РОССИЙСКАЯ АКАДЕМИЯ НАУК
ИНСТИТУТ ПРИКЛАДНОЙ АСТРОНОМИИ

На правах рукописи

**НЕЦВЕТАЕВА**
**Галина Анатольевна**

# АВТОМАТИЗАЦИЯ ИЗДАНИЯ АСТРОНОМИЧЕСКИХ ЕЖЕГОДНИКОВ

Специальность 01.03.01. —
Астрометрия и небесная механика

**АВТОРЕФЕРАТ**
**диссертации на соискание ученой степени**
**кандидата физико-математических наук**

Санкт-Петербург
1999

**Работа выполнена в Институте прикладной астрономии Российской академии наук (г. Санкт-Петербург).**

**Научный руководитель:**
член-корреспондент РАН, доктор технических наук,
профессор С. С. Лавров

**Официальные оппоненты:**
доктор физико-математических наук, профессор В. А. Брумберг,
доктор физико-математических наук Б. К. Мартыненко

**Ведущая организация:**
Главная астрономическая обсерватория РАН

**Защита диссертации состоится 16 декабря 1999 г. в 15 часов на заседании Диссертационного совета Д 200.06.01 по присуждению учёной степени кандидата физико-математических наук при Институте прикладной астрономии РАН по адресу: 191187, г. Санкт-Петербург, наб. Кутузова, д. 10.**

С диссертацией можно ознакомиться в библиотеке ИПА РАН.

Автореферат разослан «          »                    1999 г.

Учёный секретарь
Диссертационного совета Д 200.06.01
д.ф.-м.н                    А. Т. Байкова

## Общая характеристика работы

**Актуальность темы.** Параметры астрономических объектов используются в профессиональной и любительской астрономии, а также в смежных областях науки и техники. Собранные в таблицы, они вместе с сопровождающим материалом входят в состав астрономических ежегодников, издаваемых научными организациями многих стран. Одной из важнейших задач Института прикладной астрономии РАН является подготовка регулярных табличных изданий фундаментального характера, таких, как "Астрономический ежегодник" и "Эфемериды малых планет".

"Астрономический ежегодник" используется для планирования, проведения, анализа астрономических и геодезических наблюдений, космических проектов и других исследований в области астрономии, геодинамики и геодезии. По сравнению с аналогичными изданиями других стран, он содержит наиболее полную и точную информацию о движении Солнца, Луны, больших планет, звезд, о различных астрономических явлениях.

Роль ежегодника "Эфемериды малых планет" как единственного в мире издания, содержащего наиболее полные сведения об элементах орбит и координатах занумерованных малых планет, была официально закреплена решением Генеральной Ассамблеи Международного астрономического союза (МАС) в 1951 г. В комплектации ежегодника участвуют представители различных астрономических учреждений мира.

Для решения определенных задач используются компьютерные версии ежегодников. Однако книги (печатное представление ежегодников) попрежнему остаются незаменимым, ценным пособием как для наблюдателей, так и для астрономов-исследователей, профессионалов и любителей.

Подготовка таких изданий является трудоемкой и ответственной задачей. Астрономические ежегодники должны быть готовы вовремя, представляемые в них данные должны быть актуальными и достоверными, качество полиграфического оформления — современным, технология подготовки — не обременительной.

Преобладание табличного материала, вычисляемого программами, вносит в процесс подготовки ежегодников особую специфику. Основные проблемы — это извлечение данных, их ввод и компоновка в нужном представлении, оптимальное распределение материала по страницам, согласованное полиграфическое оформление издания.

Основные технологии подготовки сложного табличного издания в наше время таковы.

1. Создание для конкретного издания программы, использующей фиксированную организацию данных и генерирующей представление оригинал-макета на языке, который интерпретируется современными принтерами (например, язык PostScript) или конвертируется в интерпретируемое представление (например, язык TeX).

2. Использование настольных издательских систем (Corel Ventura, Adobe PageMaker, QuarkXPress for Windows и т.п.), текстовых процессоров (Microsoft Word, Corel WordPerfect) или систем электронных таблиц (Microsoft Excel).

