Abbreviated Path Expressions With Iterated Wild Cards: WYSIWYG Semantics and Efficient Recognition

Ahmed Abdelmeged  Karl Lieberherr
College of Computer & Information Science Northeastern University
Boston, MA 02115
{mohsen,lieber}@ccs.neu.edu

ABSTRACT
Abbreviating paths with iterated wild cards is an abstraction mechanism common to Adaptive Programming (AP), eXtensible Markup Language (XML) document processing, and Aspect Oriented Programming (AOP). Recognition of abbreviated paths is used to navigate in both AP and XML, and to decide upon advice execution in AOP. Finite state automata have been used for efficient recognition of abbreviated paths. In this paper, we introduce cover automata for abbreviated path recognition. Cover automata have significantly lower state complexity than automata used in previous approaches. One contribution of this paper is an algorithm for constructing a cover automaton for abbreviated path recognition. We also prove the correctness of our algorithm. A second contribution of this paper is a succinct formal semantics for abbreviated paths based on regular language theory which has greatly simplified our proofs.

1. INTRODUCTION
Path expressions are used to specify a set of paths pertaining to some task at hand. Path expressions are common to Object Oriented Programming (OOP), Adaptive Programming (AP), eXtensible Markup Language (XML) document processing, and Aspect Oriented Programming (AOP). In OOP, path expressions are used to retrieve information from object graphs. In AP, strategies are a form of path expression used to navigate object graphs. In XML document processing, XPath expressions are used to specify elements in XML documents for both retrieval and update. Finally, in AOP, pointcut designators are used to select a set of join points during the course of program execution.

Abbreviating path expressions is an abstraction mechanism that allows developers to write generic programs by abstracting over irrelevant structural details. Abbreviated path expression formalisms fall into two categories: explicit and implicit. Explicit formalisms provide developers with wild cards to replace the abbreviated path components. Wild cards can also be iterated to replace multiple consecutive abbreviated path components. Examples of explicit formalisms is XPath where developers can use “/” as form of iterated wild card and the AspectJ point cut language that provides the “cflow” construct as a form of iterated wild cards. The “.*” in regular expressions is a third example of iterated wild cards.

Implicit formalisms do not provide developers with wild cards. Instead, all paths are translated into an explicit form by inserting wild cards into certain places defined by what we call an expansion semantics. Implicit formalisms save the developers from writing too many wild cards to make their methods flexible. An example of implicit formalisms is the regular-expression-like strategy language in AP.

Naive recognition of path expressions with iterated wild cards is inefficient. For example, retrieval of nodes specified by the XPath expression “//para” involves a traversal of the entire XML tree and might visit a large number of nodes that can never lead to a “para” element. Schema information can be used to optimize the execution of path expressions with wild cards so that it visits nodes from which a “para" element is reachable. Several constructions [8, 5, 12] exist that are based on the idea of intersecting two automata one representing paths in the query and another representing paths in the schema. The number of states in the intersection automaton is proportional to the product of number of states in both automata [14].

Observing the fact that at runtime the intersection automaton is used to check paths in a document that conform to the schema. A cover automaton for the intersection can be used. A cover automaton can have more paths than the intersection automaton as long as these extra paths are illegal according to the schema. The benefit of using a cover automaton is that it can have a significantly lower state complexity than intersection automaton and therefore it can improve the overall runtime performance. One contribution of this paper is an algorithm for constructing a cover automaton for abbreviated path recognition. We also prove the correctness of our algorithm.

In AP, the classical interpretation of wild cards as place holders for “anything” leads to modularity and ambiguity problems. In this paper, we argue that a slightly restricted interpretation of wild cards, called WYSIWYG, can solve both problems. Furthermore, it can improve the efficiency of recognizing abbreviated paths with wild cards. We give a succinct formal semantics for abbreviated paths based on regular language theory which has greatly simplified our proofs.

The rest of this paper is organized as follows: In section

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...$10.00.
2, we introduce our notation. In section 3, we give a succinct formal semantics for abbreviated paths according to the WYSIWYG interpretation of wild cards. In section 4, we show that it is possible to improve the efficiency of path recognition. In section 5, we give a construction of an efficient recognizer for abbreviated paths with iterated wild cards. In section 6, we discuss some of the related work. Section 7 concludes this paper. It is worth mentioning that sections 3 and 4 are independent of each other and can be read separately.

