differing views of the schedule. For example, the LEWIS system specifies four different levels of detail within a sche-