

*Руководство пользователя  
IBM SPSS Modeler 16*

**IBM**

**Примечание**

Прежде чем использовать эту информацию и продукт, описанный в ней, прочтите сведения в разделе “Уведомления” на стр. 173.

**Информация о продукте**

Это издание применимо к версии 16, выпуску 0, модификации 0 IBM(r) SPSS(r) и ко всем последующим выпускам и модификациям до тех пор, пока в новых изданиях не будет указано иное.

# Содержание

## Глава 1. О программе IBM SPSS

### Modeler . . . . . 1

Продукты IBM SPSS Modeler . . . . .	1
IBM SPSS Modeler . . . . .	1
сервер IBM SPSS Modeler . . . . .	1
IBM SPSS Modeler Administration Console . . . . .	2
IBM SPSS Modeler Batch . . . . .	2
IBM SPSS Modeler Solution Publisher . . . . .	2
Адаптеры сервер IBM SPSS Modeler для IBM SPSS Collaboration and Deployment Services . . . . .	2
Выпуски IBM SPSS Modeler . . . . .	2
Документация IBM SPSS Modeler . . . . .	3
Документация SPSS Modeler Professional . . . . .	3
Документация SPSS Modeler Premium . . . . .	4
Примеры прикладных программ . . . . .	4
Папка demos . . . . .	5

## Глава 2. Новые возможности в IBM SPSS Modeler 16 . . . . . 7

## Глава 3. Обзор IBM SPSS Modeler . . . 11

Начинаем работу . . . . .	11
Запуск IBM SPSS Modeler . . . . .	11
Запуск из командной строки . . . . .	11
Соединение с сервер IBM SPSS Modeler . . . . .	12
Изменение временного каталога . . . . .	14
Запуск нескольких сеансов IBM SPSS Modeler . . . . .	14
Беглый взгляд на интерфейс IBM SPSS Modeler . . . . .	15
Холст потока IBM SPSS Modeler . . . . .	15
Палитра узлов . . . . .	16
Менеджеры IBM SPSS Modeler . . . . .	17
Проекты IBM SPSS Modeler . . . . .	19
Панель инструментов IBM SPSS Modeler . . . . .	19
Настройка панели инструментов . . . . .	20
Настройка окна IBM SPSS Modeler . . . . .	21
Изменение размера значка для потока . . . . .	22
Использование мыши в IBM SPSS Modeler . . . . .	22
Использование клавиш быстрого вызова . . . . .	22
Печать . . . . .	24
Автоматизация IBM SPSS Modeler . . . . .	24

## Глава 4. Что такое исследование данных . . . . . 25

Обзор исследования данных . . . . .	25
Оценка данных . . . . .	26
Стратегия исследования данных . . . . .	27
Модель процесса CRISP-DM . . . . .	28
Типы моделей . . . . .	29
Примеры исследования данных . . . . .	34

## Глава 5. Создание потоков. . . . . 35

Обзор создания потока . . . . .	35
Построение потоков данных . . . . .	35
Работа со узлами . . . . .	35

Работа с потоками . . . . .	41
Описания потока . . . . .	52
Выполнение потоков . . . . .	53
Работа с моделями . . . . .	54
Добавление комментариев и аннотаций к узлам и потокам . . . . .	54
Сохранение потоков данных . . . . .	59
Загрузка файлов . . . . .	61
Отображение потоков данных . . . . .	61
Советы и клавиши быстрого вызова . . . . .	63

## Глава 6. Обработка пропущенных значений . . . . . 65

Обзор пропущенных значений . . . . .	65
Обработка пропущенных значений . . . . .	66
Обработка записей с пропущенными значениями . . . . .	66
Обработка полей с пропущенными значениями . . . . .	66
Импутация, или заполнение пропущенных значений . . . . .	67
Функции CLEM для обработки пропущенных значений . . . . .	67

## Глава 7. Построение выражений CLEM . . . . . 69

О программе CLEM . . . . .	69
Примеры CLEM . . . . .	69
Типы значений и данных . . . . .	71
Выражения и условия . . . . .	71
Параметры потока, сеанса и надузла . . . . .	72
Работа со строками . . . . .	73
Обработка пустых и пропущенных значений . . . . .	73
Работа с числами . . . . .	74
Работа с датами и временем . . . . .	74
Сведение воедино нескольких полей . . . . .	75
Работа с данными множественных ответов . . . . .	76
Построитель выражений . . . . .	77
Доступ к строителю выражений . . . . .	77
Создание выражений . . . . .	77
Выбор функций . . . . .	77
Выбор полей, параметров и глобальных переменных . . . . .	78
Просмотр или выбор значений . . . . .	78
Проверка выражений CLEM . . . . .	79
Найти и заменить . . . . .	79

## Глава 8. Языковые предпочтения CLEM . . . . . 83

Справочный обзор CLEM . . . . .	83
Типы данных CLEM . . . . .	83
Целые числа . . . . .	83
Действительные числа . . . . .	84
Символы . . . . .	84
Строки . . . . .	84
Списки . . . . .	84
Поля . . . . .	85

Даты . . . . .	85
Время . . . . .	86
Операторы CLEM . . . . .	86
Справка по функциям . . . . .	88
Соглашения в описаниях функций . . . . .	89
Информационные функции . . . . .	90
Функции преобразования . . . . .	90
Функции сравнения . . . . .	91
Логические функции . . . . .	93
Числовые функции . . . . .	93
Тригонометрические функции . . . . .	95
Вероятностные функции . . . . .	95
Побитовые целочисленные операции . . . . .	96
Случайные функции . . . . .	97
Функции для работы с текстовыми значениями . . . . .	97
Функции SoundEx . . . . .	102
Функции даты и времени . . . . .	102
Функции последовательности . . . . .	106
Глобальные функции . . . . .	111
Функции для обработки пустых значений и значений Null . . . . .	112
Специальные поля . . . . .	113

## Глава 9. Использование IBM SPSS Modeler совместно с репозиторием . 115

О программе IBM SPSS Collaboration and Deployment Services Repository . . . . .	115
Сохранение и внедрение объектов репозитория . . . . .	116
Соединение с репозиторием . . . . .	117
Ввод регистрационных данных для репозитория . . . . .	117
Просмотр содержимого репозитория . . . . .	117
Запись объектов в репозиторий . . . . .	118
Задание свойств объекта . . . . .	118
Запись потоков . . . . .	120
Запись проектов . . . . .	120
Запись узлов . . . . .	121
Запись объектов вывода . . . . .	121
Запись моделей и палитр моделей . . . . .	121
Получение объектов из репозитория . . . . .	122
Выбор получаемого объекта . . . . .	122
Выбор версии объекта . . . . .	123
Поиск объектов в репозитории . . . . .	123
Изменение объектов репозитория . . . . .	124
Создание, переименование и удаление папок . . . . .	124
Блокировка и разблокировка объектов репозитория . . . . .	124
Удаление объектов репозитория . . . . .	125
Управление свойствами объектов репозитория . . . . .	125
Просмотр свойств папки . . . . .	125
Просмотр и редактирование свойств объекта . . . . .	126
Работа с метками версий объекта . . . . .	127
Внедрение потоков . . . . .	127
Опции внедрения потока . . . . .	128
Ветвь оценки . . . . .	130

## Глава 10. Экспорт во внешние прикладные программы . 133

Об экспорте во внешние прикладные программы . . . . .	133
Открытие потока в IBM SPSS Modeler Advantage . . . . .	133
Импорт и экспорт моделей в виде PMML . . . . .	134

Типы моделей, поддерживающие PMML . . . . .	134
---	-----

## Глава 11. Проекты и отчеты . 137

Введение в проекты . . . . .	137
Представление CRISP-DM . . . . .	137
Представление классов . . . . .	138
Построение проекта . . . . .	138
Создание нового проекта . . . . .	138
Добавление в проект . . . . .	138
Передача проектов в IBM SPSS Collaboration and Deployment Services Repository . . . . .	139
Задание свойств проекта . . . . .	140
Аннотирование проекта . . . . .	140
Свойства объекта . . . . .	141
Закрытие проекта . . . . .	141
Генерирование отчета . . . . .	141
Сохранение и экспорт сгенерированных отчетов . . . . .	143

## Глава 12. Настройка IBM SPSS Modeler . 145

Настройка опций IBM SPSS Modeler . . . . .	145
Настройка опций IBM SPSS Modeler . . . . .	145
Системные опции . . . . .	145
Задание каталогов по умолчанию . . . . .	146
Задание опций пользователя . . . . .	146
Настройка палитры узлов . . . . .	149
Настройка менеджера палитры . . . . .	150
Изменение представления вкладки палитры . . . . .	152
Управление узлами CEM1 . . . . .	152

## Глава 13. Вопросы производительности для потоков и узлов . 153

Порядок узлов . . . . .	153
Кэши узлов . . . . .	154
Производительность: узлы обработки . . . . .	155
Производительность: узлы моделирования . . . . .	156
Производительность: выражения CLEM . . . . .	157

## Глава 14. Средства доступности в IBM SPSS Modeler . 159

Обзор средств доступности в IBM SPSS Modeler . . . . .	159
Типы поддержки средств доступности . . . . .	159
Специальные возможности для людей с ограничениями по зрению . . . . .	159
Специальные возможности для слепых пользователей . . . . .	160
Средства доступности через клавиатуру . . . . .	160
Использование программы чтения экрана . . . . .	167
Советы по использованию . . . . .	168
Конфликты с другими программами . . . . .	169
JAWS и Java . . . . .	169
Использование диаграмм в IBM SPSS Modeler . . . . .	169

## Глава 15. Поддержка Unicode . 171

Поддержка Unicode в IBM SPSS Modeler . . . . .	171
--	-----

## Уведомления . 173

Товарные знаки . . . . .	174
--------------------------	-----

**Индекс . . . . . 177**



---

## Глава 1. О программе IBM SPSS Modeler

IBM® SPSS Modeler - это комплект инструментов исследования данных, при помощи которого можно быстро разрабатывать прогнозные модели, использующие деловые знания и опыт, и внедрять их в деловые операции для усовершенствования процесса принятия решений. Разработанный на основе модели промышленного стандарта CRISP-DM, IBM SPSS Modeler поддерживает весь процесс исследования данных, от обработки исходных данных до получения лучших деловых результатов.

IBM SPSS Modeler предлагает ряд методов моделирования, взятых из таких областей, как обучение машин, искусственный интеллект и статистика. При помощи методов, доступных на палитре Моделирование, можно извлечь новую информацию из данных и разработать прогнозные модели. У каждого из методов есть свои сильные стороны и типы задач, для решения которых он лучше всего подходит.

SPSS Modeler можно приобрести как отдельный продукт или использовать как клиент в сочетании с сервер SPSS Modeler. Кроме того, доступен ряд дополнительных возможностей, сводка которых дается в следующих разделах. Дополнительную информацию смотрите в разделе <http://www.ibm.com/software/analytics/spss/products/modeler/>.

---

### Продукты IBM SPSS Modeler

В семейство продуктов IBM SPSS Modeler и связанные с этим семейством программы входят следующие продукты:

- IBM SPSS Modeler
- сервер IBM SPSS Modeler
- IBM SPSS Modeler Administration Console
- IBM SPSS Modeler Batch
- IBM SPSS Modeler Solution Publisher
- Адаптеры сервер IBM SPSS Modeler для IBM SPSS Collaboration and Deployment Services

### IBM SPSS Modeler

SPSS Modeler - это полнофункциональная версия продукта, устанавливаемая и запускаемая на персональном компьютере. SPSS Modeler можно запустить в локальном режиме, как автономный продукт, или в распределенном режиме вместе с сервер IBM SPSS Modeler, чтобы повысить производительность на больших наборах данных.

Используя SPSS Modeler, можно быстро и интуитивно строить точные прогнозные модели, не прибегая к программированию. Используя уникальный визуальный интерфейс, можно легко визуализировать процесс анализа данных. В продукт встроены расширенные функции аналитики, при поддержке которых можно обнаруживать в данных скрытые структуры и тенденции. Можно моделировать результаты и выяснять, какие факторы на них влияют, чтобы полностью использовать деловые возможности и ограничивать риски.

SPSS Modeler доступен в двух версиях: SPSS Modeler Professional и SPSS Modeler Premium. Дополнительную информацию смотрите в разделе “Выпуски IBM SPSS Modeler” на стр. 2.

### сервер IBM SPSS Modeler

SPSS Modeler пользуется архитектурой клиент - сервер, чтобы распределять требования ресурсоемких операций по мощным серверным программам, что повышает производительность для больших наборов данных.

сервер SPSS Modeler - это отдельно лицензируемый продукт, который непрерывно работает в режиме распределенного анализа на хосте сервера совместно с одной или несколькими установками IBM SPSS Modeler. При этом сервер SPSS Modeler обеспечивает высокую производительность для больших наборов данных, поскольку ресурсоемкие операции можно выполнять на сервере без скачивания данных на компьютер клиента. Кроме того, сервер IBM SPSS Modeler обеспечивает поддержку для возможностей оптимизации SQL и моделирования в базе данных, что дает дополнительный выигрыш в производительности и автоматизации.

## **IBM SPSS Modeler Administration Console**

Modeler Administration Console - это графическая программа для управления многочисленными опциями конфигурации сервер SPSS Modeler, который также можно конфигурировать посредством файла опций. Эта прикладная программа содержит консольный пользовательский интерфейс для отслеживания и конфигурирования установок сервер SPSS Modeler installations, and is available free-of-charge сервер SPSS Modeler. Эту прикладную программу можно установить только на компьютерах Windows; однако она может управлять сервером на любой поддерживаемой платформе.

## **IBM SPSS Modeler Batch**

Хотя обычно исследование данных - интерактивный процесс, можно также запустить SPSS Modeler из командной строки, не открывая графический интерфейс. Например, у вас могут быть продолжительные или повторяющиеся задачи, которые желательно выполнить без участия пользователя. SPSS Modeler Batch - это особая версия продукта, предоставляющая поддержку всех аналитических возможностей SPSS Modeler без вызова обычного пользовательского интерфейса. Для использования SPSS Modeler Batch требуется лицензия сервер SPSS Modeler.

## **IBM SPSS Modeler Solution Publisher**

SPSS Modeler Solution Publisher - это инструмент, при помощи которого можно создать пакетную версию потока SPSS Modeler; такую версию можно запускать внешним механизмом времени выполнения или встроить во внешнюю прикладную программу. Этим способом можно публиковать и внедрять полные потоки SPSS Modeler для использования в средах, где SPSS Modeler не установлен. SPSS Modeler Solution Publisher распространяется в составе службы IBM SPSS Collaboration and Deployment Services - Scoring, для которой требуется отдельная лицензия. С этой лицензией вы получаете модуль времени выполнения SPSS Modeler Solution Publisher, при помощи которого можете запускать опубликованные потоки.

## **Адаптеры сервер IBM SPSS Modeler для IBM SPSS Collaboration and Deployment Services**

Для IBM SPSS Collaboration and Deployment Services доступен ряд адаптеров, при посредстве которых SPSS Modeler и сервер SPSS Modeler могут взаимодействовать с репозиторием IBM SPSS Collaboration and Deployment Services. При этом поток SPSS Modeler, внедренный в репозиторий, доступен для совместного использования несколькими пользователями или для обращения из прикладной программы IBM SPSS Modeler Advantage тонкого клиента. Адаптер устанавливается в той системе, в которой находится репозиторий.

---

## **Выпуски IBM SPSS Modeler**

SPSS Modeler доступен в следующих выпусках.

### **SPSS Modeler Professional**

SPSS Modeler Professional содержит все инструменты, необходимые для работы с большинством типов структурированных данных, таких как трассировка поведения и взаимодействия в системах CRM, демографии, поведения покупателей и данных о продажах.

### **SPSS Modeler Premium**



SPSS Modeler Premium - это отдельно лицензируемый продукт, расширяющий SPSS Modeler Professional для работы с такими специальными данными, как данные в аналитике объектов или социальных сетях, и с неструктурированными текстовыми данными. SPSS Modeler Premium состоит из следующих компонентов.

**IBM SPSS Modeler Entity Analytics** добавляет дополнительное измерение к прогностической аналитике IBM SPSS Modeler. Прогностическая аналитика пытается предсказать будущее поведение данных из прошлого, а объектная аналитика направлена на улучшение связности и согласованности текущих данных посредством устранения конфликтов идентичности в самих записях. Идентичность может относиться к индивидууму, организации, а также к любому другому объекту, для которого возможна неоднозначность. Разрешение идентичности может оказаться крайне необходимым для ряда полей, в том числе для управления отношениями с клиентами, обнаружения мошенничества, противодействия отмыванию денег или для национальной и международной безопасности.

**IBM SPSS Modeler Social Network Analysis** преобразует информацию о взаимосвязях в поля, характеризующие социальное поведение отдельных лиц и групп. При помощи данных, описывающих взаимосвязи, в основе которых лежат социальные сети, IBM SPSS Modeler Social Network Analysis определяет социальных лидеров, влияющих на поведение других участников сети. Кроме того, вы можете определить, какие люди наиболее подвержены влиянию других участников сети. Сочетая полученные результаты с результатами других измерений, можно создать исчерпывающие профили отдельных лиц, на которых будут основаны ваши прогнозные модели. Модели, содержащие эту социальную информацию, выполняются лучше моделей, которые ее не содержат.

**IBM SPSS Modeler Text Analytics** использует новейшие лингвистические технологии и обработку естественного языка (NLP) для быстрой обработки самых разнообразных неструктурированных текстовых данных, для извлечения и организации ключевых понятий и группирования этих понятий в категории. Извлеченные понятия и категории можно сочетать с существующими структурированными данными, такими как демографические, и применять к моделированию при помощи полного комплекта инструментов исследования данных IBM SPSS Modeler для получения более качественных и специализированных решений.

---

## Документация IBM SPSS Modeler

Документация в формате электронной справки доступна через меню Справка SPSS Modeler. Сюда входит документация по SPSS Modeler, сервер SPSS Modeler и SPSS Modeler Solution Publisher, а также Руководство по прикладным программам и другие сопроводительные материалы.

Полная документация по каждому продукту (включая указания по установке) доступна в формате PDF в подпапках \Documentation каждого продукта DVD. Документы по установке также можно скачать с веб-сайта по адресу <http://www-01.ibm.com/support/docview.wss?uid=swg27038316>.

Кроме того, документация в обоих этих форматах доступна в Информационном центре SPSS Modeler по адресу <http://publib.boulder.ibm.com/infocenter/spssmodl/v16r0m0/>.

## Документация SPSS Modeler Professional

В комплект документации SPSS Modeler Professional (включая указания по установке) входят:

- **Руководство пользователя IBM SPSS Modeler.** Общее введение в использование SPSS Modeler, в том числе о создании потоков данных, обработке пропущенных значений, построению выражений CLEM, работе с проектами и отчетами и составлению пакетов потоков для внедрения в IBM SPSS Collaboration and Deployment Services, прогнозирующие прикладные программы или IBM SPSS Modeler Advantage.
- **Узлы источников, обработки и вывода IBM SPSS Modeler.** Описания всех узлов, служащих для чтения, обработки и вывода данных в различных форматах. По существу это все узлы, кроме узлов моделирования.
- **Узлы моделирования IBM SPSS Modeler.** Описания всех узлов, служащих для создания моделей исследования данных. IBM SPSS Modeler предлагает ряд методов моделирования, взятых из таких областей, как обучение машин, искусственный интеллект и статистика.

- **Руководство по алгоритмам IBM SPSS Modeler.** Описание математических основ методов моделирования, используемых в IBM SPSS Modeler. Это руководство доступно только в формате PDF.
- **Руководство по прикладным программам IBM SPSS Modeler.** Примеры в этом руководстве служат кратким специализированным введением к тем или иным методам и технологиям моделирования. Это руководство доступно также в электронном виде в меню Справка. Дополнительную информацию смотрите в разделе “Примеры прикладных программ”.
- **Сценарии и автоматизация IBM SPSS Modeler.** Информация об автоматизации системы путем создания сценариев, включая сценарии свойств, которые могут использоваться для работы с узлами и потоками.
- **Руководство по внедрению IBM SPSS Modeler .** Информация о выполнении IBM SPSS Modeler потоков и сценариев как шагов обработки заданий под управлением IBM SPSS Collaboration and Deployment Services Deployment Manager.
- **Руководство разработчика IBM SPSS Modeler CLEF .** CLEF предоставляет возможности интеграции с программами других производителей, таких как подпрограммы обработки данных или алгоритмы моделирования, как с узлами в IBM SPSS Modeler.
- **Руководство по исследованию данных в базе данных IBM SPSS Modeler.** Информация о том, как использовать мощности вашей базы данных для повышения производительности и расширения диапазона возможностей анализа с привлечением алгоритмов от сторонних производителей.
- **Руководство администратора и руководство по производительности сервер IBM SPSS Modeler .** Информация о том, как сконфигурировать и администрировать сервер IBM SPSS Modeler.
- **Руководство пользователя по консоли администратора IBM SPSS Modeler .** Информация об установке и использовании пользовательского интерфейса консоли для мониторинга и конфигурирования сервер IBM SPSS Modeler. Консоль реализована как подключаемый модуль прикладной программы Диспетчер развертывания .
- **Руководство по CRISP-DM IBM SPSS Modeler.** Пошаговое руководство к использованию методологии CRISP-DM для исследования данных SPSS Modeler.
- **Руководство пользователя IBM SPSS Modeler Batch .** Полное руководство по использованию IBM SPSS Modeler в пакетном режиме, включая подробности выполнения в пакетном режиме и аргументы командной строки. Это руководство доступно только в формате PDF.

## Документация SPSS Modeler Premium

В комплект документации SPSS Modeler Premium (включая указания по установке) входят:

- **IBM SPSS Modeler Entity Analytics Руководство пользователя.** Информация об использовании аналитики объектов совместно с SPSS Modeler, в том числе по установке и конфигурированию репозитория, узлам аналитики объектов и задачам управления.
- **IBM SPSS Modeler Social Network Analysis Руководство пользователя.** Руководство по выполнению анализа социальной сети совместно с SPSS Modeler, включая анализ групп и анализ распространения.
- **Руководство пользователя SPSS Modeler Text Analytics .** Информация об использовании аналитики текстов совместно с SPSS Modeler, в том числе по узлам исследования текстов, интерактивной инструментальной среде, шаблонам и другим ресурсам.
- **Руководство пользователя по консоли администратора IBM SPSS Modeler Text Analytics.** Информация об установке и использовании пользовательского интерфейса консоли для мониторинга и конфигурирования сервер IBM SPSS Modeler для использования совместно с SPSS Modeler Text Analytics . Консоль реализована как подключаемый модуль прикладной программы Диспетчер развертывания .

---

## Примеры прикладных программ

Инструменты исследования данных в SPSS Modeler помогают разрешить широкий спектр деловых и организационных проблем, а примеры прикладных программ предоставляют краткие, целевые введения в конкретные методы и способы моделирования. Используемые здесь наборы данных намного меньше огромных складов данных, которыми управляют некоторые исследователи данных, но применяемые понятия и методы должны масштабироваться до реальных прикладных программ.

Обратиться к примерам можно, выбрав **Примеры прикладных программ** в меню Справка в SPSS Modeler. Файлы данных и потоки примеров устанавливаются в папке *Demos* в каталоге установки продукта. Дополнительную информацию смотрите в разделе “Папка demos”.

**Примеры моделирования баз данных.** Смотрите эти примеры в руководстве *IBM SPSS Modeler In-Database Mining Guide*.

**Примеры сценариев.** Смотрите эти примеры в руководстве *IBM SPSS Modeler Scripting and Automation Guide*.

---

## Папка demos

Файлы данных и примеры потоков, используемые с примерами прикладных программ, устанавливаются в папке *Demos* в каталоге установки продукта. К этой папке можно также обратиться из группы программ IBM SPSS Modeler в меню Пуск Windows или, щелкнув по *Demos* в списке недавно использовавшихся каталогов в диалоговом окне Открыть файл.



---

## Глава 2. Новые возможности в IBM SPSS Modeler 16

IBM SPSS Modeler добавляет следующие возможности в этом выпуске.

**Моделирование Монте-Карло.** Новый узел Источник имитации обеспечивает простой способ создания синтетических данных с нуля, используя широкий выбор статистических распределений. Другой вариант - новый узел Подгонка может автоматически построить предварительно сконфигурированный узел источника, отображающий распределение взаимосвязей между хронологическими переменными. Узел Оценка имитации - это конечный узел, предназначенный для оценки полей, получившихся из потока анализа имитации, и предоставляющий полезные диаграммы распределений и корреляций.

**Сценарии Python.** Сценарии в IBM SPSS Modeler, используемые для автоматизации процессов в пользовательском интерфейсе, теперь могут использовать язык Python; продолжается также поддержка старого режима написания сценариев. Python - это хорошо известный популярный язык написания сценариев, обеспечивающий богатый набор возможностей, в том числе разнообразный и лаконичный синтаксис, обработку ошибок и мощные дополнительные модули. **Примечание:** Вкладка **Сценарий** в меню **Инструменты > Свойства потока** теперь называется **Выполнение**.

**Выполнение в цикле и выполнение с условиями.** Это новые опции, допускающие выполнение потоков в цикле и с условиями без необходимости кодирования сценариев. Найдите эти новые опции в меню **Инструменты > Свойства потока > Выполнение** или щелкнув правой кнопкой мыши в узле и выбрав опцию выполнения **Циклическое/С условиями**.

**Узел пространственно-временных диапазонов (Space-Time-Boxes, STB).** Создайте интервалы данных положений и отметок времени для поддержки более сложного анализа. В режиме аттрактора этот узел идентифицирует также моменты времени и пространственные положения, в которых задерживаются объекты. Дополнительные функции построения выражений поддерживают выделение центроид STB, а также использование географической сетки.

**Усовершенствования Entity Analytics.** Кроме разрешения отдельных объектов теперь можно идентифицировать n-мерные взаимосвязи между объектами. Предоставляется дополнительная поддержка для изменений в реальном времени через узел потоков, декомпозиция разрешенных объектов на компоненты (Особый узел) и для анонимной формы данных, когда они передаются в репозиторий объектов. **Примечание:** преедшествующая локальная база данных SolidDB была заменена на DB2.

**Новый тип диаграмм узла оценки Receiver Operating Characteristic (ROC) и показатели площади под кривой (Area Under the Curve, AUC) и Джина для узла Анализ.** Поддерживает двоичные объекты назначения.

**Новая опция Особого узла для создания составной записи.** Позволяет задать способ агрегации для каждого группируемого поля (первое значение, последнее значение, значения конкатенации и так далее) вместо отбрасывания дублированных записей. В то время как узел Агрегат обычно используется для суммирования данных на верхний уровень, эта новая опция используется для декомпозиции дубликатов на компоненты; например, тех дубликатов, которые были обнаружены при разрешении объектов.

**Узлы Экспорт и Источник TM1.** Разрешает доступ к представлениям куба TM1 через узел источника TM1 и скоринг данных обратно на существующий куб TM1 при помощи узла Экспорт TM1.

**Выражения агрегации и агрегирующие оконные функции.** Вы можете создавать пользовательские выражения агрегации на узле Агрегат, включая встроенные агрегирующие функции (MEAN, SUM и так далее) и/или пользовательские агрегирующие функции базы данных. В этих выражениях можно получить поля, которые требуют взвешенных оконных агрегирующих функций (например, скользящих средних). Доступны встроенные и предоставляемые базой данных агрегирующие функции.

Усовершенствование анализа данных в базе данных **IBM Netezza Analytics**. Новый двухшаговый алгоритм Netezza, дополнительная опция Вспомогательной прикладной программы для управления аналитическими моделями Netezza (удаление, переименование и так далее), поддержка средств просмотра моделей для деревьев регрессии, деревья решений, K-средние и TwoStep. **Примечание:** Для новых возможностей требуется INZA 3.0.

**Новые R-узлы и построитель пользовательских диалоговых окон для R.** Кроме узла построения R-моделей и слепков моделей, введенных в SPSS Modeler 15 Fix Pack 2, в этом выпуске добавлено два новых узла - R-процесс и R-вывод. При помощи узла R-процесс можно взять данные из потока SPSS Modeler и применить преобразования к данным, используя R-сценарии. С помощью узла R-вывод можно использовать собственные пользовательские R-сценарии для выполнения анализа данных и сведения результатов скоринга модели.

Для вашего анализа можно создать текстовый и графический вывод результатов. Эти выходные данные можно направить в файл или просмотреть в браузере узла R-вывода. Построитель пользовательских диалоговых окон позволяет создавать типы узлов построения, обработки и вывода моделей, а также слепки моделей, включающие в себя средство выбора полей, текстовые, числовые и радиокнопки и управляющие элементы диалоговых подокон для включения обобщения и параметризации R-программ. До установки узла или разрешения совместного использования узла другим пользователям SPSS Modeler можно выбрать тип узла, палитру назначения и значок узла. Пользовательский построитель диалоговых окон запускается в меню Инструменты. **Примечание:** Для использования этой возможности необходимо установить SPSS Modeler - Essentials for R.

**R в базе данных.** Поддержка обратного переноса SQL для R-узлов; для Netezza, SAP Hana и Oracle используются их средства поддержки R. **Примечание:** В базах данных должны быть установлены соответствующие предоставленные поставщиками расширения R.

**Узел обработки потоковых временных рядов.** Построение и оценка моделей временных рядов за один шаг для обеспечения внедрения в реальном времени через IBM InfoSphere Streams, IBM SPSS Collaboration and Deployment Services Scoring Service или IBM SPSS Modeler Solution Publisher.

**Кнопка Предварительный просмотр.** При использовании с источником базы данных предварительный просмотр позволяет визуализировать обратный перенос SQL. При использовании совместно с источником данных Analytic Server нажмите кнопку Предварительный просмотр, чтобы получить информацию о потенциальном перемещении больших объемов данных.

**Новые опции Analytic Server на узлах автоклассификации, автонумерации и автокластеризации.** При выполнении потока для IBM SPSS Analytic Server можно выбирать между его запуском с включенным расщеплением (используйте эту опцию, если включена возможность моделей расщеплений) и опциями очень больших объемов данных (расщепления игнорируются и цели моделирования задаются для больших данных). Теперь скоринг поддерживается для моделей автоклассификации, автонумерации и автокластеризации, которые можно встроить в сервер SPSS Modeler.

**Расширенная поддержка адаптера скоринга.** Поддержка скоринга моделей анализа текста через адаптеры скоринга для базы данных. Кроме этого, новые адаптеры скоринга для баз данных DB2 LUW и Oracle.

**Узел источника представления данных.** Узел представления данных позволяет читать данные из представления аналитических данных в поток SPSS Modeler. Представление аналитических данных - это новый способ создания единого представления данных в IBM SPSS Collaboration and Deployment Services 6.