В последнем случае в рамках используемой системы осуществляется специфическое для нее программирование шаблона издания (макетов страниц и стилей) и процедур (макросов), обеспечивающих правильный ввод и распределение данных, погружение их в нужный контекст. Нужная для работы макросов организация данных обеспечивается специальными программами. Таким образом, и эта технология является программированием "под издание". При этом управление программами, вставка текстовых и графических элементов, форматирование особых элементов возлагаются на человека.

Главный недостаток указанных технологий состоит в том, что при изменении издательской среды, структуры или представления издания требуется модификация всех программ, а значит, их постоянная поддержка. При модификации программ, вычисляющих данные, могут измениться представление и структура данных. В этом случае необходима коррекция программы "под издание" или реорганизация данных в соответствии с ее требованиями. В результате задача подготовки каждого ежегодника распадается на подзадачи, каждая из которых решается своими методами (программированием того или иного рода, ручной версткой, редактированием). Такая организация подготовки изданий неэффективна и ненадежна.

Астрономические институты нуждаются в **единой** для всех ежегодников технологии, отделяющей задачу спецификации табличного издания от конкретного воплощения ее решения в виде программ на том или ином языке. Основой технологии должна быть система с входным языком, дающим астроному возможность представить свои данные в нужном для научного анализа виде, не прибегая к помощи профессионалов в области программирования и полиграфии. Оригинал-макет издания (электронная форма представления, непосредственно предшествующая типографскому воспроизводению ) должен создаваться автоматически. При наличии такой системы задача воспроизводства ежегодника сводится к модификации его описания.

С практической стороны актуальность темы определяется необходимостью снижения издержек на подготовку ежегодников и обеспечения достоверности публикации вычисляемого материала.

**Целью работы** являются:

— разработка и реализация автоматической системы подготовки астрономических табличных изданий, данные для которых вычисляются программами;

— разработка новой технологии подготовки астрономических ежегодников, не требующей профессиональных знаний в области программирования и полиграфии.

**Научная новизна работы** состоит в следующем.

1. На основе анализа современных редакционно-издательских средств и технологий, а также всего спектра проблем, связанных с подготовкой и верификацией астрономических ежегодников, выявлены и формализованы основные требования к системе автоматической подготовки табличных изданий.

2. Разработан язык декларативного типа, позволяющий создавать независимые спецификации структуры астрономического ежегодника, стилевого окружения и используемых данных. По описанию ежегодника и данных система автоматически создает оригинал-макет издания, не требующий дополнительной обработки.

3. Разработан новые методы подготовки ежегодников, позволяющие использовать характерную для астрономии терминологию и не требующие привлечения профессионалов-программистов.

4. Разработана архитектура системы автоматической подготовки табличных изданий (СВИТА).

5. Разработан механизм представления оригинал-макета астрономического ежегодника в форме, допускающей возможность создания автоматической процедуры конвертирования в современные стандартные форматы.

**Практическая ценность работы.** С помощью системы СВИТА с 1995 года издаются "Астрономический ежегодник", "Эфемериды малых планет", "Ежедневные эфемериды избранных малых планет". При подготовке не используются ручной ввод и верстка издания, не производится программная предобработка исходных данных. Система также используется при необходимости спецификаций больших таблиц в статьях и прочих изданиях.

**Апробация работы.** Основные результаты работы докладывались на семинарах Института прикладной астрономии РАН, кафедры математического обеспечения мат.-мех. факультета С.-Петербургского государственного университета и Главной астрономической обсерватории РАН.

**Публикации.** Основные результаты по теме диссертации опубликованы в пяти научных работах, перечисленных в конце автореферата.

**Структура и объем работы.** Диссертация состоит из введения, пяти глав и заключения. Она изложена на 120 страницах, содержит пять приложений. Список литературы включает 40 библиографических ссылок.

## Содержание работы

**Во введении** дана характеристика основных астрономических ежегодников, выпускаемых Институтом прикладной астрономии РАН, обрисован круг проблем, связанных с их подготовкой. Сформулированы основные требования к новой технологии производства ежегодников.

