Demonstration: A Knowledge-Based Approach to Interactive Workflow Composition

Jihie Kim, Yolanda Gil, and Marc Spraragen
University of Southern California/Information Sciences Institute
Marina del Rey, CA 90292 USA
+1 310 448 8769
{jihie, gil, marcs}@isi.edu

ABSTRACT

Complex applications in many areas, including scientific computations and business-related web services, are created from collections of components to form computational workflows. In many cases end users have requirements and preferences that depend on how the workflow unfolds, and that cannot be specified beforehand. Workflow editors therefore need to be augmented with intelligent assistance in order to help users in several key aspects of the task, namely: 1) keeping track of detailed constraints across selected components and their connections; 2) accommodating flexibly different strategies to construct workflows; e.g., from general knowledge of necessary tasks, from desired results, or from available data; and 3) taking partial or incomplete descriptions of workflows and understanding the steps needed for their completion. We have developed a system called CAT (Composition Analysis Tool) that analyzes workflows and generates error messages and suggestions in order to help users compose complete and consistent workflows. Our approach combines knowledge bases, which have rich representations of components and constraints, together with planning techniques that can track the relations and constraints among individual components. We have formalized our approach based on AI planning principles, allowing us to formulate claims about the underlying algorithms as well as the resulting workflows.

Keywords
workflow composition, descriptive logic, interactive approach

APPROACH

We have developed the CAT (Composition Analysis Tool) system [1] as an approach to interactive workflow composition that incorporates 1) *knowledge-rich descriptions* of the individual components and their constraints; 2) *a formal algorithmic understanding of partial workflows*, based on AI planning techniques. Using this approach, CAT can analyze a partial workflow composed by the user, notify the user of issues to be resolved in the current workflow, and suggest to the user what actions could be taken next.

Users may design workflows using a variety of strategies, including: 1) top-down selection of components, starting from abstract types of components and then replacing them with more specific versions; 2) working from desired data to select models that can generate those results; 3) situation-based selection of components, working from the initial data available to select components whose constraints are consistent with those data.

Supported user actions include: add/delete a component, initial data or desired results to/from the workflow; specialize an abstract component to a more specific one; establish/remove a link (between the output of one component or initial data, and the input of another component or result).

Given the goals and actions just described, CAT users can drive the composition process. At any time, the system's ErrorScan algorithm may be invoked to analyze the workflow, and to suggest useful next actions to the user, to guide the user towards composing a correct workflow.

In order to support the kinds of interactions described above, CAT's knowledge base defines the components that can be used in a workflow and their input and output parameters (The details are in [1]). Note that because the component ontology describes abstract component types as well as specific components, users can start from a high-level description of what they want without knowing the details of what actual components are available.

The analysis of partial workflows created by the user is done using an AI planning framework. While automatic planning systems can explore the space of plans systematically and guarantee that the final plans are correct, interactive workflow composition requires an approach that lets the user decide what parts of the search space to explore and that can handle incorrect partial workflows.

We have developed a domain-independent algorithm, ErrorScan, to support mixed-initiative workflow creation that assists the user by ensuring that the workflow is complete and correct. Specifically, the final workflow must comply with a set of desirable properties [1], such as *satisfied*(Workflow W), which means that for all components in W, all input parameters are satisfied by links from output parameters. For each property, CAT notifies the user of each workflow element that does not comply.

The ErrorScan algorithm generates specific suggestions to the user for how to fix each error found. In addition to checking the workflow for desirable properties as mentioned above, the algorithm consults the knowledge base to check domain-based properties (e.g., the consistency of a link is based on the parameter type definitions in the ontologies), and to generate specific suggestions (e.g., if an input parameter is not satisfied, ErrorScan will return a list of component types that have outputs that subsume the type of that parameter).

EXAMPLE

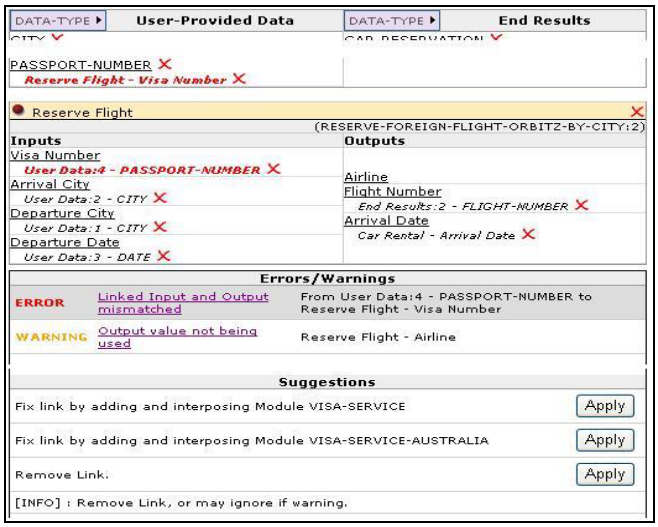

Figure 1: CAT detects error in workflow.
CAT home page: http://www.isi.edu/ikcap/cat/

Figure 1 shows the CAT interface during workflow composition (travel domain). The pieces of the interface shown in the figure are, from top to bottom: user-provided data in workflow (Passport Number in this case), end results, tasks in workflow (Reserve Flight), errors detected by CAT, and suggestions for fixing a selected error. Each task contains lists of input and output parameters (e.g., Visa Number is an input parameter of Reserve Flight). Italicized text below a parameter represents another parameter, to which the first parameter is linked. In this scenario, CAT has detected that the user has linked Passport Number to Reserve Flight:Visa Number, which is incorrect because Visa and Passport numbers are not the same data type. The default suggestion for fixing this type of error is simply to remove the incorrect link. However, CAT has also noted that the tasks Visa-Service and Visa-Service-Australia can be added to the workflow and interposed into the link, creating two correct links: 1) from Passport Number to Visa Service:Passport Number, and 2) from Visa Service:Visa Number to Reserve Flight: Visa Number. The old, incorrect link is automatically removed during this fix.

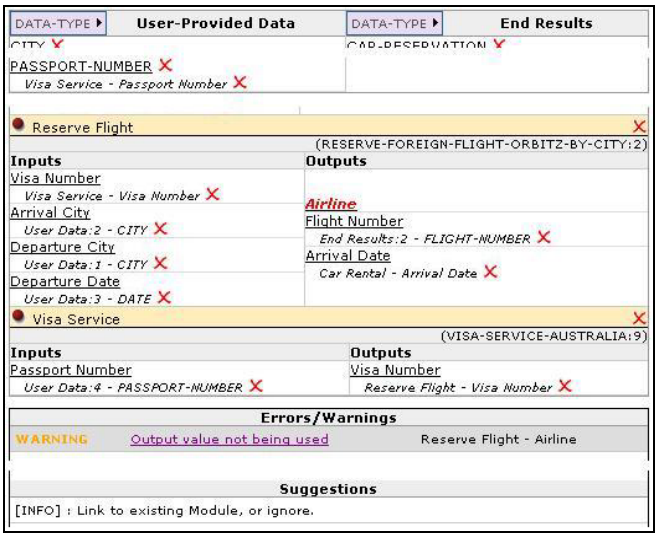

Figure 2: User applies a fix suggested by CAT.

In Figure 2, the workflow has changed; the user has taken CAT's suggestion to interpose Visa Service Australia to fix the incorrect link. Note that Visa Service Australia has been added to the list of tasks in the workflow, and Visa Service's parameters are linked to the ones that formerly were incorrectly linked to each other: user-provided data Passport Number, and Reserve Flight: Visa Number.

SUMMARY

We presented a flexible approach to interactive workflow composition that combines knowledge-based representations of components, together with planning techniques that can track the relations and constraints among components, no matter the order of the user's actions in specifying the workflow. Our approach led us to develop a system of formal properties of workflows, and the ErrorScan algorithm as implemented in CAT. This approach and implementation presents intelligent assistance for users in composing complete and correct workflows.

REFERENCE

1. J. Kim, Y. Gil, and M. Spraragen. A Knowledge-Based Approach to Interactive Workflow Composition. *To appear in Workshop on Planning and Scheduling for Grid and Web Services, at International Conference on Automated Planning and Scheduling (ICAPS-2004)*, 2004.