2. NOTATION

We use
- uppercase Greek letters to denote alphabets (e.g. $\Sigma_C$),
- lowercase Greek letters to denote strings,
- uppercase Latin letters to denote regular languages (e.g. $C$),
- lowercase Latin letters to denote symbols (e.g. $a$) as well as functions (e.g. $meta$),
- $\Sigma^*$ to denote the free monoid on an alphabet $\Sigma$,
- $f^*$ to denote a homomorphism $f^* : \Sigma_A^* \rightarrow \Sigma_B^*$ constructed by extending the function $f : \Sigma_A \rightarrow \Sigma_B$ the usual way,
- $C$ to denote the state complexity of a regular language $C$. The state complexity of a regular language is the number of states in the minimal deterministic finite automaton that accepts it,
- $R^c$ to denote the prefix closure of a regular language $R$. Formally, $R^c = \{\omega | \exists \sigma \in R : \omega \subseteq \sigma\}$,
- $P(S)$ to denote the power set of some set $S$.

3. STATIC SEMANTICS OF ABBREVIATED PATHS WITH ITERATED WILDCARDS

Given:
- A regular set $C \subseteq \Sigma_C^*$ of concrete paths.
- A set of abbreviated paths $A \subseteq (\Sigma_A \cup \diamond)^*$ where $\Sigma_A \subseteq \Sigma_C$ and $\diamond \notin \Sigma_C$ is a distinguished wild card symbol.
- All occurrences of $\diamond$ are iterated (i.e. $\diamond$ only shows under the Kleene star). Formally, $\forall \alpha, \beta \in (\Sigma_A \cup \diamond)^* : \alpha \cdot \diamond \cdot \beta \in A \Rightarrow \alpha \cdot \diamond^* \cdot \beta \subseteq A$.

The meaning of a set of abbreviated paths $A$ with respect to a set of concrete paths $C$ according to the WYSIWYG semantics, denoted $\text{WWG}(A, C)$, is a set containing all paths in $C$ that are obtainable from some path in $A$ by replacing all wild cards in it with symbols from $\Sigma_C \setminus \Sigma_A$. Given a concrete path $\omega$, there can be at most one corresponding abbreviated path $\alpha$. Furthermore, $\alpha$ can be obtained from $\omega$ by replacing all occurrences of symbols not in $\Sigma_A$ in $\omega$ by wild cards. This observation enables us to have the following succinct formal definition for $\text{WWG}(A, C)$:

$$\text{WWG}(A, C) = C \cap \{\alpha \in \Sigma_C^* | f^*(\alpha) \in A\}.$$  

Where:

$$f(\alpha) = \begin{cases} a, & a \in \Sigma_A, \\ \diamond, & \text{otherwise}. \end{cases}$$

Throughout the rest of this section we shall contrast $\text{WWG}(A, C)$ to the classic interpretation $\text{CLASSIC}(C, A)$ of abbreviated paths in which wild cards can be replaced by symbols from $\Sigma_C$.

3.1 Modularity

The purpose of using abbreviated paths $A$ in module $M$ to refer to some subset of concrete paths $C$ in some structure defined in another module $N$ is to lower the coupling between $M$ and $N$.

In the context of AP, code is attached to nodes in paths defined in $A$. Therefore, the order and frequency in which these nodes occur those concrete paths selected by paths in $A$ affects the overall behavior of the program. Ideally, it is desirable to guarantee that the order and frequency these nodes appear in the set of selected paths is the same as the strategy.

For example, consider the “Bus Route Class Graph” shown in Figure 1. $C$ contains all paths in the graph shown in Figure 1. Suppose that we are using the abbreviated path expression $A = \text{BusRoute} \cdot \diamond^* \cdot \text{Passenger}$ to select paths in the class graph. Furthermore, suppose that we have a code block $\text{BCode}$ attached to $\text{BusRoute}$ and the block $\text{PCODE}$ attached to $\text{Passenger}$. $\text{WWG}(A, C) = \text{BusRoute} \cdot \text{LoB} \cdot \text{LoB*} \cdot \text{Bus} \cdot \text{LoP} \cdot \text{LoP*} \cdot \text{Passenger}$. $\text{CLASSIC}(A, C) = \text{BusRoute} \cdot \text{LoB} \cdot \text{LoB*} \cdot \text{Bus} \cdot \text{LoP} \cdot \text{LoP*} \cdot \text{Passenger} \cdot \text{Passenger} \cdot \text{Passenger}$. All paths in $\text{WWG}(A, C)$ have exactly one $\text{BusRoute}$ followed by a one $\text{Person}$. $\text{CLASSIC}(A, C)$ contains all paths in $\text{WWG}(A, C)$ in addition to some extra paths that contain more than one $\text{BusRoute}$ followed by more than one $\text{Passenger}$. In order to exclude these paths, we need to enumerate them. This increases the coupling between $A$ and $C$ as $A$ would have to mention more irrelevant structural details than it “needs” and furthermore, as $C$ is evolved, $A$ might require updates to exclude extraneous paths.

