Information Systems 40 (2014) 115–127

Contents lists available at ScienceDirect

Information Systems

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

Similarity assessment and efficient retrieval of semantic workflows

Ralph Bergmann [a,*], Yolanda Gil [b]

[a] University of Trier, Department of Business Information Systems II, Universitätsring, 54286 Trier, Germany
[b] Information Sciences Institute, University of Southern California, Marina del Rey, CA 90292, United States

ARTICLE INFO

Available online 9 October 2012

Keywords:
Business workflows
Scientific workflows
Workflow similarity
Retrieval
Case-based reasoning

ABSTRACT

In the recent years, the use of workflows has significantly expanded from its original domain of business processes towards new areas. The increasing demand for individual and more flexible workflows asks for new methods that support domain experts to create, monitor, and adapt workflows. The emergent field of process-oriented case-based reasoning addresses this problem by proposing methods for reasoning with workflows based on experience. New workflows can be constructed by reuse of already available similar workflows from a repository. Hence, methods for the similarity assessment of workflows and for the efficient retrieval of similar workflows from a repository are of core importance. To this end, we describe a new generic model for representing workflows as semantically labeled graphs, together with a related model for knowledge intensive similarity measures. Further, new algorithms for workflow similarity computation, based on A* search are described. A new retrieval algorithm is introduced that goes beyond traditional sequential retrieval for graphs, interweaving similarity computation with case selection. We describe the application of this model and several experimental evaluations of the algorithms in the domain of scientific workflows and in the domain of business workflows, thereby showing its broad applicability.
© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Business workflow management is an established area that aims at "the automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules" [29]. In the recent years, the use of workflows has significantly expanded from the original domain of business processes towards new areas, as the modeling capabilities and the execution support that workflows provide are widely applicable. For example, in e-science *scientific workflows* are executable descriptions of automatable scientific processes such as computational science simulations and data analyses [44]. In medical healthcare, workflows can be used to support the execution of *medical guidelines*, i.e. standardized treatment processes of a certain disease [35,33]. Further, workflows can be used to represent and execute search [19] and information integration processes [24] in the context of decision support systems. Even in cookery, workflows can be used as a means to represent the cooking instructions within a recipe [36] in order to provide step-by-step guidance during cooking.

Such new applications of workflows typically deal with a number of new difficulties, particularly due to

- an increasing number of specific workflows potentially relevant to a domain,
- an increasing complexity of workflows,
- an increased demand for more flexibility, resulting in *agile workflows*, and
- the need to enable non-IT staff to create workflows and to control their execution.

These new challenges in workflow management ask for new methods and tools that support domain experts to

* Corresponding author.
E-mail addresses: bergmann@uni-trier.de (R. Bergmann), gil@isi.edu (Y. Gil).

0306-4379/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.is.2012.07.005

perform workflow-related tasks, particularly workflow modeling, composition, adaptation, analysis, and optimization.

1.1. Process-oriented case-based reasoning

Case-based reasoning (CBR) [8,6,1] has recently demonstrated its high potential for this purpose [45,34,9,30,20,16,37]. The emergent field of process-oriented case-based reasoning (POCBR) particularly focusses on the integration of process-oriented information systems with case-based reasoning. The particular appeal of CBR comes from the fact that it supports and partially automates experience-based problem solving. Process-oriented case-based reasoning aims at addressing the problem of creating new workflows as an experience-based activity. To this end, new workflows can be constructed by reuse of already available workflows that have to be adapted for new purposes and circumstances. A repository of successful workflows reflecting best-practice in a domain is the core of a POCBR approach. In case-based reasoning terminology such a repository is called a *case-base*, i.e. a collection of cases, each of which describes a particular workflow addressing a particular goal. Users can query the repository with a specification of important properties of the workflow s/he wants to create in order to retrieve potentially reusable workflows. One particular characteristic of CBR is that it allows to find cases that do not match exactly the user's query, but which are at least similar in some respect. This is based on the core assumption of CBR that similar problems have similar solutions. So, even if the repository does not contain a workflow which is immediately addressing the user's problem, a similar workflow could be available as a good starting point. Such similar workflows can then be adapted or at least provide some inspiration and guidance for the user during workflow construction.

1.2. Contributions of this paper

To implement such a POCBR approach, methods for the similarity assessment of workflows and for the efficient retrieval of similar workflows from a repository are of core importance. This paper addresses both problems in the following ways.

It is well known in CBR that the notion of similarity is crucial and that similarity must be modeled according to the particular domain at hand in order to ensure a high retrieval quality [8,31,6,13]. In line with this observation, we propose a new general framework for workflow representation and related similarity modeling. Workflows are represented as semantically annotated graphs, extending previous proposals for graph-based workflow representations [38,18,37]. The well-known local/global principle to modeling similarity measures [13,6] is extended to these graph representations, providing a flexible means for workflow similarity modeling.

Similarity-based retrieval in CBR is computationally difficult if the repository is getting large and if a domain-specific similarity measure should be used. Particularly for graph-based representations, the well-known indexing approaches [31] cannot be applied because they cannot cope with the graph structure. Even, the computational complexity of a single similarity assessment causes problems due to the fact that two graph structures must be compared. As a second contribution, this paper addresses this problem by developing several new algorithms for similarity computation and retrieval. With experimental implementations we analyzed the retrieval performance and quality of these algorithms. The most promising algorithm has then been implemented as a core component within the process-oriented CBR system CAKE [9,38]. It interweaves similarity assessment and retrieval and thereby enables efficient retrieval even for larger repositories.

As we are interested in generic methods, we demonstrate our methods using two workflow domains with very different characteristics: we address traditional business processes, which are control-flow oriented and scientific workflows, which strongly focus on the dataflow [32].

2. Foundations and related work

This section provides a focused introduction to the relevant foundations of (semantic) workflow representation and the notion of similarity in CBR. We also survey relevant related work in the area of workflow similarity and retrieval.

2.1. Workflow representation

Workflow representations typically reflect the dataflow and/or the control flow structure among a set of tasks of a process. That is, the workflow represents the partial ordering of task that are part of the overall process. Today, various workflow representation formats are used, depending on the kind of workflow. Representation approaches for business workflows have a strong focus on the control flow, usually implementing (some of) the workflow patterns proposed by van Aalst [17]. Typical control flow patterns are *sequence*, *and-split*, *and-join*, *xor-split*, *xor-join*, and possibly *loops*. Fig. 1a shows an example of a business workflow within a University administration, according to the representation[1] used in the CAKE project [9,38] at the University of Trier. The workflow describes a simple set of activities required to book a room in a University. First, one has to search for an available room in the online database. Then, one can either send a plain email with a booking request to the facility manager or alternatively one can fill in a room request form and send it. Then, the facility manager assigns the room and sends a confirmation. The graphical representation of this workflow just shows the six tasks involved, which are organized using sequences and one xor-split/join control-flow pattern. The flow of data, such as the filled form, is neglected in this representation.

Unlike business workflows, scientific workflows have a strong focus on the dataflow, typically restricting the control flow to a partial ordering of the tasks [32]. Such a simple control structure offers several advantages and has been sufficient to support a variety of applications

[1] The graphical representation was inspired by UML activity diagrams.

دریافت فوری ← متن کامل مقاله

ISIArticles
مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات