Lessons on Converting Batch Systems to Support Interaction

Experience Report

Robert DeLine **Gregory Zelesnik** **Mary Shaw**
Computer Science Department
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA, USA 15213-3891
+ 1 412 268 1298
{rdeline, gz, mary.shaw}@cs.cmu.edu

ABSTRACT
Software often evolves from batch to interactive use. Because these two usage styles are so different, batch systems usually require substantial changes to support interactive use. Specific issues that arise during conversion include assumptions about duration of system execution, incremental and partial processing, scope of processing, unordered and repeated processing, and error handling. Addressing these issues affects the implementation in the areas of memory management, assumptions and invariants, computational organization, and error handling. We use as a working example our conversion of the batch processor for the UniCon architecture description tool into an interactive architecture development tool. To capture the lessons for practitioners undertaking this type of conversion, we summarize with a checklist of design and implementation considerations.

Keywords
software evolution, interactive systems, batch systems, re-engineering

INTRODUCTION
Successful software evolves over time. One of the most common evolutionary changes is migration from batch to interactive use. This minimally requires adding capability to invoke the system's operations one at a time and to provide visibility into the intermediate results. In addition, it often involves support for finer granularity of operations, user-driven order of intermediate operations (including incremental re-computation and error checking), interactive construction of the information to be manipulated, graphical depiction of that information, new strategies for error handling, and multiple concurrent computations.

The batch implementation may not accommodate these additional capabilities gracefully. For example, batch systems can predetermine the order of computational steps, so there is little need for the clean decoupling of steps and no need for the ability to incrementally repeat individual steps. Response time and concurrent independent jobs are usually not of concern, so a single thread of control is often sufficient. All these are to some degree incompatible with interactive use.

We encountered many of these problems when we converted the tools for our UniCon architecture description language [9] from batch to interactive use. This batch processor accepts a textual description of a system's architecture, performs architecture-specific analyses, and outputs a set of instructions to build the system. In addition to these capabilities, our interactive editor also allows the user to create and modify architecture descriptions depicted as diagrams and provides useful feedback during the editing operations. This paper describes the issues that generally arise when converting a batch system to interactive use, their implications for the implementation, and the lessons we can pass along. To make the discussion concrete, we use the UniCon tool set as our working example throughout the paper.

BATCH AND INTERACTIVE SOFTWARE
To understand the issues that arise when converting a batch system to an interactive one, it is useful to compare the external observable behavior of the two kinds of systems.

Batch systems accept one or more complete inputs, produce one or more complete outputs, and terminate. The inputs and outputs have state, for example as files or text streams. Batch systems may use and perhaps update other on-line information; they may also produce intermediate results that are intended to be transient. Human intervention is neither expected nor accepted, except perhaps for special error handling. Start-up flags may tune the underlying computation, but once the computation has started, it is not influenced externally.

Interactive systems are of many kinds, differing in such properties as the locus of control (internal or external), the interaction model, and the coupling between the internal state and the external state. They share the expectation that the human user participates in the computation.

Interactive systems with internal control query the user explicitly for input and ignore or queue input provided at other times. Interactive systems with external control allow the user to choose the order and timing of the inputs. Systems with internal control often have a single control thread; systems with external control often dedicate a thread to the user interface or else poll frequently for input. To illustrate the difference, consider two systems that gather information about the user, such as name, address, and phone number. A typical system with internal control would prompt in turn for each piece of information, with no way to go back and change previously entered data. A typical system with external control, on the other hand, would provide a form to be filled out (for example, in a dialog box). The user can enter the information in any order and can change previously entered information at any time until the entire form is submitted to the system.

Interaction models vary widely. At one end of the spectrum are systems where the user answers questions at system prompts; at the other end are full object-manipulation models in which the user creates the "input" for the system dynamically, incrementally, non-monotonically, and often graphically. The more complex models allow users to create and operate on portions of the system data alternately with the system; "input" is therefore not an accurate label for the description they create. We use the term "work product" to refer to this shared definition that evolves from the "input" to the "output" through cooperation of the user and the system. The more complex models also require a representation for the user's view of the computation state as well as an internal data structure.

We are chiefly interested in object-manipulation systems with graphical interfaces and external control. Some of our observations may not apply to simpler systems.

CONVERSION ISSUES

To accomplish the change from batch to interactive, several issues must be addressed:

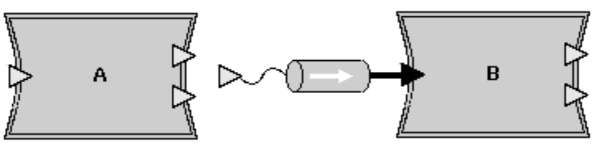
- *Assumptions about execution duration*. A batch system is invoked for a single job; one invocation of an interactive system may be used to process many individual jobs, successively or concurrently. The execution duration of an invocation of a batch system is therefore determined by its computation and the size of its input; it is usually relatively short (seconds, minutes, or hours). In contrast, the user determines the execution duration of an invocation of an interactive system. Since users can leave sessions running indefinitely on modern workstations, an interactive system may run for hours, days, or even weeks.
- *Incremental processing*. As Ambriola and Notkin [1] observe, with a batch system no internal data structures persist between multiple invocations; even small changes to the input require complete re-computation. An interactive system allows incremental editing of a work product and can therefore allow incremental update of the underlying data structures.
- *Partial processing*. Because a batch system's input is fixed at invocation, the input must be complete. The system's computation typically expects complete input and reports incompleteness as an error. An interactive system is used incrementally to create and modify a work product. Its computations must be tolerant of incomplete and missing input. These computations may be required to work sensibly on well-formed, but globally incomplete, substructures.
- *Scope of processing*. The batch system typically provides a fixed computation over a narrow range of input. For example, many compilers process one file of source code at a time. Interactive systems, however, often allow multiple work products to be active at once. For example, desktop publishing systems allow multiple unrelated documents to be edited concurrently. Since a given interactive operation should apply to only one of the work products, the scope of the underlying computations must be carefully controlled.
- *Unordered and repeated processing*. Because a batch system's computation is fixed, the computation can often be structured into sub-computations or phases, where each phase fits into a context in which certain assumptions hold. Such an assumption might be that the phase will be performed exactly once or that some other computation has previously been performed. Because an interactive system's operations are driven by the end user, the computations behind these operations typically cannot make strong assumptions about ordering and repetition.
- *Allowed operations*. A batch system's input is fixed during a given execution, while an interactive system's work product changes incrementally over a given execution. Because of this, the interactive system's data structures may be required to support operations that the batch system's do not (e.g., modification, deletion of portions of the work product).
- *Error prevention versus error detection*. A batch system can only report errors that are found in its input. Because an interactive system controls the production of the work product, the user interface can be designed to prevent many errors by disallowing the construction of ill-formed constructs.
- *Error reporting*. In a batch system, errors are usually reported as they are encountered in the input, and information about the error is reported in an error message that includes information about context, such as line numbers. In an interactive system, errors may not be reported immediately upon detection, but at a time more appropriate to the particular user interaction. Additionally, information about an error may need to be displayed graphically to supplement the text messages (if text messages are even applicable). The contextual information in error messages in an interactive system is also different; line numbers, for example, may not be suitable.
- *System control.* In a batch system, control proceeds based on the computation over the input. If at any time a severe error is encountered that would prevent successful completion of the computation, a batch system can choose to terminate. In an interactive system, however, control is in the hands of the user. When a severe error occurs, an interactive system must recover gracefully from the error and return control to the user. This is especially critical when an interactive system allows simultaneous work on multiple work products. A fatal error triggered by operations on one work product should not affect other users of the system.

When performing your own conversion of a batch system to support interactive use, you may discover, as we did, that these issues do not lead to independent modifications of the implementation. As a result, the issues should not be addressed in isolation. Sound software engineering practice suggests considering of all the changes to the batch system together since one change may impact another.

EXAMPLE: CONVERTING UNICON

Our original tool for the UniCon architecture description language was a batch system that accepted textual architecture descriptions and produced Odin* building instructions and a log of warnings and errors. The processor was organized much like a classical compiler.

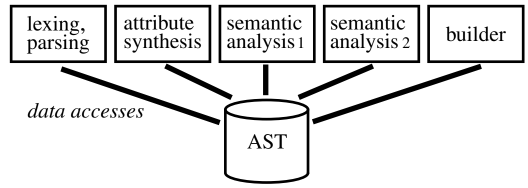
Since software architects frequently describe their systems with diagrams, we converted the original batch tool to support interactive graphical development of system descriptions. The graphical notation is interchangeable with the textual notation. During interactive development of a definition, the user adds and deletes architectural elements, edits properties of elements, and rearranges the way the elements are composed, all through a direct-manipulation user interface. An example of a diagram under construction, a description of a simple Unix pipe–filter system, is shown in the following picture. This partial snapshot from our tool



shows two filter components, A and B, along with a pipe connector between then. At this point in the editing, the pipe has been hooked up only to B.

The editor provides both immediate and deferred feedback about the design under construction. An example of immediate feedback is a kind of type check performed as the user hooks up a connector to a component. For example, to hook up the above pipe to A, the user would drag the triangle dangling off to the left of the pipe and drop it on one of the triangles on the right side of A. While the user is dragging, the editor provides immediate feedback about the legality of the potential connection. If the user tries to make an illegal connection despite this feedback, the connection is rejected and the triangle will animate back to its dangling position. In this way, the diagram is always syntactically correct, if incomplete, much like a program developed in a syntax-directed editor [3][12]. Other feedback in our editor is deferred. For example, checking whether each component is connected to at least one other component is deferred until the user "compiles" the diagram into a set of Odin instructions.

