

УДК 519.688

О ПРИМЕНЕНИИ R-ТЕХНОЛОГИИ ДЛЯ АВТОМАТИЗАЦИИ
РЕДАКЦИОННО-ИЗДАТЕЛЬСКИХ РАБОТ

Ю.Г.Косарев, А.А.Москвитин, Н.А.Чужанова

Опыт применения R-технологического комплекса для построения трансляторов [1], кодирования, контроля и коррекции текстовой информации [2-4] показал, что R-язык и R-технология обладают рядом свойств, которые позволяют надеяться на успешное применение их и для автоматизации редакционно-издательских работ.

Среди этих свойств в первую очередь можно назвать следующие:

- наглядность представления структуры данных и алгоритмов;
- удобство описания семантических свойств объектов (правил, конструкций языка и т.п.);
- конструктивность R-грамматик (программа на R-языке задает структуру входной информации, а не рассматривает ее как нечто независимое, подлежащее обработке);
- наличие библиотеки стандартных семантик для анализа и преобразования символьных последовательностей;
- наращиваемость библиотек грамматик, семантик;
- возможность использования стандартных синтермов и введения новых;
- возможность изменения управляющей программы.

Проектирование системы ПАРИС в R-технологии осуществлялось в два этапа.

На первом этапе (абстрактное проектирование) система описывалась на R-языке. В описаниях семантик использовались термины исследуемой предметной области (рис.1).

Второй этап (рабочее проектирование) - представление введенных ранее семантик через стандартные процедурные операции R-технологического комплекса (рис.2). При реализации второго этапа

Используемые стандартные семантики

Мнемоника	Ф у н к ц и и
R/USOUT(R1, ..., RN)	Определить входные регистры анализа - тора
R/ANUL(R,C)	Обнулить регистр, счетчик
R/RED(R,I)	Отредактировать регистр пробелами
R/REP(R,N,I [1, R])	Отредактировать регистр символами
R/SUBSTR(R,K,N,R1)	Переслать часть регистра
R/LENR(I,R)	Выдать текущую длину регистра в счет- чик
R/PRINTF	Печать файла на АППУ
R/OUTF	Вывод в файл
R/INF	Входная строка определена как стан- дартный файл
R/TRD(T,R,...)	Чтение из табличной памяти
R/TWR(T,R,...)	Запись в табличную память
R/TSQ(T,R,...)	Последовательный поиск в таблице
R/TSN(T,R)	Поиск поля в строке
R/USCHAR(C)	Использовать символы
WE(GCORR)	Определить имя коррективовочной грам- матики
R/MOVR(R,R1, ..., RN)	Переслать в регистр R из остальных
R/MOVRZ(R,R1, ..., RN)	Переслать в регистр R с последующим обнулением
R/COPYR(R,R1, ..., RN)	Копировать R в остальные регистры
R/COPYRZ(R,R1, ..., RN)	Копировать R с последующим обнулением
R/SETRC(R,C)	Установить текущую длину регистра
R/USIN(R,P)	Определить входную строку
R/SINTBC	Определить таблицы синтермов
R/SINTBG	Определить таблицы синтермов
R/BCRG	Сравнить регистры

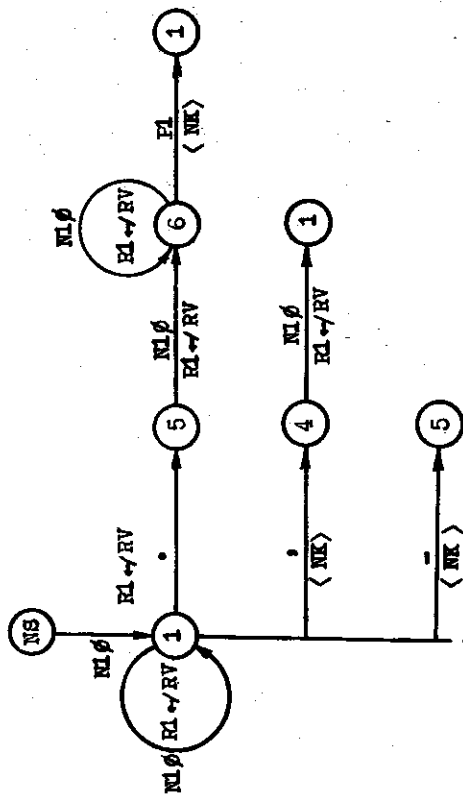
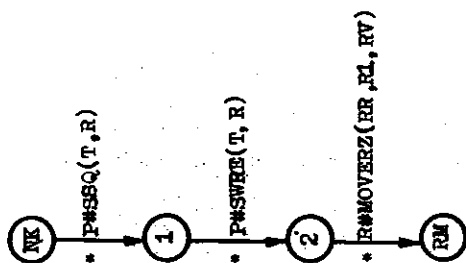
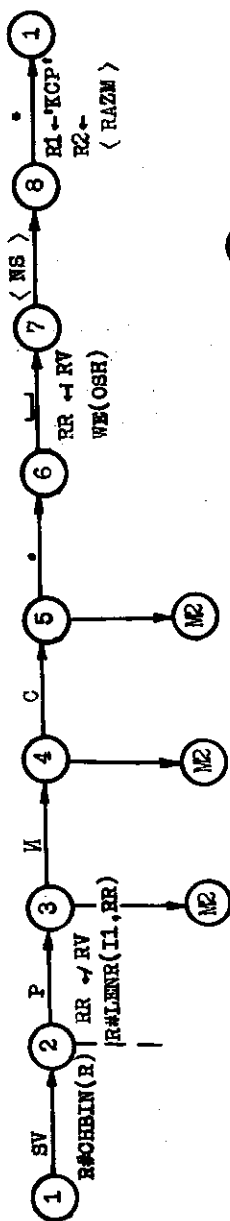


