

EXTRACT/AIDE

Extensible TRansformation And Compiler Technology

Worcester Polytechnic Institute

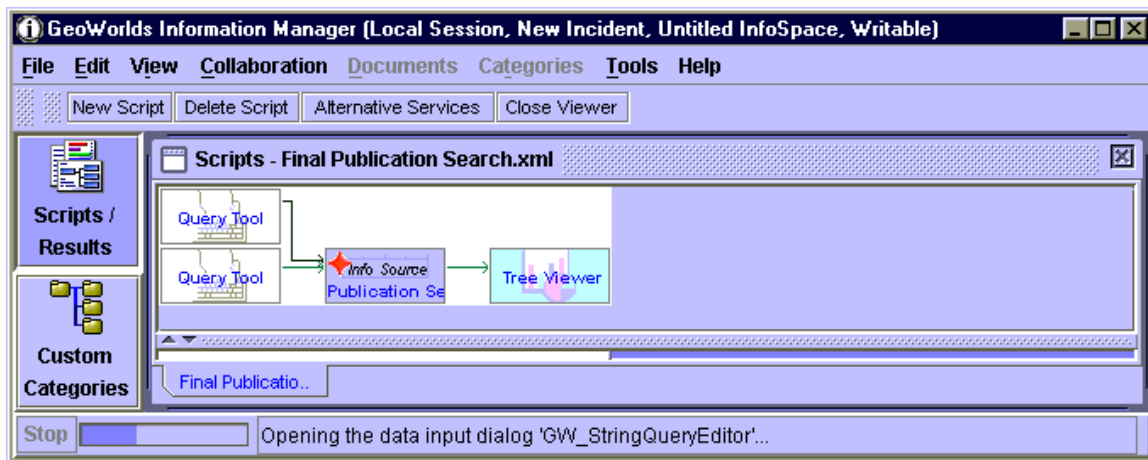
Computer Science Department

Worcester, MA 01609

George T. Heineman (heineman@cs.wpi.edu)

Overview

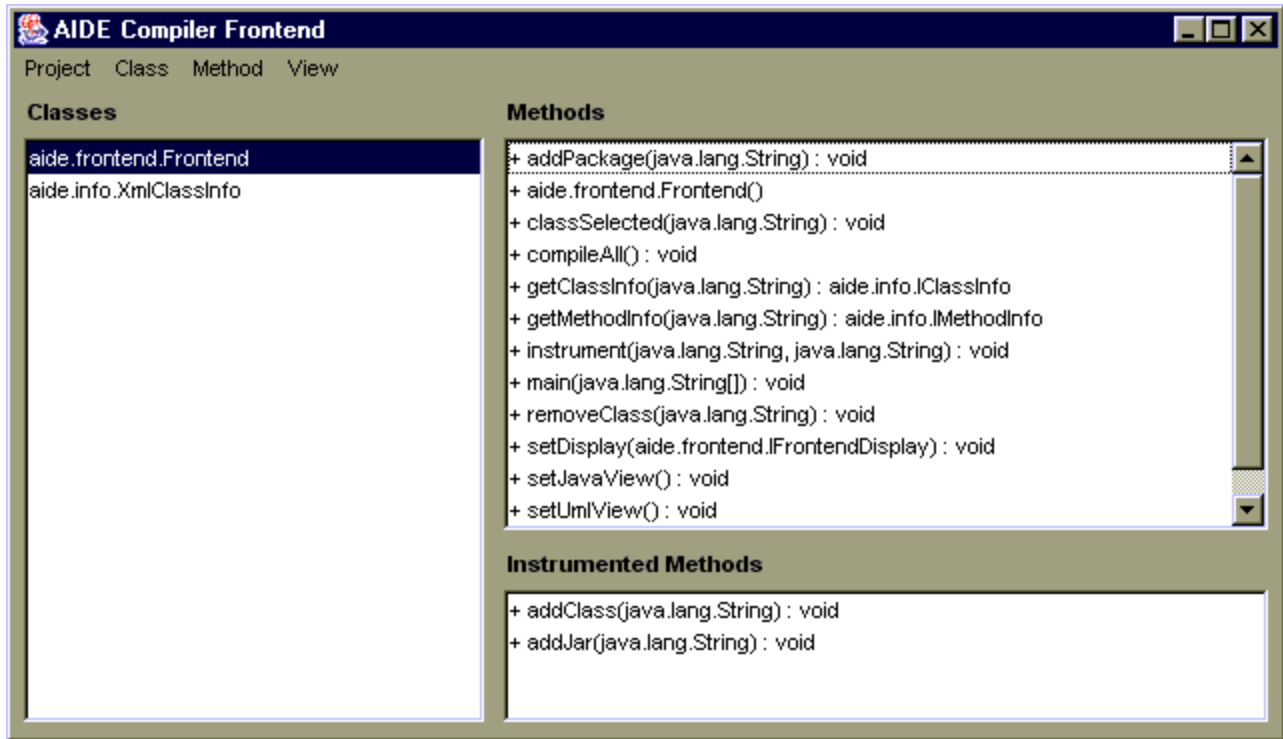
This demonstration shows how EXTRACT/AIDE can be used to troubleshoot an external service in *GeoWorlds*, a geospatial data analysis program developed at the Information Sciences Institute of the University of Southern California. *GeoWorlds* allows developers to create services that retrieve information from the World Wide Web. Pertinent information is “scraped” from the retrieved web pages using pattern matching. This approach to data retrieval is fragile. If the format or layout of a web page is modified, the patterns used to extract the data are no longer valid. Consequently, the service will return no data, leading the user to believe that their search returned no results when this may not be the case.