![Figure 1: Bus Route Class Graph](image)

3.2 Ambiguity

Ambiguity is not a problem for recognition. It becomes a problem when events during the recognition process are observed. In AP, we associate behavior with paths in an abbreviated path expression. Therefore, confusion can occur when one concrete path can match more than one abbreviated path. As mentioned earlier, with the WYSIWYG semantics, there can be at most one abbreviated path corresponding to some concrete path. With the CLASSIC semantics, ambiguity can occur. For example, consider the set $A = a \cdot \diamond^* \cdot b \cdot \diamond^* \cdot d \cup a \cdot \diamond^* \cdot c \cdot \diamond^* \cdot d$, and the concrete path $\omega = a \cdot b \cdot c \cdot d$. According to the CLASSIC semantics,
ω matches both $a \cdot b \cdot c \cdot d$, and $a \cdot b \cdot c \cdot d$. According to WYSIWYG semantics, $\omega$ matches neither.

### 3.3 Efficiency

Another, important property of the set $\operatorname{WWG}(A, C)$ is that it has a state complexity of $A \ast C$. As we shall see in the following section, the state complexity of a set of selected paths is directly related to the efficiency of their recognition.

**Theorem 3.1 (efficiency).** Let $W = \operatorname{WWG}(A, C)$, $W \leq A \ast C$.

**Proof.** Let $AA = (Q, \Sigma, \delta, q_0, F)$ be the minimal DFA that recognizes $A$. From automata Theory, the DFA $\operatorname{EXP}(AA, \Sigma_C) = (Q, \Sigma_C, \gamma, q_0, F)$, where $\gamma(q_0, a) = \delta(q_0, f(a))$, recognizes $\{ \alpha \in \Sigma_C \mid f^a(\alpha) \in A \}$. Furthermore, $\operatorname{EXP}(AA, \Sigma_C)$ has the same number of states as $AA$ which is $A$. Therefore, the state complexity of the set $\{ \alpha \in \Sigma_C \mid f^a(\alpha) \in A \}$ is $A$. Therefore, the state complexity of $\operatorname{WWG}(A, C)$ is $A \ast C$ [14].

A similar construction for \textsc{Classic}(A, C) results in a nondeterministic finite automaton. Therefore, the state complexity of \textsc{Classic}(A, C) can be exponentially larger than $\operatorname{WWG}(A, C)$. An example illustrating this exponential complexity is given in [8].

### 4. Dynamic Semantics

**Given:**

- A schema modeled as a language $C \subseteq \Sigma_C^*$. We call $\Sigma_C$ the set of classes.
- A set of selected paths modeled as another regular language $S \subseteq C$. Typically, $S$ is the meaning of a set of abbreviated paths with respect to $C$.
- A set of object paths modeled as a regular language $O \subseteq \Sigma^*_O$. We call $\Sigma_O$ the set of objects. Typically, $O$ is the set of all paths that were once active during some traversal of some graph defined over the set of objects.
- A function $\text{meta}: \Sigma_O \rightarrow \Sigma_C$ that maps objects to classes. We say that $\pi \in \Sigma^*_O$ is an instance of $\text{meta}^*(\pi)$. We also define a function $\text{slice}(L, O) = \{ \pi \in O \mid \text{meta}^*(\pi) \in L \}$ to return a subset of object paths that are instances of some path in $L$.
- $O$ conforms to $C$, denoted $\text{conforms}(O, C)$. Formally, $\text{conforms}(O, C) = O = O$.