**В первой главе** детально рассмотрены проблемы, связанные с подготовкой табличного издания, возможности и недостатки существующих технологий. Обоснована необходимость и возможность создания автоматизированной системы с входным языком декларативного типа, исключающей из процесса подготовки астрономических табличных изданий дополнительное программирование, ручной ввод данных, верстку, редактирование и корректуру. Конкретизированы требования к такой системе.

При разработке системы автоматической подготовки издания важно учитывать современные тенденции развития средств описания, представления и обработки документов. В наше время электронная форма представления документа должна допускать преобразование в один из международно признанных стандартных форматов или содержать в себе формальную основу для создания автоматической конвертирующей процедуры. Это необходимое условие успешного разрешения проблемы представления ежегодников в Интернет. Поэтому при выборе способа представления порождаемых продуктов системы важно определить перспективный стандарт, на который нужно ориентироваться при создании основы для конвертирования.

Эволюция редакционно-издательской деятельности, представленная как принципиальная смена одного вида разметки документов другим, показывает постепенно возрастающую роль отделения спецификации **структуры** документа от прочих аспектов его описания. На основе знания структуры и человек, и автоматическая процедура могут осуществить форматирование документа, вывод его с помощью различных устройств и в разных средах, индексирование, контроль за правильностью, конвертирование в другие форматы, эффективный поиск, нелинейный просмотр и другие виды обработки текста.

Формализация и стандартизация средств спецификации элементов структуры, их соотношений и атрибутов лежит в основе языков обобщенной разметки документов (GML). Международным стандартом таких языков стал язык SGML (Standart Generalised Markup Language), окончательно утвержденный ISO в 1988 г. Он предоставляет средства создания специализированных GML.

Языки обобщенной разметки позволяют специфицировать документы так, что к ним могут применяться технологии для обработки формально определенных объектов.

Наиболее известным приложением SGML является язык HTML, неотъемлемая часть World Wide Web. На смену ему приходит язык XML (Extensible Markup Language), предоставляющий не только возможность описания **аспектов представления** фрагментов документа, но и более удобные, чем в SGML, средства спецификации грамматик специализированных языков. Язык XML в скором времени будет утвержден как основной международный стандарт.

Перспективность системы в смысле конвертируемости результатов ее работы определяется:

— наличием в окончательном представлении издания разметки, идентифицирующей его структуру;
— наличием способа сопоставления элементам структуры определенного полиграфического оформления;
— возможностью сопоставить модели структурных понятий системы формальную модель XML.

**Во второй главе** содержится общее описание разработанной в ИПА РАН системы СВИТА и ее входного языка.

Система СВИТА предназначена для автоматизации набора и верстки изданий, содержащих таблицы, данные для которых вычисляются внешними программами. Декларативный язык системы позволяет описать основные аспекты ежегодника: распределение материала по страницам; структуру, содержимое и способ формирования таблиц; исходные данные; стилевое оформление. По заданию на входном языке система генерирует представления издания или его части на языке TeX в рамках макропакета PlainTex.

Основная конструкция языка — модуль, состоящий из простых и составных определений объектов. Определяемый объект получает имя и может использоваться в других определениях. Последовательность определений несущественна. Задание (модуль) оформляется как текстовый файл. Параметром модуля может быть список внешних модулей. При этом модуль интерпретируется как содержащий определения из всех внешних.

Гибкость и выразительная мощность языка обеспечивается наличием структурных и семантических конструкторов.

С помощью **семантических** конструкторов, имена которых зафиксированы в лексике языка, определяются основные порождаемые и используемые объекты: издание, раздел, страница, таблица, ячейка, источник данных, элемент данных, стиль и т.д. Каждый такой объект имеет строго определенную семантику и фиксированный ряд характеристик, задаваемых явно или по умолчанию в качестве аргументов соответствующего конструктора.

**Структурные конструкторы** — мультиобъект, выбор, итератор и повторитель — предназначены для описания объектов сложной структуры. Их наличие обусловлено необходимостью спецификаций регулярных структур: разбиения страницы на боксы (прямоугольные непересекающиеся области), вертикальной и горизонтальной структуры таблицы, данных. Данные вычисляются программами и поэтому имеют регулярную структуру.

