

Методика и информационная технология проектирования грамматик DSL на основе онтологического описания предметной области

Сагайда П. И., Тютюнник Ю. П.

Донбасская государственная машиностроительная академия, paulsagayda@ukr.net

A method and information technology of formal grammars design is developed for the domain-specific languages on the basis of ontological description of subject domain.

ВВЕДЕНИЕ

Как показал анализ, разработка и реализация предметно-ориентированных языков программирования (Domain-Specific Language, DSL) эффективны для применения в сложных предметных областях (ПО), которые удовлетворяют ряду критериев. В таких ПО для обработки данных, моделирования и прогнозирования их работы используется большое количество разнородных методов, реализуемых многорежимными алгоритмами, с различными вариантами подготовки и предобработки данных, и интерпретации результатов. При этом необходимо использовать слабо формализуемые предпочтения аналитиков по определению перечня этапов вычислительного процесса и особенностей их реализации. Выполнение такого критерия для ПО безусловно требует предоставления аналитику соответствующего инструмента в виде DSL, обоснованного с точки зрения его знаний о ПО и удобного для программного описания задачи.

Примерами таких ПО являются следующие: интеллектуальный анализ данных (Knowledge Data Discovery), в частности его подраздел Data Mining; экспертное оценивание альтернатив, в частности его подразделы «Обработка и свертка частных оценок» и «Многомерное шкалирование результатов попарного сравнения»; имитационное моделирование на основе семантических сетей, и многие другие ПО. Все эти ПО отличают большое количество существующих задач обработки данных и высокая сложность рассматриваемых организационно-технических систем. Необходима разработка решений по упрощению процесса проектирования и реализации DSL для таких ПО.

Целью данной работы является разработка методики и информационной технологии проектирования грамматик DSL на основе онтологического описания ПО.

ОСНОВНОЙ ТЕКСТ

По результатам исследований методов онтологического моделирования и дискретной математики предлагается следующая информационная технология.

На основании онтологий задач, объектов и процессов ПО [1] строятся π -сети [2], при этом полюса таких сетей генерируются с учетом концептов онтологии процессов, а ребра – с учетом концептов онтологии объектов и задач. Двухполюсники π -сети позволяют затем в автоматическом режиме получить все возможные пути между полюсами, и таким образом определить синтаксические правила контекстно-свободной (КС) грамматики следующего вида [3]:

$$G = \langle N, \Sigma, P, S \rangle, \quad (1)$$

где N – алфавит нетерминальных символов; Σ – непересекающийся с N алфавит терминальных символов; P – конечное множество правил вывода вида $A \rightarrow \alpha$, где $A \in N$, α – цепочка символов из $N \cup \Sigma$; S – выделенный символ из N , именуемый начальным символом.

КС-язык, порождаемый грамматикой G , может быть представлен следующим образом:

$$L(G) = \{x \mid S \Rightarrow_G x\}, \quad (2)$$

где запись $\alpha \Rightarrow_G \beta$ означает, что цепочка β непосредственно выводима из цепочки α в грамматике G [4].

Использование π -сетей предполагает последовательно-параллельную организацию сети, позволяющую отобразить многовариантность решения отдельных задач обработки данных в ПО, состоящих из различных этапов, однако объединяемых на основе результатов онтологического моделирования в один нетерминальный символ разрабатываемой грамматики.

Таким образом, построение грамматики G с помощью предложенной методики позволяет реализовать соответствующий язык, выразительность которого обосновывается полнотой и корректностью используемых онтологий ПО.

Следующим этапом является генерация парсера, реализующего распознавание предложений на разработанном языке (например, с помощью инструментальных средств yacc и lex), и подготовка библиотеки функций, реализующих все требуемые использованными онтологиями возможности по обработке данных, для интерпретации предложений на DSL.

Для усовершенствования предложенного подхода к разработке и реализации DSL предлагается также построить для полученной КС-грамматики автомат с магазинной памятью, и на основе логики его работы генерировать наборы взаимосвязанных диалоговых окон. Такие окна с соответствующими элементами управления и вариантами выбора позволят автоматизировать и упростить формирование предложений на DSL, корректных не только с синтаксической, но и семантической точки зрения. Соответственно, можно добиться двух дополнительных целей: выполнить интерактивную отладку предложений на DSL с функциями ограничения и завершения предложений; расширить грамматику DSL до контекстно-зависимой формы. Важность применения такой формы очевидна, например, в DSL для описания процедур интеллектуального анализа данных, в которых отдельные алгоритмы обработки данных (терминалы и нетерминалы) могут повторяться в одной и той же задаче анализа и зависеть от контекста предложения, построенного с помощью лексем DSL.

Выводы

Предложена информационная технология проектирования DSL для сложных предметных областей с большим количеством задач обработки данных и высокой сложностью рассматриваемых организационно-технических систем. Использование математического аппарата и графической нотации π -сетей позволяет обеспечить формальный переход от модели знаний о ПО разрабатываемого DSL в виде онтологии к КС-грамматике, являющейся основой языка.

ЛИТЕРАТУРА

- [1] Палагин А.В. Системная интеграция средств компьютерной техники / А.В. Палагин, Ю.С. Яковлев – Винница: «УНІВЕРСУМ-Вінниця», 2005. – 680 с.
Яблонский С.В. Введение в дискретную математику. – М.: Наука, 1986. – 384 с.

Ахо А., Ульман Дж. Теория синтаксического анализа, перевода и компиляции. – М.: Мир, 1978. – 616 с.
Соловьев С.Ю. Структура контекстно-свободных языков // Информационные процессы. – 2011. – т. 11. – №1. –
С. 161-178.
Свердлов С.З. Языки программирования и методы трансляции. – СПб.: Питер, 2007. – 638 с.