The dynamic semantics of a set of paths $S$ with respect to an object graph $O$, denoted $\text{goal}(S, O)$, is the set of object graph paths that can be legally completed to an instance of some path in $S$. Formally, $\text{goal}(S, O) = \text{slice}(S^O, O)$. At runtime, paths in $O$ are checked against a recognizer for $S^O$ after applying the $\text{meta}^*$ function to them. It is worth mentioning that the state complexity of $S^O$ is the same as the state complexity of $S$ because a DFA recognizing $S^O$ can be constructed from a DFA recognizing $S$ by turning every state leading to a final state into a final state.

### 4.1 Cover Languages

It is desirable to construct the smallest possible recognizer for $S^O$ in order to improve the overall performance at runtime. One approach to reduce its size is to construct a nondeterministic finite state recognizer. This approach was adopted in [8], and it is possible when the language $S^O$ has a minimal nondeterministic finite state recognizer that is smaller than the minimal deterministic finite state recognizer. This is the case with the classic interpretation of wild cards adopted there. However, this is not the case in general and certainly it is not the case when the WYSIWYG interpretation of wild cards is adopted.

Fortunately, there is another approach. We can rely on the fact that only legal object graph paths are going to be checked against $S^O$ at runtime, and use a recognizer for a cover language of $S^O$ that also contains some illegal paths. The state complexity of a cover language can be significantly lower than its base language. For example, if $C = S^O$, then $C = S^O \cap C$ meaning that all legal paths are selected. Since at runtime we are going to check only legal paths, a recognizer for the cover language $\Sigma_C^*$ (whose state complexity is 1) can be used.

A cover language $\text{cover}(L, C)$ of the language $L$ with respect to $C$ is formally defined as: $\text{cover}(L, C) = \{ X \subseteq \Sigma^*_C \mid X \cap C = L \}$. An automata recognizing $\text{cover}(L, C)$ is called a cover automata for $L$ with respect to $C$. We now prove that a cover automaton $\text{cover}(S, C)$ can be used for recognizing paths in $\text{goal}(S, O)$.

**Theorem 4.1.** $\forall O, C, L \subseteq C : \text{conforms}(O, C) \Rightarrow \text{slice}(\text{cover}(L, C), O) = \text{slice}(L, O)$.

**Proof.**

\[ \text{slice}(\text{cover}(L, C), O) = \{ \omega \in O \mid \text{meta}^*(\omega) \in \text{cover}(L, C) \} \]

\[ = \{ \omega \in O \mid \text{meta}^*(\omega) \in \text{cover}(L, C) \cap C \} \]

\[ = \{ \omega \in O \mid \text{meta}^*(\omega) \in L \} \]

\[ = \text{slice}(L, O). \]

### 5. Recognizing Abbreviated Path Expressions with Cover Automata

**Given:**

- A class graph modeled as a pair $C = (\Sigma_C, \Sigma_C)$ where $\Sigma_C \neq \emptyset$ is a non empty set of nodes, called classes, and $\Sigma_C \subseteq \Sigma \times \Sigma$ is a set of edges. Let $C$ be the regular language of all paths in $C$. By its definition, $C$ has the following two properties:
  - $C^C = C$
  - $\forall \alpha, \beta \in \Sigma^*_C, \exists x \in \Sigma_C : \alpha \cdot x \in C \land x \cdot \beta \in C \Rightarrow \alpha \cdot x \cdot \beta \in C$.

- A DFA $S = (Q, \Sigma^*_D \cup \circ, \delta, q_0, F)$ representing a set of abbreviated paths. We require $S$ to have the following four properties:
  - $S$ has only stuck state denoted $q_\perp \notin F$. Formally, $\forall \alpha \in \Sigma^*_D \cup \circ \cdot \delta(q_\perp, a) = q_\perp \land \forall \alpha \in Q^\prime \cdot q_\perp : \exists a \in \Sigma^*_D \cup \circ \cdot \delta(q_\perp, a) \neq q_\perp$
  - All wild card symbols appear on loops. $\forall q_\perp : \delta(q_\perp, \circ) \neq q_\perp \Rightarrow \exists \beta \in \Sigma^*_C \land \forall \alpha, \beta \in C \land \delta^*(q_\perp, f^*(a \cdot \beta)) \in F$. 


We show how to construct a automata for recognizing a cover language for WWG(\(L(S), C\)). We also prove the constructed automata correct and show that the constructed automata has the same number of states as \(S\).

**Lemma 5.1.** \(\forall q_i \in Q, a \in \Sigma_C : \eta(q_i, a) \neq q_\perp \Rightarrow \exists \beta \in \Sigma_C^* \text{ s.t. } a \cdot \beta \in C \land \delta^*(q_i, f^*(a \cdot \beta)) \in F\).