*Повторитель* предназначен для задания определенного или неопределенного числа повторений подобъектов в сложных объектах. Неопределенный повторитель обозначается “$*$”.

Конструктор *выбора* служит для описания альтернативных ситуаций. Синтаксически он обозначается квадратными скобками, в которых перечисляется список альтернатив.

*Мультиобъект* описывает структуру объекта как список компонент, синтаксически заключаемый в круглые скобки. Компонентой списка может быть любой объект: простое значение, семантический или структурный конструктор (в том числе один, несколько или неопределенное число мультиобъектов). Таким образом, мультиобъект может описывать объект сложной иерархической структуры.

Если в структуре объекта одна компонента повторяется определенное или неопределенное число раз, и в качестве разделителя повторений всегда выступает компонента другого типа, то такая ситуация специфицируется с помощью *итератора*. Итератор обозначается фигурными скобками, внутри которых указывается итерируемая и итерирующая компоненты.

Возможность рекурсии в мультиобъекте вместе с использованием итераций, повторений и выбора делает его инструментом, эквивалентным по выразительной мощности регулярным выражениям.

Структура оригинал-макета издания описывается как последовательность разделов. Разделы специфицируются с помощью семантических конструкторов *Страницы*, *Поток*, *Содержание*.

Конструктор *Страницы* определяет последовательность страниц или пар страниц (разворотов) фиксированной табличной структуры, возможно, с текстовыми фрагментами. Его аргументом является выражение (мультиобъект) над объектами типа «Спецификация структуры страницы», задающими по определенным правилам разбиение страницы на боксы. Боксы могут вкладываться друг в друга, базовые боксы содержат фрагменты текста, страничные фрагменты таблиц и рисунки. Боксу сопоставляется спецификация его содержимого: исходный материал, необходимые способы его преобразования и, явно или по умолчанию, стиль оформления.

Конструктор *Поток* используется при порождении многостраничного табличного материала сложной структуры с возможными текстовыми вставками. Его аргументом является мультиобъект над спецификациями таблиц и текстов.

Конструктор *Содержание* определяет оформление оглавления, которое строится автоматически на основе собранных в процессе генерации заголовков.

Конструктор *Таблица* определяет все аспекты представления таблицы — важнейшей компоненты ежегодника. Таблица астрономического ежегодника характеризует значения определенных параметров некоторого небесного тела (или совокупности тел) в заданной системе координат в зависимости от момента времени (в той или иной временной системе). Каждый столбец идентифицируется своим порядковым номером, ему можно сопоставить все возможные варианты представления и оформления соответствующего параметра. Типичная строка таблицы представляет последовательность (кортеж) значений элементов отдельных столбцов, определенным образом связанных друг с другом. На экране или бумаге кортеж выводится горизонтально, слева направо. Поэтому спецификация столбцов и их характеристик называется описанием горизонтальной структуры таблицы.

Мультиобъект горизонтальной структуры таблицы определяет характеристики каждого столбца и преобладающую в таблице пролинейку между столбцами.

Вертикальная структура печатного представления таблицы отличается от логического представления этой структуры в виде однородной последовательности кортежей. Типичные кортежи объединяются в группы для удобства визуального анализа или по семантическому признаку, группирование может быть иерархическим. Группы могут разделяться интервалами, линейками, заголовками, строками с дополнительной информацией другого типа (сюда можно отнести шапки, итоговую информацию по группам или страницам строк).

Вертикальная структура таблицы описывается мультиобъектом над спецификациями строк. Спецификация строки — это либо задание с помощью соответствующего семантического конструктора одной из типичных строк, горизонтальной линейки, интервала или шапки ("толстой" строки особого вида), либо явное выражение на языке TeX.

Определение типичной строки задает разбиение ее на клетки. Клетка — это одна или несколько соседних ячеек строки. Для клетки указывается источник материала (в основном элемент данных, т.е. строка, поставляемая определенной функцией над данными или их композицией), и шаблон.