**Прочие усовершенствования.**

- Опция Teradata Query Banding в пресетах соединения с базой данных.
- Измененные опции узла GLMM.
- Поддержка зашифрованных и сжатых файлов .sav.

- Поддержка единой регистрации (single sign-on, SSO) сервер SPSS Modeler, не требующей IBM SPSS Collaboration and Deployment Services.
- Поддержка единой регистрации на уровне базы данных.
- SQL Pushback для узла Выборка для zDB2.
- Теперь R-узлы устанавливаются как часть базовой установки SPSS Modeler.
- Адаптеры средства моделирования для IBM SPSS Collaboration and Deployment Services устанавливаются через IBM Installation Manager.
- Изменение моделей расщепления.
- Поддержка локализации на русском языке.





---

## Глава 3. Обзор IBM SPSS Modeler

---

### Начинаем работу

Как прикладная программа исследования данных, IBM SPSS Modeler предлагает стратегический подход к поиску полезных взаимосвязей в больших наборах данных. В отличие от более традиционных статистических методов, здесь от вас не требуется заранее знать, что вы ищите. Вам дается возможность изучать данные, пробуя различные модели и исследуя разного рода взаимосвязи, пока вы не найдете полезную информацию.

---

### Запуск IBM SPSS Modeler

Чтобы запустить прикладную программу, выберите:

Пуск > [Все] программы > IBM SPSS Modeler 16 > IBM SPSS Modeler 16

Через несколько секунд откроется главное окно.

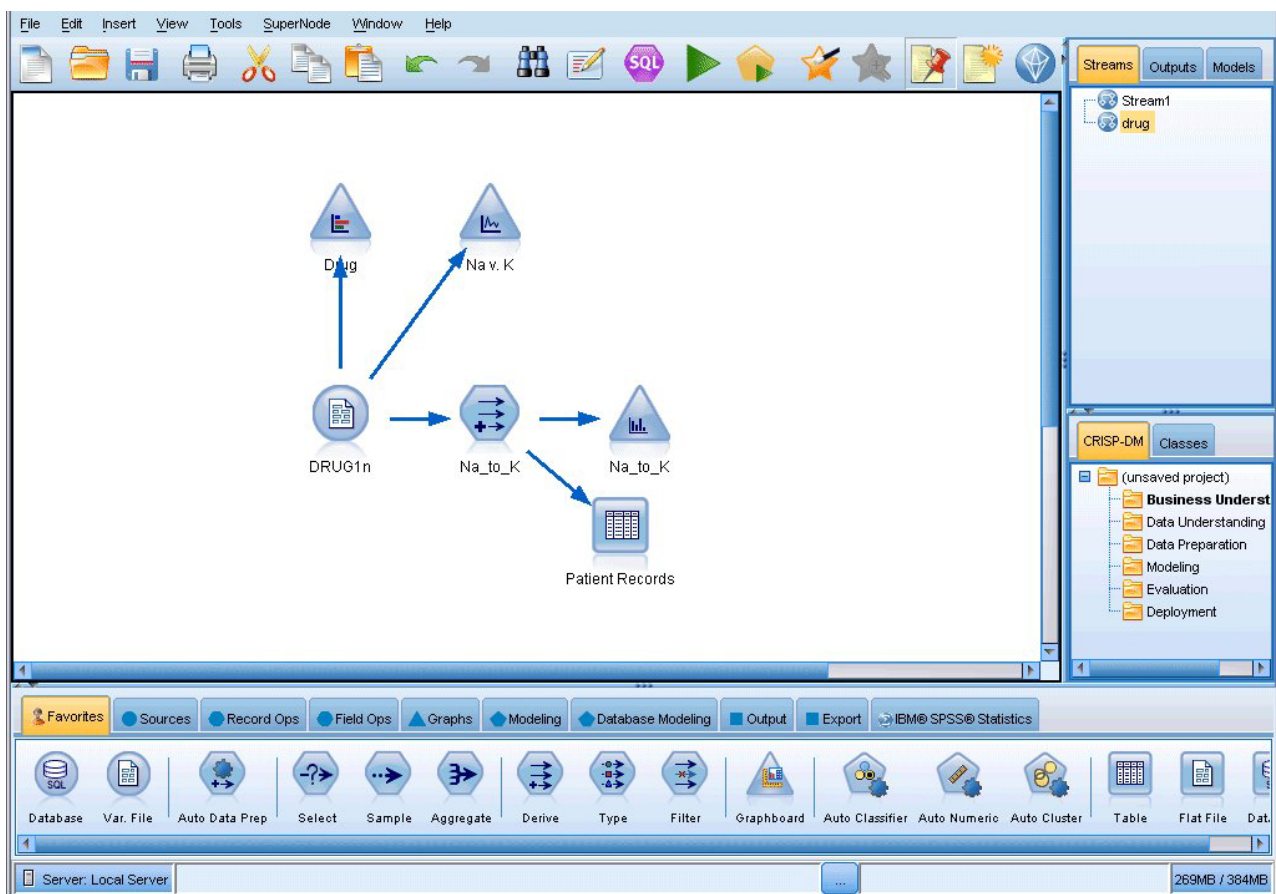


Рисунок 1. Главное окно прикладных программ IBM SPSS Modeler

### Запуск из командной строки

Для запуска IBM SPSS Modeler можно использовать командную строку операционной системы, как описано ниже.

1. Откройте окно DOS (окно командной строки) на компьютере с IBM SPSS Modeler.
2. Для запуска интерфейса IBM SPSS Modeler в интерактивном режиме введите команду `modelerclient` с нужными аргументами, например:

```
modelerclient -stream report.str -execute
```

Доступные аргументы (флаги) позволяют подключаться к серверу, загружать потоки, выполнять сценарии и указывать при необходимости прочие параметры выполнения.

## Соединение с сервер IBM SPSS Modeler

IBM SPSS Modeler можно запускать как автономную прикладную программу или как клиент, подключенный непосредственно к сервер IBM SPSS Modeler или к сервер IBM SPSS Modeler или кластеру сервера через подключаемый модуль координатора процессов в IBM SPSS Collaboration and Deployment Services. В нижней части окна IBM SPSS Modeler слева будет выведено текущее состояние соединения.

Каждый раз, когда требуется соединиться с сервером, можно вручную ввести имя сервера, к которому вы хотите подключиться, или выбрать заранее заданное вами имя. Однако если используется IBM SPSS Collaboration and Deployment Services, можно выполнить поиск в списке серверов или кластеров сервера в диалоговом окне Регистрация на сервере. Возможность просмотра через службы Statistics, работающие в сети, обеспечивается координатором процессов.

Чтобы соединиться с сервером:

1. В меню Инструменты выберите **Регистрация на сервере**. Откроется диалоговое окно Регистрация на сервере. Другой вариант - дважды щелкните по области состояния соединения в окне IBM SPSS Modeler.
2. В диалоговом окне задайте опции соединения с компьютером локального сервера или выберите соединение в таблице.
  - Нажмите кнопку **Добавить** или **Изменить**, чтобы добавить или изменить соединение. Дополнительную информацию смотрите в разделе “Добавление и изменение соединений с сервер IBM SPSS Modeler” на стр. 13.
  - Нажмите кнопку **Поиск** для доступа к серверу или кластеру сервера в координаторе процессов. Дополнительную информацию смотрите в разделе “Поиск серверов в IBM SPSS Collaboration and Deployment Services” на стр. 13.

**Таблица серверов.** Эта таблица содержит набор заданных соединений с серверами. В таблице выводятся соединение по умолчанию, имя сервера, описание и номер порта. Можно вручную добавить новое соединение, а также выбрать или найти существующее соединение. Чтобы задать конкретный сервер в качестве соединения по умолчанию, включите переключатель в столбце По умолчанию таблицы этого соединения.

**Путь к данным по умолчанию.** Задайте путь для данных на компьютере сервера. Нажмите кнопку с многоточием (...) для просмотра нужного положения.

**Задать параметры авторизации.** Оставьте этот переключатель выключенным, чтобы разрешить возможность **единой регистрации**, которая пытается зарегистрироваться на сервере, используя локальное имя пользователя и пароль на вашем компьютере. Если единая регистрация невозможна или отключена этим переключателем (например, для входа в учетную запись администратора), для ввода параметров регистрации будут доступны следующие поля.

**ID пользователя.** Введите имя пользователя для входа на сервер.

**Пароль.** Укажите пароль, связанный с указанным имени пользователя.

**Домен.** Задайте домен, который будет использоваться для входа на сервер. Доменное имя требуется, только если компьютер сервера находится в другом домене Windows по отношению к компьютеру клиента.

3. Нажмите кнопку **ОК**, чтобы завершить подключение.

Чтобы отсоединиться от сервера:

1. В меню Инструменты выберите **Регистрация на сервере**. Откроется диалоговое окно Регистрация на сервере. Другой вариант - дважды щелкните по области состояния соединения в окне IBM SPSS Modeler.
2. В диалоговом окне выберите Локальный сервер и нажмите кнопку **ОК**.

## Добавление и изменение соединений с сервер IBM SPSS Modeler

Можно вручную изменить или добавить соединение с сервером в диалоговом окне Регистрация на сервере. При нажатии кнопки **Добавить** открывается пустое диалоговое окно **Добавить/Изменить сервер**, в которое можно ввести подробности соединения с сервером. После выбора существующего соединения и нажатия кнопки **Изменить** в диалоговом окне Подключение к серверу откроется диалоговое окно **Добавить/Изменить сервер** со сведениями об этом соединении, где можно внести необходимые изменения.

*Примечание:* Соединение с сервером, добавленное из IBM SPSS Collaboration and Deployment Services, нельзя изменить, поскольку имя, номер порта и другая информация заданы в IBM SPSS Collaboration and Deployment Services.

Чтобы добавить соединения с сервером

1. В меню Инструменты выберите **Регистрация на сервере**. Откроется диалоговое окно Регистрация на сервере.
  2. В этом диалоговом окне нажмите кнопку **Добавить**. Откроется диалоговое окно **Добавить/Изменить сервер**.
  3. Введите сведения о соединении с сервером и нажмите кнопку **ОК**, чтобы сохранить это соединение и вернуться в диалоговое окно Регистрация на сервере.
- **Сервер.** Задайте доступный сервер или выберите его из списка. Компьютер сервера задается алфавитно-цифровым именем (например, *myserver*) или IP-адресом, назначенным компьютеру сервера, (например, 202.123.456.78).
  - **Порт.** Укажите номер порта, на котором сервер ожидает сообщений. Если номер по умолчанию не работает, попросите у администратора системы правильный номер порта.
  - **Описание.** Введите необязательное описание для этого соединения с сервером.
  - **Обеспечить защищенное соединение (использовать SSL).** Укажите, нужно ли использовать подключение по протоколу SSL (**Secure Sockets Layer**). SSL - обычно используемый протокол для защиты данных, пересылаемых по сети. Для доступа к этой функции необходимо разрешить SSL на сервере, где запущен сервер IBM SPSS Modeler. При необходимости уточните подробности у вашего администратора.

Чтобы изменить соединения с сервером

1. В меню Инструменты выберите **Регистрация на сервере**. Откроется диалоговое окно Регистрация на сервере.
2. В этом диалоговом окне выберите соединение, которое нужно изменить, и нажмите кнопку **Изменить**. Откроется диалоговое окно **Добавить/Изменить сервер**.
3. Измените сведения о соединении с сервером и нажмите кнопку **ОК**, чтобы сохранить это соединение и вернуться в диалоговое окно Регистрация на сервере.

## Поиск серверов в IBM SPSS Collaboration and Deployment Services

Вместо того, чтобы вводить соединение с сервером вручную, можно выбрать в сети сервер или кластер серверов через координатор процессов, доступный в IBM SPSS Collaboration and Deployment Services. Кластер серверов - это группа серверов, внутри которой координатор процессов определяет сервер, лучше всего подходящий для ответов на запрос обработки.

Хотя можно и вручную добавлять серверы в диалоговом окне Подключение к серверу, поиск имеющихся серверов позволяет соединяться с серверами без необходимости знания правильного имени сервера и номера порта. Эта информация предоставляется автоматически. Однако при этом необходима правильная информация об учетной записи, такая как имя пользователя, домен и пароль.

*Примечание:* Если у вас нет доступа к функции координатора процессов, можно вручную ввести имя сервера с которым вы хотите соединиться, или выбрать имя, заданное вами ранее. Дополнительную информацию смотрите в разделе “Добавление и изменение соединений с сервер IBM SPSS Modeler” на стр. 13.

Чтобы найти серверы и кластеры:

1. В меню Инструменты выберите **Регистрация на сервере**. Откроется диалоговое окно Регистрация на сервере.
2. В диалоговом окне нажмите кнопку **Поиск**, чтобы открыть диалоговое окно Поиск серверов. Если вы еще не вошли в IBM SPSS Collaboration and Deployment Services при попытке просмотра координатора процессов, вам предложат сделать это.
3. Выберите сервер или кластер серверов в списке.
4. Нажмите кнопку **ОК**, чтобы закрыть диалоговое окно и добавить это соединение в диалоговое окно Подключение к серверу.

## Изменение временного каталога

Для некоторых операций, выполняемых сервер IBM SPSS Modeler, требуется создавать временные файлы. По умолчанию IBM SPSS Modeler создает временные файлы в каталоге временных файлов системы. Вы можете изменить положение каталога временных файлов, как описано ниже.

1. Создайте новый каталог с именем *spss* и подкаталог с именем *servtemp*.
2. Откройте для редактирования файл *options.cfg* в подкаталоге */config* в каталоге установки IBM SPSS Modeler. Отредактируйте в этом файле параметр *temp\_directory*, задав: *temp\_directory*, "C:/spss/servtemp".
3. После этого нужно перезапустить службу сервер IBM SPSS Modeler. Для этого можно щелкнуть по вкладке **Службы** на Панели управления Windows. Просто остановите эту службу и запустите ее еще раз, чтобы активировать внесенные изменения. Кроме того, служба перезапустится при перезапуске компьютера.

Теперь все временные файлы будут записываться в этот новый каталог.

*Примечание:* Самая частая ошибка при этой операции - использование не того вида кривой черты; нужна дробная черта, а не обратная.

## Запуск нескольких сеансов IBM SPSS Modeler

Если нужно запускать несколько сеансов IBM SPSS Modeler одновременно, требуются некоторые изменения в настройке IBM SPSS Modeler и Windows. Например, это нужно, если у вас две серверных лицензии и вы хотите выполнять два потока на двух серверах с одного компьютера клиента.

Чтобы разрешить несколько сеансов IBM SPSS Modeler:

1. Выберите:  
**Пуск > [Все] Программы > IBM SPSS Modeler 16**
2. Щелкните правой кнопкой по ярлыку (значку) IBM SPSS Modeler 16 и выберите **Свойства**.
3. В текстовом поле **Назначение** добавьте *-noshare* в конце строки.
4. В Проводнике Windows выберите:  
**Сервис > Свойства папки...**
5. На вкладке Типы файлов выберите пункт Поток IBM SPSS Modeler и нажмите кнопку **Дополнительно**.
6. В диалоговом окне Изменений свойств типа файлов выберите пункт Открыть в IBM SPSS Modeler и нажмите кнопку **Изменить**.
7. В текстовом поле **Приложение, исполняющее действие** добавьте *-noshare* перед аргументом **-stream**.

---

## Беглый взгляд на интерфейс IBM SPSS Modeler

В каждой точке процесса исследования данных удобный интерфейс IBM SPSS Modeler приглашает вас применить свой бизнес-опыт. При этом мощные и точные модели обеспечиваются применением таких алгоритмов моделирования, как прогнозные алгоритмы, а также алгоритмы классификации, сегментации и обнаружения связей. Результаты моделей легко можно встроить и передать в базы данных, IBM SPSS Statistics и широкий круг других прикладных программ.

Работа с IBM SPSS Modeler включает в себя три этапа работы с данными.

- Во-первых, данные передаются в IBM SPSS Modeler.
- Во-вторых, данные пропускаются через последовательность операций.
- Наконец, данные отправляются в объект назначения.

Эта последовательность операций называется **поток данных**, поскольку данные запись за записью текут потоком из источника, проходят через все операции и, наконец, стекают в объект назначения -- модель или один из типов вывода данных.

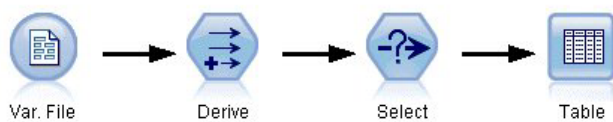


Рисунок 2. Простой поток

## Холст потока IBM SPSS Modeler

Холст потока - самая большая область окна IBM SPSS Modeler, в которой строят потоки данных и работают с ними.

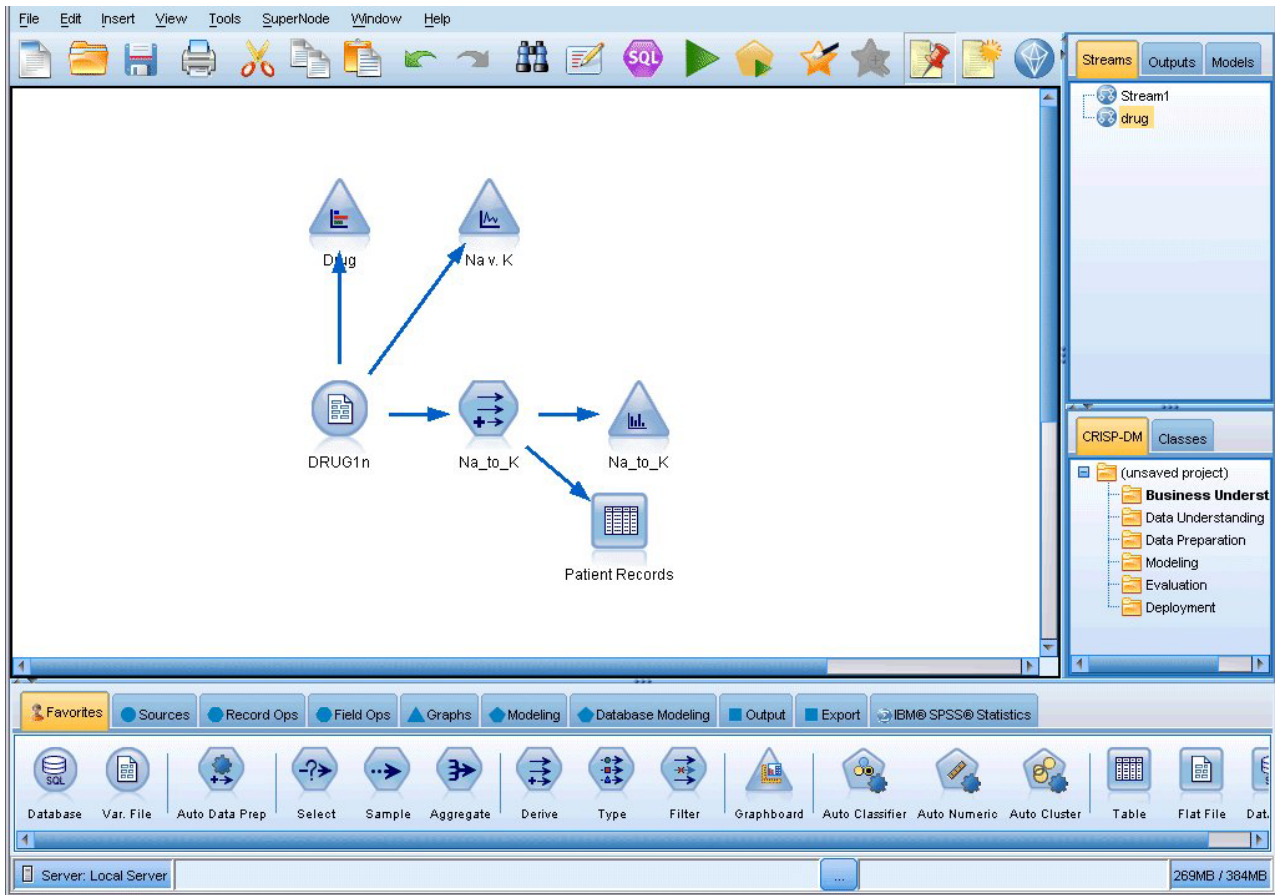


Рисунок 3. Рабочее пространство IBM SPSS Modeler (представление по умолчанию)

Потоки создаются на холсте интерфейса путем построения диаграмм, содержащих нужные операции над данными. Каждая операция представлена значком, или **узлом**, и узлы соединены стрелками в **поток**, представляющий путь данных через все операции.

В IBM SPSS Modeler можно одновременно работать с несколькими потоками, как на одном холсте, так и открыв новый холст потока. В течение сеанса потоки хранятся в менеджере потоков в верхнем правом углу окна IBM SPSS Modeler.

## Палитра узлов

Большая часть инструментов данных и моделирования в IBM SPSS Modeler располагаются на **палитре узлов** вдоль нижней части окна, под холстом потока.

Например, вкладка **Операции над записями** палитры узлов содержит узлы, которые можно использовать для таких операций над **записями** данных, как отбор, объединение и присоединение.

Чтобы добавить узлы на холст, щелкайте дважды по значкам на палитре узлов или перетаскивайте их на холст. Затем соедините их, чтобы создать **поток**, по которому будут передаваться данные.

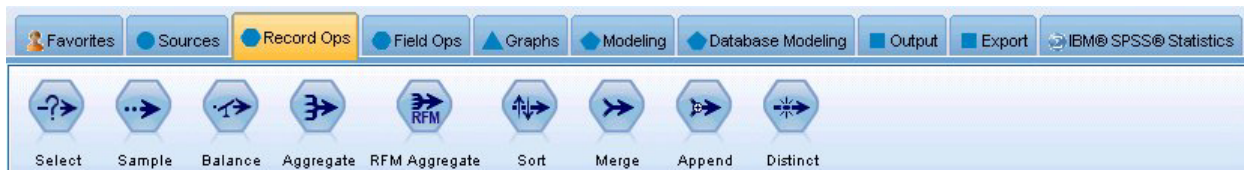


Рисунок 4. Вкладка Операции над записями на палитре узлов

Каждая вкладка на палитре узлов содержит подборку узлов, применяемых на том или ином этапе потока, а именно:

- **Источники.** Эти узлы поставляют данные в IBM SPSS Modeler.
- **Операции над записями.** Эти узлы выполняют такие операции над **записями** данных, как отбор, объединение и присоединение.
- **Операции над полями.** Эти узлы выполняют такие операции над **полями** данных, как фильтрация, вычисление новых полей и выяснение типа измерений для данных полей.
- **Диаграммы.** Эти узлы выводят наглядные представления данных до и после моделирования. Доступны такие диаграммы, как графики, гистограммы, Web-узлы и диаграммы оценки.
- **Моделирование.** Эти узлы используют алгоритмы моделирования, доступные в IBM SPSS Modeler, такие как нейросети, деревья решений, алгоритмы кластеризации и секвенирование данных.
- **Моделирование в базе данных.** Эти узлы используют алгоритмы моделирования, доступные в базах данных Microsoft SQL Server, IBM DB2, Oracle и Netezza.
- **Вывод.** Эти узлы создают различные представления данных, диаграмм и результатов моделирования, которые можно просматривать в IBM SPSS Modeler.
- **Экспорт.** Эти узлы создают различные представления для просмотра во внешних прикладных программах, таких как IBM SPSS Data Collection или Excel.
- **IBM SPSS Statistics.** Эти узлы импортируют или экспортируют данные, обмениваясь ими с IBM SPSS Statistics, а также запущенными процедурами IBM SPSS Statistics.

Познакомившись ближе с IBM SPSS Modeler, вы можете настроить содержимое палитры, как вам удобнее.

Ниже Палитры узлов находится панель отчетов, которая показывает ход выполнения различных операций, например, чтения данных в поток данных. Также ниже Палитры узлов находится панель состояния, содержащая информацию о том, чем в настоящее время занята прикладная программа и требуется ли ввод информации от пользователя.

## Менеджеры IBM SPSS Modeler

В верхней правой части окна расположена панель менеджеров. Она содержит три вкладки для управления потоками, выводом и моделями.

На вкладке Потоки можно открывать, переименовывать, сохранять и удалять потоки, созданные в сеансе.

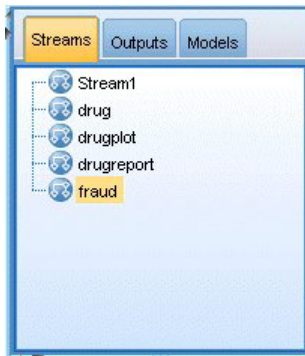


Рисунок 5. Вкладка Поток

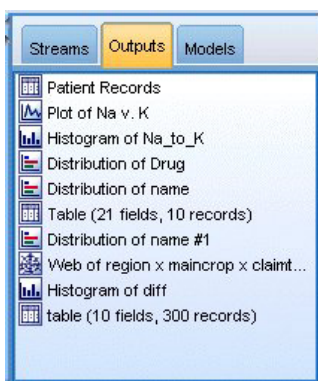


Рисунок 6. Вкладка Вывод

Вкладка Вывод содержит различные файлы, такие как диаграммы и таблицы, созданные операциями потоков в IBM SPSS Modeler. Перечисленные на вкладке таблицы, диаграммы и отчеты можно выводить на экран, сохранять, переименовывать и закрывать.



Рисунок 7. Вкладка моделей, содержащая слепки моделей

Из трех вкладок менеджеров вкладка моделей - самая богатая по арсеналу средств. Эта вкладка содержит все **слепки** моделей по всем моделям, сгенерированным в IBM SPSS Modeler в текущем сеансе. Эти модели можно просматривать непосредственно со вкладки моделей, а можно добавлять в поток на холсте.



## Проекты IBM SPSS Modeler

У правого края окна внизу расположена панель проектов, на которой можно создавать **проекты** исследования данных (группы файлов, связанных с некоторой задачей исследования данных) и управлять ими. Есть два способа просматривать проекты, создаваемые в IBM SPSS Modeler — в представлении Классы и в представлении CRISP-DM.

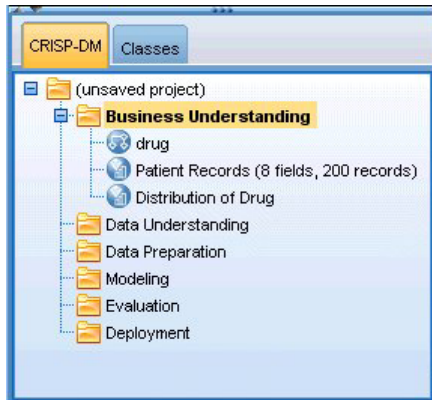


Рисунок 8. Представление CRISP-DM

Вкладка CRISP-DM обеспечивает один из способов организовать проекты - по испытанной на практике, непроприетарной методологии Cross-Industry Standard Process for Data Mining. Как опытным, так и начинающим исследователям данных инструмент CRISP-DM поможет лучше организовать и скоординировать работу.

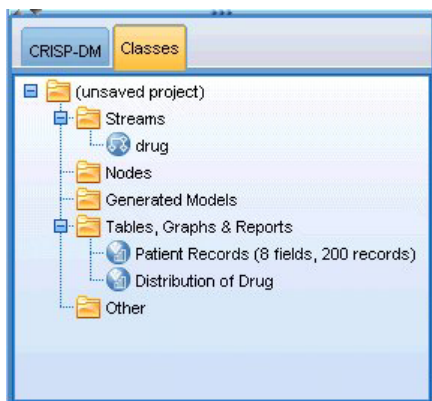


Рисунок 9. Представление классов

Вкладка Классы обеспечивает один из способов организовать работу в IBM SPSS Modeler по категориям — по типам создаваемых объектов. Это представление полезно при инвентаризации данных, потоков и моделей.

## Панель инструментов IBM SPSS Modeler






















В верхней части окна IBM SPSS Modeler находится панель со значками, выполняющими ряд полезных действий. Ниже описываются значки панели инструментов и их назначение.



Создать новый поток



Открыть поток

	Сохранить поток		Вывести на печать текущий поток
	Вырезать & переместить в буфер обмена		Копировать в буфер обмена
	Вставить выделенное		Отменить последнее действие
	Повторить		Искать узлы
	Правка свойств потока		Предварительный просмотр сгенерированного SQL
	Запустить текущий поток		Запустить выбранную часть потока
	Остановить поток (значок активен, только если поток запущен)		Добавить надузел
	Увеличить (только на надузлах)		Уменьшить (только на надузлах)
	Без разметки в потоке		Вставить комментарий
	Скрыть разметку потока (если есть)		Показать скрытую разметку потока
	Открыть поток в IBM SPSS Modeler Advantage		

Разметка потока состоит из комментариев к потоку, ссылок на модели и оценочных показателей ветвей.

Ссылки на модели описаны в руководстве *Узлы моделирования IBM SPSS*.

## Настройка панели инструментов

Можно изменить различные аспекты панели инструментов, такие как:

- Показывать ли панель инструментов
- Доступны ли подсказки при значках
- Крупные значки или мелкие

Чтобы показать и скрыть панель инструментов:

1. Выберите в главном меню:

**Вид > Панель инструментов > Показать**

Чтобы изменить настройку подсказок или размер значков:

1. В главном меню выберите:

**Вид > Панель инструментов > Настроить**

Выберите **Показывать подсказки** или **Крупные кнопки**, как это потребуется.

## Настройка окна IBM SPSS Modeler

Используя разделители между различными частями интерфейса IBM SPSS Modeler, вы можете изменять размеры и закрывать инструменты, как вам будет удобнее. Например, если вы работаете с большим потоком, вы можете, нажимая на маленькие стрелки, которые есть на каждом разделителе, закрыть палитру узлов, панель менеджеров и панель проектов. Это увеличит размер холста потока, освободив место для работы над большим потоком или несколькими потоками.

Другой вариант - в меню Вид выбрать **Палитра узлов**, **Менеджеры** или **Проект**, чтобы включить или выключить эти элементы.