**Proof.** \(\Rightarrow\) direction:
- case \(a \in \Sigma_S\): Immediate, from the definition of \(\eta\) and the compatibility condition.
- case \(a \in \Delta_q\): Immediate, from the definition of \(\Delta_q\).
- case otherwise: from the definition of \(\eta\), \(\eta(q_i, a) = q_\perp\).

\(\Leftarrow\) direction:
- By the definition of \(\Delta_q\) and the compatibility condition, \(a \in \Sigma_S \cup \Delta_q\). Therefore, from the definition of \(\eta\), \(\eta(q_i, a) = \delta(q_i, f(a))\).
- But, \(\delta(q_i, f(a)) \neq q_\perp\), because \(\delta(q_i, f^*(a \cdot \beta)) = \delta^*(\delta(q_i, f(a)), f^*(a \cdot \beta)) \in F\) and by definition of \(q_\perp\) and \(q_\perp \notin F\) and \(\forall a \in \Sigma_C^* : \eta^*(q_i, a) = \delta^*(q_i, f^*(a))\).

**Lemma 5.2.** \(\forall q_i \in Q, a \in \Sigma_C : \eta^*(q_i, a) \neq q_\perp \Rightarrow \eta^*(q_i, a) = \delta^*(q_i, f^*(a))\).

**Proof.** Immediate, by simple induction on \(|a|\).

**Theorem 5.3 (Correctness).** \(L(RR(S, C)) \subseteq \text{cover}(WWG(L(S), C)\circ, C)\).

**Proof.** We show that:
1. \(L(RR(S, C)) \cap C \subseteq WWG(L(S), C)\circ C\).
2. \(WWG(L(S), C)\circ C \subseteq L(RR(S, C)) \cap C\).

Which reduces to:
- \(L(RR(S, C)) \cap C \subseteq (C \cap \{\omega \in \Sigma_C^* | \delta^*(q_0, f^*(\omega)) \in F\})\circ\)
- Which, by the definition of prefix closure, reduces to: \(L(RR(S, C)) \cap C \subseteq \{\alpha \in \Sigma_C^* \text{ s.t. } \exists \beta \in \Sigma_C^* \text{ s.t. } a \cdot \beta \in C \land \delta^*(q_0, f^*(\omega)) \in F\}\)
- Which further reduces to: \(\forall \alpha \in \Sigma_C^* \exists \eta^*(q_0, a) \neq q_\perp \land \alpha \in C \land \exists \beta \in \Sigma_C^* \text{ s.t. } a \cdot \beta \in C \land \delta^*(q_0, f^*(a \cdot \beta)) \in F\)
- We proceed by induction on \(|\alpha|\):
  - Base case: \(|\alpha| = 0\)
    - \(\alpha = \varepsilon\) and therefore, \(\eta^*(q_0, \varepsilon) = q_0\). By the definition of \(S\), \(\exists \beta \in \Sigma_C^* \text{ s.t. } a \cdot \beta \in C \land \delta^*(q_0, f^*(\omega)) \in F\). Therefore, \(\exists \beta \in C \land \delta^*(q_0, f^*(a \cdot \beta)) \in F\).
  - Induction step: Let \(\alpha = \omega \cdot a\)
    - Therefore, \(\eta^*(q_0, \omega \cdot a) \neq q_\perp\). But, \(\eta^*(q_0, \omega \cdot a) = \eta(q_0, a)\), where \(q_i = \eta(q_0, \omega)\). Therefore, \(q_i \neq q_\perp\), and by Lemma 5.2, \(\delta^*(q_0, f^*(\omega)) \in F\). Furthermore, \(\eta(q_0, a) \neq q_\perp\). Therefore, by Lemma 5.1, \(\exists \beta \in \Sigma_C^* \text{ s.t. } a \cdot \beta \in C \land \delta^*(q_0, f^*(a \cdot \beta)) \in F\).

But, \(\omega \cdot a \in C\), therefore, by the definition of \(C\), \(\omega \cdot a \cdot \beta = a \cdot \beta \in C\). And, \(\delta^*(q_0, f^*(a \cdot \beta)) = \delta^*(\delta^*(q_0, f^*(\omega)), f^*(a \cdot \beta)) = \delta^*(q_0, f^*(\omega \cdot a \cdot \beta)) = \delta^*(q_0, f^*(a \cdot \beta))\), therefore, \(\delta^*(q_0, f^*(a \cdot \beta)) \in F\).