The UniCon batch processor has a conventional compiler architecture, shown below. The computation is staged in phases, where each phase manipulates a central data structure, called the abstract syntax tree (AST). The lexing and



parsing phases create the AST. The attribute synthesis phase fleshes out the AST by caching search results as synthesized attributes in specific tree nodes. Two semantic analysis phases ensure that language-specific properties hold over the AST; the earlier phase checks local properties, while the later phase checks global properties. Finally the building phase produces the processor's output, the set of Odin instructions, from the AST. We implemented the processor in C, using the standard tools Lex [6] and Yacc [4] to produce the lexing and parsing phases.

ARCHITECTURAL ALTERNATIVES

When converting a batch system to an interactive one, the major strategic decisions address the overall structure, or architecture, of the modified system. These decisions include the kind of user interface to add and the way the user interface will interact with the computations to be retained from the batch system, which we'll call the computational subsystem.

Since we are chiefly interested in interactive graphical manipulation, we will consider only the case in which the user interface is a distinct subsystem of the interactive system. Further, we will assume that the user interface needs to invoke operations of the computational subsystem incrementally and on partially-formed results. This rules out the option of complete separation, in which a graphical editor works independently of the computational subsystem and sends its final result off for batch processing.

Defining the interaction between the user interface and the computational subsystem requires defining the abstractions, representations, and control to be used [5]. The most significant of these is the abstraction, especially the major choice between interacting via shared state and interacting via discrete actions—that is, whether the computations of the two components will be driven by the *state* of the shared data or whether it will be driven by messages, events, or other actions that announce *changes* in the shared data. This choice leads to decisions on representations and control.

Strategy 1: Shared State

One class of abstractions for the relation between the user interface and the computational subsystem is state sharing. In a shared-state system, both components can manipulate and access the data of the computation. Neither takes the initiative, and neither is responsible for notifying the other of changes. In general, the tighter the coupling between internal and external system state, the more likely it is that shared state will be preferred.

* Odin is a make-like utility for system construction, which computes complete dependency information automatically. For more information, contact Geoff Clemm (geoff@atria.com).

The representation decisions deal with the data *per se*. This can either be accomplished within a single name space by sharing data structures or across multiple name spaces through a repository. The control issues center on the number of threads of control and synchronization.

Strategy 1.1: Shared Data Structures
This strategy tightly couples the user's and system's views of the state of the computation. In the final system, both the interactive subsystem and the computational subsystem have direct access to the data structures that represent the work product. The interactive subsystem updates these data structures as the user edits the work product. The computational subsystem computes over the data structures, as it did before the addition of the interactive interface. Given that both parts must access the data structures, a natural architectural choice is for both to access the same data structures stored in a shared memory. If a single thread of control is to be used, you now need to consider how the thread passes from one subsystem to the other. If multiple threads of control will be used (e.g. separate processes), you need to worry about synchronizing access to the shared data.

Strategy 1.2: Shared Repository
If numerous components of a system should all share the same state but run as independent processes, it is possible to share the state through an external repository. Because of interprocess overheads, response time will be slower than for shared data structures.

Integration of new tools into a set that processes the same data is often accomplished via a shared repository architecture. The data to be manipulated is given a structure in a repository. Each tool that works on the data contains an importer and an exporter which allow access to the data in its persistent form in the repository. The importer and exporter for a given tool convert the data between the repository form and a form natural for the performance of its own task. Multiple tools can communicate with each other through the data in the repository [7]. Integration of new tools into the set requires, at most, implementing both a new importer and a new exporter. Software development environments that are systems of CASE tools are often implemented as shared repository systems.

Strategy 2: Discrete Actions
An alternative class of strategies for the relation between the interactive subsystem and the computational subsystem is interaction via discrete actions such as procedure calls, messages, or events. In a discrete-action system, components act in response to changes in the environment or of the shared state; the discrete actions are the means of announcing the changes. This permits a looser coupling between the two components, especially their representations [8].

The representation decisions under these strategies must deal with the information carried by the actions. The control issues are tied chiefly to the actions.

Strategy 2.1: API for Discrete Actions
The most common discrete-action organization is the use of an application programming interface (API) to define a set of procedure calls that can be used to invoke the capabilities of a system. Note that this usually provides an essentially asymmetric relation: one component provides services to be invoked by another.

The X windows system [10] provides an example of this type of asymmetric loose coupling. An X-based application consists of a client that provides application-specific computations and a server that is in charge of the display. The client and server are connected by a communication channel that is governed by an API. The client calls this API both to send update events, such as drawing commands, to the server and to request input events, such as mouse clicks, from the server.

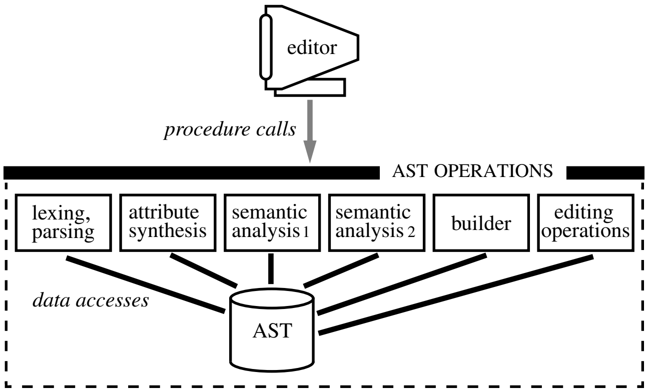
Symmetric procedural interfaces can also be defined; this requires joint design of the cooperating components. Systems in which messages or events are used as triggers are more likely to be symmetric than those based on APIs.