*Шаблон* — это последовательность символов, определяющая одновременно и условие применимости данных, и способ их преобразования в нужное представление.

Второй важнейшей компонентой подготовки издания является описание используемых для заполнения таблиц данных, которые вычисляются внешними, не связанными с системой СВИТА программами. Следовательно, система должна допускать описание любого формата данных и обеспечивать использование данных в исходном виде, без коррекции и подгонки под систему.

Описание данных основано на следующих предпосылках:
— спецификация данных не должна зависеть от описания таблиц;
— данные представляются одним или несколькими текстовыми файлами, файлы состоят из записей, каждую запись можно представить как последовательность семантически значимых элементов (лексем), их смысл понятен пользователю и может быть обозначен соответствующими терминами;
— в файлах можно выделить несколько различных типов записей, записи одного типа образуют табличную структуру (начальные позиции лексем одни и те же);
— можно определить конечный набор спецификаторов лексем, определяющих способ их анализа.

Описание данных определяет список используемых файлов и структуру данных **по записям**. Структура по записям — это мультиобъект над конструкторами *Формат*, *Комментарий* и *Выбор*. Вложенность мультиобъектов, итераторов и повторителей не ограничена. Конструктор *Формат* описывает лексическую структуру логической записи данных в терминах системных спецификаторов лексем. Спецификаторам можно дать имена, идентифицирующие соответствующие лексемы. Конструктор *Комментарий* описывает структуру одной записи данных, трактуемой как комментарий.

Достаточный набор спецификаторов лексем и наличие структурных конструкторов предоставляют широкие возможности спецификации вычисленных программами данных — от детального соответствия структуре соответствующих таблиц до полной от нее независимости.

**Третья глава** посвящена наиболее важным особенностям системы и языка. В ней описан алгоритм интерпретации данных, особенности спецификаций таблиц, механизм элементарных вычислений, средства разбиения табличного материала на страницы, идентификация имен и неявная параметризация объектов, стилевой механизм.

Самый существенный в работе системы СВИТА аспект — это процесс интерпретации астрономических данных: их анализ, отбор, правильное представление в таблицах. Наиболее часто используемый и одновременно самый общий вариант спецификации данных, позволяющий использовать одни и те же данные для спецификации различных таблиц, представляет описание данных в файлах как неопределенной длины последовательность записей, каждая из которых имеет несколько вариантов формата. При наличии сходных по формату, но по-разному интерпретируемых записей, при сложной структуре формируемой таблицы данные описываются более сложным мультиобъектом.

На семантическом уровне алгоритм занесения данных в таблицы удобно представить в виде двух взаимодействующих процессов:

— процесса синтеза строк таблицы в соответствии с последовательностью шаблонов, определяемой описанием структуры таблицы по строкам (шаблоном строки является совокупность шаблонов ее клеток с учетов возможных вариантов);

— процесса анализа записей данных в соответствии с последовательностью форматов, определяемой описанием структуры данных по записям (формат записи — это определенная последовательность спецификаторов лексем);

Процесс анализа создает значения лексем, процесс синтеза использует эти значения. Оба процесса являются безвозвратными: очередная запись входного файла читается один раз, сгенерированные строки записываются в выходной файл один раз. Таким образом, предполагается, что фактически используемые записи данных находятся в том же относительном порядке, что и порождаемые строки таблицы. Относительный порядок означает, что среди записей данных могут быть неиспользуемые (комментарии), а среди строк таблицы не все формируются из внешних данных.

Шаблоны и форматы могут быть "жесткими" (т.е. в таблице должна быть строка заданного шаблона, в файле должна быть запись заданного формата), и могут быть "мягкими" за счет наличия в мультиобъектах альтернатив (конструкторов выбора) и неопределенных повторителей (*).

Файлы в процессе генерации таблицы иногда анализируются параллельно: значения некоторых столбцов таблицы могут определяться по данным из вспомогательного файла, ключ для которых извлекается из основного файла. В этом случае файлы логически можно представить объединенными в один посредством операции, похожей на эквисоединение реляционных таблиц: к определенным записям основного файла приклеиваются соответствующие записи из вспомогательных. На поведении алгоритма интерпретации данных это не сказывается.

