

Chapter 8

Semantic Matching, Propagation and Transformation for Composition in Component- Based Systems

Eric Bouillet
IBM Research, USA

Mark Feblowitz
IBM Research, USA

Zhen Liu
IBM Research, USA

Anand Ranganathan
IBM Research, USA

Anton Riabov
IBM Research, USA

ABSTRACT

Composition of software applications from component parts in response to high-level goals is a long-standing and highly challenging goal. We target the problem of composition in flow-based information processing systems and demonstrate how application composition and component development can be facilitated by the use of semantically described application metadata. The semantic metadata describe both the data flowing through each application and the processing performed in the associated application code. In this paper, we explore some of the key features of the semantic model, including the matching of outputs to input requirements, and the transformation and the propagation of semantic properties by components.

Information analysts and decision-makers in many application domains face a daunting task. In order to quickly arrive at and continually update the business intelligence that informs their advice or decisions, they must collect data from many di-

verse sources, adapt and integrate that data and apply a variety of analytic models. For such users, the need for timely acquisition and analysis is not adequately addressed by either fixed, monolithic applications nor by queries against fixed-schema data stores. As new sources are discovered and new or evolving analyses are needed, new ap-

DOI: 10.4018/978-1-4666-0261-8.ch008

plication requirements trigger new development cycles, resulting in delayed decisions and requiring analysts to resort to manual acquisition and analysis in the interim.

Such circumstances have given rise to an important area in end-user driven computing that Cherbakov, et al, describe as *situational applications* (SAs) (Cherbakov, Bravery, & Pandya, 2007): "...situational applications describes applications built to address a particular situation, problem, or challenge. The development life cycle of these types of applications is quite different from the traditional IT-developed, SOA-based solution. SAs are usually built by casual programmers using short, iterative development life cycles that often are measured in days or weeks, not months or years."

How is it that the end user can assemble such applications? Would it be reasonable to expect that, to arrive at a desired outcome, they could identify the components needed (from a library of hundreds or thousands), know how to assemble them, and know how to prevent the seemingly sensible assembly of components into incorrect compositions? While developers can manually craft code modules and compose them into applications, how can the less knowledgeable application end users do so, without acquiring deep knowledge of the code, interface compatibility, the rules of assembly, etc.? Mashup tools such as Yahoo Pipes (Yahoo Pipes, 2007) and IBM Damia (Altinel et al, 2007) simplify the logistics of assembling situational applications, but they are limited to a relatively small and inextensible set of components. Such tools might encode interface compatibility rules, but they provide little visibility into the sometimes subtle semantics of the components' capabilities. We present a method of separately specifying *processing goals* for situational applications and the *functional capabilities* of the components to be used in addressing the goals. These specifications are used by composition tools that either

automatically provide a near-optimal solution or support the more knowledgeable user in the formulation of such a solution. We present a means of expressing processing goals using semantic expressions (graph expressions) built using terminology from the end users' problem domains. Descriptions of components' capabilities are similarly expressed. This approach is supported by SAWMILL, our semantics-based automatic flow composition middleware, as first introduced in (Bouillet et al., 2007).

The general problem of application composition is extremely challenging. To address a more tractable problem space we limit our focus to those applications described as *flow-based* information processing applications (Morrison, 1971). Flow-based applications are component assemblies arranged in a directed acyclic graph (*flow composition*) of black-box components connected by data flow links. Systems like Yahoo Pipes and IBM DAMIA support the creation of data mashups as flow-based applications. Stream processing systems like System S (Jain et al., 2006) also support flow-based applications that continuously process streaming data. Acyclic workflows in service-oriented systems can also be viewed as flow-based applications, where different services exchange messages either directly or through a coordinator service (e.g., a BPEL workflow).

In this article we describe composition of flow-based applications, describe the formal model underlying the semantic descriptions and describe how these descriptions support the composition of flow-based applications. In particular, we focus on key features of the semantic model that are particularly useful in facilitating composition. These include the matching of outputs of components to input requirements of other components, modeling the propagation of certain semantic properties by components from the inputs to the outputs, and modeling the transformation of other semantic properties by the components.

18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/semantic-matching-propagation-transformation-composition/65126

Related Content

Uncertain Static and Dynamic Analysis of Imprecisely Defined Structural Systems

S. Chakraverty and Diptiranjana Behera (2014). *Mathematics of Uncertainty Modeling in the Analysis of Engineering and Science Problems* (pp. 357-382).

www.irma-international.org/chapter/uncertain-static-dynamic-analysis-imprecisely/94520

Entropy Quad-Trees for High Complexity Regions Detection

Rosanne Vetro, Dan A. Simovici and Wei Ding (2013). *Advances in Abstract Intelligence and Soft Computing* (pp. 299-316).

www.irma-international.org/chapter/entropy-quad-trees-high-complexity/72788

Cancer Classification From DNA Microarray Using Genetic Algorithms and Case-Based Reasoning

Lilybert Machacha and Prabir Bhattacharya (2021). *International Journal of Software Science and Computational Intelligence* (pp. 17-37).

www.irma-international.org/article/cancer-classification-from-dna-microarray-using-genetic-algorithms-and-case-based-reasoning/266226

Neuro-Imaging Machine Learning Techniques for Alzheimer's Disease Diagnosis

Gehad Ismail Sayed and Aboul Ella Hassanien (2017). *Handbook of Research on Machine Learning Innovations and Trends* (pp. 522-540).

www.irma-international.org/chapter/neuro-imaging-machine-learning-techniques-for-alzheimers-disease-diagnosis/180959

P3 Process for Object-Relational Data Migration to NoSQL Document-Oriented Datastore

Aicha Aggoune and Mohamed Sofiane Namoune (2022). *International Journal of Software Science and Computational Intelligence* (pp. 1-20).

www.irma-international.org/article/p3-process-for-object-relational-data-migration-to-nosql-document-oriented-datastore/309994