Strategy 2.2: Encapsulation of Shared State
A special case of an API is the encapsulation of a data structure into an object or abstract data type (ADT). Here the operations of the ADT provide the programming interface, but the essential abstraction is very similar to a shared data structure. This strategy can resolve the tension between a data-sharing abstraction and an implementation that hinders direct application of the abstraction. That is, you can approximate data sharing by adding a component that encapsulates the data and exports operations symmetrically to the user interface and the computational subsystem. Taylor and Johnson note that a key benefit of this encapsulation is its insulation of the computational parts of the system from changes in the user interface [11].

Architectural Decisions for UniCon
Since we wanted our editor to give immediate feedback about the architecture description under construction, we realized that both our user interface (interactive subsystem) and UniCon processor (computational subsystem) would need low-latency access to the data structures representing the description. Hence, Strategy 1.1 was attractive. Applied to our system, this strategy meant having the user interface and the UniCon processor both access the AST in shared memory.

Unfortunately, this architecture had two disadvantages. First, for reasons not discussed here, the user interface and UniCon processor were to be written in different programming languages, which makes implementing shared data tricky. Furthermore, the AST is a complicated data structure with many representation invariants. If the user interface code were to directly modify this data structure, it would be required to maintain these invariants.

Because of the AST's invariants, Strategy 2.2 was the right choice for our interactive editor. The AST and its invariants were encapsulated as a set of abstract data types. To update the AST, the user interface calls the operations on the abstract data types. Since the code for the existing batch processor already upheld the data structures invariants, the code for the operations on the abstract data types was scavenged from the batch processor's code and reused. In addition, some new operations were added to accommodate the interactive editing features of the editor. The architecture of the interactive editor is shown below.



IMPLEMENTATION CONSIDERATIONS

The conversion issues previously discussed lead to a number of requirements on the implementation. Some implementations may already satisfy some or all of the requirements, but most will require revisions of one kind or another. For narration purposes, it would be convenient if each of the conversion issues had an independent impact on the implementation. Unfortunately the reality of conversion is not so simple. To structure the discussion, we cluster the required revisions into four main areas of concern: memory management; assumptions and invariants on internal data structures; organization of computation; and error handling. The table below summarizes which conversion issues impact which implementation concerns, thereby establishing a correspondence between the discussions of conversion issues (rows) and implementation considerations (columns).

	memory management	assumptions & invariants	computational organization	error handling
assumptions about execution duration	√			
incremental processing		√	√	√
partial processing		√	√	
scope of processing			√	√
unordered and repeated processing		√	√	
allowed operations	√	√	√	
error prevention versus detection				√
error reporting			√	√
system control				√

For each implementation consideration, we discuss several types of revisions. For each type of revision, we first describe the general impact on the implementation of any batch system. Then, as a concrete illustration, we show the specific impact on our working example.

Memory Management

Assumptions about execution duration and the allowed operations on system state both have an impact on the design of memory management functionality in a system. Unlike many batch systems, interactive systems are long-lived and often support operations that allow memory to be reclaimed. Because of this, interactive systems have both the opportunity and obligation to manage memory carefully.

Memory Usage Patterns

In a batch system, memory is typically allocated to build one or more data structures that are needed throughout the computation and hence cannot be reclaimed until the end. As a result, memory management is usually simple because there is no need to track and release any unused data structures. Many interactive systems, on the other hand, provide operations to the user (e.g., delete operations) that render previously allocated memory reclaimable. Furthermore, interactive systems are often long-lived; if unneeded memory were not reclaimed, the system could eventually run out of free memory.

In the UniCon batch tool, memory is primarily allocated to construct the AST. The AST must persist through all phases of processing up to the last one. Releasing this memory is unnecessary since the operating system frees all of the system's memory when the system terminates. Hence, the batch processor contained no memory deallocation code. As new operations that remove parts of the AST were added to support the interactive editor, memory deallocation code was also added to release the allocated memory.

Memory Leaks

Not all memory that a batch system consumes is required to persist until the end of its computation. Temporary space may be required, for example, for buffers or transient copies of data. Because a batch system is relatively short-lived, it is often safe to allocate new memory for temporary space and then abandon it, rather than carefully tracking and releasing it when possible. In a language, like C, where memory management must be done by hand, tracking and releasing unneeded memory can be tricky to get right. Because the batch system's short lifetime can make memory leaks tolerable, simplifying memory management may be an appropriate means for lowering the cost of system development. In an interactive system, however, such a cost-saving method is not appropriate: over its longer life, the interactive system could easily run out of free memory.