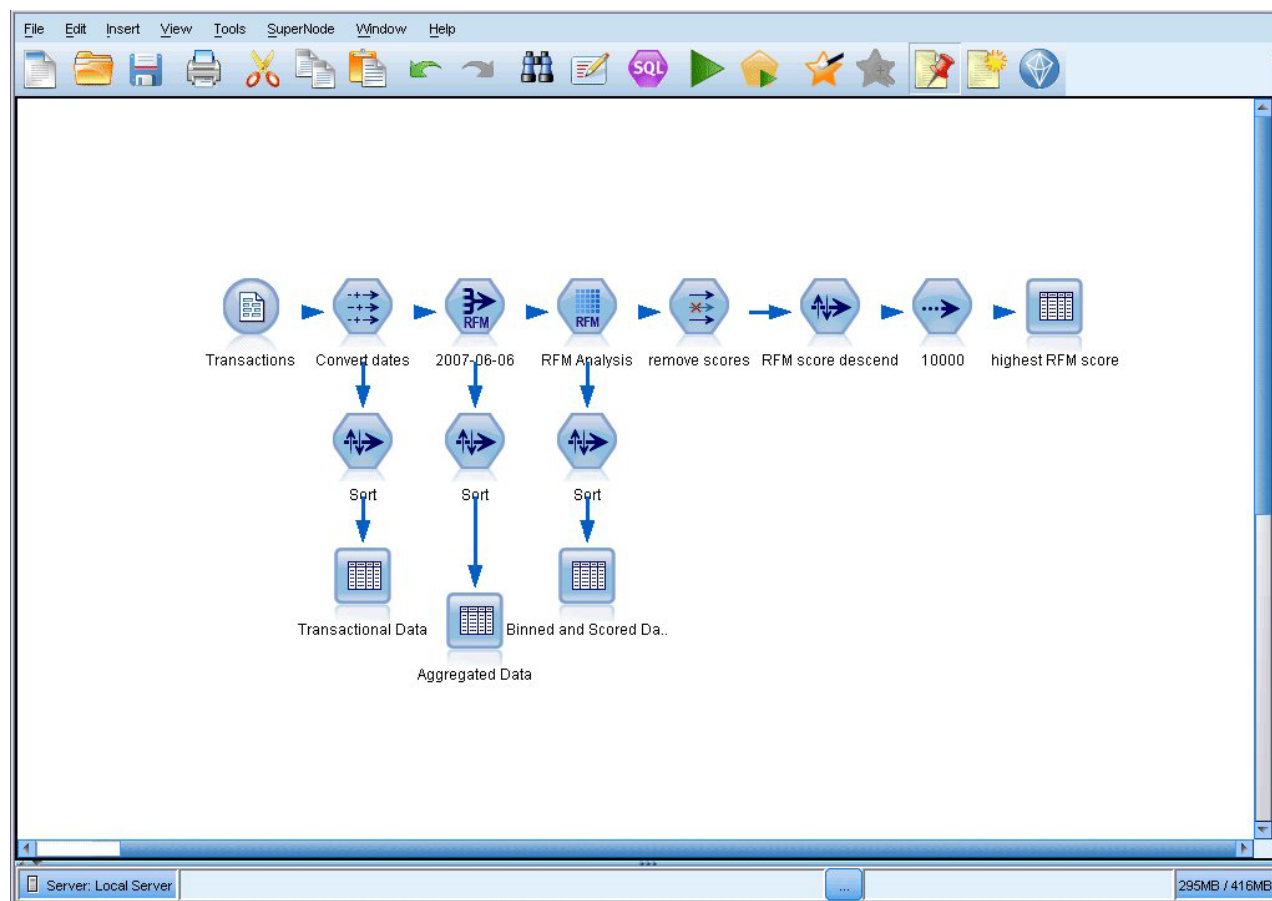


Рисунок 10. Развернуть холст потока

Вместо того, чтобы закрыть палитру узлов и панели менеджеров и проекта, можно пользоваться холстом потока как страницей с прокруткой, перемещаясь по холсту при помощи вертикальной и горизонтальной полос прокрутки сбоку и внизу окна IBM SPSS Modeler.

Кроме того, можно управлять выводом разметки, к которой относятся комментарии к потоку, ссылки на модели и индикаторы ветви оценки. Чтобы показать или скрыть эти подробности, выберите:

**Вид > Разметка потока**

## Изменение размера значка для потока

Размер значков потока можно изменять следующими способами.

- Через настройку свойств потока
- Через всплывающее меню в потоке
- При помощи клавиатуры

Представление всего потока можно масштабировать до ряда размеров от 8% до 200% стандартного размера значка.

Чтобы масштабировать весь поток (используя свойства потока)

1. В главном меню выберите **Инструменты > Свойства потока > Опции > Макет**.
2. Выберите нужный размер в меню **Размер значка**.
3. Нажмите кнопку **Применить**, чтобы увидеть результат.
4. Нажмите кнопку **ОК**, чтобы сохранить изменение.

Чтобы масштабировать весь поток (используя меню)

1. Щелкните правой кнопкой по фоновому участку холста потока.
2. Выберите пункт меню **Размер значка** и выберите нужный размер значка.

Чтобы масштабировать весь поток (используя клавиатуру)

1. Нажмите **Ctrl + [-]** на основной клавиатуре, чтобы уменьшить размер до предыдущего в ряду допустимых размеров.
2. Нажмите **Ctrl + [+]** на основной клавиатуре, чтобы увеличить размер до следующего в ряду допустимых размеров.

Эта возможность особенно полезна для получения общего обзора сложного потока. Кроме того, с ее помощью можно минимизировать число необходимых страниц при печати потока.

## Использование мыши в IBM SPSS Modeler

Основные способы использования мыши в IBM SPSS Modeler включают следующее:

- **Один щелчок.** Правая или левая кнопка мыши выбирает пункт в меню, открывает всплывающее меню и выбирает обычные элементы контроля и опции. Щелчок и удержание кнопки служит для перетаскивания узлов.
- **Двойной щелчок.** Щелкните дважды левой кнопкой, чтобы поместить узел на холст потока и отредактировать существующий узел.
- **Щелчок средней кнопкой.** Нажмите среднюю кнопку и перетащите указатель, чтобы соединить узлы на холсте потока. Щелкните средней кнопкой дважды, чтобы отсоединить узел. Если у вас нет трехкнопочной мыши, эту возможность можно симитировать, нажимая клавишу **Alt** во время щелчка и перетаскивания мышью.

## Использование клавиш быстрого вызова

Со многими операциями визуального программирования в IBM SPSS Modeler связаны клавиатурные сокращения. Например, можно удалить узел, щелкнув по этому узлу и нажав клавишу **Delete** на клавиатуре.

Другой пример - можно быстро сохранить поток, если нажать клавишу S, удерживая нажатой клавишу Ctrl. Подобные управляющие команды записываются как сочетание клавиши Ctrl с другой клавишей -- например, как Ctrl+S.

Ряд клавиш быстрого вызова используется при обычных операциях Windows, например, Ctrl+X, чтобы вырезать в буфер обмена. IBM SPSS Modeler поддерживает такие клавиатурные сокращения наряду с приведенными ниже клавишами быстрого вызова специально для прикладной программы.

*Примечание:* Старые клавиши быстрого вызова, использовавшиеся в IBM SPSS Modeler, в ряде случаев конфликтуют со стандартными клавиатурными сокращениями Windows. Эти старые сочетания клавиш теперь поддерживаются с добавлением клавиши Alt. Например, для включения и выключения кэша теперь можно использовать Ctrl+Alt+C.

Таблица 1. Поддерживаемые клавиши быстрого вызова

Клавиши	Функция
Ctrl+A	Выделить все
Ctrl+X	Вырезать
Ctrl+N	Новый поток
Ctrl+O	Открыть поток
Ctrl+P	Печатать
Ctrl+C	Скопировать
Ctrl+V	Вставить
Ctrl+Z	Откат
Ctrl+Q	Выделить все узлы вниз по направлению потока от выделенного узла
Ctrl+W	Снять выделение со всех узлов вниз по направлению потока (работает совместно с Ctrl+Q как включение-выключение)
Ctrl+E	Выполнить выделенный узел
Ctrl+S	Сохранить текущий поток
Alt+клавиши со стрелками	Переместить выделенные узлы на холсте потока в направлении стрелки на клавише
Shift+F10	Открыть всплывающее меню для выделенного узла

Таблица 2. Поддерживаемые клавиатурные сокращения вместо старых горячих клавиш

Клавиши	Функция
Ctrl+Alt+D	Дублировать узел
Ctrl+Alt+L	Загрузить узел
Ctrl+Alt+R	Переименовать узел
Ctrl+Alt+U	Создать узел пользовательского ввода
Ctrl+Alt+C	Включить-выключить кэш
Ctrl+Alt+F	Очистить кэш
Ctrl+Alt+X	Развернуть надузел
Ctrl+Alt+Z	Детализировать или свернуть
Delete	Удалить узел или соединение

---

## Печать

Приведенные ниже объекты можно напечатать в IBM SPSS Modeler:

- Диаграммы потока
- Диаграммы
- Таблицы
- Отчеты (с узла отчетов и из окна Отчеты по проекту)
- Сценарии (из диалоговых окон свойств потока, отдельного сценария или сценария надузла)
- Модели (из браузера моделей, выбранных вкладок диалогового окна, просмотра деревьев)
- Аннотации (используя для вывода вкладку аннотаций)

Чтобы напечатать объект:

- Для печати без предварительного просмотра нажмите кнопку Печать на панели инструментов.
- Чтобы настроить страницу перед печатью, выберите **параметры страницы** в меню Файл.
- Для предварительного просмотра перед печатью выберите **Предварительный просмотр печати** в меню Файл.
- Чтобы открыть обычное диалоговое печати с опциями выбора принтера и настройками стиля выберите **Печать** в меню Файл.

---

## Автоматизация IBM SPSS Modeler

Поскольку исследование данных может быть сложным и подчас продолжительным процессом, IBM SPSS Modeler содержит в себе несколько типов поддержки кодирования и автоматизации.

- Язык **CLEM** (Control Language for Expression Manipulation, управляющий язык для преобразования выражений) - это язык для анализа и обработки данных в потоках IBM SPSS Modeler. Исследователи данных широко используют CLEM в операциях потока для выполнения таких простых задач, как вычисление прибыли по данным затрат и доходов, и таких сложных, как преобразование данных Web-журнала в набор пригодных для использования полей и записей.
- **Сценарии** - мощный инструмент для автоматизации процессов в интерфейсе пользователя. Сценарии выполняют некоторые типы действий, которые пользователи могут выполнять при помощи мыши или клавиатуры. Кроме того, можно задавать вывод и выполнять операции над сгенерированными моделями.

---

## Глава 4. Что такое исследование данных

---

### Обзор исследования данных

Используя ряд технологий, **исследование данных** находит в теле данных самородки полезных сведений. Исследование данных извлекает сведения таким образом, что они применимы в таких областях, как поддержка решений, прогнозирование и оценка. Часто имеется огромный объем малоценных данных, от которых в необработанном виде мало толку. Ценность представляют сведения, скрытые в этих данных.

При исследовании данных успех приходит от сочетания ваших (и ваших экспертов) знаний о данных и современных активных технологий анализа, когда компьютер распознает глубинные взаимосвязи и особенности данных. В процессе исследования из данных хронологии генерируются модели, которые затем служат для прогнозирования, обнаружения структур и прочего. Технология построения таких моделей называется **машинным обучением** или **моделированием**.

#### Технологии моделирования

IBM SPSS Modeler содержит ряд технологий машинного обучения и моделирования, которые можно грубо разбить на несколько групп по типу решаемых задач.

- К методам прогнозного моделирования относятся деревья решений, нейросети и статистические модели.
- Модели кластеризации сосредоточены на обнаружении групп сходных записей и разметке записей согласно группе, к которой отнесена запись. К методам кластеризации относятся кластеризация Коонена, *k*-средние и двухшаговый анализ.
- Правила связывания устанавливают зависимость некоторого заключения (например, приобретения некоторого конкретного продукта) с набором условий (приобретением ряда других продуктов).
- Экранирование моделей может использоваться для экранирования данных с целью найти поля и записи, потенциально интересные для моделирования, и выявить выбросы, которые могут не удовлетворять известным шаблонам. Доступные методы: отбор показателей и обнаружение аномалий.

#### Работа с данными и обнаружение данных

IBM SPSS Modeler также включает в себя ряд утилит, которые дают вам возможность проявить свой опыт в отношении данных:

- **Работа с данными.** Строятся новые элементы данных, вычисленные по существующим; данные разбиваются на осмысленные поднаборы. Данные из различных источников можно объединять и фильтровать.
- **Просмотр и визуализация.** Выводит на экран различные аспекты данных, используя узел аудита для первоначального аудита, включая диаграммы и статистику. Расширенная визуализация включает в себя интерактивные диаграммы, которые можно экспортировать, чтобы добавлять в отчеты по проекту.
- **Статистика.** Подтверждает предполагаемые взаимосвязи между переменными в данных. Кроме того, статистику из IBM SPSS Statistics можно использовать в IBM SPSS Modeler.
- **Проверка гипотез.** Строятся и затем проверяются модели поведения данных.

Обычно эти утилиты служат для выявления многообещающих наборов атрибутов в данных. Затем эти атрибуты можно передать технологиям моделирования, чтобы попробовать выявить стоящие за ними правила и взаимозависимости.

#### Типичные применения

Типичные применения исследования данных включают в себя следующее:

**Прямая рассылка.** Выяснить, у каких демографических групп самый высокий уровень отклика. Использовать эту информацию для максимизации отклика на будущие рассылки.

**Кредитная оценка.** Использовать личную кредитную историю для принятия решений о кредите.

**Кадры.** Познакомиться с практиками приема на работу в прошлом и создать решающие правила для оптимизации процесса приема на работу.

**Медицинские исследования.** Создать решающие правила, предлагающие прибегать к процедурам на основе медицинских данных.

**Анализ рынка.** Выявите переменные, например, география, цена и характеристики посетителей, которые связаны с продажами.

**Контроль качества.** Проанализируйте данные о производстве продукта и выявите переменные, влияющие на дефекты продукта.

**Политические исследования.** Используйте данные опросов, чтобы выработать политику, применяя решающие правила для выбора наиболее важных переменных.

**Здравоохранение.** Опросы пользователей и клинические данные можно использовать совместно для обнаружения переменных, вносящих вклад в здоровье.

Терминология

Термины **атрибут**, **поле** и **переменная** означают элемент данных, который встречается во всех рассматриваемых наблюдениях. Набор значений различных атрибутов из одного наблюдения называется **записью**, **примером** или **наблюдением**.

---

## Оценка данных

К данным применяются определенные критерии, без выполнения которых трудно ожидать полезных результатов от исследования данных. Важные аспекты данных и их применения, которые вам следует учитывать, представлены в приведенных ниже разделах.

Убедитесь, что данные доступны

Возможно, это покажется вам очевидным, но даже доступные вроде бы данные иногда оказываются в неудобной для использования форме. IBM SPSS Modeler может импортировать данные из баз данных (через ODBC) или из файлов. Однако иногда данные хранятся на некотором компьютере в иной форме и прямой доступ к ним невозможен. Чтобы использовать такие данные, сначала их надо скачать или вывести в приемлемой форме. Иногда данные нужно собрать из нескольких баз данных и источников. Иногда они не доступны в сети. Если данные существуют только на бумаге, для их исследования потребуется сначала выполнить ввод данных.

Проверьте, что данные охватывают нужные атрибуты

Цель исследования данных - выявить нужные атрибуты, и на первый взгляд эта проверка кажется странной. Однако на самом деле полезно посмотреть на доступные данные и прикинуть, какие потенциально важные факторы остались не записаны. Например, если вы пытаетесь прогнозировать продажи мороженого и имеете большой массив информации о точках продажи или хронологии продаж, но нет сведений о погоде и температуре воздуха, это может быть важным упущением. Хотя даже в отсутствии важных атрибутов исследование данных может принести полезные результаты, пропущенные атрибуты скажутся на точности прогнозов.



Быстрый способ оценить ситуацию - выполнить всесторонний аудит данных. Прежде чем двигаться дальше, подсоедините к источнику данных узел аудита данных и выполните его, чтобы сгенерировать полный отчет.

Имейте в виду опасность зашумленных данных

Нередко данные содержат ошибки и субъективные, а потому изменчивые суждения. Общее название для таких явлений - **шум**. Иногда присутствие шума в данных считается нормальным. Даже если действуют специальные правила, они предотвращают шум не во всех случаях.

Обычно чем больше шума в данных, тем труднее получить точные результаты. Однако методы машинного обучения IBM SPSS Modeler способны работать с зашумленными данными и иногда успешно работают даже с такими наборами данных, которые содержат до 50% шума.

Убедитесь, что у вас достаточно данных

При исследовании данных необязательно важен объем набора данных. Намного важнее *представительность* набора данных, а также охват возможных исходов и сочетаний переменных.

Обычно чем больше атрибутов учитывается, тем больше нужно записей для репрезентативного охвата.

Если данные репрезентативны и действуют обычные правила, то вполне возможно, что выборка данных размером в несколько тысяч записей (а иногда и всего лишь несколько сотен) дает результаты не хуже, чем миллион -- и при том быстрее.

Найдите экспертов по данным

Во многих случаях вы будете работать со своими собственными данными, где вам знакомы все атрибуты и их смысл. Но если вы работаете над данными другого отдела вашей организации или данными клиента, желательно иметь контакт с экспертами, которым эти данные знакомы. Эксперты помогут вам опознать важные атрибуты и интерпретировать результаты исследования данных, отделяя самородки информации от обманок, артефактов, порожденных аномалиями в наборах данных.

---

## Стратегия исследования данных

Как и во многих других сферах бизнеса, плановый, систематический подход к делу существенно повышает эффективность исследования данных. Даже при использовании новейших инструментов исследования данных, таких как IBM SPSS Modeler, для большей части работы по исследованию данных требуется аналитик с опытом в данном бизнесе, чтобы процесс не сбился с пути. Планируя исследование, ответьте себе на следующие вопросы:

- Какую существенную проблему вы хотите решить?
- Какие источники данных вам доступны, и какие части данных имеют отношение к текущей проблеме?
- Какая предварительная обработка и очистка данных потребуется перед собственно исследованием данных?
- Какие технологии исследования данных вы будете использовать?
- Как вы оцените результаты анализа после исследования данных?
- Как получить максимум результатов из информации, полученной при исследовании данных?

Обычный процесс исследования данных может быстро усложниться. Нужно отслеживать много факторов -- комплексные бизнес-проблемы, множественные источники данных, различие в качестве данных с разных источников, массив технологий исследования данных, различные способы измерения успешности исследования данных и так далее.

Чтобы не сбиться с пути, помогает четко определенная модель процесса исследования данных. Модель процесса помогает отвечать на вопросы из приведенного выше списка и не упускать из виду важные моменты. Она служит дорожной картой исследования данных, помогая не запутаться в дебрях ваших данных.

Рекомендуемый процесс исследования данных для использования совместно с SPSS Modeler называется CRISP-DM (Cross-Industry Standard Process for Data Mining, межотраслевой стандартный процесс исследования данных). Как видно из названия, эта модель разработана как общая модель, применимая к широкому кругу отраслей и проблем бизнеса.

---

## Модель процесса CRISP-DM

Общая модель процесса CRISP-DM состоит из шести этапов, решающих основные задачи исследования данных. Вместе шесть этапов составляют циклический процесс, предназначенный для внедрения исследования данных в более крупную бизнес-практику.

Эти шесть этапов такие:

- **Понимание бизнеса.** Возможно, это самый важный этап исследования данных. Понимание бизнеса включает в себя определение целей бизнеса, оценку ситуации, определение целей исследования данных и создание проекта.
- **Понимание данных.** Данные - это сырье для исследования данных. На этом этапе нужно понять, каковы источники данных и характеристики этих источников. Сюда входит сбор начальных данных, описание данных, изучение данных и проверка их качества. Незаменимый инструмент для изучения данных - узел аудита данных, доступный на палитре узлов вывода.
- **Подготовка данных.** Составив каталог своих ресурсов данных, следует подготовить эти данные к исследованию. К подготовке относятся отбор, очистка, построение, интеграция и форматирование данных.
- **Моделирование.** Это, конечно, самая яркая часть исследования данных, когда сложные методы анализа извлекают из ваших данных скрытые сведения. К этому этапу относятся выбор технологий моделирования, генерирование тестовых структур и, наконец, построение и оценка моделей.
- **Оценка.** Выбрав модели, вы можете оценить, насколько полезными оказались результаты исследования данных для бизнес-целей. К этому этапу относятся оценка результатов, обзор процесса исследования данных и определение последующих действий.
- **Внедрение.** Теперь можно пожинать урожай всех вложенных усилий. Этот этап посвящен интеграции новых знаний в дежурные бизнес-процессы для решения первоначально поставленных бизнес-задач. Этот этап включает в себя внедрение плана, мониторинг и обслуживание, составление итогового отчета и обзор проекта.

О такой модели процесса нужно знать несколько важных моментов. Во-первых, хотя обычно процесс протекает примерно в той последовательности, как описано в предыдущих абзацах, есть также ряд мест, где этапы взаимно влияют друг на друга с нарушением линейного порядка. Например, подготовка данных обычно действительно предшествует моделированию. Но на этапе моделирования принимаются решения и собираются сведения, которые порой заставляют пересмотреть этап подготовки данных, после чего, возможно, наткнуться на новые вопросы с моделированием. Эти два этапа могут обмениваться результатами, пока не будет достигнуто адекватное решение задач обоих этапов. Аналогичным образом этап оценки может привести к переоценке первоначального понимания бизнеса, в результате чего вы, возможно, решите, что искали ответ не тот вопрос. В этот момент вы можете ревизовать понимание бизнеса и снова пройти по этапам процесса с лучшим пониманием его цели.

Второй важный момент - исследование данных по своей природе итеративно. Очень редко, если вообще хоть когда-либо, удастся просто составить проект исследования данных, выполнить его, упаковать данные и пойти отдыхать. Чтобы ответить на чаяния ваших клиентов, исследование данных потребует продолжительных усилий. Знания, приобретенные в одном цикле исследования данных, почти наверняка приведут к новым вопросам, проблемам и возможностям для обнаружения и удовлетворения потребностей

ваших клиентов. Обычно в ответ на эти новые вопросы, проблемы и возможности следует повторить исследование данных. Такой процесс исследования и обнаружения новых возможностей должен стать частью вашего восприятия своего дела и краеугольным камнем всей вашей бизнес-стратегии.

Это вступление - лишь краткий взгляд на модель процесса CRISP-DM. Во всех подробностях с этой моделью можно познакомиться по следующим ссылкам:

- *Руководство по CRISP-DM*, которое наряду с другой документацией доступно в папке \Documentation на установочном диске.
- Справочная система CRISP-DM, доступная в меню Пуск или как **Справка по CRISP-DM** в меню Справка в IBM SPSS Modeler.

---

## Типы моделей

IBM SPSS Modeler предлагает ряд методов моделирования, взятых из таких областей, как обучение машин, искусственный интеллект и статистика. При помощи методов, доступных на палитре Моделирование, можно извлечь новую информацию из данных и разработать прогнозные модели. У каждого из методов есть свои сильные стороны и типы задач, для решения которых он лучше всего подходит.

*Руководство по прикладным программам IBM SPSS Modeler* предоставляет примеры многих таких методов вместе с общей информацией о процессе моделирования. Это руководство доступно и в виде оперативного учебного пособия, и в формате PDF. Дополнительную информацию смотрите в разделе “Примеры прикладных программ” на стр. 4

Методы моделирования разделены на три категории:

- Классификация
- Взаимосвязь
- Сегментация

### Модели классификации

*Модели классификации* используют значения одного или нескольких **входных** полей, чтобы предсказать значения одного или нескольких выходных полей, или полей **назначения**. Некоторые примеры таких способов следующие: деревья решений (алгоритмы дерева C&R, QUEST, CHAID и C5.0), регрессии (линейная, логистическая, обобщенная линейная и Кокса), нейронные сети, модели опорных векторов и Байесовские сети.

Модели классификации позволяют организациям предсказать известные результаты, например, купит ли покупатель товар или уйдет без покупки, или похожа ли транзакция на известный шаблон мошенничества. Способы моделирования включают в себя компьютерное обучение, вывод правил по индукции, идентификацию подгрупп, статистические способы и генерирование нескольких моделей.

### Узлы классификации



Узел автоклассификации создает и сравнивает несколько различных моделей для двоичных выходных данных (да или нет, уйдет клиент или останется и так далее), что позволяет выбрать лучший подход для данного анализа. Поддерживается несколько алгоритмов моделирования, что делает возможным выбор желательных для использования способов, конкретных опций для каждого из них и критериев сравнения результатов. Этот узел генерирует набор моделей на основе заданных опций и ранжирует лучших кандидатов в соответствии с заданными вами критериями.



Узел автоматической нумерации оценивает и сравнивает модели для выходных данных в количественном числовом диапазоне при помощи нескольких разных способов. Этот узел работает аналогично другим узлам автоклассификации, допуская выбор алгоритмов для использования и экспериментирование с несколькими комбинациями опций при одном проходе моделирования. Поддерживаемые алгоритмы включают в себя нейросети, дерево C&R, CHAID, линейную регрессию, обобщенную линейную регрессию и механизмы опорных векторов (support vector machines, SVM). Модели можно сравнивать на основе корреляции, относительной ошибки или числа используемых переменных.



Узел дерева классификации и регрессии (Classification and Regression, C&R) генерирует дерево решений, позволяющее предсказывать или классифицировать будущие наблюдения. Этот метод использует рекурсивное разделение, чтобы расщепить обучающие записи на сегменты, на каждом шаге минимизируя неоднородность, причем узел дерева считается «чистым», если все 100% наблюдений в узле попадают в конкретную категорию поля назначения. Входные поля и поля назначения могут быть из числового диапазона или категориальными (номинальными, порядковыми или флагами); все расщепления бинарны (только две подгруппы).



Узел QUEST предоставляет метод бинарной классификации для построения деревьев решений, разработанный для уменьшения времени обработки, требуемого для анализа больших деревьев C&R, при одновременном подавлении обнаруженного в способах деревьев классификации предпочтения входных полей, допускающих больше расщеплений. Входные поля могут быть в числовом диапазоне (количественными), но поле назначения должно быть категориальным. Все расщепления бинарные.



Узел CHAID генерирует деревья решений, используя статистику хи-квадрат для определения оптимальных расщеплений. В отличие от узлов дерева C&R и QUEST, CHAID может генерировать не только бинарные деревья, то есть у некоторых расщеплений может быть больше двух ветвей. Входные поля и поле назначения могут быть количественными (числовой диапазон) или категориальными. Исчерпывающий CHAID - это модификация метода CHAID, при котором прорабатывается более тщательная работа по изучению всех возможных расщеплений для каждого предиктора, но это требует больше времени для вычислений.



Узел C5.0 строит или дерево решений, или набор правил. Эта модель работает, разделяя выборку на основании значения в поле, дающего максимальный информационный выигрыш на каждом уровне. Поле назначения должно быть категориальным. Разрешено несколько разделений на подгруппы, и таких подгрупп может быть больше двух.



Узел списка решений определяет подгруппы или сегменты, которые показывают более высокое или более низкое правдоподобие для данного бинарного результата по сравнению с полной совокупностью. Например, вы могли бы искать клиентов с низкой вероятностью оттока или с высокой вероятностью отклика на кампанию. Вы можете включить свои знания о бизнесе в модель, добавляя свои собственные пользовательские сегменты и параллельно просматривая альтернативные модели, чтобы сравнить результаты. Модели списка решений состоят из списка правил, в котором каждое правило имеет условие и следствие. Правила применяются по очереди, и первое подходящее правило определяет результат.



Модели линейной регрессии предсказывают значения непрерывного целевого поля на основе линейных взаимосвязей между целевым полем и одним или несколькими предикторами.



Узел PCA/фактора предоставляет мощные средства сокращения числа данных для уменьшения сложности ваших данных. Анализ главных компонент (principal components analysis, PCA) находит линейные комбинации входных полей, которыми главным образом определяются изменения в целом наборе полей, где компоненты ортогональны друг другу. Факторный анализ направлен на выявление скрытых факторов, объясняющих структуру корреляций в наборе наблюдаемых полей. Цель обоих подходов - найти небольшое количество производных полей, которые эффективно суммируют информацию исходного набора входных полей.



Узел выбора возможностей изучает входные поля на возможность удаления, основываясь на наборе критериев (таких как процентная доля пропущенных значений); затем этот узел ранжирует важность оставшихся полей по отношению к заданному полю назначения. Например, если у набора данных сотни потенциальных входных полей, какие из них потенциально наиболее полезны при моделировании исхода лечения пациента?



Дискриминантный анализ делает более строгие предположения, чем логистическая регрессия, но он может быть ценной альтернативой или дополнением к анализу логистической регрессии, когда эти предположения оказываются правильными.



Логистическая регрессия - это статистический метод для классификации записей на основании значений входных полей. Она аналогична линейной регрессии, но логистическая регрессия использует категориальные поля назначения вместо численных.



Обобщенная линейная модель расширяет общую линейную модель, так что зависимая переменная считается линейно связанной с факторами и ковариатами через заданную функцию связи. Более того, модель допускает наличие у зависимой переменной распределения, отличающегося от нормального. Она включает в себя функциональные возможности большого количества статистических моделей, в том числе линейной регрессии, логистической регрессии, логлинейных моделей для количества данных и интервал-цензурированных моделей выживания.



Обобщенная линейная смешанная модель (generalized linear mixed model, GLMM) обобщает линейную модель таким образом, что у значений назначения может быть отличное от нормального распределение и оно будет линейно связано с факторами и ковариатами через задаваемую функцию связи, так что наблюдения могут быть скоррелированными. Обобщенные линейные смешанные модели включают широкий набор моделей, начиная от простой линейной регрессии и кончая сложными многоуровневыми моделями для не нормально распределенных данных с повторными измерениями.



Узел регрессии Кокса позволяет построить модель дожития для данных времени-до-события в присутствии цензурируемых записей. Эта модель создает функцию дожития, которая предсказывает вероятность, что изучаемое событие произойдет в данное время ( $t$ ) для данных значений входных переменных.



Узел механизма опорных векторов (Support Vector Machine, SVM) позволяет классифицировать данные по одной или двум группам без переобучения. SVM хорошо работает с широкими наборами данных, в частности, в случае очень большого числа входных полей.



Узел Байесовская сеть позволяет построить вероятностную модель, комбинируя наблюдаемые и записанные сведения с очевидными с точки зрения здравого смысла данными, чтобы установить правдоподобие возникновения событий. Этот узел в основном работает с усиленным деревом наивными байесовскими сетями (Tree Augmented Naïve Bayes, TAN) и полными марковскими сетями, которые изначально используются для классификации.



Узел Самообучаемая модель откликов (Self-Learning Response Model, SLRM) позволяет построить модель, в которой одно новое наблюдение или всего несколько наблюдений могут быть использованы для повторной оценки модели без необходимости повторного обучения модели с использованием всех данных.



Узел временных рядов оценивает экспоненциальное сглаживание, а также одномерные и многомерные модели авторегрессии и проинтегрированного скользящего среднего (Autoregressive Integrated Moving Average, ARIMA) для временных рядов и создает прогнозы будущего выполнения. Предшественником узла временных рядов всегда должен быть узел Интервалы времени.