Следующий пример показывает изменение спецификаций таблицы и данных в зависимости от требований оформления и усиления контроля над правильностью данных.

Пусть таблица представляет группы строк, каждая группа состоит из некоторого числа строк типа A, за которым следует некоторое число строк типа B, затем — типа C. Число строк указанного типа в каждой группе различно, число групп также зависит от исходного набора данных. В этом случае спецификация таблицы по строкам может задаваться мультиобъектом $(*[PA, PB, PC])$, где $PA$, $PB$, $PC$ — шаблоны строк, описывающие в деталях конкретное представление строк, [ ] — конструктор выбора. Если каждая группа должна окружаться особым оформлением, то мультиобъект усложняется, например, так:

$(*(<\ ...\ >, (*PA, *PB, *PC), <\ ...\ >))$,

где $<\ ...\ >$ специфицирует нужное оформление — интервалы, комментарии, прографку.

Спецификация данных может иметь вид мультиобъекта $*[FA, FB, FC]$, где $FA$, $FB$ и $FC$ — форматы записей.

Пусть семантика данных такова, что в каждой группе обязательно должна иметься по крайней мере одна запись каждого из типов A, B и C. Контроль над этим можно отразить в мультиобъекте описания данных:

$*(FA, *FA, FB, *FB, FC, *FC)$.

Детальный мультиобъект спецификации данных может включать указания всех возможных альтернатив лексем, места и состава игнорируемых записей. Тем самым можно обеспечить достаточно глубокий контроль над правильностью исходных данных.

Иногда таблица, формируемая на основе внешних данных, должна представлять результаты некоторых вычислений над ними. Например, в верхней части страницы может указываться максимальное из расположенных на странице значений некоторого столбца. Специальная программная предобработка данных не всегда возможна, так как количество строк таблицы на странице может зависеть от разнообразия типов и, соответственно, способов представления строк.

Кроме функций, поставляющих данные, в системе СВИТА предусмотрен расширяемый набор функций, предназначенных для вычисления определенных значений в процессе обработки определенной совокупности записей данных.

Мультиобъект структуры таблицы по строкам определяет не только порядок строк в таблице, но и последовательность их генерации, что делает возможным процедурное, алгоритмическое осмысление спецификации структуры по строкам. На этом основано применение аналога переменных — *накопителей.* Конструкторы перевычисления и обновления накопителей размещаются в нужных местах спецификации структуры по строкам. Место расположения очередного значения накопителя произвольно, в структуре таблицы оно может предшествовать спецификации самих вычислений. В генерируемом системой ТеХ-файле в этом месте располагается оператор чтения очередной записи из текстового файла, который формируется системой в процессе генерации таблиц и содержит последовательность результатов вычислений, т.е. значений накопителей после обработки определенного числа записей. В результате трансляции ТеХ-файла готовый для вывода DVI-файл содержит всю информацию в правильном представлении.

Существенную роль в работе системы СВИТА играет стилевой механизм, позволяющий обеспечить согласованность полиграфического оформления издания.

Стиль издания представлен набором основных характеристик оформления издания — размером страниц, полей, набором шрифтов и т.д. , а также совокупностью различных стилей текста и стилей таблиц. Стиль таблицы задает используемый шрифт, расположение таблицы внутри бокса или страницы, наличие и вид обрамления, отступы от границ столбцов, характерные интервалы и высоты строк. Пользователь может как полагаться на встроенное в систему стилевое окружение (при этом таблицам и текстам стили назначаются по умолчанию), так и экспериментировать с различными стилями при отладке своего издания. При спецификации объекта стиль или его отдельные параметры можно переопределить. Пользователь может определить собственные стили таблиц или текста, используя соответствующие конструкторы.

На основе анализа используемых стилей транслятор системы вычисляет многие численные характеристики элементов издания, которые используются при оценке размещаемости материала и отладке без реальных данных.