The UniCon batch tool followed this simplification strategy and never released the memory used for temporary space, even when that space became unreachable. While converting the processor to support an interactive editor these memory leaks were found and corrected.

Assumptions and Invariants

When converting a batch system to an interactive one, many of the batch computation's invariants and assumptions will need to be revisited. Interactivity often requires incremental modifications to the data structures and toleration of incomplete input. These may invalidate many of the assumptions the computation makes about the state of the data structures. The interactive portions of the system need to respect the batch computation's assumptions, while the batch computation may need to weaken its assumptions to accommodate interactivity.

Invariants and Transition Assumptions

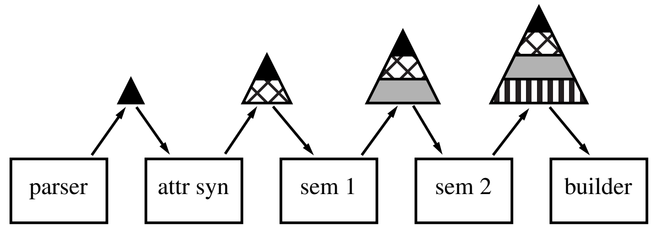
In a typical batch system, no data structures persist between invocations. Any change to the input requires complete re-computation of the results. In contrast, in an interactive system, incremental changes can be made to the work product within a given session. Incremental changes to the work product lead to incremental changes to the underlying data structures. These changes must preserve any invariants on the data structures required for the safe re-computation of the results.

In addition to invariants on data structures, computations within an implementation make other assumptions about the status of the data when they begin computation. In a batch system with a fixed order of computation, these assumptions are naturally established by the ordinary sequence of events.

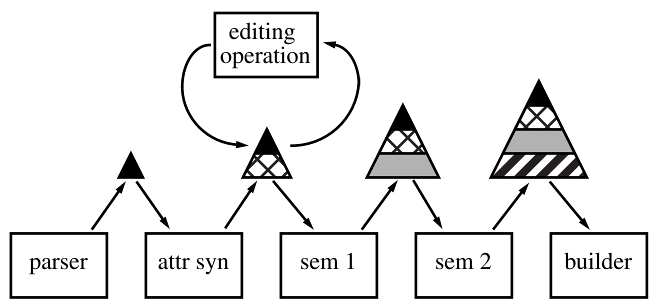
However, incremental re-computation and processing of partial data structures can easily lead to invocation of code in new contexts, and the entry assumptions made by existing code may no longer apply. For example, computation *X* may establish a relationship *R* within a data structure that computation *Y* requires. (*R* need not be an invariant of the structure, just a precondition for *Y*.) If *Y* follows *X* in the batch version, all is well. However, an interactive change might invoke computation *Z*, which invalidates *R*, and then another change might invoke *Y* without first passing through *X* or any other operation that re-establishes *R*.

In the UniCon batch processor, the invariants and transition assumptions all focus on the central data structure, the AST. The data in this tree are called attributes and fall into two categories: lexical and synthesized attributes*. For a given language construct, the lexical attributes are the constituent parts of the construct, as given by the grammar; the synthesized attributes are the cached results of expensive operations, such as tree traversals.

The AST's invariant property is that the lexical attributes are present and correct. The first phase, the lexing and parsing phase, establishes this property, and every subsequent phase upholds it. Further, each phase establishes additional properties that the subsequent phases require as a precondition. For example, the attribute synthesis phase establishes the property that the synthesized attributes are present and correct, which the next phase—the local semantic analysis phase—requires. This process of building up properties from phase to phase can be pictured as below. The triangles represents the proven properties of the AST, with each kind of hashing representing a different property. The arrows indicate the phases that ensure and require these properties.



For the interactive tool to be able to provide immediate feedback, the user interface code must use some of the computations from the semantic analysis phases of the UniCon processor. A precondition of these computations is that the AST's synthesized attributes are present and correct. To ensure that each user interface interaction upholds this condition and to localize the knowledge required to manipulate the AST's synthesized attributes, the abstract data types that represent the AST export a set of operations that correspond to editing commands in the editor. Each of these editing operations is responsible for upholding the condition that the synthesized attributes are present and correct. Thus, after any series of these operations, a semantic analysis operation may be called since its precondition is met. A revision of the previous illustration shows how these editing operations affect the AST properties. As shown by the cycle, each operation both requires and upholds the postcondition of the attribute synthesis phase.



Weakening Assumptions

A batch system's input must be complete, since there is no opportunity for the input to be updated before the computation takes place. If the input to a batch system is incomplete, the system reports the error and terminates. An interactive system, on the other hand, is used to create and edit the work product. Although it may withhold some operations until the work product has reached a certain level of completeness, an interactive system must tolerate incompleteness.