For our demonstration, we developed a service that performs a search of publications given a set of keywords and authors (the Info Source in the above screen shot). The service submits the query to a web and performs a series of regular expression queries on the resulting web page to extract the matching publications. If this external service changes its format, this *GeoWorlds* script will return no documents. Our demonstration adds value to *GeoWorlds* by enabling script developers to develop probes that ensure properties of external services.

Using our technology, developers can create Embedded Code Sensors (ECS). An ECS is a probe that can emit events at a variety of user-defined points in a program's execution: for example, when an object is instantiated, a class attribute is accessed, a user-specified assertion fails, a reflective method invocation occurs, or an exception is thrown. In our demonstration, we instrumented the *GeoWorlds* interface to the external service using the EXTRACT/AIDE compiler. An ECS was inserted to examine data being passed to the pattern matching method of our service; if the format of the remote service changes unexpectedly, a probe event is emitted. A generic *GeoWorlds* script gauge detects such events and alerts the user of changes in the external environment.

EXTRACT (Extensible TRansformation And Compiler Technology) is a tool currently in development that allows users to perform general-purpose transformations on Java code. It supports a variety of code transformations including insertion, removal, restructuring, and the embedding of ECS.



Technical Details

EXTRACT performs code transformations on Java source code rather than byte code. The source code is parsed into an abstract syntax tree (AST) and manipulated using graph-rewriting techniques. By directly accessing the parse tree, EXTRACT transformations can easily be specified and carried out.

EXTRACT allows users to specify general code transformations that are self-contained, modular, extensible, and explicitly declared as atomic steps. We are using an XPath-like language to allow users to select nodes from the AST and define transformations over those selections. There are two types of transformations that are supported. First, we support node attribute changes (e.g., changing a method declaration from public to protected). Second, we support child insertion and replacement (e.g., removing, rewriting, and reinserting an AST sub-tree). These transformations are expressed in a language similar to XSLT. Transformation scripts can be analyzed and compiled into Java code that performs the modifications to the AST.

We are currently working on analyzing combinations of transformations. Based on conflict serializability theory, we will develop a scheduler that can guarantee consistency after the application of a number of transformations EXTRACT requires JDK 1.3 and uses the OpenJava MetaObject Protocol for AST operations.

Software Specifications

Platform: Java 2 (JDK 1.3). Tested on Solaris and Microsoft Windows platforms

URL: <http://www.cs.wpi.edu/~heineman/dasada>

Availability: EXTRACT Beta Release 1 Scheduled for July 10, 2002; existing AIDE tool now available