Рис.2. Пример рабочей программы обработки ссылки на рисунок.

Операции, полученные модификацией имеющихся семантик

Мнемоника	Ф у н к ц и и
R*RAZM	Разметка текста внутренними кодами
R*CHBIN	Модификация оператора "Ключ"
R*SCHK	Формирование строки по концу конструкции
R*SCH	Длина строки в типографских знаках
R*FILER	Вычисление длины последовательности в типографских знаках и символах
R*PAZ	Печать в разрядку
R*REM	Резервирование места
R*FORM	Формирование трафаретов
R*PTK	Печать текста по колонкам
R*OWTS	Определить номер корректируемой строки
R*ROT	Подчеркнуть текст
R*ADD	Вычислить адрес
R*BOT	Просмотреть последовательность символов в обратном порядке
R*STAND	Извлечь содержимое архива
R*SKKD	Сформировать стандарт сокращения
R*VST	Вставить часть текста

правильно, малоэффективно. В то же время почти все семантики третьей группы типичны не только для проектируемой системы, но и для других задач, в частности, задач прикладной лингвистики. К этим семантикам относятся (табл. 3):

Обработка списковых структур. В ходе работы системы создаются вспомогательные средства: списки ссылок на рисунки и таблицы, словари терминов, конструкций и т.п. с единым списковым механизмом доступа. Программирование при этом сводится к заданию или замене таблицы, словаря. Создание вспомогательных средств позволит контролировать работу системы, в некоторых случаях исключать повторный просмотр текста рукописи, выявлять разного рода ошибки и сообщать о них автору или редактору.

Магазинные памяти для обработки символьной информации. Эти памяти в данном проекте необходимы, главным образом, для генерации символов текста в обратном порядке.

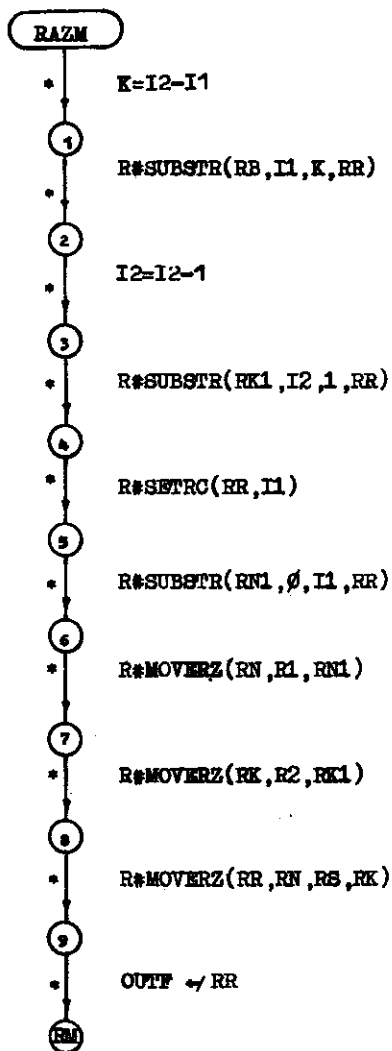


Рис.3. Разметка конструкций внутренними кодами.