2. \(WWG(L(S), C)\circ C \subseteq L(RR(S, C)) \cap C\).

Which reduces to:
- \((C \cap \{\omega \in \Sigma_C^* | \delta^*(q_0, f^*(\omega)) \in F\})\circ \subseteq L(RR(S, C)) \cap C\)
- Which, by the definition of prefix closure, reduces to: \(\{\alpha \in \Sigma_C^* \text{ s.t. } \exists \beta \in \Sigma_C^* \text{ s.t. } a \cdot \beta \in (C \cap \{\omega \in \Sigma_C^* | \delta^*(q_0, f^*(\omega)) \in F\}) \subseteq L(RR(S, C)) \cap C\)
- Which further reduces to: \(\forall \alpha \in \Sigma_C^* \exists \beta \in \Sigma_C^* \text{ s.t. } \exists \beta \in C \land \delta^*(q_0, f^*(a \cdot \beta)) \in F\)
- By induction hypothesis, \(q_i \neq q_\perp\). By Lemma 5.2, \(\delta^*(q_0, f^*(\omega)) = q_i\). But, \(\delta^*(q_0, f^*(a \cdot \beta)) = \delta^*(q_0, f^*(a \cdot \beta)) = \delta^*(\delta^*(q_0, f^*(\omega)), f^*(a \cdot \beta)) = \delta^*(q_0, f^*(a \cdot \beta))\), therefore, \(\delta^*(q_0, f^*(a \cdot \beta)) \in F\). And \(a \cdot \beta \in C\), therefore, \(a \cdot \beta \in C\). By Lemma 5.1, we can conclude \(\eta(q_0, a) \neq q_\perp\). Hence, \(\eta^*(q_0, a) \neq q_\perp\)

**Theorem 5.4 (Efficiency).** \(RR(S, C)\) has the same number of states as \(S\).

**Proof.** Immediate from the definition of \(RR(S, C)\)

Although, it is possible to search for a cover automata for WWG(L(S), A) with a smaller state complexity, \(RR(S, C)\) has another attractive feature in the context of AP. There is a one-to-one correspondence between the states of \(S\) and the states of \(RR(S, C)\). Therefore, code blocks attached to the states of \(S\) can be easily be attached to the states of \(RR(S, C)\).

6. RELATED WORK

The WYSIWYG expansion idea was introduced from a complexity perspective in [9] under the name of pure paths. WYSIWYG semantics was further explored in [13, 2] which provide a first glimpse at the treatment provided in this paper. But the proofs are long. The solution provided in this paper is clearly better and provides a nice demonstration of the power of regular languages.

Disambiguation techniques for matching a single string against a regular expression are discussed in [6, 3]. In this work, we deal with a set of strings. Mendelzon and Wood analyzed the complexity of finding real paths in graphs [10]. They showed that finding simple regular paths in a graph is NP-complete problem while finding regular paths is a polynomial time problem. Sereni and de Moor study the static determination of cflow pointcuts in AspectJ [12].

They reason also in terms of sets of paths, and they use regular expressions as pointcut language. They model pointcut designators as automata. They do whole program analysis on the call graph of the program and try to determine whether a potential join point fits into one of the following three cases: (1) it always matches a cflow pointcut; (2) it never matches a cflow pointcut; (3) it maybe matches a cflow
pointcut. In case (3), there is still a need to have dynamic matching code.

Automata are widely used to evaluate XPath Queries on XML documents [7, 4]. Some use schemata to speed up the processing [5] while others do not [4]. Simple, but very effective techniques, such as the Stream Index [7] are used to skip over irrelevant document parts. None of the papers in the XML literature use our idea of cover automata to significantly simplify the deterministic automata. Our notion of cover languages is more general than than [1] as we deal with infinite regular languages.

7. CONCLUSION

This paper brings a line of work, called the Theory of Traversals for Adaptive Programming [11, 8, 2, 13] to a natural conclusion. The paper simplifies the model to its essence which allows a very elegant derivation of provably correct algorithms for implementing basic tools useful for numerous applications, e.g., XML, AOP and AP. Correctness arguments that used to fill many pages are reduced to just a few lines.

8. REFERENCES