Узел  $k$  ближайших соседей (k-Nearest Neighbor, KNN) связывает новое наблюдение с категорией или значением  $k$  объектов, ближайших к нему в пространстве предикторов, где  $k$  - это целое число. Подобные наблюдения близки друг к другу, а непохожие наблюдения, наоборот, удалены друг от друга.

## Модели связывания

*Модели связывания* находят структуры в ваших данных, в которых один или несколько объектов (таких как события, покупки или атрибуты) связаны с одним или несколькими другими объектами. Модели конструируют наборы правил, определяющие эти взаимосвязи. Здесь поля среди данных могут быть и входными полями, и полями назначения. Вы могли бы найти эти связи вручную, но алгоритмы правил связывания делают это гораздо быстрее и могут изучить более сложные структуры. Модели Априори и SARMA - это примеры использования таких алгоритмов. Еще один тип модели связывания - это модель обнаружения последовательностей, которая находит последовательные шаблоны в структурированных по времени данных.

Модели связывания наиболее полезны при предсказании нескольких выходных значений, например, покупатель, купивший товар X, купил также Y и Z. Модели связывания связывают конкретное следствие (например, решение что-либо купить) с набором условий. Преимущество алгоритмов правил связывания перед более стандартными алгоритмами дерева решений (C5.0 и дерева C&R) состоит в том, что связывание может существовать между любыми атрибутами. Алгоритм дерева решений построит правила с возможностью одного следствия, в то время как алгоритмы связывания стараются найти несколько правил, у каждого из которых может быть отдельное следствие.

### Узлы связывания



Узел Априори извлекает набор правил из данных, выделяя правила с наибольшим информационным содержанием. Узел Априори предлагает пять различных способов выбора правил и использует сложные схемы индексирования для эффективной обработки больших наборов данных. Для больших задач узел Априори обычно быстрее при обучении; у него нет произвольного ограничения количества правил, которые можно сохранить, и он может обрабатывать правила с количеством предварительных условий до 32. Для узла Априори требуются категориальные входные и выходные поля, он был оптимизирован для полей такого типа и показывает с ними высокую производительность.



Модель SARMA извлекает из данных набор правил, не требуя, чтобы вы задавали входные или выходные поля. В отличие от узла Априори, узел SARMA предлагает параметры построения для поддержки правил (поддержка относится и к antecedentes, и к консеквентам), а не только для поддержки antecedentes. Это означает, что сгенерированные правила можно использовать в более широком наборе прикладных программ, например, чтобы найти список продуктов или услуг (antecedents), консеквент которых - это товар, который вы хотите продвигать в этом летнем сезоне.



Узел последовательности обнаруживает правила связывания для последовательных или зависящих от времени данных. Последовательность - это список наборов элементов с тенденцией появления в предсказуемом порядке. Например, покупатель, который приобрел лезвие и лосьон после бритья, с большой вероятностью в следующий раз купит крем для бритья. Узел последовательности основан на алгоритме правил связывания CARMA, использующем эффективный двухпроходный способ обнаружения последовательностей.

## Модели сегментации

*Модели сегментации* делят данные на сегменты, или кластеры, записей с одинаковыми структурами входных полей. Так как эти модели работают только с входными полями, у них нет отношения к выходным полям (полям назначения). Примеры моделей сегментации - это сети Коонена, кластеризация К-средних, двухшаговая кластеризация и выявление аномалий.

Модели сегментации (их называют также "моделями кластеризации") полезны в тех случаях, когда конкретный результат неизвестен (например, при идентификации нового шаблона мошенничества или интересующих вас групп в базе данных клиентов). Модели кластеризации уделяют главное внимание идентификации групп сходных записей и присвоению меток записям в соответствии с группами, к которым они принадлежат. Это делается без использования преимуществ предварительного знания о группах и их характеристиках, что отличает модели кластеризации от других способов моделирования, то есть отсутствие предварительно определенного выходного поля (поля назначения), значение в котором предсказывалось бы моделью. У этих моделей нет правильных или неправильных ответов. Их ценность в способности захватывать интересные группировки данных и представлять полезные описания этих группировок. Модели кластеризации часто используются для создания кластеров или сегментов, которые используются в качестве входных данных при последующем анализе (например, при разделении потенциальных покупателей по однородным подгруппам).

### *Узлы сегментации*



Узел автоматической кластеризации оценивает и сравнивает модели кластеризации, идентифицирующие группы записей со сходными характеристиками. Этот узел работает аналогично другим узлам автоматического моделирования, допуская экспериментирование с несколькими комбинациями опций при одном проходе моделирования. Модели можно сравнивать при помощи базовых показателей, пытаясь фильтровать и ранжировать с их использованием полезность моделей кластеризации и предоставить показатель на основе важности конкретных полей.



Узел К-средних кластеризует набор данных в отдельные группы (или кластеры). Этот метод определяет фиксированное количество кластеров, итерационно распределяет записи по кластерам и настраивает центры кластеров, пока дальнейшие уточнения более не улучшают модель. Вместо попытки предсказать выходное значение  $k$ -средние используют процесс, называемый неконтролируемым обучением, чтобы обнаружить структуры в наборе входных полей.



Узел Коонена генерирует тип нейросети, которую можно использовать для кластеризации набора данных в отдельные группы. Когда сеть полностью обучена, похожие записи должны быть близко друг от друга на выходной карте, а отличающиеся записи должны быть сильно разделены. По количеству наблюдений, захваченных каждым нейроном в слепке модели, можно определить сильные нейроны. Это может дать представление об оправданном количестве кластеров.



Узел Двухшаговый использует метод двухшаговой кластеризации. На первом шаге проводится первый проход по данным, при котором необработанные входные данные сжимаются в управляемый набор подкластеров. На втором шаге используется способ иерархической кластеризации для все большего слияния подкластеров в крупные и еще более крупные кластеры. У двухшагового метода есть преимущество автоматической оценки оптимального числа кластеров для обучающих данных. Он может эффективно обрабатывать поля смешанных типов и большие наборы данных.



Узел выявления аномалий определяет необычные наблюдения, или выбросы, которые не соответствуют структуре “нормальных” данных. При помощи этого узла можно находить выбросы даже в том случае, если они не подходят ни под какие ранее известные шаблоны или вы точно не уверены, что именно ищете.

## Модели исследования данных в базе данных

IBM SPSS Modeler поддерживает интеграцию с инструментами исследования данных и моделирования, доступных от поставщиков баз данных, в том числе Oracle Data Miner, IBM DB2 InfoSphere Warehouse и Microsoft Analysis Services. Построение моделей, их скоринг и сохранение в базе данных - все эти операции возможны в прикладной программе IBM SPSS Modeler. Полную информацию смотрите в *Руководстве по исследованию данных в базах данных IBM SPSS Modeler*, доступном на установочном носителе.

## IBM SPSS Statistics Модели

Если есть копия IBM SPSS Statistics, установленная и лицензированная на вашем компьютере, можно получить доступ к определенным подпрограммам IBM SPSS Statistics из IBM SPSS Modeler, чтобы построить и оценить модели.

## Дополнительная информация

Доступна также подробная документация об алгоритмах моделирования. Дополнительную информацию смотрите в *Руководстве по алгоритмам IBM SPSS Modeler*, доступном на установочном носителе продукта.

---

## Примеры исследования данных

Лучший способ познакомиться с исследованием данных на практике - это запустить пример. Несколько примеров прикладных программ доступны в *Руководстве по прикладным программам IBM SPSS Modeler*, которое предоставляет краткое и сконцентрированное на выделенной задаче введение в конкретные способы и приемы моделирования. Дополнительную информацию смотрите в разделе “Примеры прикладных программ” на стр. 4.



---

## Глава 5. Создание потоков

---

### Обзор создания потока

Исследование данных при помощи IBM SPSS Modeler - это прежде всего последовательная обработка данных на ряде узлов, называемая **потоком**. Узлы представляют операции над данными, а стрелки между узлами - направление передачи данных. Обычно поток данных служит для считывания данных в IBM SPSS Modeler, ряда последовательных операций обработки данных и отправка результатов в объект назначения, такой как таблица или средство просмотра.

Например, пусть вам нужно открыть источник данных, добавить новое поле, выбрать записи с учетом значений новом поле и вывести результаты в таблице. В этом случае поток данных будет состоять из четырех узлов:



Узел файла переменных, который служит для считывания данных из источника данных.



Узел вычислений, который служит для добавления в набор данных нового, производного поля.



Узел выбора, который служит для задания критериев извлечения записей из потока данных.



Узел таблицы, который служит для вывода результатов обработки на экран.

---

### Построение потоков данных

В уникальном интерфейсе IBM SPSS Modeler можно исследовать данные, работая с наглядными диаграммами потоков данных. В общих чертах поток строится так:

- Добавьте узлы на холст потока.
- Соедините узлы в поток.
- Задайте опции узлов и потока.
- Выполните поток.

В этом разделе приводится более подробная информация о работе с узлами при создании более сложных потоков данных. Также рассматриваются опции и настройки для узлов и потоков. Пошаговую инструкцию по созданию потока с использованием данных в комплекте поставки IBM SPSS Modeler (в подпапке *Демонстрационные материалы* папки установки программы) смотрите в разделе “Примеры прикладных программ” на стр. 4.

### Работа со узлами

Узлы служат в IBM SPSS Modeler для наглядного изучения данных. Различные узлы в рабочем пространстве представляют различные объекты и действия. Палитра в нижней части окна IBM SPSS Modeler содержит все узлы, используемые при построении потока.

Существует несколько типов узлов. **Узлы источника** передают данные в поток; эти узлы расположены на вкладке Источники на палитре узлов. **Узлы обработки** выполняют операции с отдельными записями и полями; эти узлы находятся на палитре на вкладках Операции и Операции с полями. **Узлы вывода** создают различные представления данных, диаграмм и результатов моделирования; их вкладки на палитре узлов называются Диаграммы, Вывод и Экспорт. **Узлы моделирования** используют статистические алгоритмы для создания слепков моделей; они расположены на палитре узлов на вкладке Моделирование, а также на вкладке Моделирование в базе данных, если последняя активна. Дополнительную информацию смотрите в разделе “Палитра узлов” на стр. 16.

Узлы соединяются друг с другом, образуя потоки, которые выполняются, что наглядно представить взаимосвязи и вывести заключения. Потоки аналогичны сценариям -- их можно сохранять и повторно использовать с другими файлами данных.

Исполняемый узел, обрабатывающий данные потока, называется **конечным узлом**. Узел моделирования или узел вывода - это конечный узел, если расположен в конце потока или ветви потока. К конечному узлу нельзя присоединить дальнейшие узлы.

*Примечание:* палитру Узлы можно настраивать. Дополнительную информацию смотрите в разделе “Настройка палитры узлов” на стр. 149.

## Добавление узлов в поток

Есть несколько способов добавить узлы в поток из палитры узлов:

- Дважды щелкните по узлу на палитре. *Примечание:* двойной щелчок по узлу автоматически соединяет его с текущим потоком. Дополнительную информацию смотрите в разделе “Соединение узлов в потоке”.
- Перетащите узел из палитры и отпустите его на холсте потока.
- Щелкните по узлу на палитре, затем щелкните по холсту потока.
- Выберите нужную опцию в меню Вставка IBM SPSS Modeler.

Добавив узел на холст, дважды щелкните по этому узлу, чтобы открыть его диалоговое окно. Доступные опции зависят от типа добавляемого узла. Для получения информации о конкретных элементах управления в диалоговом окне, нажмите кнопку **Справка** при этом элементе.

### Удаление узлов

Чтобы удалить узел из потока данных, щелкните по нему и либо нажмите клавишу Delete, либо щелкните правой кнопкой и выберите **Удалить** в меню.

## Соединение узлов в потоке

Узлы, добавленные на холст потока, образуют поток данных только после того, как будут соединены. Соединения между узлами показывают направление передачи данных с одной операции на следующую. Есть ряд способов задать соединение узлов, чтобы образовать поток: при помощи двойного щелчка, средней кнопки мыши или вручную.

### Добавить и соединить узлы при помощи двойных щелчков

Самый простой способ образовать поток - это щелкать дважды по узлам на палитре. Этот метод автоматически соединяет новый узел с выделенным узлом на холсте потока. Например, если на холсте есть узел базы данных, то вы можете выделить его, а затем дважды щелкнуть по следующему узлу на палитре, например, узлу вычислений. Узел вычислений будет автоматически соединен с существующим узлом базы данных. Продолжая этот процесс, вы дойдете до конечных узлов, таких как узел гистограммы и табличный узел, после чего все новые узлы начнут соединяться с последним неконечным узлом, расположенным выше по потоку.

### Соединить узлы при помощи средней кнопки мыши

На холсте потока можно щелкнуть и протащить от одного узла до другого средней кнопкой мыши. (Если у вашей мыши нет средней кнопки, можно вместо этого нажимать клавишу Alt во время перетаскивания от одного узла до другого.)

Соединить узлы вручную

Если у вас нет средней кнопки мыши и вы предпочитаете соединять узлы вручную, можете во всплывающем меню узла выбрать соединение с другим узлом, уже существующем на холсте.

1. Щелкните правой кнопкой по узлу, с которого хотите начать соединение. Откроется меню узла.
2. В меню выберите **Соединить**.
3. Значок соединения появится на начальном узле и на указателе. Щелкните по второму узлу на холсте, и два узла будут соединены.

Несколько советов при соединении узлов. Вы получите сообщение об ошибке, если попытаетесь выполнить любой из следующих типов соединения:

- Соединение, ведущее в узел источника
- Соединение, исходящее из конечного узла
- Превышение максимально допустимого числа входящих соединений для некоторого узла
- Соединение уже соединенных узлов
- Заикливание (передача данных на узел, через который они уже проходили)

## Обход узла в потоке

При обходе узла в потоке данных все его входящие и исходящие стрелки заменяются на стрелки, которые ведут непосредственно из узлов в началах его входящих стрелок в узлы на концах его исходящих стрелок. Если у узла не было ни одной входящей или ни одной исходящей стрелки, то при обходе этого узла все его стрелки не перенаправляются, а просто удаляются.

Пусть, например, поток вычисляет новое поле, фильтрует поля и затем выводит обзор результатов в гистограмме и в таблице. Если теперь нужно показать соответствующую диаграмму и таблицу с данными до фильтрации полей, можно добавить в поток узел Гистограмма или Таблица, а можно обойти узел Фильтр. Если обойти узел Фильтр, стрелки, которые приходят в диаграмму и таблицу, пройдут непосредственно от узла вычислений. Узел Фильтр с потоком рассоединяется.

Чтобы обойти узел

1. На холсте потока дважды щелкните средней кнопкой мыши по узлу, который нужно обойти. Другой вариант - использовать Alt+двойной щелчок.

*Примечание:* это действие можно отменить командой **Откат** в меню Правка или нажатием Ctrl+Z.

## Выключение узлов в потоке

В потоках можно выключать узлы обработки с одним входом; в результате выключенный узел игнорируется при выполнении потока. Экономятся усилия на удаление или обход узла; узел и его связи с другими узлами не удаляются. По-прежнему можно открыть и отредактировать параметры узла, хотя внесенные изменения подействуют только после того, как вы снова включите этот узел.

Пусть, например, поток фильтрует несколько полей, а затем строит модели по уменьшенному набору данных. Если, кроме того, нужно построить такие же модели *без фильтрации* полей, чтобы посмотреть, улучшает ли фильтрация результаты моделей, можно просто выключить узел фильтра. После того, как вы выключите узел фильтра, данные потекут по узлам моделирования непосредственно с узла вычислений на узел типа.

Чтобы выключить узел

1. На холсте потока щелкните правой кнопкой по узлу, который нужно выключить.

2. Во всплывающем меню выберите **Выключить узел**.

Другой вариант - в меню Правка выбрать **Узел > Выключить узел**. Когда будет нужно вернуть узел в поток, выберите **Включить узел** аналогичным образом.

*Примечание:* это действие можно отменить командой **Откат** в меню Правка или нажатием **Ctrl+Z**.

## Добавление узлов в существующих соединениях

Можно добавить новый узел между двумя соединенными узлами путем перетаскивания соединительной стрелки.

1. Пользуясь средней кнопкой мыши, нажмите на соединительную стрелку, которую хотите разбить узлом, и начните ее перетаскивать. Вариант без средней кнопки мыши - выполнить нажатие и перетаскивание, удерживая нажатой клавишу **Alt**.
2. Перетащите соединительную стрелку на узел, которым хотите ее разбить, и отпустите кнопку мыши.

*Примечание:* Вы можете отказаться от разбивающего узла и восстановить прямое соединение, выполнив **обход узла**.

## Удаление соединений между узлами

Чтобы удалить соединение между двумя узлами:

1. Щелкните правой кнопкой по стрелке соединения.
2. В меню выберите **Удалить соединение**.

Чтобы удалить все входящие и исходящие соединения узла, сделайте одно из следующего:

- Выберите узел и нажмите **F3**.
- Выберите узел и в главном меню выберите:

**Правка > Узел > Отсоединить**

## Задание опций для узлов

Создав и соединив узлы, вы можете настроить для них ряд опций. Щелкните правой кнопкой по узлу и выберите один из пунктов меню.

- Выберите **Правка**, чтобы открыть диалоговое окно для выбранного узла.
- Выберите **Соединить**, чтобы вручную соединить один узел с другим.
- Выберите **Отсоединить**, чтобы удалить все входящие и исходящие связи узла.
- Выберите **Переименовать и аннотировать**, чтобы открыть вкладку Аннотации в диалоговом окне редактирования.
- Выберите **Новый комментарий**, чтобы добавить комментарий, относящийся к узлу. Дополнительную информацию смотрите в разделе “Добавление комментариев и аннотаций к узлам и потокам” на стр. 54.
- Выберите **Отключить узел**, чтобы “скрыть” узел при обработке. Чтобы снова сделать узел видимым при обработке, выберите **Включить узел**. Дополнительную информацию смотрите в разделе “Выключение узлов в потоке” на стр. 37.
- Выберите **Вырезать** или **Удалить**, чтобы удалить один или несколько выбранных узлов с холста потока.  
*Примечание:* Если выбрать **Вырезать**, то эти узлы можно потом вставить; после опции **Удалить** этого сделать нельзя.
- Выберите **Копировать узел**, чтобы создать копию узла без соединений. Скопированное можно добавить в новый или существующий поток.
- Выберите **Загрузить узел**, чтобы открыть ранее сохраненный узел и загрузить его опции в текущий выбранный узел. *Примечание:* Эти узлы должны быть одного типа.
- Выберите **Получить узел**, чтобы получить узел из подсоединенного IBM SPSS Collaboration and Deployment Services Repository.

- Выберите **Сохранить узел**, чтобы сохранить подробности узла в файл. Загрузить подробности узла можно только в другой узел того же типа.
- Выберите **Записать узел**, чтобы записать выбранный узел в подсоединенный IBM SPSS Collaboration and Deployment Services Repository.
- Выберите **Кэш**, чтобы раскрыть меню с опциями кэширования выбранного узла.
- Выберите **Отображение данных**, чтобы раскрыть меню с опциями отображения данных на новый источник или задания обязательных полей.
- Выберите **Создать надузел**, чтобы раскрыть меню с опциями создания надузла в текущем потоке.
- Выберите **Сгенерировать узел ввода пользователя**, чтобы заменить существующий узел. Примеры, сгенерированные этим узлом, будут содержать те же поля, что и текущий узел.
- Выберите **Запустить отсюда**, чтобы выполнить все конечные узлы ниже по потоку от выбранного узла.

## Опции кэширования для узлов

Чтобы оптимизировать выполнение потока, для любого неконечного узла можно задать **кэш**. После того, как для узла задан кэш, при следующем выполнении потока данных этот кэш будет заполнен прошедшими через этот узел данными. С этого момента данные будут считываться из этого кэша (который хранится на диске во временном каталоге), а не из источника данных.

Кэширование особенно полезно после операции, которая требует много времени, такой как сортировка, объединение или суммирование. Например, пусть у вас есть узел источника, сконфигурированный для чтения данных о продаже из базы данных, и узел суммирования, который вычисляет сводку продаж по положению. Имеет смысл задать кэш для узла суммирования, а не для узла источника, чтобы кэшировать сводные данные, а не весь набор данных.

*Примечание:* Кэширование на узлах источника, то есть простое копирование исходных данных, прочитанных в IBM SPSS Modeler, в большинстве случаев не повышает производительность.

Узлы, для которых включено кэширование, помечаются значком документа в верхнем правом углу. Если в кэше узла есть данные, этот значок документа зеленый.

Чтобы включить кэш

1. На холсте потока щелкните правой кнопкой по узлу и в меню выберите **Кэш**.
2. В подменю кэширования выберите **Включить**.
3. Выключить кэш можно, щелкнув правой кнопкой по узлу и выбрав **Выключить** в подменю кэширования.

Кэширование узлов в базе данных

Для потоков, выполняемых в базе данных, возможно кэширование промежуточных данных во временную таблицу в базе данных, а не в файловой системе. В сочетании с оптимизацией SQL возможен значительный выигрыш в производительности. Например, можно кэшировать и повторно использовать вывод потока, объединяющего несколько таблиц для создания производной таблицы исследования данных. Можно дополнительно повысить производительность, автоматически сгенерировав SQL для всех узлов ниже по потоку.

При использовании кэширования в базе данных для строковых значений длиннее 255 символов нужно либо убедиться, что выше по потоку есть узел типа и значения поля считаны, либо задать длину строки в параметре `default_sql_string_length` в файле `options.cfg`. Тогда для соответствующего столбца во временной таблице будет задана ширина, достаточная для таких строковых значений.

Чтобы воспользоваться преимуществами кэширования в базе данных, нужно разрешить и оптимизацию SQL, и кэширование в базе данных. Обратите внимание на то, что параметрами оптимизации сервера переопределяются параметры клиента. Дополнительную информацию смотрите в разделе “Задание опций оптимизации для потоков” на стр. 44.

Если кэширование в базе данных разрешено, то просто щелкните правой кнопкой по любому неконечному узлу, чтобы кэшировать данные в этой точке, и кэш автоматически будет создан непосредственно в базе данных при следующем выполнении потока. Если же кэширование в базе данных или оптимизация SQL не разрешены, кэш будет записан в файловую систему.

*Примечание:* Временные таблицы для кэширования поддерживаются в следующих базах данных: DB2, Netezza, Oracle, SQL Server и Teradata. В остальных базах данных при кэшировании используются обычные таблицы. Код SQL можно оптимизировать под конкретную базу данных - за помощью обратитесь в службу поддержки.

Чтобы очистить кэш

Значок белого документа на узле показывает, что его кэш пуст. Когда кэш заполнен, значок документа закрашен зеленым. Чтобы изменить содержимое кэша, сначала нужно очистить кэш, а затем снова запустить поток данных, который заполнит этот кэш.

1. На холсте потока щелкните правой кнопкой по узлу и в меню выберите **Кэш**.
2. В подменю кэширования выберите **Очистка**.

Чтобы сохранить кэш

Содержимое кэша можно сохранить как файл данных IBM SPSS Statistics (\*.sav). После этого можно перезагрузить файл как кэш или сконфигурировать узел, для которого файл кэша будет источником данных. Кроме того, можно загрузить кэш, сохраненный из другого проекта.

1. На холсте потока щелкните правой кнопкой по узлу и в меню выберите **Кэш**.
2. В подменю кэширования выберите **Сохранить кэш**.
3. В диалоговом окне Сохранить кэш перейдите в положение, куда нужно сохранить файл кэша.
4. Введите имя в текстовом поле Имя файла.
5. Убедитесь, что в списке Тип файлов выбрано \*.sav, и нажмите кнопку **Сохранить**.

Чтобы загрузить кэш

Если вы сохранили файл кэша прежде, чем удалить его из узла, этот кэш можно загрузить повторно.

1. На холсте потока щелкните правой кнопкой по узлу и в меню выберите **Кэш**.
2. В подменю кэширования выберите **Загрузить кэш**.
3. В диалоговом окне Загрузить кэш перейдите в положение файла кэша, выберите его и нажмите кнопку **Загрузить**.

## Предварительный просмотр данных на узлах

Чтобы убедиться, что при построении потока данные изменяются так, как ожидалось, вы можете на всех важных этапах пропустить данные через узел Таблица. Чтобы сэкономить эти усилия, есть возможность сгенерировать с каждого узла предварительный просмотр, содержащий выборку создаваемых данных, что уменьшает время, затраченное на построение каждого узла.

Для узлов выше по потоку от слепка модели предварительный просмотр содержит входные поля; для слепка модели или узлов ниже по потоку от слепка (кроме конечных узлов) предварительный просмотр содержит входные и сгенерированные поля.

По умолчанию выводится 10 строк таблицы, но это число можно изменить в свойствах потока. Дополнительную информацию смотрите в разделе “Задание общих опций потоков” на стр. 42.

В меню Сгенерировать можно создать несколько типов узлов.

## Блокировка узлов

Чтобы не позволить другим пользователям вносить коррективы в настройки одного или нескольких узлов в потоке, можно поместить этот узел или узлы в особый тип под названием надузел, а затем заблокировать надузел, защитив его паролем.

## Работа с потоками

После того, как вы соединили на холсте потока узлы источника, обработки и конечные узлы, вы создали поток. Потоки как собрания узлов можно сохранять, аннотировать и добавлять в проекты. Можно также задавать многочисленные опции для потоков, такие как оптимизация, настройка даты и времени, параметры и сценарии. Эти свойства рассматриваются в приведенных ниже разделах.

В IBM SPSS Modeler можно использовать и модифицировать несколько потоков данных в одном сеансе IBM SPSS Modeler. Справа в главном окне находится панель менеджеров, которая помогает в навигации по открытым в настоящее время потокам, выводам и моделям. Если вы не видите панель менеджеров, выберите **Менеджеры** в меню Вид, а затем щелкните по вкладке **Потоки**.

На этой вкладке можно:

- Открывать потоки.
- Сохранять потоки.
- Сохранять потоки в текущий проект.
- Закрывать потоки.
- Открывать новые потоки.
- Записывать и получать потоки в репозитории IBM SPSS Collaboration and Deployment Services (если доступно на вашей площадке). Дополнительную информацию смотрите в разделе “О программе IBM SPSS Collaboration and Deployment Services Repository” на стр. 115.

Для доступа к этим возможностям щелкните правой кнопкой по потоку на вкладке Потоки.

### Задание опций для потоков

Можно задать ряд опций, применяемых в текущем потоке. Кроме того, можно сохранить эти опции как значения по умолчанию, чтобы применять ко всем вашим потокам. Доступны следующие опции.

- **Основные.** Различные опции, такие как символьные и текстовые обозначения в потоке. Дополнительную информацию смотрите в разделе “Задание общих опций потоков” на стр. 42.
- **Дата/Время.** Опции, связанные с форматом выражений даты и времени. Дополнительную информацию смотрите в разделе “Задание опций даты и времени для потоков” на стр. 43.
- **Форматы чисел.** Опции, управляющие форматом числовых выражений. Дополнительную информацию смотрите в разделе “Задание опций формата чисел для потоков” на стр. 43.
- **Оптимизация.** Опции для оптимизации производительности потока. Дополнительную информацию смотрите в разделе “Задание опций оптимизации для потоков” на стр. 44.
- **Запись в журнал и вывод состояния.** Опции, управляющие записью SQL в журнал и выводом состояния обработки записей. Дополнительную информацию смотрите в разделе “Задание опций записи SQL в журнал и выводом состояния обработки записей для потоков” на стр. 45.
- **Разметка.** Опции, относящиеся к представлению потока на холсте. Дополнительную информацию смотрите в разделе “Задание опций представления потоков” на стр. 46.

Чтобы задать опции потока

1. В меню Файл выберите **Свойства потока** (или выберите поток на вкладке потоков на панели менеджеров, щелкните правой кнопкой и выберите во всплывающем меню **Свойства потока**).
2. Щелкните по вкладке **Параметры**.

Другой вариант - в меню Инструменты выберите:

## Свойства потока > Опции

**Задание общих опций потоков:** Общие опции - это набор различных опций, применяемых к различным аспектам текущего потока.

**Десятичный разделитель.** Выберите запятую (,) или точку (.) как десятичный разделитель.

**Символ группировки.** Для форматов вывода чисел можно выбрать символ для разбиения цифр на тройки (например, пробел в такой записи, как 3 000.00). Доступные опции - не разбивать на тройки, точка, запятая, пробел и выбор по национальной версии (в этом случае используется значение по умолчанию из текущей национальной версии).

**Кодировка.** Задайте метод потока по умолчанию для кодировки текста. (*Примечание:* Применяется только к узлу исходного файла переменных и узлу экспорта в плоский файл. Остальные узлы эту опцию не используют; в большинстве файлов есть встроенная информация о кодировке.) Можно выбрать значение системы по умолчанию или UTF-8. Значение по умолчанию задается на Панели управления Windows или, при работе в распределенном режиме, на компьютере сервера. Дополнительную информацию смотрите в разделе “Поддержка Unicode в IBM SPSS Modeler” на стр. 171.

**Оценка набора правил.** Определяет, как оцениваются модели набора правил. По умолчанию наборы правил проводят **Голосование**, чтобы объединить предсказания отдельных правил в итоговое предсказание. Чтобы наборы правил пользовались по умолчанию методом первого совпадения, выберите **Первое совпадение**. Имейте в виду, что эта опция не применяется к моделям списка решений, где согласно алгоритму всегда используется метод первого совпадения.

**Максимальное количество строк для вывода при предварительном просмотре данных.** Задайте число строк, выводимых при предварительном просмотре данных узла. Дополнительную информацию смотрите в разделе “Предварительный просмотр данных на узлах” на стр. 40.

**Максимальное число элементов для номинальных полей.** Выберите, чтобы задать максимальное число элементов для номинальных полей (полей набора значений), после которого тип данных поля становится **Без типа**. Эта опция полезна при работе с громоздкими наборами номинальных полей. *Примечание:* Когда для поля задается тип измерений **Без типа**, автоматически задается роль поля **Никакой**. Это значит, что такие поля недоступны для моделирования.

**Ограничить размер набора для моделирования Коонена и k-средних.** Выберите, чтобы задать максимальное число элементов для номинальных полей, используемых в моделировании сетей Коонена и *k*-средних. По умолчанию размер набора 20; после этого числа поле игнорируется и генерируется предупреждение с информацией о проблемном поле.

Заметим, что ради обратной совместимости эта опция применяется также к старому узлу нейросети, который был заменен начиная с IBM SPSS Modeler версии 14; этот узел может встретиться в унаследованных потоках.