Табличные па-
мяти с ассоциатив-
ным доступом. В систе-
ме используется большое число
перекодировочных и других таб-
лиц, поиск в которых удобно ор-
ганизовать путем непосредствен-
ного использования кода в каче-
стве относительного адреса.

Ассоциативный
поиск входа в R-
грамматику. В R-техно-
логическом комплексе предусмотрен
последовательный просмотр
правил из заданного комплекса.
Это неудобно для комплексов с
большим числом правил, что имеет
место в данной системе, напри-
мер, при организации словарей.
Для сокращения времени выбора
нужного правила из комплекса
удобно использовать семантику,
позволяющую по коду входного
символа определить имя очеред-
ного комплекса правил.

Адресный до-
ступ. При обращении к неко-
торым фиксированным областям па-
мяти удобно использовать адреса
(например, при настройке на
входные параметры системы).

Применение R-языка позво-
ляет удобно описывать связи меж-
ду блоками системы, жестко оп-
ределить последовательность их
подключения. Однако для систем
с динамически перестраиваемой
архитектурой, каковой является
система ПАРИС, вопрос о переда-

Вновь созданные семантики

Мнемоника	Ф у н к ц и и
R* RDA	Читать по адресу
R* WRA	Писать по адресу
R* WRIA	Формировать ячейку связи управляющего поля памяти
R* WORK	Настроить очередной блок
R* SRD	Читать строку списка
R* SRDE	Читать элемент строки списка
R* SWR	Писать строку в список
R* SWRE	Писать элемент в строку списка
R* SSQ	Найти и запомнить элемент или строку списка
R* BWRR	Писать в магазин
R* BRD	Читать из магазина без разрушения
R* BRDR	Читать из магазина с разрушением
R* PSV	Переадресовать строку списка
R* VVT	Перевести высоту включения в типографские знаки
R* SHVT	Перевести ширину включения в типографские знаки
R* FIND	Вычислить длину последовательности в пунктах

че управления той или иной программой, блоку решается на основе анализа информации в процессе работы других блоков. Например, блок анализа паспорта статьи (рис.4) в зависимости от того, имеются добавления к стандарту или нет, передает управление либо блоку добавления к стандарту, а из него в блок формирования стандарта, либо сразу блоку формирования стандарта. Наличие добавлений к списку сокращений требует передачи управления от блока формирования списка сокращений к блоку добавлений к списку сокращений.

Для оперативного управления архитектурой системы вся настройка и управляющая информация внесена в управляющий массив. Последовательность подключения блоков и частей системы определяется путем изменения конфигурации системы в управляющем массиве, где каждому блоку однозначно поставлена в соответствие ячейка связи. По окончании работы блока передача управления осуществляется по адресу, записанному в соответствующей блоку ячейке связи. Работа с ячейками связи ведется с помощью двух семантик R***WRIA** и R***WORK**.

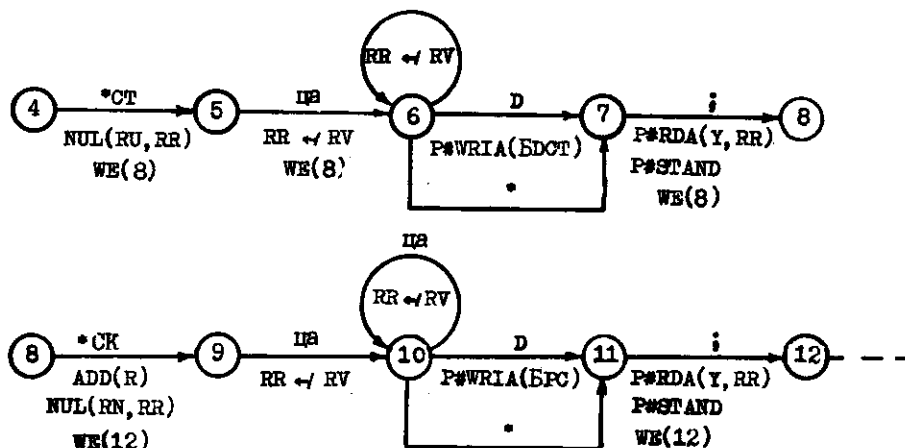


Рис.4. Фрагмент блока анализа паспорта статьи (БАПСТ).

Для случаев, когда последовательность подключения блоков жестко определена, связь между ними осуществляется по правилам, реализованным в R-технологическом комплексе для операционной системы ЕС.