The original code for the UniCon processor's phases was written with completeness in mind. For example, the phases that occur after the attribute synthesis phase were written to assume that all synthesized attributes are present and correct. Since the interactive editor allows the incremental creation of the AST, at any given time some parts of an architecture description—hence some nodes in the AST and

* Abstract syntax trees, as described in canonical compiler texts [1], also contain inherited attributes, which the UniCon batch processor does not use.

their attributes—may be missing. These assumptions about the presence of AST nodes and attributes had to be weakened, and the phases with these newly weakened preconditions had to be updated to be more robust.

The processor's parsing phase is an interesting case. Since the interactive editor must be able both to save an architecture description at any time and to load that description, the parser and unparser must be able to accommodate incompleteness. In particular, in order to save an incomplete design, the unparser must be able to record a potentially incomplete AST in a textual form; for the editor to load that definition back in, the parser must produce an incomplete AST from the text the unparser produced. What makes the parsing phase unique is that it is implemented not purely as source code, but as a grammar (Yacc script) annotated with code fragments.

Two steps were needed to weaken the parsing phase's completeness assumptions. First, the grammar was updated to allow some constructs to be missing. For example, a place in the grammar that requires a list of one or more elements was changed to require a list of zero or more elements. The code to check whether an instance of the construct in fact does contain at least one element was added to a later semantic analysis phase.

We could have continued modifying the grammar until it allowed all the intended kinds of incompleteness. However, certain language constructs were deemed too important to allow their absence from an architecture description; without these constructs, the description would simply be too hard to read and understand. With respect to these constructs, we left the grammar unmodified. In the editor, if the user saves an architecture description that is missing an instance of one of these constructs, the unparser instead writes a "dummy" version of the construct. (This use of dummy constructs is rather like inserting stub procedures into source code for testing purposes. The source code language insists that for every procedure call there be a corresponding procedure definition. In the case where no "real" procedure definition yet exists, a "dummy" version is temporary inserted.) The parser recognizes these dummy constructs as stand-ins for missing information and produces the AST accordingly. Architecture descriptions containing these dummy constructs are readable, and their incompleteness is readily visible.

Computational Organization

A batch system often makes strong assumptions about both the completeness and range of input it will accept. A typical batch compiler, for example, handles a single complete source code file per invocation. Because of this, a batch systems' computations can be expressed in monolithic pieces. An interactive system, on the other hand, not only handles incrementally created (and therefore sometimes incomplete) work products, but also typically allows more than one work product to be edited at once. The computations underlying the interactive system therefore need to be expressed as finer grained pieces, and those pieces need to handle multiple data.

Granularity of invocable computations

Even if the steps of the batch computation can be identified and their assumptions made explicit, further modifications may be required to support incremental updates and processing of partial inputs. As noted above, an interactive system is often expected to handle partial definitions gracefully—perhaps not performing complete computations, but at least providing display, feedback, and localized checking. In order to do this, the batch computation must be organized so that individual steps can be invoked individually.

The UniCon processor's semantic analysis phases were written as two monolithic computations: one for checking local properties and one for checking global ones. Because some of the checks that were embedded in these monolithic computations were to be called interactively from the editor, the individual checks needed to be excised and made available as individual computations. In particular, each of these checks became a separate operation in our abstract data types.

The parsing phase is again an interesting case. We used the Lex and Yacc standard tools to produce our parsing phase. These tools are designed for producing a monolithic parser: a single parsing routine that parses languages from a single grammar. As part of its design, the user interface needed not only a parser and unparser for entire architecture descriptions, but also a parser and unparser for instances of one particular UniCon language construct. In a better world, we would have broken up the monolithic parsing phase into separately callable routines, as we did with the semantic analysis phases. Instead, to suit the tools, we duplicated part of the grammar as a separate Yacc script and produced a separate "mini" parser.

Multiple Concurrent Computations

In a batch system, the input to the system is typically a single coherent set of data to be processed. In an interactive system, multiple, concurrent activities or work products are often concurrently maintained in the system. In such a system, the state for each work product must be kept separately. Global data in such a system must be designed carefully to avoid interference among independent computations. Code should be re-entrant wherever multiple independent computations might invoke the same code paths in such a way that state assumptions might be at risk. Synchronization may also be required in some cases.

The scope of computation in the UniCon batch processor is limited to the set of input language definitions, which usually describes a single system (or a small, finite set of systems). The processor builds a single AST from this set of input definitions. Since each phase of the processor operates on this same tree, the root of the tree was made into a global variable so that it would be accessible to every phase without having to pass it around as a function parameter.

An architecture editor that allows only one system to be edited at a time would be limited and clumsy. Hence adding an interactive interface to the batch processor imposed a new requirement that the interactive system build and maintain a set of ASTs (one for each active editing session) rather than a single AST. The global tree root in the batch processor was changed to become a local variable in the main program, and every function in the processor that accessed the tree root directly was changed to take a tree root as a function parameter. This allowed the interactive interface to parse, unparse, and process any tree in its set of trees by simply passing it to the processor routines as a parameter.