**Обновить источники данных при выполнении.** Выберите, чтобы автоматически обновить все узлы источников при выполнении потоков. Это действие аналогично нажатию кнопки **Обновить** на узле источника, только этой опцией автоматически обновляются все узлы источников текущего потока, кроме узлов ввода пользователя.

*Примечание:* Выбор этой опции очищает кэши узлов ниже по потоку, даже если данные не изменились. Но очистка выполняется только один раз при каждом выполнении потока, так что кэши узлов ниже по потоку все равно могут послужить полезным временным хранилищем в пределах одного выполнения. Например, пусть вы задали кэш в середине потока после операции сложного вычисления и у вас есть несколько диаграмм и отчетов, расположенных ниже по потоку от узла вычислений. При выполнении потока кэш узла вычислений будет очищен и заново заполнен, но только для первой диаграммы или отчета. остальные конечные узлы будут считывать данные из кэша узла вычислений.



**Показать метки полей и значений в выводе.** Выводит метки полей и значений в таблицах, на диаграммах и в других объектах вывода. Если метки не существуют, вместо них выводятся имена полей и значения данных. По умолчанию метки выключены, но их можно включать и выключать по отдельности во всех остальных местах IBM SPSS Modeler. Можно также задать вывод меток в окне вывода, нажав кнопку переключателя на панели инструментов.



Рисунок 11. Значок на панели инструментов для переключения меток полей и значений

**Показать число выполнений.** После выполнения потока выводит на вкладке Число выполнений число выполнений по отдельности для всех узлов потока. Дополнительную информацию смотрите в разделе “Просмотр числа выполнений узла” на стр. 47.

**Сохранить для использования по умолчанию.** Заданные опции применяются только к текущему потоку. Нажмите эту кнопку, чтобы задать эти опции как опции по умолчанию для всех потоков.

**Задание опций даты и времени для потоков:** Эти опции задают формат различных выражений даты и времени в текущем потоке.

**Импортировать дату/время как.** Выберите, обрабатывать ли поля даты-времени как тип данных даты-времени или импортировать такие поля как строковые переменные.

**Формат даты.** Выберите формат даты для хранения полей даты или для обработки строк как значений даты в функциях даты CLEM.

**Формат времени.** Выберите формат времени для хранения полей времени или для обработки строк как значений времени в функциях времени CLEM.

**Дни/минуты разворачивания.** Для форматов времени выберите, интерпретировать ли отрицательные значения разности времени как момент в предыдущем дне или часе.

**Базовый уровень дат (1 янв).** Выберите годы базового уровня (день всегда 1 января), которые будут использоваться функциями даты CLEM, работающими с одной датой.

**Двузначное обозначение года, начиная с.** Задайте крайний год для добавления цифр столетия к годам, заданных двузначными числами. Например, если крайний год - 1930, то в дате 05/11/02 год будет интерпретироваться как 2002. Соответственно будут интерпретироваться и годы двадцатого века после 30; следовательно, дата 05/11/73 будет отнесена к 1973 году.

**Сохранить для использования по умолчанию.** Заданные опции применяются только к текущему потоку. Нажмите эту кнопку, чтобы задать эти опции как опции по умолчанию для всех потоков.

**Задание опций формата чисел для потоков:** Эти опции задают формат различных числовых выражений в текущем потоке.

**Формат вывода чисел.** Можно выбрать стандартный формат вывода чисел (####.###), экспоненциальный (#####E+###) или формат валюты (\$###.##).

**Число десятичного знаков (в стандартном формате, экспоненциальном, в формате валюты).** Для ряда форматов вывода задает число десятичных знаков при выводе на экран или на печать действительных чисел. Эта опция задается отдельно для каждого формата вывода.

**Вычисления в.** Выберите радианы или градусы как единицу измерения в тригонометрических выражениях CLEM. Дополнительную информацию смотрите в разделе “Тригонометрические функции” на стр. 95.

**Сохранить для использования по умолчанию.** Заданные опции применяются только к текущему потоку. Нажмите эту кнопку, чтобы задать эти опции как опции по умолчанию для всех потоков.

**Задание опций оптимизации для потоков:** С помощью параметров оптимизации можно оптимизировать производительность потока. Обратите внимание на то, что параметры производительности и оптимизации на сервер IBM SPSS Modeler (если они используются) переопределяют аналогичные параметры на клиенте.

*Примечание:* Моделирование баз данных и оптимизация SQL требуют, чтобы на компьютере IBM SPSS Modeler была включена возможность соединения с сервером IBM SPSS Modeler. При включенной возможности соединения можно обращаться к алгоритмам баз данных, выполнять обратный перенос SQL непосредственно с клиента сервером IBM SPSS Modeler и обращаться к серверу IBM SPSS Modeler. Проверьте текущее состояние лицензии, для чего в меню клиента сервером IBM SPSS Modeler выберите:

**Справка > О программе > Дополнительные подробности**

Если возможность соединения включена, на вкладке Состояние лицензии вы увидите опцию **Разрешение для сервера**.

Дополнительную информацию смотрите в разделе “Соединение с сервер IBM SPSS Modeler” на стр. 12.

*Примечание:* Поддержка обратного переноса SQL и оптимизации зависит от типа используемой базы данных. Наиболее свежие сведения о базах данных и драйверах ODBC, поддерживаемых и проверенных для использования с IBM SPSS Modeler 16, смотрите на сайте технической поддержки <http://www.ibm.com/support>.

**Включить перезапись в потоке.** Выберите эту опцию, чтобы разрешить перезапись потока в IBM SPSS Modeler. Доступны четыре типа перезаписи, можно выбрать один или оба. Перезапись потока изменяет порядок узлов в потоке в фоновом режиме, обеспечивая более эффективную работу без изменения семантики потока.

- **Оптимизировать построение SQL.** Эта опция позволяет изменять порядок узлов в потоке, чтобы передавать больше операций при помощи генерирования SQL для выполнения в базу данных. При обнаружении узла, который не может быть переведен в SQL, оптимизатор будет искать далее нисходящие узлы, которые можно перевести в SQL и безопасно переставить в положение перед проблемным узлом, не влияя на семантику потока. База данных может выполнять операции эффективнее, чем IBM SPSS Modeler, а кроме того, такие переносы помогают сократить размер набора данных, возвращаемого в IBM SPSS Modeler для обработки. Это, в свою очередь, может сократить сетевой трафик и ускорить операции потока. Обратите внимание на то, что оптимизация SQL не окажет влияния, если не включить переключатель **Генерировать SQL**.
- **Оптимизировать выражение CLEM.** Эта опция позволяет оптимизатору искать выражения CLEM, которые можно предварительно обработать до запуска потока, чтобы увеличить скорость работы. Простой пример - если у вас есть выражение типа  $\log(\text{salary})$ , оптимизатор вычислит фактическое значение salary и передаст его на обработку. Это можно использовать, чтобы улучшить обратный перенос SQL и повысить производительность сервера IBM SPSS Modeler.
- **Оптимизировать выполнение синтаксиса.** Этот метод перезаписи потока повышает эффективность операций, в которых участвует несколько узлов, содержащих синтаксис IBM SPSS Statistics. Оптимизация достигается путем объединения команд синтаксиса в одну операцию вместо выполнения каждой операции по отдельности.
- **Оптимизировать другое выполнение.** Этот метод перезаписи потока повышает эффективность операций, которые не могут быть делегированы базе данных. Оптимизация достигается путем максимально раннего сокращения объема данных в потоке. Поток перезаписывается для перестановки операций ближе к источнику данных при одновременном поддержании целостности данных, уменьшая таким образом нисходящий поток данных для таких дорогостоящих операций, как объединения.

**Включить параллельную обработку.** При работе на компьютере с несколькими процессорами эта опция позволяет системе распределять нагрузку по этим процессорам, что позволяет повысить производительность. Параллельная обработка может быть выгодной при использовании нескольких узлов

или отдельных указанных ниже узлов: C5.0, Слияние (по ключевому слову), Сортировка, Интервал (ранговый и мозаичный методы) и Агрегирование (с использованием одного или нескольких ключевых полей).

**Генерировать SQL.** Выберите эту опцию, чтобы включить генерирование SQL, что позволит передать потоковые операции обратно в базу данных при помощи кода SQL и сгенерировать процессы выполнения, которые могут повысить производительность. Для дальнейшего повышения производительности можно выбрать **Оптимизировать генерацию SQL**, чтобы максимизировать число операций, переданных обратно в базу данных. Когда операции для узла будут переданы обратно в базу данных, узел после запуска потока будет выделен фиолетовым.

- **Кэширование баз данных.** Для потоков, генерирующих SQL, который будет выполняться в базе данных, данные могут попутно кэшироваться во временную таблицу в базе данных, а не в файловую систему. В сочетании с оптимизацией SQL это может привести к существенному выигрышу в производительности. Например, вывод потока, объединяющего несколько таблиц для создания представления исследования данных, может кэшироваться и при необходимости использоваться повторно. Если включено кэширование базы данных, просто щелкните правой кнопкой по любому неконечному узлу для кэширования данных в этой точке, и при следующем запуске потока кэш будет создан автоматически непосредственно в базе данных. Это позволяет генерировать SQL для нисходящих узлов, еще более повышая производительность. При необходимости эту опцию можно отключить, например, если правила политики или разрешения не допускают записи данных в базу данных. Если кэширование базы данных или оптимизация SQL не разрешены, кэш будет записан в файловую систему. Дополнительную информацию смотрите в разделе “Опции кэширования для узлов” на стр. 39.
- **Использовать упрощенное преобразование.** Эта опция включает преобразование данных из строк в числа или из чисел в строки, если они сохранены в подходящем формате. Например, если данные хранятся в базе данных в виде строки, но фактически содержат значимое число, они могут быть преобразованы для использования при обратном переносе.

*Примечание:* Ввиду небольших различий в реализации SQL, потоки, запущенные в базе данных, могут вернуть результаты, несколько отличные от возвращаемых при запуске в IBM SPSS Modeler. По аналогичным причинам эти различия могут также варьировать в зависимости от поставщика базы данных.

**Сохранить для использования по умолчанию.** Заданные опции применяются только к текущему потоку. Нажмите эту кнопку, чтобы задать эти опции как опции по умолчанию для всех потоков.

**Задание опций записи SQL в журнал и выводом состояния обработки записей для потоков:** Эти настройки включают в себя различные опции, управляющие выводом операторов SQL, сгенерированных потоком, и выводом числа записей, обработанных потоком.

**Выводить SQL в журнале сообщений во время выполнения потока.** Задаёт, передавать ли в журнал сообщений операторы SQL, сгенерированные при выполнении потока.

**Выводить подробности построения SQL в журнале сообщений во время подготовки потока.** При использовании предварительного просмотра потока задаёт, передавать ли в журнал сообщений предварительные версии операторов SQL.

**Показывать SQL.** Задаёт, должны ли все операторы SQL в журнале содержать собственные функции SQL или стандартные функции ODBC в формате {fn FUNC(...)}, которые генерируются IBM SPSS Modeler. Последний вариант опирается на функции драйвера ODBC, которые, возможно, не реализованы. Например, эта опция никак не повлияет на SQL Server.

**Переформатировать SQL для упрощения чтения.** Задаёт, нужно ли отформатировать SQL в журнале для лучшей читаемости.

**Выводить состояние записей.** Задаёт, когда следует сообщать о получении записей на конечных узлах. Задайте число, чтобы обновлять состояние каждые  $N$  записей.

**Сохранить для использования по умолчанию.** Заданные опции применяются только к текущему потоку. Нажмите эту кнопку, чтобы задать эти опции как опции по умолчанию для всех потоков.

**Задание опций представления потоков:** Эти настройки дают ряд возможностей оптимизировать вид холста потока и удобство работы с потоком.

**Минимальная ширина холста потока.** Задайте минимальную ширину холста потока в пикселях.

**Минимальная высота холста потока.** Задайте минимальную высоту холста потока в пикселях.

**Скорость прокрутки потоков.** Задайте скорость, с которой холст потока начинает прокручиваться при перетаскивании узла по холсту. Большие числа задают большую скорость прокрутки.

**Максимальная длина имен значков.** Задайте предельное число символов в именах узлов на холсте потока.

**Размер значков.** Выберите вариант, чтобы масштабировать все представление потока до одного из размеров от 8% до 200% стандартного размера значка.

**Размер ячеек сетки.** Выберите в списке размер ячейки сетки. Это число используется для выравнивания узлов на холсте потока по невидимой сетке. Размер ячейки сетки по умолчанию - 0,25.

**Привязать к сетке.** Выберите, чтобы выравнивать значки по невидимой сетке (по умолчанию выбрано).

**Размещение сгенерированных значков.** Выберите, в какое место на холсте помещаются значки узлов, сгенерированных из слепков модели. По умолчанию - сверху слева.

**Сохранить для использования по умолчанию.** Заданные опции применяются только к текущему потоку. Нажмите эту кнопку, чтобы задать эти опции как опции по умолчанию для всех потоков.

**Свойства потока Analytic Server (сервер аналитических служб):** Эти параметры обеспечивают несколько опций для работы Analytic Server (сервер аналитических служб).

#### **Максимальное число записей для обработки вне Analytic Server (сервер аналитических служб)**

Задайте максимальное количество записей, которые будут импортироваться на сервер SPSS Modeler с источника данных Analytic Server (сервер аналитических служб).

#### **Уведомление в случае невозможности обработки узла в Analytic Server (сервер аналитических служб)**

Этот параметр определяет, что произойдет, когда поток, переданный в Analytic Server (сервер аналитических служб), будет содержать узел, который нельзя обработать в Analytic Server (сервер аналитических служб). Укажите действия: выпустить предупреждение и продолжить обработку потока или остановить обработку и сообщить об ошибке.

#### **Параметры хранения моделей расщепления**

##### **Сохранить модели расщепления по ссылке в Analytic Server (сервер аналитических служб), когда превышен размер модели (МБайт)**

Слепки моделей обычно хранятся как часть потока. Модели расщеплений с множеством расщеплений могут создать большие слепки, и перемещение слепка от потока в Analytic Server (сервер аналитических служб) и обратно может повлиять на производительность. В качестве решения, когда модель расщеплений превосходит заданный размер, она хранится в Analytic Server (сервер аналитических служб), а слепок в SPSS Modeler содержит ссылку на модель.

##### **Папка по умолчанию для хранения моделей по ссылке в Analytic Server (сервер аналитических служб) после завершения выполнения**

Задайте путь по умолчанию, где вы хотите хранить модели расщеплений в Analytic Server (сервер аналитических служб). Этот путь должен начинаться с допустимого имени проекта Analytic Server (сервер аналитических служб).

### Папка для хранения постоянных моделей

Задайте путь по умолчанию, где вы хотите хранить "постоянные" модели. Постоянная модель не вычищается после завершения сеанса SPSS Modeler.

## Просмотр сообщений об операциях с потоком

Сообщения, касающиеся операций с потоком, таких как выполнение, оптимизация и время, затраченное на построение и оценку, удобно просматривать на вкладке Сообщения в диалоговом окне свойства потока.

Кроме того, в этой таблице приведены сообщения об ошибках.

Чтобы просмотреть сообщения о потоке

1. В меню Файл выберите **Свойства потока** (или выберите поток на вкладке потоков на панели менеджеров, щелкните правой кнопкой и выберите во всплывающем меню **Свойства потока**).
2. Щелкните по вкладке **Сообщения**.

Другой вариант - в меню Инструменты выберите:

### Свойства потока > Сообщения

Помимо сообщений, касающихся операций с потоком, здесь приводятся сообщения об ошибках. Когда выполнение потока прерывается из-за ошибки, это диалоговое окно открывается на вкладке сообщений, где видно сообщение об ошибке. Кроме того, узел с ошибками выделен красным на холсте потока.

Если оптимизация SQL и запись в журнал разрешены в диалоговом окне Опции пользователя, то выводится также информация о сгенерированных SQL. Дополнительную информацию смотрите в разделе “Задание опций оптимизации для потоков” на стр. 44.

Показанные здесь сообщения потока можно сохранить, выбрав **Сохранить сообщения** в выпадающем списке кнопки Сохранить (слева, под выступом вкладки Сообщения). Кроме того, можно очистить все сообщения для данного потока, выбрав **Очистить все сообщения** в списке кнопки Сохранить.

## Просмотр числа выполнений узла

На вкладке сообщений можно выбрать показ времени выполнения, где представлено время выполнения отдельно по каждому узлу потока, выполняемого на сервере IBM SPSS Modeler. Отметим, что время может быть неточным для потоков, выполняемых в других областях, таких как R или Analytic Server (сервер аналитических служб).

*Примечаний:* Чтобы эта возможность работала, должен быть включен переключатель **Показать время выполнения** в **Общих** настройках на вкладке **Опции**.

Таблице времени выполнения узлов содержит приведенные ниже столбцы. Щелкните по заголовку столбца, чтобы отсортировать записи в порядке возрастания или убывания (например, чтобы увидеть узлы с наибольшими временами выполнения).

**Конечный узел.** Идентификатор ветви, которой принадлежит узел. Идентификатором служит имя узла, которым оканчивается ветвь.

**Метка узла.** Имя узла, для которого приводится время выполнения.

**ID узла.** Уникальный идентификатор узла, для которого приводится время выполнения. Этот идентификатор генерируется системой при создании узла.

**Время выполнения.** Время в секундах, затраченное на выполнение этого узла.

## Задание параметров потока и сеанса

Параметры можно определить для использования в выражениях CLEM и в сценариях. На самом деле, это пользовательские переменные, которые сохранены и сохраняются с текущим потоком, сеансом или надузлом, и к которым можно обратиться из пользовательского интерфейса, а также из сценария. Например, если вы сохраняете поток, сохраняются и все параметры, заданные для этого потока. (Это отличает их от переменных локального сценария, которые можно использовать только в сценарии, в котором они объявлены). Параметры часто используются в сценариях для управления поведением сценария, задавая информацию о полях и значениях, которую не требуется жестко указывать в коде сценария.

Область использования параметра зависит от того, где он задан:

- Параметры потока можно задать в сценарии потока или в диалоговом окне свойств потока, и они доступны на всех узлах в потоке. Эти параметры выводятся в списке Параметры построителя выражений.
- Параметры сеансов можно задать в автономном сценарии или в диалоговом окне параметров сеанса. Они доступны для всех потоков, используемых в текущем сеансе (все потоки, перечисленные на вкладке Потоки панели менеджера).

Параметры можно задать также для надузлов, и в этом случае они видны только на узлах, инкапсулированных в этом надузле.

Чтобы задать параметры потока и сеанса в пользовательском интерфейсе

1. Чтобы задать параметры потока, выберите в главном меню:

**Инструменты > Свойства потока > Параметры**

2. Чтобы задать параметры сеанса, выберите **Задать параметры сеанса** в меню Инструменты.

**Вывести приглашение?.** Включите этот переключатель, если во время выполнения нужно предложить пользователю ввести значение этого параметра.

**Имя.** Здесь перечислены имена параметров. Введя имя в этом поле, можно создать новый параметр. Например, чтобы создать параметр для минимальной температуры, можно ввести `minvalue`. Не включайте префикс \$P-, обозначающий параметр в выражениях CLEM. Это имя используется также для вывода в Построителе выражений CLEM.

**Длинное имя.** Выводит описательное имя для каждого созданного параметра.

**Хранение.** Выберите тип хранения файла из списка. Хранение указывает способ хранения значений данных в параметре. Например, при работе со значениями, содержащими начальные нули, которые вы хотите сохранить (например, 008), в качестве типа хранения следует выбрать **Строка**. В противном случае нули будут убраны из значения. Доступно хранение типов строка, целое, действительное число, время, дата и отметка времени. Для параметров дат следует учитывать, что значения нужно задавать при помощи нотации стандарта ISO, как показано в следующей теме.

**Значение.** Выводит текущее значение для каждого созданного параметра. Настройте параметр нужным вам образом. Имейте в виду, что для параметров дат значения должны задаваться в нотации стандарта ISO (то есть: ГГГГ-ММ-ДД). Даты, заданные в других форматах, не принимаются.

**Тип (необязательно).** Если планируется внедрение потока во внешнюю прикладную программу, выберите в списке шкалу измерений. В противном случае рекомендуется оставить столбец *Тип* как есть. Если вы хотите задать для параметра ограничения значений, такие как верхняя и нижняя границы для числового диапазона, выберите в списке **Задать**.

Имейте в виду, что полное имя, опции хранения и типа можно задать для параметров только через пользовательский интерфейс. С помощью сценариев эти опции задать нельзя.

Для перемещения выбранного параметра вверх или вниз по списку доступных параметров используйте кнопки со стрелками в правой части окна. Для удаления выбранного параметра используйте кнопку удаления (с меткой X).

### **Задание приглашений для ввода значений параметров во время выполнения**

Если в некотором потоке в разных случаях требуется вводить различные значения одного и того же параметра, можно задать приглашения, в которых во время выполнения нужно будет ввести значение для одного или нескольких параметров потока или сеанса.

**Анализ важности независимых переменных.** (Необязательно) Можно ввести значение параметра, а можно оставить значение по умолчанию, если оно есть.

**Выключить эти приглашения.** Включите этот переключатель, если не хотите выводить эти приглашения при выполнении потока. Вы можете снова задать их вывод, включив переключатель **Выводить приглашение?** в диалоговом окне свойств потока или свойств сеанса, в котором эти параметры определяются. Дополнительную информацию смотрите в разделе “Задание параметров потока и сеанса” на стр. 48.

### **Задание ограничений значений для типа параметров**

При внедрении потока можно сделать ограничения значений для параметров доступными внешним прикладным программам, которые читают потоки данных моделирования. В этом диалоговом окне можно задать значения, доступные внешнему пользователю, выполняющему этот поток. В зависимости от типа данных ограничения значений в этом диалоговом окне динамически изменяются. Показанные здесь опции такие же, как те, что доступны для значений из узла типа.

**Тип.** Содержит тип измерений, выбранный в настоящее время. Вы можете изменить это значение согласно назначению параметра в IBM SPSS Modeler.

**Хранение.** Содержит хранимый тип данных, если известен. Типы данных не зависят от типа измерений (непрерывного, номинального или флага), выбираемых для работы в IBM SPSS Modeler. Хранимый тип данных можно изменить на главной вкладке параметров.

Нижняя часть этого диалогового окна динамически изменяется в зависимости от типа измерений, выбранного в поле **Тип**.

Непрерывные типы измерений

**Нижний.** Задайте нижний предел для значений параметра.

**Верхний.** Задайте верхний предел для значений параметра.

**Метки.** Можно задать метки для любых значений в поле диапазона. Нажмите кнопку **Метки**, чтобы открыть отдельное диалоговое окно для задания меток значений.

Номинальные типы измерений

**Значения.** В этой опции можно задать значения параметра, который будет использоваться как номинальное поле. Эти значения не будут навязываться в потоке IBM SPSS Modeler, но будут доступны в выпадающем списке при внешнем внедрении. Кнопками со стрелками и кнопкой удаления можно изменять существующие значения, расставлять их в новом порядке и удалять.

Типы измерений флага

**True.** Задайте значение флага для параметра, когда условие выполнено.

**False.** Задайте значение флага для параметра, когда условие не выполнено.

**Метки.** Можно задать метки для значений поля флага.

## Опции внедрения потока

На вкладке внедрения в диалоговом окне свойств потока можно задать опции внедрения потока как сценария в IBM SPSS Collaboration and Deployment Services с целью обновлять модели, автоматизировать расписание заданий или использовать в дальнейшем в IBM Analytical Decision Management. Для всех потоков перед внедрением сначала нужно указать ветвь оценки; дополнительные требования и опции зависят от типа внедрения. Дополнительную информацию смотрите в разделе “Сохранение и внедрение объектов репозитория” на стр. 116.

## Создание циклов выполнения для потоков

Используя вкладку Выполнение в диалоговом окне свойств потока, можно задать условия циклов для автоматизации повторяющихся задач в текущем потоке.

После задания этих условий их можно использовать как вводную часть написания сценария, когда в его окне вводятся главные элементы сценария для вашего потока, которые затем можно изменять, например, использовать как основание для построения лучших сценариев. Дополнительную информацию смотрите в разделе “Глобальные функции” на стр. 111.

Задать цикл для потока

1. В меню Файл выберите **Свойства потока** (или выберите поток на вкладке потоков на панели менеджеров, щелкните правой кнопкой и выберите во всплывающем меню **Свойства потока**).
2. Щелкните по вкладке **Выполнение**.
3. Выберите режим выполнения **Выполнение в цикле / с условиями**.
4. Перейдите на вкладку **Цикл**.

Другой вариант - в меню Инструменты выберите:

**Свойства потоков > Выполнение**

Еще один вариант - щелкнуть правой кнопкой мыши по узлу и в контекстном меню выбрать:

**Выполнение в цикле / с условиями > Изменить параметры цикла**

**Итерация.** Изменить значение номера этой строки нельзя, но можно добавить, удалить или переместить итерацию вверх или вниз, используя кнопки справа от таблицы.

**Заголовки таблицы.** Здесь выводятся ключ итерации и все созданные при конфигурировании цикла переменные итерации.

## Просмотр глобальных значений для потоков

На вкладке глобальных значений в диалоговом окне свойств потока можно просмотреть глобальные значения, заданные для текущей системы. Глобальные значения создаются при помощи узла Задать глобальные параметры, чтобы определить для выбранных полей статистические показатели, такие как среднее, сумма или стандартное отклонение.

После выполнения узла Задать глобальные параметры эти значения доступны для различных применений в операциях потока. Дополнительную информацию смотрите в разделе “Глобальные функции” на стр. 111.

Чтобы просмотреть глобальные значения для потока

1. В меню Файл выберите **Свойства потока** (или выберите поток на вкладке потоков на панели менеджеров, щелкните правой кнопкой и выберите во всплывающем меню **Свойства потока**).
2. Щелкните по вкладке **Глобальные параметры**.

Другой вариант - в меню Инструменты выберите:



## Свойства потока > Глобальные параметры

**Доступные глобальные.** В этой таблице содержатся доступные глобальные параметры. Значения здесь нельзя редактировать, но можно очистить все глобальные значения для потока кнопкой Очистить все значения справа от таблицы.

## Поиск узлов в потоке

Для поиска узлов в потоке доступен ряд критериев, таких как имя узла, категория и идентификатор. Эта возможность особенно полезна при работе со сложными потоками, содержащими большое число узлов.

Чтобы найти узлы в потоке

1. В меню Файл выберите **Свойства потока** (или выберите поток на вкладке потоков на панели менеджеров, щелкните правой кнопкой и выберите во всплывающем меню **Свойства потока**).
2. Щелкните по вкладке **Поиск**.

Другой вариант - в меню Инструменты выберите:

## Свойства потока > Поиск

Чтобы уточнить поиск, можно задавать сразу несколько опций, кроме поиска ID узла (при помощи поля **ID равен**), при котором остальные опции исключаются.

**Содержание метки узла.** Включите этот переключатель и введите целиком или частично метку искомого узла. Поиск не учитывает регистр букв, а несколько слов обрабатываются как единый фрагмент текста.

**Категория узла.** Включите этот переключатель и выберите категорию в списке, чтобы искать узел определенного типа. **Узел обработки** - это узел с вкладки Операции с записями или Операции с полями на палитре узлов; **Узел применения модели** - это узел слепка модели.

**Ключевые слова включают.** Включите этот переключатель и введите одно или несколько ключевых слов для поиска узлов, содержащих такой текст в поле Ключевые слова на вкладке аннотаций в диалоговом окне узла. Введенный текст ключевого слова должен точно соответствовать искомому. Можно задать несколько ключевых слов через точку с запятой, чтобы найти разные варианты (например, ввести протон;нейтрон, чтобы найти все узлы, содержащие хотя бы одно из этих ключевых слов. Дополнительную информацию смотрите в разделе “Аннотации” на стр. 58.

**Содержание аннотации.** Включите этот переключатель и введите одно или несколько слов для поиска узлов, содержащих такой текст в основном текстовом поле на вкладке аннотаций в диалоговом окне узла. Поиск не учитывает регистр букв, а несколько слов обрабатываются как единый фрагмент текста. Дополнительную информацию смотрите в разделе “Аннотации” на стр. 58.

**Генерирует поле с именем.** Включите этот переключатель и введите имя генерируемого поля (например, \$Медикамент\_С). При помощи этой опции можно искать узлы моделирования, генерирующие конкретное поле. Введите только одно имя поля, которое должно точно соответствовать искомому.

**ID эквивалентен.** Включите этот переключатель и введите ID искомого узла, чтобы найти конкретный узел с таким идентификатором (при выборе этой опции прошлые опции становятся недоступными). ID узлов назначаются системой при создании узла и могут использоваться для ссылки на узел в сценариях и средствах автоматизации. Введите только один ID узла, который должен точно соответствовать искомому. Дополнительную информацию смотрите в разделе “Аннотации” на стр. 58.

**Поиск в надузлах.** Этот переключатель включен по умолчанию и задает поиск узлов как вне, так и внутри надузлов. Выключите этот переключатель, если нужен поиск только среди узлов вне надузлов, то есть на верхнем уровне потока.

**Найти.** Задав все нужные опции, нажмите эту кнопку, чтобы начать поиск.

Узлы, соответствующие заданным опциям, выводятся в нижней части диалогового окна. Выберите узел в этом списке, чтобы выделить его на холсте потока.

## Переименование потоков

На вкладке аннотаций в диалоговом окне свойств потока можно добавить описательные аннотации для потока и создать для него пользовательское имя. Эти варианты особенно полезны при генерировании отчетов для потоков, добавленных на панель проектов. Дополнительную информацию смотрите в разделе “Аннотации” на стр. 58.

## Описания потока

Для каждого созданного вами потока IBM SPSS Modeler генерирует описание потока, содержащее информацию о содержимом потока. Это может быть полезно, когда вы пытаетесь понять, что делает этот поток, но у вас не установлен IBM SPSS Modeler, например, когда вы открываете поток в IBM SPSS Collaboration and Deployment Services.

Описание потока выводится в формате документа HTML, содержащего ряд разделов.

Общая информация о потоке

Этот раздел содержит имя потока, а также подробности и времени создания и последнего сохранения потока.

Описание и комментарии

Этот раздел включает в себя любое из следующего:

- Аннотации потока (смотрите раздел “Аннотации” на стр. 58)
- Комментарии, не связанные с отдельными узлами
- Комментарии, связанные с отдельными узлами в ветвях моделирования и оценки потока

Информация об оценках

Этот раздел содержит информацию под различными заголовками, относящимися к ветви оценок потока.