В целом языковые средства системы соответствуют уровню требований, выдвинутых в начале разработки проекта. Все проблемы, связанные с изменением представления исходных данных, структуры и представления таблиц, общего оформления издания переносятся на уровень корректировки спецификаций на входном языке. Для построения одной и той же таблицы издания могут быть использованы данные в различном представлении и по-разному описанные. На основе одного и того же потока данных могут быть сформированы различные по составу и представлению таблицы. Таким образом, нет необходимости в специальной программной подготовке данных “под издание”.

При одном и том же описании данных можно варьировать спецификацию таблицы: менять местами или удалять столбцы, регулировать интервалы между строками, отсеивать ненужные записи (тем самым осуществлять выборку данных).

Система допускает спецификацию таблиц, автоматически распределяемых по страницам (конструктор *Поток*). На это разбиение можно достаточно тонко влиять с помощью конструктора разрывающей строки (**eject**). Он специфицирует обычную строку. Если совокупность лексем очередной физической записи накладывается на шаблон строки, то, кроме формирования строки, происходит разрыв страницы. При этом осуществляются действия, определенные семантикой конструктора и значениями его аргументов. Таким образом, можно определить семантически обусловленный переход на новую страницу. В спецификации строки могут быть условные клетки (не отображающиеся в таблице). Их шаблон может сопоставляться не только с лексемой данных, но и со значением накопителя. Так можно определить разрыв в зависимости от вычисляемого условия.

**Четвертая глава** посвящена основным принципам реализации системы СВИТА: представлению данных, трансляции описания издания, генерации его представления на языке TeX.

Система может функционировать в одном из трех режимов отладки и в нормальном режиме генерации кода. Каждый из режимов отладки представляет собой циклический процесс. При отладке пользователь может задать требуемый уровень автоматизации размещения материала, соответствующий одному из трех следующих режимов работы:

— полный расчет; в этом случае на основе предвычисленной при трансляции информации делается полная оценка размещаемости материала в горизонтальном и вертикальном направлениях и попытка оптимизировать размещение, не используя реальные данные и TeX-транслятор;

— вертикальный расчет, при котором оценка производится только в вертикальном направлении: полный расчет невозможен без наличия реальных данных или просто нежелателен пользователю;

— прямая генерация, когда анализ размещаемости производится самим пользователем на основе просмотра страниц, порожденных генератором с использованием реальных данных.

Транслятор переводит описание издания во внутреннее представление, формирует в зависимости от стилевых установок и параметров спецификаций множество готовых фрагментов генерируемого кода, а также в зависимости от режима отладки определяет нужные для оценки размещения физические размеры объектов (с помощью транслятора TeX как вычислителя).

Генератор осуществляет:

— проверку соответствия реальных данных их описанию (форматам и шаблонам в описании издания);
— оптимизацию размещения материала на полосе издания;
— необходимую численную обработку табличных данных;
— собственно генерацию объектного кода оригинал-макета.

Полоса издания формируется из физических боксов, имеющих две характеристики: высоту и ширину бокса. Основная задача оценщика размещаемости — вычисление размеров базовых боксов для объектов типа таблица. В целом работа алгоритма размещения объектов на полосе издания заключается в следующем:

— вычисление не заданных размеров боксов, если это возможно;
— проверка размещаемости содержимого в бокс заданного размера и выдача рекомендаций пользователю по изменению параметров стилей;
— оптимизация размещения боксов на полосе издания с учетом параметров сжатия материала.

**В пятой главе** обсуждается возможность конвертирования продуктов системы СВИТА в стандартные форматы, проблема альтернативной генерации представления издания в другой среде, перспективы использования системы.

Система СВИТА порождает представление издания в такой форме, которая допускает конвертирование в стандартный формат. Результат работы системы является размеченным документом, так как представляет собой программу на языке TeX. Кроме того, генерация в системе сопровождается внесением в порождаемый код дополнительной структурной разметки (если задан соответствующий режим работы). Это обеспечивает полную анализируемость его структуры. Разметка формируется на основании семантической иерархии конструкторов с использованием механизма стилей и указаний пользователя. Для любого фрагмента текста издания (таблицы, заголовка, колонтитула, сноски, просто текста) при его описании явно или по умолчанию задается статус, т.е. место в иерархии элементов структуры издания. Этот статус является существенной смысловой частью маркирующего комментария. Статус таблицы или текста является атрибутом назначаемого явно или по умолчанию стиля. В целях обеспечения большей гибкости статус также является необязательным аргументом конструктора таблицы или текста.