Error Handling

Three aspects of error handling in a batch system are affected by a conversion to support interactive use: error detection, error reporting, and system control. In batch systems, error handling consists solely of error detection while interactive systems can also prevent some errors during the creation of the work product. Error reporting in a batch system consists of outputting error messages upon detection of errors. Interactive systems have different error reporting requirements, especially if the interface is graphical in nature. Lastly, batch systems may choose to terminate when a serious error is encountered in the input. In interactive systems termination is always controlled by the user. The interactive system must recover from errors gracefully and return control to the user.

Error Prevention versus Error Detection

The batch approach to error handling is essentially different from the interactive approach. A batch system can easily analyze its input with the expectation that it will be complete and then report any discrepancies. Batch-style error handling, therefore, is error detection and reporting. Errors are reported in a log style, typically as text, and the location of an error in the input must be determined from the text of the message and any other information such as line numbers.

Interactive systems can both prevent and detect errors. For example, a command-driven interface can prompt the user to enter commands and detect when the user has entered an invalid command. This interface could be changed to a menu of commands, thereby preventing the user from entering an invalid one. Error prevention tends to be tied to those operations that allow incremental creation and updating of the work product. The system monitors the user's input during these tasks and checks for correctness. If the input is incorrect, feedback can be given and the input rejected. Error detection may still be appropriate in an interactive system for those operations that have batch characteristics such as those that process or analyze the work product as a whole.

The original UniCon tool analyzed the input for correctness and completeness before it attempted to build the system from the description. It reported any incompleteness or incorrectness that it detected during this analysis as a log of error messages.

Since the interactive UniCon tool supports creation and editing of architecture descriptions, it has the opportunity to prevent some errors. For example, since the editor allows an architectural description to be edited as high-level pictures, many kinds of syntactic errors are not possible. This is the graphical equivalent of the way syntax-directed editors insert language constructs all in once piece. Another example of error prevention occurs when the end user interactively makes a connection. The user interface provides immediate feedback about the legality of that connection and will not allow an illegal connection to be made. The user interface prevents as many errors as it can while the user creates and edits language constructs. The capability is limited to checks that depend on attributes that are guaranteed to be in place. When the user later builds the system as a whole, any remaining semantic checks are performed and errors detected and reported.

Hence, the editor's error handling is a mixture of error prevention and error detection. As part of designing the editor's user interface, the semantic checks in the batch processor were categorized into those to be performed immediately and those to be deferred until the user builds the system. This decision hinged on the appropriateness of making a particular semantic check interactive as well as on whether it could be performed at an interactive speed. Several semantic checks in the batch processor were redesigned and/or had their performance tuned as a result.

Error Reporting

When interactive interfaces are graphical rather than textual, batch error handling must be redesigned to support reporting of semantic errors in a form consistent with the user interface—that is, graphically. In a batch system, an error report usually consists of an error message containing the location of the error (line number), severity indicator (error, warning, or informational), and informative text. A graphical interface, however, has different information requirements for errors. First, line numbers do not adequately identify the locations of the errors. Additionally, informative text, though applicable, is not adequate since it may be difficult to convert the error information to a meaningful graphic form.

The interactive interface must be able to determine the context in which the error occurred in order to graphically display the error information and location to the user. In a batch system, however, this information is embedded in the text of the error message. To determine the context of an error, the interactive interface may require access to the portion of the internal data structures in which the error occurred (and possibly the surrounding portions as well). From this context, it can determine where the error occurred and indicate this in the interface graphically. For example, the interface might highlight the graphical representation of the construct in which the error occurred with color and then display the informative text along with it.

To support these requirements for error reporting in the Unicon editor, the batch processor was modified to bundle contextual information with the original error message for each error. The extra contextual information consists of pointers to nodes in the AST. The interactive editor uses this contextual information to find the erroneous part of the architecture description under construction and to annotation it with an error indicator that includes the error message.

Control Issues

In a batch system, the flow of control through the system is determined by the computation and proceeds based on the makeup of the input. On encountering an error, the batch system can choose between continuing the computation and terminating the job. In an interactive system, on the other hand, the flow of control through the system is most determined by the user's selection of operations. Termination of the interactive session is typically the user's choice and not the system's. Further, many interactive systems must support multiple simultaneous work products. An error in a single work product must not affect the others.

These different expectations about the locus of control can create conflicts when a batch system is converted to support interactive use. The code in the batch system may have to be redesigned to relinquish control to users and to recover gracefully from errors that would otherwise cause the system to terminate. Furthermore, it must restore legal state.

The UniCon batch processor was designed to terminate when it determined that errors in the input or in the run-time environment were too severe to continue. Rather than return control to the main program on error detection, the processor was designed to terminate immediately after reporting the error.

To support interactive use, when an error is detected, the code was redesigned to report it and then return an error status to the main program. This type of graceful recovery allows the user to address the error and continue the editing session. The redesign essentially gave control of the system to the interactive interface, and ultimately the user.