- **Комментарии.** Включает в себя комментарии, связанные только с узлами в ветви оценок.
- **Поля ввода.** Содержит поля ввода, а также их типы хранения (например, строковое, целое, действительное и так далее).
- **Поля вывода.** Содержит поля вывода, включая дополнительные поля, сгенерированные узлом моделирования, а также их типы хранения.
- **Анализ важности независимых переменных.** Содержит все параметры, относящиеся к ветви оценок потока и поддерживающие просмотр и редактирование при каждой оценке модели. Эти параметры выводятся при нажатии кнопки **Параметры оценок** на вкладке **Внедрение** в диалоговом окне свойств потока.
- **Узел модели.** Содержит имя и тип модели (например, нейросеть, дерево C&R и так далее). Это слепок модели, выбранный для поля **Узел модели** на вкладке **Внедрение** в диалоговом окне свойств потока.
- **Подробности модели.** Содержит подробности слепка модели, указанного под предыдущим заголовком. Где это возможно, показаны также важность предикторов и диаграммы оценок для модели.

Информация о моделировании

Содержит информацию, относящуюся к ветви моделирования потока.

- **Комментарии.** Содержит все комментарии или аннотации, соединенные с узлами на ветви моделирования.
- **Поля ввода.** Содержит поля ввода, а также их роль на ветви моделирования (в формате значения поля роли, например, Ввод, Назначение, Расщепление и так далее).

- **Анализ важности независимых переменных.** Содержит все параметры, относящиеся к ветви моделирования потока и поддерживающие просмотр и редактирование при каждом обновлении модели. Эти параметры выводятся при нажатии кнопки **Параметры построения модели** на вкладке **Внедрение** в диалоговом окне свойств потока.
- **Узел моделирования.** Содержит имя и тип узла моделирования, используемые при генерировании или обновлении модели.

## Предварительный просмотр описаний потока

Для просмотра содержимого описания потока в браузере щелкните по опции в диалоговом окне свойств потока. Содержимое описание зависит от опций, заданных на вкладке **Внедрение** в этом диалоговом окне. Дополнительную информацию смотрите в разделе “Опции внедрения потока” на стр. 128.

Чтобы просмотреть описание потока:

1. Выберите в главном меню IBM SPSS Modeler:  
**Инструменты > Свойства потока > Внедрение**
2. Задайте тип внедрения, узел оценки и, если нужно, параметры оценки.
3. Если тип внедрения - Обновление модели, то можно выбрать:
  - Узел моделирования и, если нужно, параметры построения модели
  - Слепок модели на ветви оценок потока
4. Нажмите кнопку **Предварительный просмотр описания потока**.

## Экспорт описаний потока

Содержимое описания потока можно экспортировать в файл HTML.

Чтобы экспортировать описание потока:

1. Выберите в главном меню:  
**Файл > Экспортировать описание потока**
2. Введите имя для файла HTML и нажмите кнопку **Сохранить**.

## Выполнение потоков

Задав нужные опции потоков и соединив нужные узлы, вы можете выполнить поток, обработав данные на узлах потока. Есть несколько способов выполнить поток в IBM SPSS Modeler. Варианты:

- В меню **Инструменты** выберите **Выполнить**.
- На панели инструментов нажмите одну из кнопок **Выполнить...** При помощи этих кнопок можно выполнить весь поток или только выделенный конечный узел. Дополнительную информацию смотрите в разделе “Панель инструментов IBM SPSS Modeler” на стр. 19.
- Чтобы выполнить один поток данных, щелкните правой кнопкой по конечному узлу и во всплывающем меню выберите **Выполнить**.
- Чтобы выполнить часть потока данных, щелкните правой кнопкой по любому неконечному узлу и во всплывающем меню выберите **Выполнить с этой точки**. Будут выполнены только операции после выбранного узла.

Чтобы остановить выполняемый поток, можете нажать красную кнопку **Стоп** на панели инструментов или выбрать **Остановить выполнение** в меню **Инструменты**.

Если поток выполняется дольше трех секунд, открывается диалоговое окно **Обратная связь выполнения**, в котором показан ход выполнения.

У некоторых узлов есть свои окна с дополнительной информацией о выполнении потока. Они открываются при выборе соответствующей строки в диалоговом окне. Первая строка выбрана автоматически.

## Работа с моделями

Если поток содержит узел моделирования (то есть один из узлов со вкладки Моделирование или Моделирование баз данных на палитре узлов), то при выполнении потока создается **слепок модели**. Слепок модели - это контейнер для **модели**, содержащий набор правил, формул и уравнений, при помощи которых генерируются прогнозы по тем или иным исходным данным, и служащий ядром механизма прогностической аналитики.



Рисунок 12. Образец модели

После успешного выполнения узла моделирования на холст потока помещается соответствующий слепок модели; значок слепка - золотой ромбик (от которого происходит английский термин *nugget*, самородок). Слепок можно открыть и просмотреть его содержимое, чтобы узнать подробности модели. Чтобы просмотреть прогнозы, присоедините и выполните один или несколько конечных узлов, которые выведут прогнозы в удобочитаемом виде.

Типичный поток моделирования содержит следующие две ветви. **Ветвь моделирования** содержит узел моделирования, а также предшествующие ему узлы источников данных и предварительной обработки. **Ветвь оценок** создается при выполнении узла моделирования и содержит слепок модели и один или несколько конечных узлов, при помощи которых можно просматривать прогнозы.

Дополнительные сведения смотрите в руководстве *Узлы моделирования IBM SPSS Modeler*.

## Добавление комментариев и аннотаций к узлам и потокам

Иногда нужно описать поток для других сотрудников организации. Для вашего удобства к потокам, узлам и слепкам моделей можно присоединять пояснительные комментарии.

После этого другие сотрудники смогут просматривать эти комментарии на экране, и можно будет напечатать изображение потока, включающее в себя комментарии.

Можно вывести список всех комментариев для потока или надузла, изменить порядок комментариев в списке, отредактировать текст комментария и изменить цвет текста и фона комментария. Дополнительную информацию смотрите в разделе “Вывод комментариев к потоку” на стр. 57.

Кроме того, можно добавлять заметки в формате аннотаций к потокам, узлам и слепкам, используя для этого вкладку Аннотации в диалоговом окне свойств потока, в диалоговом окне узла или в диалоговом окне слепка модели. Эти аннотации видны только при открытой вкладке Аннотации, кроме тех аннотаций потока, которые возможно вывести также как экранные комментарии. Дополнительную информацию смотрите в разделе “Аннотации” на стр. 58.





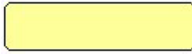

## Комментарии

Комментарии выглядят как текстовые поля, в которые можно ввести любой объем текста, и вы можете добавить любое число комментариев. Комментарий может быть отдельным (не соединенным с каким-либо объектом потока) или быть соединенным с одним или несколькими узлами или слепками моделей в потоке. Отдельные комментарии обычно используются для описания назначения потока в целом; соединенные комментарии описывают свой узел или слепок. Можно присоединить несколько комментариев к одному узлу или слепку и можно создать любое число отдельных комментариев для потока.

*Примечание:* Можно также в качестве экранных комментариев показать аннотации к потоку, но их нельзя соединять с узлами и слепками. Дополнительную информацию смотрите в разделе “Преобразование аннотаций в комментарии” на стр. 58.

Как описано в следующей таблице, вид текстового окна изменяется, показывая текущий режим комментария (или аннотации, показанной как комментарий).

Таблица 3. Режимы текстового окна комментария или аннотации

Текстовое окно комментария	Текстовое окно аннотации	Режим	Показывает	Получается, если...
		Правка	Комментарий открыт для редактирования.	создавать новый комментарий (или аннотацию), либо щелкнуть дважды по существующему.
		Последн. выбранн.	Комментарий можно переместить, изменить по размерам или удалить.	щелкнуть по фону потока после редактирования или один раз щелкнуть по существующему комментарию (или аннотации).
		Просм.	Редактирование завершено.	щелкнуть по другому узлу, комментарию или аннотации после редактирования.

Когда вы создаете новый отдельный комментарий, он сначала выводится на холсте потока в верхнем левом углу.

Если вы присоединяете комментарий к узлу или слепку, он сначала выводится над своим объектом потока.

Текстовое окно имеет белый фон в знак того, что можно вводить текст. Введя текст, щелкните вне текстового окна. Фон комментария станет желтым в знак того, что ввод текста завершен. Комментарий останется выделенным, и его можно переместить, изменить по размеру или удалить.

Если щелкнуть еще раз, граница станет сплошной линией в знак того, что редактирование завершено.

Двойной щелчок по комментарию переводит текстовое окно в режим правки -- фон становится белым и текст комментария можно редактировать.

Комментарии можно также присоединять к надузлам.

**Операции с комментариями:** Над комментариями можно выполнять ряд операций. Варианты:

- Добавить непривязанный комментарий
- Привязать комментарий к узлу или слепку
- Отредактировать комментарий
- Изменить размер комментария
- Переместить комментарий
- Отвязать комментарий
- Удалить комментарий
- Показать или скрыть все комментарии для потока

Чтобы добавить непривязанный комментарий

1. Убедитесь, что в потоке ничего не выбрано.
2. Сделайте одно из следующего:
  - В главном меню выберите: **Вставить > Новый комментарий**
  - Щелкните правой кнопкой по фону потока и во всплывающем меню выберите **Новый комментарий**.

- Нажмите кнопку **Новый комментарий** на панели инструментов.
3. Введите текст комментария (или вставьте текст из буфера обмена).
  4. Щелкните по узлу в потоке, чтобы сохранить комментарий.

Чтобы привязать комментарий к узлу или слепку

1. Выберите один или несколько узлов или слепков на холсте потока.
2. Сделайте одно из следующего:
  - В главном меню выберите:  
**Вставить > Новый комментарий**
  - Щелкните правой кнопкой по фону потока и во всплывающем меню выберите **Новый комментарий**.
  - Нажмите кнопку **Новый комментарий** на панели инструментов.
3. Введите текст комментария.
4. Щелкните по другому узлу в потоке, чтобы сохранить комментарий.  
Другой способ:
5. Вставьте непривязанный комментарий (смотрите предыдущий раздел).
6. Сделайте одно из следующего:
  - Выберите комментарий, нажмите F2, затем выберите узел или слепок.
  - Выберите узел или слепок, нажмите F2, затем выберите комментарий.
  - (Только для трехкнопочной мыши) Поместите указатель мыши на комментарий, нажмите среднюю кнопку и, удерживая ее нажатой, перетащите указатель на узел или слепок, где отпустите кнопку мыши.

Чтобы привязать комментарий к дополнительному узлу или слепку

Если комментарий уже привязан к узлу или слепку или находится на уровне потока и нужно привязать этот комментарий к дополнительному узлу или потоку, сделайте одно из следующего:

- Выберите комментарий, нажмите F2, затем выберите узел или слепок.
- Выберите узел или слепок, нажмите F2, затем выберите комментарий.
- (Только для трехкнопочной мыши) Поместите указатель мыши на комментарий, нажмите среднюю кнопку и, удерживая ее нажатой, перетащите указатель на узел или слепок, где отпустите кнопку мыши.

Чтобы отредактировать существующий комментарий

1. Сделайте одно из следующего:
  - Щелкните дважды по текстовому блоку комментария.
  - Выберите текстовый блок и нажмите клавишу Enter.
  - Щелкните правой кнопкой по текстовому блоку, чтобы открыть его меню, и выберите Правка.
2. Отредактируйте текст комментария. При редактировании можно использовать стандартные сочетания клавиш Windows, например, Ctrl+C для копирования текста. Другие возможности при редактировании есть во всплывающем меню комментария.
3. Щелкните вне текстового блока один раз, чтобы вывести элементы управления размером, и еще раз, чтобы завершить комментарий.

Чтобы изменить размер текстового блока комментария

1. Выберите комментарий, чтобы вывести элементы управления размером.
2. Щелкните по элементу управления и перетащите его, чтобы изменить размер блока.
3. Щелкните вне текстового блока, чтобы сохранить изменение.

Чтобы переместить существующий комментарий

Если вы хотите переместить комментарий, но не объекты, к которым он привязан (если такие есть), сделайте одно из следующего:

- Поместите указатель мыши на комментарий и, удерживая нажатой левую кнопку мыши, перетащите комментарий на новое место.
- Выберите комментарий и, удерживая нажатой клавишу Alt, перетащите комментарий клавишами со стрелками.

Если вы хотите переместить комментарий вместе со всеми узлами и слепками, к которым он привязан:

1. Выберите все объекты, которые хотите переместить.
2. Сделайте одно из следующего:
  - Поместите указатель мыши на один из этих объектов и, удерживая нажатой левую кнопку мыши, перетащите объекты на новое место.
  - Выберите один из этих объектов и, удерживая нажатой клавишу Alt, перетащите объекты клавишами со стрелками.

Чтобы отвязать комментарий от узла или слепка

1. Выберите один или несколько комментариев, которые хотите отвязать.
2. Сделайте одно из следующего:
  - Нажмите клавишу F3.
  - Щелкните правой кнопкой по выбранному комментарию и в его меню выберите Отвязать.

Чтобы удалить комментарий

1. Выберите один или несколько комментариев, которые хотите удалить.
2. Сделайте одно из следующего:
  - Нажмите клавишу Delete.
  - Щелкните правой кнопкой по выбранному комментарию и в его меню выберите Удалить.

Если комментарий был привязан к узлу или слепку, соединительная линия также удаляется.

Если непривязанный комментарий - это преобразованная аннотация к потоку или надузлу, при удалении комментария с холста текст комментария остается на вкладке Аннотации для потока или надузла.

Чтобы показать или скрыть все комментарии для потока

1. Сделайте одно из следующего:
  - В главном меню выберите:  
**Вид > Комментарии**
  - На панели инструментов нажмите **Показать/скрыть комментарии**.

**Вывод комментариев к потоку:** Вы можете вывести список всех комментариев к конкретному потоку или надузлу.

В этом списке можно

- Изменять порядок следования комментариев
- Редактировать текст комментариев
- Изменять цвет текста и фона тех или иных комментариев

Вывод списка комментариев

Чтобы вывести комментарии к потоку, сделайте одно из следующего:

- В главном меню выберите:

### Инструменты > Свойства потока > Комментарии

- Щелкните правой кнопкой по потоку на панели менеджеров и выберите **Свойства потока**, затем **Комментарии**.
- Щелкните правой кнопкой по фону потока на холсте и выберите **Свойства потока**, затем **Комментарии**.

**Текст.** Текст комментария. Щелкните дважды по тексту, чтобы превратить поле в текстовый блок, допускающий редактирование.

**Связи.** Имя узла, к которому привязан комментарий. Если это поле пусто, то комментарий привязан к потоку.

**Кнопки размещения.** Перемещают выбранный комментарий вверх или вниз по списку.

**Цвета комментариев.** Чтобы изменить цвет текста или фона некоторого комментария, выберите комментарий, включите переключатель **Пользовательские цвета** и выберите цвет в списке **Фон, Текст** или в обоих. Нажмите **Применить**, затем щелкните по фону потока, чтобы изменение вступило в силу. Нажмите кнопку **ОК**, чтобы сохранить изменение.

**Преобразование аннотаций в комментарии:** Аннотации к потокам или надузлам можно преобразовывать в комментарии.

Аннотации к потокам преобразуются в комментарий на холсте потока, не привязанный к тому или иному узлу.

Аннотация к надузлу преобразуется в комментарий, не привязанный к надузлу на холсте потока, но появляющийся при детализации надузла.

Чтобы преобразовать аннотацию к потоку в комментарий

1. В меню **Инструменты** выберите **Свойства потока**. (Другой вариант - щелкнуть правой кнопкой по потоку на панели менеджеров и выбрать **Свойства потока**.)
2. Щелкните по вкладке **Аннотации**.
3. Включите переключатель **Показать аннотацию как комментарий**.
4. Щелкните по **ОК**.

Чтобы преобразовать аннотацию к надузлу в комментарий

1. Щелкните дважды по значку надузла на холсте.
2. Щелкните по вкладке **Аннотации**.
3. Включите переключатель **Показать аннотацию как комментарий**.
4. Щелкните по **ОК**.

## Аннотации

Есть ряд способов аннотировать узлы, потоки и модели. Можно добавлять описательные аннотации и задавать пользовательские имена. Эти варианты особенно полезны при генерировании отчетов для потоков, добавленных на панель проектов. Кроме того, для узлов и для слепков моделей можно добавлять текст подсказки, помогающий различать сходные узлы на холсте потока.

### Добавление аннотаций

При редактировании узла или слепка модели открывается диалоговое окно, содержащее среди прочих вкладок вкладку аннотаций, на которой можно задать различные опции аннотации. Кроме того, можно открыть непосредственно вкладку аннотаций.



1. Чтобы аннотировать узел или слепок модели, щелкните правой кнопкой по узлу или слепку на холсте потока и выберите **Переименовать или аннотировать**. Появится диалоговое окно редактирования, открытое на вкладке аннотаций.
2. Чтобы аннотировать поток, в меню Инструменты выберите **Свойства потока**. (Другой вариант - щелкнуть правой кнопкой по потоку на панели менеджеров и выбрать **Свойства потока**.) Откройте вкладку Аннотации.

**Имя.** Выберите **Настроить**, чтобы подправить автоматически сгенерированное имя или создать уникальное имя для узла, как оно будет показываться на холсте потока.

**Контекстная справка по инструменту.** (Только для узлов и для слепков моделей) Введите текст, который появится как подсказка на холсте потока. Это особенно полезно при работе с большим количеством сходных узлов.

**Ключевые слова.** Задайте ключевые слова, которые послужат в отчетах по проекту и при поиске узлов в потоке, а также отслеживании объектов, записанных в репозитории (смотрите раздел “О программе IBM SPSS Collaboration and Deployment Services Repository” на стр. 115). Можно задать несколько ключевых слов через точку с запятой, например, доход; тип обрезки; величина иска. Пробелы в начале и в конце ключевого слова отсекаются, например, доход ; тип обрезки даст тот же результат, что доход;тип обрезки. (Но пробелы внутри ключевых слов не отсекаются. Например, различаются тип обрезки с одним пробелом и тип обрезки с двумя пробелами.)

В основной текстовой области можно ввести пространственные аннотации о том, какие операции выполняются на узле или какие решения здесь принимаются. Например, если поток повторно используется вами или другими пользователями, полезно делать заметки о таких решениях, как отбрасывание полей с многочисленными пробелами при помощи узла фильтра. Аннотируя узел, вы записываете этого рода информацию вместе с узлом. Кроме того, можно задать, что эти аннотации войдут в отчет по проекту, созданный на панели проекта. Дополнительную информацию смотрите в разделе “Введение в проекты” на стр. 137.

**Показать аннотацию как комментарий.** (Только для аннотаций потока и надузла) Включите этот переключатель, чтобы преобразовать аннотации в отдельный комментарий, который может быть показан на холсте потока. Дополнительную информацию смотрите в разделе “Добавление комментариев и аннотаций к узлам и потокам” на стр. 54.

**ID.** Содержит уникальный ID, который может использоваться для ссылки на узел в сценариях и средствах автоматизации. Это значение генерируется автоматически при создании узла и не может изменяться. Кроме того, имейте в виду, что во избежание конфликта с заглавной буквой "O", в этих ID не используется цифра 0. Нажмите кнопку копирования справа, чтобы скопировать ID, и затем можете вставить его в сценарий и куда угодно.

## Сохранение потоков данных

После того, как поток создан, его можно сохранить для повторного использования в будущем.

Чтобы сохранить поток

1. В меню Файл выберите **Сохранить поток** или **Сохранить поток как**.
2. В диалоговом окне Сохранить перейдите в папку, куда нужно сохранить файл потока.
3. В текстовом поле Имя файла введите имя для потока.
4. Включите переключатель **Добавить в проект**, если сохраненный поток нужно добавить в текущий проект.

Нажатие на кнопку **Сохранить** сохраняет поток как файл с расширением \*.str в указанном каталоге.

**Файлы автоматических резервных копий.** При каждом сохранении потока предыдущая сохраненная версия файла автоматически консервируется как резервная копия, для чего к имени файла добавляется дефис (например, *мойпоток.str-*). Чтобы восстановить резервную версию, просто удалите этот дефис и откройте файл еще раз.

## Сохранение состояний

Помимо потоков, можно сохранять **состояния**, которые включают в себя текущую диаграмму потока и все созданные слепки моделей (перечисленные на вкладке моделей на панели менеджеров).

Чтобы сохранить состояние

1. В меню Файл выберите:  
**Состояние > Сохранить состояние или Сохранить состояние как**
2. В диалоговом окне Сохранить перейдите в папку, куда нужно сохранить файл состояния.

Нажатие на кнопку **Сохранить** сохраняет состояние как файл с расширением *\*.cst* в указанном каталоге.

## Сохранение узлов

Кроме того, можно сохранить отдельный узел, щелкнув правой кнопкой по узлу на холсте потока и выбрав во всплывающем меню **Сохранить узел**. Как расширение файла задайте *\*.nod*.

## Сохранение нескольких объектов потока

Если при выходе из IBM SPSS Modeler остается несколько несохраненных объектов, таких как потоки, проекты или слепки модели, то выводится приглашение сохранить их перед закрытием программы. Если вы выбираете сохранение, открывается диалоговое окно, где можно сделать отдельно выбрать сохранение для каждого объекта.

1. Просто включите переключатели для всех объектов, которые хотите сохранить.
2. Нажмите кнопку **ОК**, чтобы сохранить каждый из объектов в нужном положении.

Будут открыты обычные диалоговые окна Сохранить для всех объектов. Когда вы завершите операции сохранения, прикладная программа закроется согласно первоначальной команде.

## Сохранение результатов

Таблицы, диаграммы и отчеты, сгенерированные из узлов вывода IBM SPSS Modeler, можно сохранить в формате объект вывода (*\*.cou*).

1. При просмотре вывода, который нужно сохранить, выберите в меню окна вывода:  
**Файл > Сохранить**
2. Задайте имя и положение файла вывода.
3. Если нужно, включите переключатель **Добавить файл в проект** в диалоговом окне Сохранить, чтобы включить файл в текущий проект. Дополнительную информацию смотрите в разделе “Введение в проекты” на стр. 137.

Другой вариант - щелкнуть правой кнопкой по любому объекту вывода на панели менеджеров и во всплывающем меню выбрать **Сохранить**.

## Шифрование и дешифрование информации

Сохраняя поток, узел, проект, файл вывода или слепки модели, его можно зашифровать, чтобы предотвратить несанкционированное использование. Для этого нужно включить дополнительный переключатель при сохранении и добавить пароль в сохраняемый элемент. Такое шифрование может быть задано для любого сохраняемого элемента и оно дает дополнительную защиту, отличную от шифрования SSL при передаче файлов между IBM SPSS Modeler и сервер IBM SPSS Modeler.

При попытке открыть зашифрованный файл появляется приглашение ввести пароль. После ввода правильного пароля элемент дешифруется автоматически и открывается, как обычно.

Чтобы зашифровать элемент

1. В диалоговом окне Сохранить для элемента, который нужно зашифровать, нажмите кнопку **Опции**. Откроется диалоговое окно Опции шифрования.
2. Выберите **Зашифровать этот файл**.
3. Для дополнительной защиты можно включить переключатель **Маскировать пароль**. В этом случае все вводимые вами символы будут показаны как кружочки.
4. Введите пароль. *Предупреждение:* Если вы забудете пароль, открыть файл или модель будет невозможно.
5. Если вы включили переключатель **Маскировать пароль**, введите пароль еще раз, чтобы подтвердить, что вы вводите его правильно.
6. Нажмите кнопку **ОК**, чтобы вернуться в диалоговое окно Сохранить.

*Примечание:* Если вы сохраняете копию уже защищенного паролем элемента, то элемент автоматически сохраняется в зашифрованном формате с тем же паролем, если не изменить настройки в диалоговом окне Опции шифрования.

## Загрузка файлов

В IBM SPSS Modeler можно загружать ряд сохраненных объектов, таких как:

- потоки (.str)
- состояния (.cst)
- модели (.gm)
- палитра моделей (.gen)
- узлы (.nod)
- вывод (.cou)
- проекты (.cpi)

Открытие новых файлов

Потоки можно загружать непосредственно через меню файл.

1. В меню Файл выберите **Открыть поток**.

Все остальные типы файлов можно открыть, выбрав пункты подменю в меню файл. Например, чтобы загрузить модель, выберите в меню Файл:

**Модели > Открыть модель или загрузить палитру моделей**

Как открыть последние использованные файлы

Для быстрой загрузки недавно использованных файлов можно использовать опции в нижней части меню Файл.

Выберите **Недавние потоки**, **Недавние проекты** или **Недавние состояния**, чтобы раскрыть нужный список недавно использованных файлов.

## Отображение потоков данных

При помощи утилиты отображения можно соединить новый источник данных с уже существующим потоком. Утилита отображения не только сконфигурирует соединение, но также поможет указать, каким образом полями в новом источнике будут замещаться поля в существующем потоке. Вместо того, чтобы заново создать весь поток данных для нового источника данных, вы можете просто соединить новый источник данных с существующим потоком.

При помощи утилиты отображения данных можно соединить два фрагмента потока и убедиться, что задано правильное отображение для имен всех существенных полей. Построение отображения данных имеет своим результатом создание узла фильтра, в котором нужные поля ищутся и переименовываются.

Есть два эквивалентных способа отобразить данные:

**Выбрать узел замены.** Этот метод начинается с заменяемого узла. Сначала нужно щелкнуть правой кнопкой по заменяемому узлу, затем, выбрав во всплывающем меню **Отображение данных > Выбрать узел замены**, выберите узел, на который хотите заменить первый выбранный узел.

**Цель отображения** Этот метод начинается с узла, добавляемого в поток. Сначала нужно щелкнуть правой кнопкой по добавляемому узлу, затем, выбрав во всплывающем меню **Отображение данных > Цель отображения**, выберите узел, с которым хотите объединить первый выбранный узел. Этот метод особенно полезен для отображения конечного узла. *Примечание:* Невозможно отображение на узлы слияния или присоединения. Следует просто присоединить поток к узлу слияния обычным образом.

Отображение данных тесно интегрировано с построением потока. Если попытаться протянуть соединение в узел, уже имеющий входящее соединение, появится приглашение заменить старое соединение либо построить отображение в этот узел.

## Отображение данных в шаблон

Чтобы заменить источник данных в шаблоне потока на новый узел источника и тем самым передать свои данные в IBM SPSS Modeler, нужно выбрать пункт **Выбрать заменяющий узел** во всплывающем меню **Отображение данных**. Это доступно для всех узлов, кроме узлов объединения, присоединения и всех конечных узлов. Применение средства отображения данных помогает убедиться в правильности сопоставления полей из операций существующего потока и из нового источника данных. Приведенный ниже порядок действий описывает в общих чертах процесс отображения данных.

**Шаг 1: Задайте существенные поля в исходном узле источника.** Задание существенных полей обеспечит правильное выполнение операций потока. Дополнительную информацию смотрите в разделе “Задание существенных полей” на стр. 63.

**Шаг 2: Добавьте на холст потока новый источник данных.** Выберите узел источника, который передаст в поток новые данные.

**Шаг 3: Замените узел источника в шаблоне.** Выберите пункт **Отображение данных** во всплывающем меню узла источника шаблона, щелкните по **Выбрать заменяющий узел** и выберите узел источника заменяющих данных.

**Шаг 4: Проверьте отображенные поля.** В открывшемся диалоговом окне проверьте, что программа правильно сопоставляет потоку поля источника заменяющих данных. Несопоставленные существенные поля выводятся красным цветом. Эти поля используются в операциях потока и должны заменяться на аналогичные поля в новом источнике данных, чтобы операции ниже по потоку работали правильно. Дополнительную информацию смотрите в разделе “Осмотр отображенных полей” на стр. 63.

После того, как вы проверили в диалоговом окне, что все существенные поля отображены правильно, старый источник данных отсоединяется и к потоку присоединяется новый источник данных через узел фильтра *Отображение*. Этот узел фильтра направляет в поток поля после отображения. Кроме того, на холст потока добавляется узел фильтра *Обратное отображение*. Узел фильтра *Обратное отображение* можно включить в поток, чтобы вернуться к именам полей до отображения. Однако имейте в виду, что после отмены отображения все конечные узлы ниже по потоку нужно отредактировать, заново выбрав поля и наложения.

## Отображение потоков

Как и при соединении узлов, в этом методе отображения данных не требуется заранее задать существенные поля. В этом методе вы просто соединяете поток с потоком, выбирая пункт **Цель отображения в**

выпадающем меню **Отображение данных**. Такой тип отображения данных полезен для отображения конечных узлов и копирования-вставки из потока в поток. *Примечание:* Пункт меню **Цель отображения** не может использоваться для отображения на узлы объединения, присоединения и все типы узлов источника.

Чтобы отобразить данные из потока в поток

1. Щелкните правой кнопкой по узлу, который нужно использовать для соединения с новым потоком.
2. Выберите в меню:

**Отображение данных > Цель отображения**

3. Используйте указатель для выбора нужного узла в потоке назначения.
4. В открытом диалоговом окне убедитесь, что поля сопоставлены правильно, и нажмите кнопку **ОК**.

## Задание существенных полей

При отображении в существующий поток существенные поля обычно задаются автором потока. Существенными объявляются поля, которые используются в операциях ниже по потоку. Например, пусть в существующем потоке строится модель, в которой используется поле *Паника*. Для потока поле *Паника* существенно, так как без него нельзя построить модель. Аналогичным образом есть поля, обязательные для узлов обработки, таких как узел вычислений, поскольку без них нельзя вычислить производное поле. Если такие поля заданы как существенные явным образом, это помогает вовремя убедиться, что поля нового источника данных отображаются на все нужные поля. Если какие-то обязательные поля не отображены, выводится сообщение об ошибке. Вы можете решить, что какие-то из узлов обработки и вывода больше не нужны, и удалить их из потока; тогда соответствующие поля могут быть удалены из списка существенных полей.

Чтобы задать существенные поля

1. Щелкните правой кнопкой по заменяемому узлу источника в шаблоне потока.
2. Выберите в меню:

**Отображение данных > Задать существенные поля**

3. Добавьте или удалите поля в списке, пользуясь средством выбора полей. Средство выбора полей открывается щелчком по значку справа от списка полей.

## Осмотр отображенных полей

После того, как вы выбрали точку, в которой один поток данных или источник данных будет отображен на другой, открывается диалоговое окно, чтобы выбрать в нем поля для отображения или убедиться в правильности системного отображения по умолчанию. Если нет пары для какого-то существенного поля в потоке данных или на источнике данных, оно показано красным. Все непарные поля из источника данных пройдут через узел фильтра без изменений, но имейте в виду, что несущественные поля также можно отобразить.

**Исходные.** Содержит все поля шаблона или существующего потока — все поля, присутствующие ниже по потоку. Поля из нового источника данных будут отображены в эти поля.