В данной работе оказалось удобным применять отличную от принятой в R-технологическом комплексе форму организации синтермов. В основе этой организации лежит такое перекодирование термов, при котором принадлежность данного терма к синтерму может быть определена с помощью маски. Введены следующие синтермы:

бп - буква - цифра,	ПБ - буква прописная,
б - буква,	ГЛ - гласная,
ц - цифра,	СЛ - согласная,
па - цифра арабская,	ОС - особая (ь,ъ,й),
пр - цифра римская,	Р - разделители,
ок - буква строчная,	ВК - внутренние коды.

Эти синтермы распадаются на группы по свойству вложенности друг в друга (рис.5).

Для данной задачи важно ввести автоматическую упаковку словарей в виде R-грамматик [3,4] для анализа текстов, связанного с необходимостью поиска заданных терминов, конструкций, авторских и общепринятых сокращений и т.д.

Введенные выше дополнения позволили одновременно удовлетворить всем требованиям к разработке системы ПАРИС, а именно:

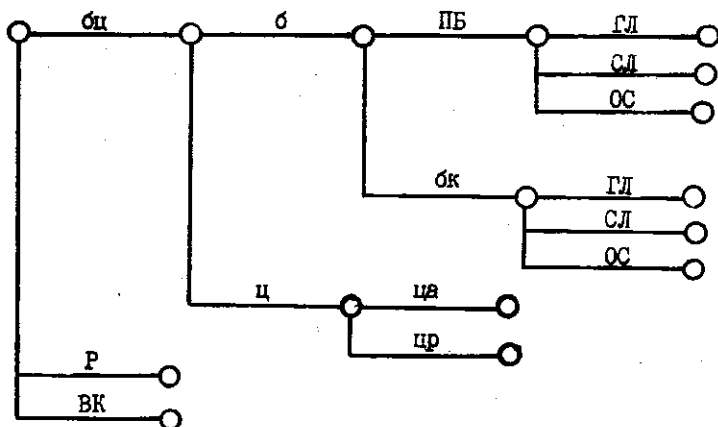


Рис. 5. Схема вложенности синтермов.

- простота настройки достигается за счет введения управляющего массива и применения средств Р-технологического комплекса для расшифровки заданий, информации пользователя и т.д.;

- наглядность структуры системы обеспечивается отделением всей управляющей информации от рабочей, а также естественностью представления структур данных и алгоритмов в Р-технологическом комплексе;

- высокое качество программирования достигается путем использования: управляющих программ, основанных на хорошо отработанных методах и написанных высококвалифицированным системным программистом; семантик, которые, как было показано выше, в большинстве случаев сводятся к стандартным; табличных памяти с ассоциативным доступом; магазинов для символической информации; ассоциативного доступа к правилам грамматики; новой организации синтермов, средств динамической перестройки архитектуры и автоматической упаковки текста в Р-грамматиках;

- простота и наглядность общения достигаются введением языка пользователя, который мало чем отличается от естественного языка; расшифровка заданий, директив, инструкций и т.п., вводимых пользователем, осуществляется средствами Р-технологического комплекса;

- простота машинной ориентации системы может быть достигнута, если входным языком будет служить Р-язык и библиотека стан-

дартных семантика R-технологического комплекса будет пополнена семантиками, описанными выше, а также предусмотрена возможность введения новой организации семантермов и управляющих программ.

Применение R-технологии с учетом указанных выше дополнений позволило за сравнительно короткое время (2,5 чел.-года) создать рабочий проект системы и приступить к ее отладке.

Л и т е р а т у р а

1. ГЛУШКОВ В.М., ВЕЛЬБИЦКИЙ И.В. Технологии программирования и проблемы ее автоматизации. - "Управляющие системы и машины", 1976, № 6, с. 75-93.
2. ХАБАРОВ В.В., КОСАРЕВ Ю.Г. Об эффективности автоматического кодирования и исправления ошибок при подготовке данных. - В кн.: Вычислительные системы. Вып. 62. Ассоциативное кодирование. Новосибирск, 1975, с. 106-118.
3. КОСАРЕВ Ю.Г., ХАБАРОВ В.В. Применение R-метаязыка для кодирования, контроля и коррекции текстовой информации. - В кн.: Теория и практика системного программирования. Киев, 1976, с. 131-139.
4. КОСАРЕВ Ю.Г., ХАБАРОВ В.В., ДУБОВСКАЯ Н.И., ШВЕДЧИКОВ С.И. Автоматизация составления средств кодирования, контроля и коррекции текстовой информации. - В кн.: Теория и практика системного программирования. Киев, 1976, с. 139-150.

Поступила в ред.-изд.отд.

6 января 1978 года