ADVICE TO SOFTWARE ENGINEERS

We summarize our lessons on adapting batch systems to accommodate interaction in the form of a checklist. This checklist is addressed to software developers or maintainers and is intended to be used when they recognize that they are confronted with a batch-to-interactive conversion problem.

Advice on External Behavior to the System

1. *Choose a style of interaction that fits your problem.*
   Here we consider one of the richer cases, in which you add a graphical editor for manipulating the work product. Before beginning your conversion, consider the nature of interaction required and consider other alternatives. Lane's thesis provides additional guidance [3].
2. *Add new operations if necessary.*
   Determine whether incremental and partial processing or the interactive style require additional operations. If so, add them in a form consistent with pre-existing operations.
3. *Replace as many error reports as possible with checks that ward off the error. Check error reports to be sure they will be meaningful in an interactive setting.*
   Find and review the error reports that the batch system can generate. For each one, determine whether the user interface can be designed to prevent it. This is most likely to be possible when the error is localized to an input item. If the error can be prevented, do so. For each remaining error report, examine the information included in the report to determine whether it will be meaningful in an interactive context; the most common problem will be the identification of the problematical piece of the work product. Restate these error reports as necessary. If new operations require new error reports, handle them the same way.

Advice on Internal Structure of the System

4. *Decide on the appropriate relation between the user interface and the former batch program.*
   Examine the tightness of the required coupling, the timing requirements and frequency of interaction between the two components, and the available mechanisms for interaction.
5. *Restructure the batch system as necessary to reduce the granularity of operations.*
   See where you need to perform operations on partial input and where you need to incrementally re-process the work product after the user modifies it. Identify the assumptions made by the code for these operations and restructure the software to allow these operations to be called individually. This will usually increase the number of procedures or other invocable parts and hence reduce the granularity of the batch system.
6. *Identify and isolate system state.*
   Global variables are not your friends. Find all the data structures that represent the state of the system and make sure they will survive reentrant use. Usually this will involve instantiating all the data for each work product. It may not be necessary for local variables of atomic operations. This step is particularly important if you have chosen a shared-data architecture.

Advice on the Code

7. *Make the assumptions of the operations explicit.*
   Identify the invariants of data structures and the ordering and context assumptions of procedures. Make them explicit so the next person will not have to work so hard. Unfortunately, this is hard to do and we do not have a recipe for how to do it.
8. *Be sure you do complete storage management.*
   Free intermediate data structures and chase down all memory leaks.

ACKNOWLEDGEMENTS

We would like to thank our colleagues in the Composable Systems research group for on-going discussions on UniCon and its tool set. Thanks to Bob Monroe for giving us feedback on the paper on short notice and to Daniel Jackson for perceptive questions. Geoff Clemm has promptly helped us with Odin on numerous occasions, for which we are grateful. This work has been supported by the Wright Laboratory, Aeronautical Systems Center, Air Force Materiel Command, USAF, and the Advanced Research Projects Agency, under grant F33615-93-1-1330 and by a grant from Siemens Corporation. This paper represents the views of the authors and not of Carnegie Mellon University or any of the sponsoring institutions.

REFERENCES

1. A.V. Aho, R. Sethi, and J. D. Ullman. "Compilers: Principles, Techniques, and Tools." Addison-Wesley Publishing Company, Reading, Massachusetts, 1986.
2. V. Ambriola and D. Notkin. "Reasoning about interactive systems," *IEEE Trans. on Software Eng*. 14(2), Feb. 1988.
3. A.N. Habermann and D. Notkin. "Gandalf: Software development environments," *IEEE Trans. on Software Eng*. 12(12): 1117-27, Dec. 1986.
4. S.C. Johnson. "Yacc - Yet Another Compiler-Compiler," Comp. Sci. Tech. Rep. No. 32, Bell Laboratories, July 1975.
5. T. Lane, "User Interface Software Structures," Report CMU/SEI-90-SR-13, May 1990.
6. M.E. Lesk. "Lex - A Lexical Analyzer Generator," Comp. Sci. Tech. Rep. No. 39, Bell Laboratories, Oct. 1975.
7. D.E. Mularz. "Pattern-based integration architecture," In *Languages of Program Design*, Addison-Wesley, 1995, pp. 441-452.
8. S.P. Reiss. "Connecting tools using message passing in the Field Environment," *IEEE Software*, 7(4):57-66, July 1990.
9. M. Shaw, R. DeLine, D. V. Klein, T. L. Ross, D. M. Young, G. Zelesnik. "Abstractions for software architecture and tools to support them," *IEEE Trans. on Software Eng*. 21(4), Apr. 1994.
10. R.W. Sheifler and J. Gettys. "The X window system," *ACM Trans. on Graphics*, 5(2):79-109, Apr. 1986.
11. R. Taylor and G. Johnson. "Separation of concerns in the Chiron-1 user interface development and management system," *Proc. INTERCHI '93*, 1993.
12. T. Teitelbaum and T. Reps. "The Cornell Program Synthesizer: A syntax-directed programming environment," *Communications of the ACM* 24(9): 563-573, Sept. 1981.