**Отображенные.** Содержит все поля, выбранные для отображения на поля шаблона. Это поля, имена которых иногда нужно изменить, чтобы сопоставить с исходными полями, используемыми в операциях потока. Щелкните в клетке таблицы, соответствующей полю, чтобы активировать список доступных полей.

Если вы не уверены, какие поля следует отобразить, иногда помогает сначала изучить данные источника. Например, на вкладке **Типы** на узле источника можно просмотреть сводку по данным источника.

---

## Советы и клавиши быстрого вызова

Для быстрой и удобной работы познакомьтесь со следующими советами и клавишами быстрого вызова:

- **Стройте потоки быстро при помощи двойных щелчков.** Двойной щелчок по узлу на палитре помещает узел в холст и соединяет с текущим потоком.

- **Используйте сочетания клавиш для выбора узлов ниже по потоку.** Нажимайте Ctrl+Q и Ctrl+W, чтобы выделять все узлы ниже по потоку и снимать это выделение.
- **Используйте сочетания клавиш для соединения и разъединения узлов.** Когда на холсте выбран узел, нажмите клавишу F2, чтобы начать соединение, перейдите к нужному узлу клавишей Tab и нажмите Shift+Пробел, чтобы завершить соединение. Нажмите клавишу F3, чтобы отсоединить все входящие и исходящие соединения выделенного узла.
- **Настройте вкладку на палитре узлов, поместив на нее часто используемые вами узлы.** В меню Инструменты выберите **Управление палитрами**, чтобы открыть диалоговое окно для добавления, удаления и перемещения узлов, показанных на палитре узлов.
- **Переименуйте узлы и добавьте подсказки.** У каждого узла в диалоговом окне есть вкладка Аннотации, на которой вы можете задать пользовательское имя для узлов на холсте, а также добавить подсказку, которая поможет организовать поток. Кроме того, вы можете создать пространственные аннотации для отслеживания хода работы, сохранять подробности процесса и делать заметки обо всех деловых решениях, которые требуется принять и которые были приняты.
- **Вставьте значения автоматически в выражение CLEM.** В построителе выражений, доступном из различных диалоговых окон (например, из диалоговых окон узлов вычислений и заполнения), можно автоматически вставлять значения полей в выражение CLEM. Нажмите кнопку значений в построителе выражений, чтобы выбрать среди существующих значений поля.



Рисунок 13. Кнопка Значения

- **Ищите файлы быстрыми способами.** При поиске файла в диалоговом окне Открыть используйте список файлов (нажмите на кнопку с желтым ромбиком в верхней части диалогового окна рядом с полем Папка), чтобы открыть ранее использованные каталоги, а также каталоги IBM SPSS Modeler по умолчанию. Нажимайте кнопки прокрутки вперед и назад для навигации по использованным каталогам.
- **Минимизируйте дебри из окон вывода.** Можно быстро закрыть и удалить окно вывода кнопкой с красной буквой X в верхнем правом углу всех окон вывода. Так вы сможете оставить лишь многообещающие и интересные результаты на вкладке Вывод на панели менеджеров.

В программе доступен полный спектр клавиш быстрого вызова. Дополнительную информацию смотрите в разделе “Средства доступности через клавиатуру” на стр. 160.

А вы знали, что можно...

- Перетаскивать и выделять мышью группы узлов на холсте потока.
- Копировать из одного потока и вставлять их в другой.
- Открывать справку из каждого диалогового окна и окна вывода.
- Получить справку о методологии CRISP-DM (Cross-Industry Standard Process for Data Mining, межотраслевой стандартный процесс исследования данных). (В меню Справка выберите **Справка CRISP-DM**.)

---

## Глава 6. Обработка пропущенных значений

---

### Обзор пропущенных значений

На этапе подготовки данных к исследованию часто требуется заменить пропущенные значения.

**Пропущенные значения** - это значения в наборе данных, которые неизвестны, не собраны или введены с ошибками. Обычно это недопустимые значения для соответствующих полей. Например, пусть поле *Пол* должно содержать только значения *М* и *Ж*. Обнаружив в этом поле значения *Д* или *Х*, вы можете смело предположить, что эти значения некорректны и должны считаться пропусками. Аналогичным образом, отрицательное значение в поле *Возраст* не имеет смысла и тоже должно интерпретироваться как пропущенное. Иногда такие очевидно неверные значения вводятся при опросе специально, чтобы сообщить об отказе от ответа. Иногда нужно изучить такие пробелы внимательнее, чтобы выяснить, не служит ли отказ от ответа, например, отказ назвать свой возраст, фактором в предсказании конкретного исхода.

Некоторые технологии моделирования хорошо справляются с пропущенными значениями, некоторые хуже. Так, C5.0 и Arpriori хорошо справляются со значениями, которые явным образом объявляются "пропущенными" на узле типа. Другие технологии моделирования испытывают трудности при обработки пропущенных значений, увеличивается время обучения, снижается точность генерируемых моделей.

Существует несколько типов пропущенных значений, распознаваемых IBM SPSS Modeler:

- **Пустые или системные пропущенные значения.** Это те нестроковые значения, которые были оставлены в виде пробельных символов в базе данных или файле источника, а не были особо определены как "пропущенные" на узле источника или типа. Системные пропущенные значения выводятся как **\$null\$**. Имейте в виду, что пустые строки не рассматриваются как пустые значения (NULL) в IBM SPSS Modeler, хотя и могут обрабатываться как пустые значения некоторыми базами данных.
- **Пустые строки и пробельные значения.** Значения пустых строк и пробельные значения (строки без видимых символов) обрабатываются не так, как пустые значения (NULL). Пустые строки обрабатываются для большинства целей как эквивалент пробельным значениям. Например, если выбрать опцию для обработки пробельных значений как пробельных символов на узле источника или типа, эта настройка будет также применена и к пустым строкам.
- **Пустые или пользовательские пропущенные значения.** Существуют значения (например, unknown, 99 или -1), которые определяются на узле источника или узле типа явным образом как пропущенные значения. Дополнительно можно выбрать опцию обработки пустых значений (NULL) и пробельных значений как пробельных символов, что позволит пометить их флагами для специальной обработки и исключить из большинства вычислений. Например, функция @BLANK позволяет обрабатывать эти значения (также, как и пропущенные значения других типов) как пробельные символы.

**Считывание смешанных данных.** Обратите внимание на то, что при чтении полей с числовыми данными (таким как целые числа, действительные числа, время, отметка времени, дата) нечисловые значения задаются как *пустые* или *пропущенные значения системы*. Это связано с тем, что, в отличие от некоторых других прикладных программ, здесь не допускается смешение разных типов данных в одном поле. Чтобы предотвратить это, поля со смешанными данными следует считывать как строки, изменив им тип на узле источника или во внешней прикладной программе.

**Чтение пустых строк из Oracle.** При чтении из базы данных Oracle или записи в нее данных следует знать, что в отличие от IBM SPSS Modeler и от большинства других баз данных Oracle обрабатывает и хранит значения пустых строк, приравнивая их к пустым значениям (NULL). Это означает, что данные, извлеченные из базы данных Oracle, могут вести себя не так, как данные, извлеченные из файла или другой базы данных, и могут вернуть другие результаты.

---

## Обработка пропущенных значений

Решение о том, как обрабатывать пропущенные значения, следует принимать с опорой на знание бизнеса или области. Чтобы уменьшить время обучения и повысить точность, лучше удалить пробелы из набора данных. С другой стороны, если оставить пустые значения, возможно открытие новых бизнес-возможностей или лучшее понимание ситуации. Выбирая лучший подход, учитывайте следующие аспекты данных:

- Размер набора данных
- Число полей, содержащих пробелы
- Объем пропущенной информации

В общих чертах, есть два подхода:

- Можно исключить поля или записи с пропущенными значениями
- Можно заполнить, заменить или подавить пропущенные значения, воспользовавшись целым рядом методов

Все эти подходы в большой степени автоматизированы при использовании узла аудита данных. Например, можно сгенерировать узел фильтра, который исключает поля, бесполезные для моделирования из-за слишком большого количества пропущенных значений, и сгенерировать надузел, который заполняет пропущенные значения для остальных полей. Здесь в полную меру реализуются возможности аудита, благодаря которым вы не только можете оценить текущее состояние данных, но и выбирать действия с учетом этой оценки.

## Обработка записей с пропущенными значениями

Если большая часть пропущенных значений сосредоточена в небольшом числе записей, эти записи можно просто исключить. Например, обычно банк хранит подробные и полные досье на своих клиентов-заемщиков. Но если банк не так скрупулезен при утверждении займов для своих сотрудников, в этих досье могут оставаться пробелы. В таком случае у вас есть два варианта, как поступить с пропущенными значениями:

- Удалить записи о сотрудниках при помощи узла отбора.
- Если набор данных большой, можно удалить все записи с пробелами.

## Обработка полей с пропущенными значениями

Если большая часть пропущенных значений сосредоточена в небольшом числе полей, вы можете решить проблему не на уровне записей, а на уровне полей. В этом случае у вас есть возможность, прежде чем выбрать тот или иной способ обработки пропущенных значений, выполнить экспериментальную проверку важности отдельных полей. Если поле имеет малую важность для моделирования, то нет смысла бороться за него, даже если число пропущенных значений не так велико.

Например, пусть маркетинговая компания собирает данные по листу опроса, содержащему 50 вопросов. Два вопроса касаются возраста и политических взглядов, и респонденты склонны не сообщать эту информацию. В таком случае поля *Возраст* и *Политические\_взгляды* содержат много пропущенных значений.

Тип измерений поля

При выборе подхода к пропущенным значениям следует учесть также тип измерений обсуждаемых полей.

**Числовые поля.** Для полей числового типа, такого как *Непрерывный*, всегда следует устранить нечисловые значения до построения модели, поскольку многие модели просто не работают, если в числовых полях встречаются пробелы.

**Категориальные поля.** Для полей категориального типа, такого как *Номинальный* и *Флаг*, замена пропущенных значений не обязательна, но она повышает точность модели. Например, модель, в которой используется поле *Пол*, будет работать и при наличии бессмысленных значений вроде *Д* и *Х*, но удаление всех значений кроме *М* и *Ж* повысит точность модели.



Экранирование или удаление полей

Есть несколько способов заэкранировать поля со слишком большим количеством пропущенных значений:

- Можно использовать узел аудита данных для фильтрации полей с учетом их качества.
- Можно использовать узел отбора показателей, чтобы заэкранировать поля, где процент пропущенных значений выше заданного, и ранжировать поля с учетом важности относительно заданной цели.
- Вместо удаления полей можно использовать узел типа, чтобы задать для поля роль **Никакая**. Тогда поля останутся в наборе данных, но будут исключены из моделирования.

---

## Импутация, или заполнение пропущенных значений

В тех наблюдениях, где пропущенных значений немного, есть смысл вставить какие-то значения вместо пробелов. Это можно сделать в отчете аудита данных, задав нужные опции для конкретных полей, чтобы затем сгенерировать надузел, который вставит значения, используя ряд методов. Это самый гибкий вариант, причем он также дает возможность обработать большое число полей на одном узле.

Для импутации пропущенных значений доступны следующие методы:

**Фиксированная.** Подставляет фиксированное значение (среднее значение поля, среднюю точку диапазона либо задаваемую вами константу).

**Переменный.** Подставляет случайное значение на основе нормального или равномерного распределения.

**Выражение.** Позволяет задать пользовательское выражение. Например, можно заменить значения на глобальную переменную, созданную узлом задания глобальных значений.

**Алгоритм.** Подставляет значение, предсказанное моделью на основе алгоритма C&RT. Для каждого поля, импутированного этим методом, будет отдельная модель C&RT наряду с узлом заполнения, заменяющим пробелы и пустые значения на значение, предсказанное моделью. Затем узел заполнения будет использован для удаления сгенерированных моделью полей предсказания.

Другой вариант - подавить значения для определенных полей, используя узел типа; убедитесь, что типы полей охватывают только допустимые значения, и задайте в столбце *Проверка* значение **Подавить** для полей, в которых нужно заменить пустые значения.

---

## Функции CLEM для обработки пропущенных значений

Имеется ряд функций для обработки пропущенных значений. Перечисленные ниже функции часто используются в узлах отбора и заполнения для отбрасывания или заполнения пропущенных значений:

- `count_nulls`(СПИСОК)
- `@BLANK`(ПОЛЕ)
- `@NULL`(ПОЛЕ)
- `undef`

Функции с символом @ можно использовать совместно с функцией @FIELD, чтобы обнаружить пустые значения в одном или нескольких полях. Поля с пустыми значениями можно просто пометить флагом, а можно заполнить заменяющими значениями или использовать в ряде других операций.

Можно подсчитать число пустых значений в списке полей, как показано ниже:

```
count_nulls(['владениекартой' 'владениекартой2' 'владениекартой3'])
```

Для функций, которые принимают на входе список полей, можно в качестве входного аргумента задавать специальные функции диапазона полей @FIELDS\_BETWEEN и выбора полей по шаблону @FIELDS\_MATCHING, как показано в следующем примере:

```
count_nulls(@FIELDS_MATCHING('card*'))
```

При помощи функции `undef` можно заполнить поля пропущенными значениями системы, которые выводятся как **\$null\$**. Так, чтобы заменить числовое значение, можно использовать условный оператор, например:

```
if not(Возраст > 17) or not(Возраст < 66) then undef else Возраст endif
```

Все значения вне диапазона будут заменены на пропущенное значение системы, которое выводится как **\$null\$**. При помощи функции `not()` можно перехватить все остальные значения, включая отрицательные. Дополнительную информацию смотрите в разделе “Функции для обработки пустых значений и значений Null” на стр. 112.

Замечание об отбрасывании записей

При использовании узла отбора для отбрасывания записей имейте в виду, что синтаксис подразумевает трехзначную логику и автоматически включает пустые значения в операторы выборки. Чтобы исключить пустые значения (пропущенные значения системы) в выражении выборки, нужно явно указать это, задав в выражении условие `and not`. Например, чтобы выбрать и включить все записи, в которых для типа рецептурного медикамента задано значение Медикамент С, используйте такой оператор выборки:

```
Медикамент = 'медикаментС' and not(@NULL(Медикамент))
```

Прошлые версии исключенных пустых значений в таких ситуациях.

---

## Глава 7. Построение выражений CLEM

---

### О программе CLEM

Язык CLEM (Control Language for Expression Manipulation, управляющий язык для преобразования выражений) - мощный язык для анализа и обработки данных в потоках IBM SPSS Modeler. Исследователи данных широко используют CLEM в операциях потока для выполнения таких простых задач, как вычисление прибыли по данным затрат и доходов, и таких сложных, как преобразование данных Web-журнала в набор пригодных для использования полей и записей.

CLEM используется в IBM SPSS Modeler для того, чтобы:

- Сравнивать и оценивать условия для полей записи.
- Вычислять значения производных полей.
- Вычислять новые значения для существующих полей.
- Анализировать последовательность записей.
- Вставлять данные из записей в отчеты.

Выражения CLEM незаменимы для подготовки данных в IBM SPSS Modeler и могут применяться на широком диапазоне узлов — от операций с записями и полями (выборка, балансировка, заполнение) до построения графиков и вывода (анализ, отчеты, таблица). Например, на узле вычислений при помощи CLEM можно создать производное поле на основе формулы, такой как отношение.

Кроме того, выражения CLEM можно использовать для операций глобального поиска и замены. Например, выражение @NULL(@FIELD) можно использовать на узле заполнения для замены **системных пропущенных значений** на целое число 0. (Чтобы заменить **системные пропущенные значения**, также называемые пробелами, пользуйтесь функций @BLANK.)

Можно создавать и более сложные выражения CLEM. Например, можно вычислить новые поля с учетом условного набора правил. Например, можно создать новую категорию при помощи таких выражений: If: CardID = @OFFSET(CardID,1), Then: @OFFSET(ValueCategory,1), Else: 'exclude'.

В этом примере при помощи функции @OFFSET провозглашается, что "Если значение поля *CardID* для данной записи такое же, как для предыдущей записи, следует вернуть значение поля с именем *ValueCategory* для предыдущей записи. В противном случае назначается строка "exclude". Тем самым, если встречается подряд две записи с одинаковым *ID карты*, им следует назначить одну и ту же категорию ценности. (Записи со строкой exclude можно в дальнейшем отсеять узлом отбора.)

---

### Примеры CLEM

Приведенные ниже примеры выражений иллюстрируют правильный синтаксис и типы возможных выражений с использованием CLEM.

Простые выражения

Формулами могут быть и совсем простые выражения, как приведенное ниже, которое вычисляет новое поле из значений полей *После* и *До*:

$(\text{После} - \text{До}) / \text{До} * 100.0$

Заметьте, что значения полей задаются как имена полей без кавычек.

Аналогичным образом следующее выражение просто возвращает логарифм для каждого значения поля *зарплата*.

```
log(зарплата)
```

### Сложные выражения

Возможны и куда более длинные, сложные выражения. Следующее выражение возвращает значение *true* (истина), если значение двух полей (*\$KX-Kohonen* и *\$KY-Kohonen*) попадают в заданные диапазоны. Заметьте, что здесь имена полей заключены в одинарные кавычки; дело в том, что эти имена содержат специальные символы.

```
(' $KX-Kohonen' >= -0.2635771036148072 and ' $KX-Kohonen' <= 0.3146203637123107  
and ' $KY-Kohonen' >= -0.18975617885589602 and  
' $KY-Kohonen' <= 0.17674794197082522) -> T
```

Для ряда функций, таких как строковые функции, требуется ввести несколько параметров с использованием правильного синтаксиса. В следующем примере функция *subscr* возвращает первый символ поля *produce\_ID*, показывающий, что позиция ТМЦ органическая, генно-модифицированная или обычная. Результаты выражения описываются опцией `-> `result``.

```
subscr(1,produce_ID) -> `c`
```

Аналогичным образом работает следующее выражение:

```
stripchar(`3`,`123`) -> `12`
```

Важно заметить, что значения символьного типа всегда заключаются в одинарные кавычки.

### Объединение функций в выражении

Нередко выражения CLEM состоят из сочетания функций. Следующее выражение совместно использует функции *subscr* и *lowertoupper* и возвращает первый символ *produce\_ID*, преобразованный в верхний регистр.

```
lowertoupper(subscr(1,produce_ID)) -> `C`
```

То же выражение может быть записано в краткой форме:

```
lowertoupper(produce_ID(1)) -> `C`
```

Другое часто используемое сочетание функций:

```
locchar_back(`n`, (length(web_страница)), web_страница)
```

Это выражение находит символ ``n`` в значениях поля *web\_страница*, читая их с конца. При помощи функции *length* это выражение динамически вычисляет длину текущего значения; если бы вместо этого задавалось статическое значение, например, 7, оно бы оказывалось недопустимым для значений короче 7 символов.

### Специальные функции

Доступны многочисленные специальные функции (начинающиеся с символа `@`). Среди наиболее широко используемых функций:

```
@BLANK('ссылочный ID') -> T
```

Нередко специальные функции используются в сочетании, например, чтобы одновременно отметить несколько полей, состоящих из одних пробелов.

```
@BLANK(@FIELD)-> T
```

Дополнительные примеры рассматриваются в документации CLEM. Дополнительную информацию смотрите в разделе “Справочный обзор CLEM” на стр. 83.

---

## Типы значений и данных

Выражения CLEM аналогичны формулам, построенным из значений, имен полей, операций и функций. Простейшее допустимое выражение CLEM - это значение или имя поля. Примеры допустимых значений:

```
3
1,79
'банан'
```

Примеры имен полей:

```
Product_ID
'$P-NextField'
```

где *Product* - имя поля из набора данных корзины покупателя, тогда как *'\$P-NextField'* - имя параметра, а значение выражения - значение именованного поля. Обычно имена полей начинаются с буквы и могут содержать также цифры и подчеркивания (`_`). Можно использовать имена, нарушающие эти правила, если помещать имя в кавычки. Значения CLEM могут быть такими:

- Строки -- например, "c1", "Тип 2", "фрагмент произвольного текста"
- Целые -- например, 12, 0, -189
- Действительные числа -- например, 12,34, 0,0, -0,0045
- Поля даты-времени -- например, 05/12/2002, 12/05/2002, 12/05/02

Возможно также использование таких элементов:

- Коды символов -- например, `a` or 3
- Списки элементов -- например, [1 2 3], ['Тип 1' 'Тип 2']

Обычно коды символов и списки не встречаются в значениях полей. Чаще всего они используются как аргументы функций CLEM.

Правила взятия в кавычки

Хотя программа гибко подходит к вычленению полей, значений, параметров и строк в выражении CLEM, при создании выражений рекомендуется соблюдать следующие прагматические правила:

- **Строки** — всегда берите строку в двойные кавычки при записи строк ("Тип 2" или "значение"). Можно использовать и одиночные кавычки, но это создает риск спутать их с взятыми в кавычки полями.
- **Символы** — Всегда используйте одиночные кавычки, вот такие ```. Например, обратите внимание на символ `d` в функции `stripchar(`d`,"drugA")`. Единственное исключение - при использовании целого числа для ссылки на определенный символ в строке. Например, обратите внимание на символ `5` в функции `lowertoupper("drugA"(5)) -> "A"`. *Примечание:* На обычной английской и американской клавиатуре клавиша одинарной кавычки (гравис, Unicode 0060) находится под клавишей Esc.
- **Поля** — в выражениях CLEM обычно указываются без кавычек (`subscr(2,arrayID)`) -> CHAR). Если есть пробелы и другие специальные символы, можно использовать одинарные кавычки ('Порядковый номер'). Поля, указанные в кавычках и не определенные в наборе данных, ошибочно принимаются за строки.
- **Параметры** — всегда берутся в одинарные кавычки ('\$P-threshold').

---

## Выражения и условия

Выражения CLEM могут возвращать результат (используется при вычислении новых значений) -- например:

```
Вес * 2.2
Возраст + 1
sqrt(Сигнал-Эхо)
```

Или могут оцениваться как *true* (истина) или *false* (ложь, используется при выборе условия) -- например:

Медикамент = "МедикаментА"  
Возраст < 16  
not(Энерговыведение) and Мощность > 2000

Операции и функции в выражениях CLEM можно использовать в произвольных сочетаниях -- например:  
`sqrt(abs(Сигнал)) * max(T1, T2) + БазовыйУровень`

Порядок вычисления выражения определяется скобками и приоритетом операций. В этом примере порядок вычисления такой:

- Вычисляется `abs(Сигнал)`, а для полученного результата вычисляется `sqrt`.
- Вычисляется `max(T1, T2)`.
- Эти два результата перемножаются: приоритет `x` выше, чем `y +`.
- Наконец, к результату прибавляется `БазовыйУровень`.

Приоритет операций, начиная с тех операций, которые выполняются первыми, такой:

- Аргументы функций
- Вызовы функций
- **xx**
- **x / mod div rem**
- + -
- > < >= <= /= == = /=

Если нужно переопределить порядок по приоритету или вы просто не хотите сомневаться, каким будет порядок вычислений, используйте скобки для явного задания порядка -- например,  
`sqrt(abs(Сигнал)) * (max(T1, T2) + БазовыйУровень)`

---

## Параметры потока, сеанса и надузла

Параметры можно определить для использования в выражениях CLEM и в сценариях. На самом деле, это пользовательские переменные, которые сохранены и сохраняются с текущим потоком, сеансом или надузлом, и к которым можно обратиться из пользовательского интерфейса, а также из сценария. Например, если вы сохраняете поток, сохраняются и все параметры, заданные для этого потока. (Это отличает их от переменных локального сценария, которые можно использовать только в сценарии, в котором они объявлены). Параметры часто используются в сценариях для управления поведением сценария, задавая информацию о полях и значениях, которую не требуется жестко указывать в коде сценария.

Область использования параметра зависит от того, где он задан:

- Параметры потока можно задать в сценарии потока или в диалоговом окне свойств потока, и они доступны на всех узлах в потоке. Эти параметры выводятся в списке Параметры построителя выражений.
- Параметры сеансов можно задать в автономном сценарии или в диалоговом окне параметров сеанса. Они доступны для всех потоков, используемых в текущем сеансе (все потоки, перечисленные на вкладке Потоки панели менеджера).

Параметры можно задать также для надузлов, и в этом случае они видны только на узлах, инкапсулированных в этом надузле.

Использование параметров в выражениях CLEM

Параметры представляются в выражениях CLEM как `$P-имя_параметра`, где `имя_параметра` - имя параметра. При использовании в выражениях CLEM параметры нужно заключать в одинарные кавычки — например, `'$P-масштаб'`.

Доступные параметры удобно просматривать в построителе выражений. Чтобы просмотреть текущие параметры:

1. В любом диалоговом окне, принимающем выражения CLEM, нажмите кнопку Построитель выражений.
2. В списке полей выберите **Параметры**.

Можно выбрать параметры из списка для вставки в выражение CLEM. Дополнительную информацию смотрите в разделе “Выбор полей, параметров и глобальных переменных” на стр. 78.

---

## Работа со строками

Для строк доступен ряд операций, в том числе:

- Преобразование строки к верхнему или нижнему регистру — `uppertolower(СИМВОЛЫ)`.
- Удаление указанных символов, например, ``ID_`` или ``$``, из строковой переменной — `stripchar(СИМВОЛЫ, СТРОКА)`.
- Выяснение длины (числа символов) для строковой переменной — `length(СТРОКА)`.
- Сравнение строк по лексикографическому порядку — `alphabefore(СТРОКА1, СТРОКА2)`.
- Удаление пробелов по краям, в начале или в конце значений — соответственно `trim(СТРОКА)`, `trim_start(СТРОКА)` или `trimend(СТРОКА)`.
- Извлечь первые или последние *n* символов из строки — `startstring(ДЛИНА, СТРОКА)` или `endstring(ДЛИНА, СТРОКА)`. Например, пусть у вас есть поле с именем *позиция\_ТМЦ*, в котором сочетается имя продукта и 4-значный код ID (КАМЕРА АКМЕ-D109). Чтобы создать новое поле, содержащее только 4-значный код, задайте в узле вычислений формулу:  
`endstring(4, позиция_ТМЦ)`
- Проверка совпадения с шаблоном — `СТРОКА matches ШАБЛОН`. Например, чтобы выбрать сотрудников, в должности которых так или иначе упоминается "рынок" или что-то "рыночное", можно задать на узле выбора следующее:  
`должность matches "*рыно*"`
- Замена всех экземпляров подстроки в строке — `replace(ПОДСТРОКА, НОВАЯПОДСТРОКА, СТРОКА)`. Например, чтобы перед исследованием данных заменить все экземпляры неподдерживаемого символа, например, заменить вертикальную черту конвейера (|) на точку с запятой, задайте на узле заполнения функцию `replace`. В разделе **Заполнить поля**: выберите все поля, в которых может встретиться этот символ. В качестве условия **Заменять**: выберите **Всегда** и задайте следующее условие в разделе **Заменить на**:  
`replace('|', ';', @FIELD)`
- Вычисление поля флага исходя из наличия указанной подстроки. Например, при помощи строковой функции в приведенном ниже выражении на узле вычислений можно сгенерировать отдельное поле для каждого ответа:

```
hassubstring(museums, "museum_of_design")
```

Дополнительную информацию смотрите в разделе “Функции для работы с текстовыми значениями” на стр. 97.

---

## Обработка пустых и пропущенных значений

Замена пустых или пропущенных значений - обычная практика при подготовке данных для программ исследования данных. CLEM предлагает ряд инструментов для автоматической обработки пустых значений. Чаще всего для работы с пробелами создается узел заполнения, хотя приведенные ниже функции можно использовать на любом узле, принимающем выражения CLEM:

- `@BLANK(ПОЛЕ)` служит для выявления записей, значения которых пусты, для конкретного поля, такого как *Возраст*.

- @NULL (ПОЛЕ) служит для выявления записей, содержащих системные пропущенные значения, для одного или нескольких указанных полей. В IBM SPSS Modeler системные пропущенные значения выводятся в виде \$null\$.

Дополнительную информацию смотрите в разделе “Функции для обработки пустых значений и значений Null” на стр. 112.

---

## Работа с числами

В IBM SPSS Modeler доступен ряд стандартных операций с числовыми значениями, таких как:

- Вычисление синуса заданного угла —  $\sin(\text{ЧИСЛО})$
- Вычисление натурального логарифма числовых полей —  $\log(\text{ЧИСЛО})$
- Вычисление суммы двух чисел —  $\text{ЧИСЛО1} + \text{ЧИСЛО2}$

Дополнительную информацию смотрите в разделе “Числовые функции” на стр. 93.

---

## Работа с датами и временем

Форматы даты и времени могут меняться в зависимости от источника данных и национальной версии. В каждом потоке можно задавать свои форматы даты и времени, это делается в диалоговом окне свойств потока. В приведенных ниже примерах показаны широко используемые функции для работы с полями даты и времени.

Вычисление прошедшего времени

При помощи семейства функций, аналогичных приведенной ниже, удобно вычислять, сколько времени прошло после базовой даты. Эта функция возвращает, как действительное число, время в месяцах от базовой даты до даты, представленной строкой ДАТА. Это приблизительная величина в предположении, что месяц - это 30,0 дней.

`date_in_months(ДАТА)`

Сравнение значений даты и времени

При помощи функций, аналогичных приведенной ниже, можно сравнивать значения полей даты или времени из разных записей. Эта функция возвращает значение *true* (истина), если строка даты ДАТА1 представляет дату, которая раньше даты, представленной строкой даты ДАТА2. В противном случае эта функция возвращает значение 0.

`date_before(ДАТА1, ДАТА2)`

Вычисление разностей

Вы также можете вычислять разности между двумя значениями времени и двумя значениями даты, используя такие функции, как:

`date_weeks_difference(Дата1, Дата2)`

Эта функция возвращает, как действительное число, время в неделях от даты, представленной строкой даты ДАТА1, до даты, представленной строкой даты ДАТА2. Предполагается, что неделя равна 7,0 дней. Если ДАТА2 раньше, чем ДАТА1, эта функция возвращает отрицательное значение.

Сегодняшняя дата



Текущую дату можно добавить в набор данных, используя функцию @TODAY. Сегодняшняя дата добавляется как строка в указанное поле или в новое поле с использованием формата даты, выбранного в диалоговом окне свойств потока. Дополнительную информацию смотрите в разделе “Функции даты и времени” на стр. 102.

---

## Сведение воедино нескольких полей

Язык CLEM содержит ряд функций, возвращающих сводную статистику по нескольким полям. Эти функции особенно полезны при анализе данных опросов, где несколько ответов на вопрос сохраняются в отдельных полях. Дополнительную информацию смотрите в разделе “Работа с данными множественных ответов” на стр. 76.