Для таблицы порождаются дополнительные комментарии, специфицирующие типы и количество столбцов, структуру шапки. Иерархическое разбиение страниц на боксы также отображается в дополнительной разметке.

Само описание издания на языке СВИТА в совокупности с описанием используемых стилей можно считать спецификацией множества изданий определенного типа. Формальная определенность структуры издания на уровне спецификаций языка СВИТА обозначает возможность альтернативной генерации в терминах другой среды. Такой средой может служить язык, разработанный на основе XML.

Любая программная система, имеющая отношение к обработке табличной информации и нуждающаяся в качественном ее выводе на экран или бумагу, может использовать систему СВИТА как инструмент вывода таблиц. Конкретным примером может быть инструментальная система программирования АстроТОП, разработанная и применяемая в Институте прикладной астрономии РАН. Она предназначена для создания прикладного программного обеспечения в областях, связанных с хранением и обработкой больших массивов астрономической информации.

## Основные результаты работы

1. Формализованы основные требования к системе автоматической подготовки табличных изданий.
2. Разработана система, автоматизирующая подготовку печатных изданий астрономических ежегодников. Система имеет входной язык декларативного типа, который позволяет описать издание, исходные данные и полиграфическое оформление. По спецификации астрономического издания система автоматически генерирует представление оригинал-макета на языке TEX.
3. Разработаны механизмы трансляции описания издания и генерации результата в зависимости от требуемого уровня автоматизации.
4. Разработаны алгоритмы оценки размещаемости и оптимизации размещения печатного материала, алгоритмы автоматического построения сложных таблиц. Алгоритмы используют транслятор языка TEX в качестве вычислителя.

5. Разработаны средства представления элементов стилевого окружения, повышающих качество и уровень согласованности спецификаций табличных изданий, а также методы их использования в расчетах при автоматическом распределении материала и оценке его размещаемости.
6. Разработан механизм представления оригинал-макета астрономического ежегодника в форме, допускающей возможность создания автоматической процедуры конвертирования в международные стандартные форматы.
7. Внедрена и практически используется замкнутая технология подготовки астрономического издания, не требующая дополнительного программирования, ручной верстки и редактирования.

## Публикации по теме работы


1. Нецветаева Г. А., Новиков Ф. А., Парийская Е. Ю. Система автоматической верстки изданий таблиц (СВИТА). Сообщение Института прикладной астрономии РАН, № 118. С.-Петербург, 1998.
2. Нецветаева Г. А., Новиков Ф. А., Парийская Е. Ю. Система автоматической верстки изданий таблиц (СВИТА): аспекты реализации. Сообщение Института прикладной астрономии РАН, № 119. С.-Петербург, 1998.
3. Нецветаева Г. А. Использование языка TeX системой автоматической верстки изданий таблиц (СВИТА). Сообщение Института прикладной астрономии РАН, № 125. С.-Петербург, 1999.
4. Нецветаева Г. А. Решение проблем подготовки табличных изданий с помощью автоматизации верстки. Сообщение Института прикладной астрономии РАН, № 126. С.-Петербург, 1999.
5. Нецветаева Г. А. Система автоматической верстки табличных изданий СВИТА. Труды ИПА РАН, вып.4, С.-Петербург, 1999.

.

Оригинал-макет подготовлен с помощью системы LaTeX

Редактор
Тех. редактор

| | | | | |
|---|---|---|---|---|
| Подписано к печати 9.11.99 | Формат $60 \times 90/16$. | | Офсетная печать. | Печ.л. 1.0 |
| Уч.-изд.л. 1.0.4 | Тираж 100 | Заказ | Индекс | Цена бесплатно |

Отпечатано в типографии ИТА РАН.

Институт теоретической астрономии РАН, 191187, С.-Петербург, наб. Кутузова, 10.