### Функции сравнения

Для сравнения значения в нескольких полях служат функции `min_n` и `max_n`, например:

```
max_n(['оплатакарты1' 'оплатакарты2' 'оплатакарты3' 'оплатакарты4'])
```

Кроме того, вы можете пользоваться рядом функций подсчета, чтобы узнать число значений, которые удовлетворяют определенным критериям, даже если эти значения хранятся в разных полях. Например, чтобы подсчитать число карт, которые оставались у держателей более пяти лет:

```
count_greater_than(5, ['владениекартой' 'владениекартой2' 'владениекартой3'])
```

Чтобы подсчитать пустые значения в том же наборе полей:

```
count_nulls(['владениекартой' 'владениекартой2' 'владениекартой3'])
```

Обратите внимание, что здесь подсчитывается число карт, а не число их держателей. Дополнительную информацию смотрите в разделе “Функции сравнения” на стр. 91.

Чтобы подсчитать, сколько раз указанное значение встречается в нескольких полях, можно использовать функцию `count_equal`. В следующем примере подсчитывается число полей в списке, содержащих значение Да.

```
count_equal("Да",[Ответ1, Ответ2, Ответ3])
```

Работа этой функции поясняется следующими примерами подсчета значений Да при разных значениях полей.

*Таблица 4. Значения функции.*

Ответ1	Ответ2	Ответ3	Количество
Да	Нет	Да	2
Да	Нет	Нет	1

### Числовые функции

Вы можете получить статистику по нескольким полям, пользуясь функциями `sum_n`, `mean_n` и `sdev_n`, например:

```
sum_n(['баланскарты1' 'баланскарты2' 'баланскарты3'])  
mean_n(['баланскарты1' 'баланскарты2' 'баланскарты3'])
```

Дополнительную информацию смотрите в разделе “Числовые функции” на стр. 93.

### Генерирование списков полей

Для функций, которые принимают на входе список полей, можно в качестве входного аргумента задавать специальные функции диапазона полей @FIELDS\_BETWEEN(начало, конец) и выбора полей по шаблону @FIELDS\_MATCHING(шаблон). Например, если порядок следования полей как показано выше в примере sum\_n, тот же результат даст следующая формула:

```
sum_n(@FIELDS_BETWEEN(баланскарты1, баланскарты3))
```

Другой вариант - подсчитать число пустых значений во всех полях, начинающихся с "баланскарты":

```
count_nulls(@FIELDS_MATCHING('баланскарты*'))
```

Дополнительную информацию смотрите в разделе “Специальные поля” на стр. 113.

---

## Работа с данными множественных ответов

При анализе данных множественных ответов можно пользоваться рядом функций сравнения, включая следующие:

- value\_at
- first\_index / last\_index
- first\_non\_null / last\_non\_null
- first\_non\_null\_index / last\_non\_null\_index
- min\_index / max\_index

Например, пусть вопрос с множественными ответами задан о первой, второй и третьей по важности причинах решения сделать конкретную покупку (например, среди таких причин, как цена, личная рекомендация, отзыв в СМИ, местный поставщик, другие). В этом случае вы можете узнать о важности цены, вычислив индекс поля, в которое она впервые вошла:

```
first_index("цена", [Причина1 Причина2 Причина3])
```

Аналогичным образом, пусть, например, вы попросили клиентов ранжировать три автомобиля в порядке вероятности покупки и закодировали ответы как три поля следующим образом:

Таблица 5. Пример ранжирования автомобилей.

ID клиента	автомобиль1	автомобиль2	автомобиль3
101	1	3	2
102	3	2	1
103	2	3	1

В этом случае вы можете узнать индекс поля самого популярного среди опрошенных клиентов автомобиля (получившего ранг #1, то есть самый низший), воспользовавшись функцией min\_index:

```
min_index(['автомобиль1' 'автомобиль2' 'автомобиль3'])
```

Дополнительную информацию смотрите в разделе “Функции сравнения” на стр. 91.

Ссылки на наборы множественных ответов

При помощи специальной функции @MULTI\_RESPONSE\_SET можно ссылаться на все поля в наборе множественных ответов. Например, если в предыдущем примере три поля *автомобиль* входят в набор множественных ответов с именем *ранжировки\_автомобилей*, тот же результат можно получить так:

```
max_index(@MULTI_RESPONSE_SET("ранжировки_автомобилей"))
```

---

## Построитель выражений

Выражения CLEM можно набирать вручную, а можно использовать построитель выражений, который содержит список всех функций и операций CLEM, а также полей данных текущего потока, и позволяет быстрее строить выражения, избавляя от необходимости запоминать точные имена полей и функций. Кроме того, элементы управления построителя автоматически заключают в кавычки нужные поля и значения, облегчая создание синтаксически правильных выражений.

*Примечание:* построитель выражений не поддерживается в сценариях и настройках параметров.

## Доступ к построителю выражений

Построитель выражений доступен на всех узлах, где используются выражения CLEM, включая узлы выбора, балансировки, вычислений, заполнения, анализа и отчетов, а также табличные узлы. Открыть его можно щелчком по значку калькулятора справа от поля формулы.

## Создание выражений

Построитель выражений предлагает не только полные списки полей, функций и операций, но также доступ к значениям данных, если создан экземпляр данных.

Чтобы создать выражение с использованием построителя выражений

1. Наберите текст в поле выражения, пользуясь списками функций и полей для справки.  
*или*
2. Выберите нужные поля и функции в списках прокрутки.
3. Чтобы добавить поле или функцию в поле выражения, щелкните дважды или нажмите кнопку с желтой стрелкой.
4. Чтобы вставить в выражение операции, используйте кнопки операндов в центре диалогового окна.

## Выбор функций

Список функций содержит все функции и операции, доступные в CLEM. Прокрутите список, чтобы найти нужную функцию, или, чтобы ускорить поиск, выберите в выпадающем списке нужный поднабор функций или операций. Доступные функции сгруппированы по категориям для удобства поиска.

Большая часть этих категорий описана в разделе Ссылки в описании языка CLEM. Дополнительную информацию смотрите в разделе “Справка по функциям” на стр. 88.

Остальные категории приведены ниже.

- **Общие функции** содержит подборку наиболее широко употребляемых функций.
- **Последние использованные** содержит список функций CLEM, использованных в текущем сеансе.
- **Функции @** содержит список всех специальных функций, имена которых начинаются с символа "@".
- **Функции базы данных.** Если поток содержит соединение с базой данных (в виде узла источника базы данных), то этот раздел содержит функции, доступные из этой базы данных, включая пользовательские функции (UDF).
- **Агрегации по базе данных.** Если поток содержит соединение с базой данных (в виде узла источника базы данных), этот раздел содержит опции агрегации, доступные в базе данных.
- **Агрегации окна базы данных.** Если поток содержит соединение с базой данных (в виде узла источника базы данных), этот раздел содержит опции агрегации окна, которые вы можете использовать в базе данных.

Эти опции называются "агрегациями", но они не предназначены для использования в узле Агрегация, они применимы скорее к таким узлам, как Производные данные или Выбор. Это объясняется тем что выходные значения - скалярные, а не настоящие агрегации, то есть они не сокращают объем выходных

данных в выводе подобно тому, как это делает узел Агрегация. Например, вы можете использовать этот сорт агрегации для вычисления скользящего среднего по строкам данных, такого как "среднее текущей строки и всех предыдущих строк".

- **Встроенные агрегации.** Содержит список возможных для использования режимов агрегации.
- **Операции** содержит все операции, которые можно использовать при построении выражений. Операции также доступны через кнопки в центра диалогового окна.
- **Все функции** содержат полный список функций, доступных в CLEM.

Выбрав группу функций, щелкните дважды, чтобы вставить функции в поле выражения в точке, где находится указатель ввода.

## Выбор полей, параметров и глобальных переменных

Поле содержит все поля, доступные в этой точке потока данных. Прокрутите список, чтобы найти нужное поле. Чтобы добавить поле в выражение, щелкните дважды или нажмите кнопку с желтой стрелкой.

Дополнительную информацию смотрите в разделе “Параметры потока, сеанса и надузла” на стр. 72.

Кроме полей, можно выбрать следующее:

**Наборы множественных ответов.** Дополнительную информацию смотрите в руководстве *Узлы источников, обработки и вывода в IBM SPSS Modeler*.

**Последние использованные** содержит список полей, наборов множественных ответов, параметров и глобальных переменных, использованных в текущем сеансе.

**Анализ важности независимых переменных.** Дополнительную информацию смотрите в разделе “Параметры потока, сеанса и надузла” на стр. 72.

**Глобальные переменные.** Дополнительную информацию смотрите в руководстве *Узлы источников, обработки и вывода в IBM SPSS Modeler*.

## Просмотр или выбор значений

В системе есть ряд мест, откуда можно просмотреть значения полей: из построителя выражений, отчетов аудита данных и при редактировании будущих значений на узле Временные интервалы. Обратите внимание на то, что для использования этой возможности экземпляр данных должен быть полностью создан на узле источника или типа, чтобы уже были известны система хранения, типы и значения.

Чтобы просмотреть значения некоторого поля из построителя выражений или с узла временных интервалов, выберите нужное поле и нажмите кнопку выбора значений, чтобы открыть диалоговое окно, содержащее выражения выбранного поля. Если выбрать одно из значений и нажать кнопку **Вставка**, оно будет вставлено в текущее выражение или список.



Рисунок 14. Кнопка выбора значений

В случае поля флага или номинального поля будут выведены все определенные значения. В случае непрерывного поля (поля числового диапазона) будут выведены минимальное и максимальное значения.

## Проверка выражений CLEM

Чтобы проверить выражение, в построителе выражений (в нижнем правом углу) выберите **Проверить**. Непроверенные выражения выводятся в красном цвете. Если найдены ошибки, выводится сообщение с указанием причины.

Проверяется следующее:

- Правильное взятие в кавычки значений и имен полей
- Правильное использование параметров и глобальных переменных
- Допустимое использование операций
- Существование полей, на которые делается ссылка
- Существование и задание глобальных объектов, на которые делается ссылка

Если у вас встречаются синтаксические ошибки, попробуйте не набирать выражения вручную, а пользоваться кнопками списков и операций. Тогда поля и значения будут браться в кавычки автоматически.

---

## Найти и заменить

Диалоговое окно Найти/заменить доступно там, где вы редактируете текст сценария или выражения, в том числе в редакторе сценариев, в построителе выражений CLEM или при определении шаблонов на узле Отчет. При изменении текста в любой из этих областей нажмите комбинацию клавиш Ctrl+F для перехода в диалоговое окно изменений и убедитесь, что указатель мыши сфокусирован в области текста. Например, при работе на узле Заполнитель можно получить доступ к диалоговому окну из любой текстовой области на вкладке Параметры или из поля текста в построителе выражений.

1. При указателе в области текста нажмите клавиши Ctrl+F для доступа к диалоговому окну Найти/заменить.
2. Введите текст для поиска или выберите из раскрывающегося списка элементы недавнего поиска.
3. Введите текст для замещения, если такой есть.
4. Нажмите кнопку **Найти далее** для запуска поиска.
5. Нажмите кнопку **Заменить** для замены текущего выбранного фрагмента или кнопку **Заменить все** для замены всех или выбранных экземпляров.
6. Диалоговое окно закрывается после каждой операции. Находясь в любой области текста, нажмите F3 для повторения последней операции поиска или Ctrl+F для повторного доступа к диалоговому окну.

Опции поиска

**Учитывать регистр.** Задаёт, учитывается ли при поиске регистр символов; например, совпадает ли *myvar* с *myVar*. Текст замещения всегда вставляется так, как он введен, независимо от этого параметра.

**Только слова целиком.** Задаёт, учитывает ли операция поиска совпадения текста внутри слов. При выборе этой опции поиск для *spider* не найдет соответствий с вариантами *spiderman* или *spider-man*.

**Регулярные выражения.** Задаёт, используется ли синтаксис регулярных выражений (смотрите следующий раздел). При выборе этой опции отключается опция **Только слова целиком** и ее значение игнорируется.

**Только выбранный текст.** Управляет областью выполнения поиска при использовании опции **Заменить все**.

Синтаксис регулярных выражений

Регулярные выражения позволяют искать специальные символы, такие как символы табуляции или перехода на новую строку, классы или диапазоны символов, такие как от *a* до *d* при любых цифровых или нецифровых параметрах и граничные положения, такие как начало или конец строки. Поддерживаются следующие типы выражений.

Таблица 6. Совпадения символов.

Символы	Совпадает
x	Символ x
\\	Символ обратной дробной черты
\0n	Символ с восьмеричным значением 0n (0 <= n <= 7)
\0nn	Символ с восьмеричным значением 0nn (0 <= n <= 7)
\0mnn	Символ с восьмеричным значением (0 <= m <= 3, 0 <= n <= 7)
\xhh	Символ с шестнадцатиричным значением 0xhh
\uhhhh	Символ с шестнадцатиричным значением 0xhhhh
\t	Символ табуляции ('\u0009')
\n	Символ новой строки (начало строки) ('\u000A')
\r	Символ возврата каретки ('\u000D')
\f	Символ перехода к новой странице ('\u000C')
\a	Символ предупреждения (звонок) ('\u0007')
\e	Символ перехода ESC ('\u001B')
\cx	Управляющий символ, соответствующий x

Таблица 7. Совпадения классов символов.

Классы символов	Совпадает
[abc]	a, b или c (простой класс)
[^abc]	Любой символ, кроме a, b или c (вычитание)
[a-zA-Z]	От a до z или от A до Z, включительно (диапазон)
[a-d[m-p]]	От a до d или от m до p (объединение). Альтернативно это можно задать как [a-dm-p]
[a-z&&[def]]	От a до z и d, e или f (пересечение)
[a-z&&[^bc]]	От a до z, кроме b и c (вычитание). Альтернативно это можно задать как [ad-z]
[a-z&&[^m-p]]	От a до z, но не от m до p (вычитание). Альтернативно это можно задать как [a-lq-z]

Таблица 8. Предварительно определенные классы символов.

Предварительно определенные классы символов	Совпадает
.	Любой символ (разделители строк могут и совпадать, и не совпадать)
\d	Любая цифра: [0-9]
\D	Не цифра: [^0-9]
\s	Пробельный символ: [ \t\n\r0B\f\r]
\S	Не пробельный символ: [^\s]
\w	Символ слова: [a-zA-Z_0-9]
\W	Символ не слова: [^\w]

Таблица 9. Совпадения границ.

Обнаружители совпадений границ	Совпадает
^	Начало строки

Таблица 9. Совпадения границ (продолжение).

Обнаружители совпадений границ	Совпадает
\$	Конец строки
\b	Граница слова
\B	Граница не слова
\A	Начало ввода
\Z	Конец ввода кроме заключительного разделителя, если он есть
\z	Конец ввода





---

## Глава 8. Языковые предпочтения CLEM

---

### Справочный обзор CLEM

В этом разделе описывается язык управления для работы с выражениями (Control Language for Expression Manipulation, CLEM), который представляет из себя мощный инструмент, используемый для анализа и обработки данных, используемых в потоках IBM SPSS Modeler. Вы можете использовать CLEM на узлах для выполнения задач самого разного свойства, от проверки выполнения условия или извлечения данных до вставки данных в отчеты.

Выражения CLEM состоят из значений, имен полей, операторов и функций. Правильно пользуясь синтаксисом, можно создавать множество мощных операций с данными.

---

### Типы данных CLEM

Типами данных CLEM могут быть любые из следующих:

- Целые числа
- Действительные числа
- Символы
- Строки
- Списки
- Поля
- Дата/Время

Правила для заключения в кавычки

Хотя у IBM SPSS Modeler весьма гибкое поведение при определении полей, значений, параметров и строк, используемых в выражении CLEM, следующие общие правила представляют список "настоятельных рекомендаций" для использования при создании выражений:

- Строки - Всегда используйте двойные кавычки при написании строк, например "Type 2". Вместо двойных можно использовать и одинарные кавычки, но существует риск путаницы с заключенными в кавычки полями.
- Поля - Используйте одинарные кавычки только при необходимости, при наличии пробелов или других специальных символов, например, 'Order Number'. Поля, заключенные в кавычки, но не определенные в наборе данных, будут ошибочно читаться как строки.
- Параметры - Всегда применяйте одинарные кавычки при использовании параметров, например, '\$P-threshold'.
- Символы - всегда используйте одинарные обратные кавычки (`), например, stripchar(`d`, "drugA").

Эти правила более подробно описываются в следующих разделах.

### Целые числа

Целые числа представлены как последовательность десятичных цифр. Дополнительно можно поместить перед целым числом знак минус (-) для обозначения отрицательного числа, например, 1234, 999, -77.

Язык CLEM обрабатывает целые числа с произвольной точностью. Максимальный размер целого числа зависит от вашей платформы. Если значения слишком большие для вывода в целочисленном поле, изменение типа поля на Действительное обычно восстанавливает значение.

## Действительные числа

Термин *Действительные числа* относятся к числу с плавающей запятой. Действительные числа представляются одной или несколькими цифрами, после которых следует запятая и одна или несколько цифр. Действительные числа в CLEM хранятся с двойной точностью.

Дополнительно перед действительным числом можно поместить знак минус (-) для обозначения отрицательного числа, например, 1,234, 0,999, -77,001. Используйте обозначения *<число>e <показатель степени>*, чтобы выразить действительное число в экспоненциальной форме, например, 1234,0e5, 1,7e-2. Когда прикладная программа IBM SPSS Modeler читает числовые строки из файлов и автоматически преобразует их в числа, принимаются и числа без цифр перед запятой или без цифр после, например, 999. или .11. Однако такая форма запрещена в выражениях CLEM.

*Примечание:* При использовании действительных чисел в выражениях CLEM в качестве десятичного разделителя нужно использовать точку независимо от параметров текущего потока или локали. Например, указывайте

```
Na > 0.6
```

а не

```
Na > 0,6
```

Такое обозначение используется, если даже запятая в качестве десятичного разделителя задана в диалоговом окне свойств потока, и соответствует общему правилу, что синтаксис кода должен быть независимым от любой конкретной локали или соглашения.

## Символы

Символы (привычно обозначаемые как CHAR) обычно используются в выражении CLEM для выполнения проверки строк. Например, можно использовать функцию `isuppercode` для определения, из верхнего ли регистра первый символ строки. Следующее выражение CLEM использует символ для указания, что нужно проверить первый символ строки:

```
isuppercode(subscrs(1, "MyString"))
```

Чтобы выразить код (а не положение) конкретного символа в выражении CLEM, используйте одинарные символы обратных кавычек в виде *`<символ>`*, например, ``A``, ``Z``.

*Примечание:* Типа хранения CHAR для поля не существует, поэтому значение в поле извлекается из выражения или заполняется выражением, которое дает в результате тип CHAR, а затем этот результат будет преобразован в строку.

## Строки

В общем случае вам нужно заключить строки в двойные кавычки. Примеры строк - это "c35product2" и "referrerID". Для обозначения специальных символов используйте обратную дробную черту, например, "\\$65443". (Для обозначения обратной дробной черты используйте двойную обратную дробную черту \\). Можно использовать и одинарные кавычки вокруг строки, но результат не будет отличаться от поля в кавычках ('referrerID'). Дополнительную информацию смотрите в разделе "Функции для работы с текстовыми значениями" на стр. 97.

## Списки

Список - это упорядоченная последовательность элементов, у которых может быть смешанный тип. Списки заключаются в квадратные скобки ([]). Примеры списков - это [1 2 4 16] и ["abc" "def"]. Списки не используются как значение полей IBM SPSS Modeler. Они используются для предоставления аргументов функциям, таких как участник и один из.

## Поля

Имена в выражениях CLEM - это не имена функций, которые предполагалось использовать в виде имен полей. Их можно записать просто как `Power`, `val27`, `state_flag` и так далее, но если имя начинается с цифры или включает в себя неалфавитные символы, например, пробелы (за исключением нижнего подчеркивания), заключите имя в одинарные кавычки, например, `'Power Increase'`, `'2nd answer'`, `'#101'`, `'$P-NextField'`.

*Примечание:* Поля в кавычках, но не определенные в наборе данных, будут ошибочно прочитаны как строки.

## Даты

Вычисления даты основаны на "базовой" дате, которая задается в диалоговом окне свойств потока. Базовая дата по умолчанию - это 1 января 1900 года.

Язык CLEM поддерживает следующие форматы даты.

Таблица 10. Форматы даты языка CLEM

Формат	Примеры
ДДММГГ	150163
ММДДГГ	011563
ГГММДД	630115
ГГГГММДД	19630115
ГГГГДДД	Четырехзначный год, после которого следует трехзначный номер дня в году, например, 2000032 представляет 32-й день 2000 года, то есть 1 февраля 2000 года.
ДЕНЬ	День недели в текущей локали, например, Понедельник, Вторник, ... в русской локали.
МЕСЯЦ	Месяц в текущей локали, например, Январь, Февраль, ....
ДД/ММ/ГГ	15/01/63
ДД/ММ/ГГГГ	15/01/1963
ММ/ДД/ГГ	01/15/63
ММ/ДД/ГГГШ	01/15/1963
ДД-ММ-ГГ	15-01-63
ДД-ММ-ГГГГ	15-01-1963
ММ-ДД-ГГ	01-15-63
ММ-ДД-ГГГГ	01-15-1963
ДД.ММ.ГГ	15.01.63
ДД.ММ.ГГГГ	15.01.1963
ММ.ДД.ГГ	01.15.63
ММ.ДД.ГГГГ	01.15.1963
ДД-МЕС-ГГ	15-ЯНВ-63, 15-январь-63, 15-Янв-63
ДД/МЕС/ГГ	15/ЯНВ/63, 15/январь/63, 15/Янв/63
ДД.МЕС.ГГ	15.ЯНВ.63, 15.январь.63, 15.Янв.63
ДД-МЕС-ГГГГ	15-ЯНВ-1963, 15-январь-1963, 15-Янв-1963
ДД/МЕС/ГГГГ	15/ЯНВ/1963, 15/январь/1963, 15/Янв/1963
ДД.МЕС.ГГГГ	15.ЯНВ.1963, 15.январь.1963, 15.Янв.1963
МЕС ГГГГ	Jan 2004

Таблица 10. Форматы даты языка CLEM (продолжение)

Формат	Примеры
к К ГГГГ	Дата, представленная цифрой (1–4), показывающая номер квартала, после чего следует буква <i>К</i> и четырехзначный год, например, 25 декабря 2004 года будет представлено как 4 К 2004.
нн НД ГГГГ	Двухзначное число, представляющее номер недели в году, после чего следуют буквы <i>НД</i> и четырехзначный год. Неделя года вычисляется в предположении, что первый день недели - это понедельник, а в первой неделе есть по крайней мере один день.

## Время

Язык CLEM поддерживает следующие форматы времени.

Таблица 11. Форматы времени языка CLEM

Формат	Примеры
ччммсс	120112, 010101, 221212
ччмм	1223, 0745, 2207
ммсс	5558, 0100
чч:мм:сс	12:01:12, 01:01:01, 22:12:12
чч:мм	12:23, 07:45, 22:07
мм:сс	55:58, 01:00
(ч)ч:(м)м:(с)с	12:1:12, 1:1:1, 22:12:12
(ч)ч:(м)м	12:23, 7:45, 22:7
(м)м:(с)с	55:58, 1:0
чч.мм.сс	12.01.12, 01.01.01, 22.12.12
чч.мм	12.23, 07.45, 22.07
мм.сс	55.58, 01.00
(ч)ч.(м)м.(с)с	12.1.12, 1.1.1, 22.12.12
(ч)ч.(м)м	12.23, 7.45, 22.7
(м)м.(с)с	55.58, 1.0

## Операторы CLEM

Доступны следующие операторы.

Таблица 12. Операторы языка CLEM.

Операция	Комментарии	Старшинство (смотрите следующий раздел)
or	Используется между двумя выражениями CLEM. Возвращает значение true, если у любого или обоих выражений значение true.	10
and	Используется между двумя выражениями CLEM. Возвращает значение true, если значения обоих выражений - true.	9
=	Используется между двумя сравнимыми элементами. Возвращает значение true, если ITEM1 совпадает с ITEM2.	7

Таблица 12. Операторы языка CLEM (продолжение).

Операция	Комментарии	Старшинство (смотрите следующий раздел)
==	Идентично =.	7
/=	Используется между двумя сравнимыми элементами. Возвращает значение true, если ITEM1 не равен ITEM2.	7
/==	Идентично /=.	7
>	Используется между двумя сравнимыми элементами. Возвращает значение true, если ITEM1 строго больше ITEM2.	6
>=	Используется между двумя сравнимыми элементами. Возвращает значение true, если ITEM1 больше или равно ITEM2.	6
<	Используется между двумя сравнимыми элементами. Возвращает значение true, если ITEM1 строго меньше ITEM2.	6
<=	Используется между двумя сравнимыми элементами. Возвращает значение true, если ITEM1 меньше или равно ITEM2.	6
&&=_0	Используется между двух целых чисел. Эквивалентно логическому выражению INT1 && INT2 = 0.	6
&&/=_0	Используется между двух целых чисел. Эквивалентно логическому выражению INT1 && INT2 /= 0.	6
+	Складывает два числа: NUM1 + NUM2.	5
><	Объединяет две строки; например, STRING1 >< STRING2.	5
-	Вычитает одно число из другого: NUM1 - NUM2. Может использоваться также перед числом: - NUM.	5
*	Используется для умножения двух чисел: NUM1 * NUM2.	4
&&	Используется между двух целых чисел. Результат - это поразрядное 'И' целых чисел INT1 и INT2.	4
&&~	Используется между двух целых чисел. Результат - это поразрядное 'И' для INT1 и поразрядного дополнения INT2.	4
	Используется между двух целых чисел. Результат - это поразрядное 'ИЛИ включительно' для INT1 и INT2.	4
~	Используется перед целым числом. Вычисляет поразрядное дополнение INT.	4
/&	Используется между двух целых чисел. Результат - это поразрядное 'ИЛИ исключительно' для INT1 и INT2.	4
INT1 << N	Используется между двух целых чисел. Вычисляет битовый образ значения INT, сдвинутый влево на N позиций.	4

Таблица 12. Операторы языка CLEM (продолжение).

Операция	Комментарии	Старшинство (смотрите следующий раздел)
INT1 >> N	Используется между двух целых чисел. Вычисляет битовый образ значения INT, сдвинутый вправо на N позиций.	4
/	Используется для деления одного числа на другое: NUM1 / NUM2.	4
**	Используется между двумя числами: BASE ** POWER. Возвращает результат возведения BASE в степень POWER.	3
rem	Используется между двумя целыми числами: INT1 rem INT2. Возвращает остаток, INT1 - (INT1 div INT2) * INT2.	2
div	Используется между двумя целыми числами: INT1 div INT2. Выполняет целочисленное деление.	2

### Старшинство операторов

Старшинство определяет синтаксический анализ сложных выражений, особенно не использующих скобок, с несколькими операторами infix. Например,

$3 + 4 * 5$

анализируется как  $3 + (4 * 5)$ , а не  $(3 + 4) * 5$ , так как относительное старшинство указывает на первоначальный анализ оператора \* до оператора +. У каждого оператора языка CLEM есть связанное с ним значение старшинства; чем меньше это значение, тем важнее оно в списке анализа, то есть соответствующий оператор будет обрабатываться до других операторов с большими значениями старшинства.

## Справка по функциям

Следующие функции CLEM доступны для работы с данными в IBM SPSS Modeler. Вы можете ввести код для этих функций во многих диалоговых окнах, например, в узлах Извлечь и Задать как флаг, или можно использовать построитель выражений, чтобы создать допустимые выражения CLEM, не запоминая списки функций или имена полей.

Таблица 13. Функции CLEM для использования с данными IBM SPSS Modeler

Тип функции	Описание
Информация	Используется для получения сведений о значениях полей. Например, функция is_string возвращает значение true для всех записей строкового типа.
Преобразования типов	Используется для построения новых полей или преобразования типа хранения. Например, функция to_timestamp преобразует выбранное поле в отметку времени.
Сравнение	Используется для сравнения значений в полях друг с другом или с заданной строкой. Например, <= используется для сравнения значений в двух полях по соотношению меньше или равно.
Логические	Используется для выполнения логических операций, таких как if, then, else.
Числовой	Используется для выполнения числовых вычислений, например, натурального логарифма значения в поле.
Тригонометрические	Используется для выполнения тригонометрических вычислений, например, арккосинуса заданного угла.

Таблица 13. Функции CLEM для использования с данными IBM SPSS Modeler (продолжение)

Тип функции	Описание
Вероятность	Возвратить вероятности на основе различных распределений, например, вероятность того, что значение из распределения Стьюдента <i>t</i> будет меньше конкретного значения.
Битовые	Используется для обращения с целыми числами как с наборами бит.
Переменный	Используется для случайного выбора элементов или генерирования случайных чисел.
Строчное	Используется для выполнения разнообразных операций со строками, например, <code>stripchar</code> позволяет удалить конкретный символ.
SoundEx	Используется для нахождения строк, когда их точное написание не известно, на основании фонетических предположений, как произносятся определенные буквы.
Дату и время	Используется для выполнения разнообразных операций с полями даты, времени отметки времени.
Последовательность	Используется для получения информации о последовательности записей в наборе данных или для выполнения операций на основании этой последовательности.
Глобальный	Используется для доступа к глобальным значениям, созданным на узле Задать глобальные значения. Например, <code>@MEAN</code> используется для ссылки на среднее для всех значений в поле по всему набору данных.
Пустые и null	Используется для доступа к пустым пользовательским значениям или системным значениям отсутствия, отметки флагом и часто для заполнения. Например, <code>@BLANK (FIELD)</code> используется для установления флага <code>true</code> записям, в которых присутствуют пустые значения.
Специальные поля	Используется для обозначения конкретных исследуемых полей. Например, <code>@FIELD</code> используется при извлечении нескольких полей.

## Соглашения в описаниях функций

Следующие соглашения используются по всему этому руководству при ссылке на элементы функции.

Таблица 14. Соглашения в описаниях функций.

Соглашение	Описание
<i>BOOL</i>	Логическое значение, или флаг, например, <code>true</code> или <code>false</code> .
<i>NUM</i> , <i>NUM1</i> , <i>NUM2</i>	Любое число.
<i>REAL</i> , <i>REAL1</i> , <i>REAL2</i>	Любое действительное число, такое как 1,234 или -77,01.
<i>INT</i> , <i>INT1</i> , <i>INT2</i>	Любое целое число, такое как 1 или -77.
<i>CHAR</i>	Код символа, такой как <code>`A`</code> .
<i>STRING</i>	Строка, такая как <code>"referrerID"</code> .
<i>LIST</i>	Список элементов, таких как <code>["abc" "def"]</code> .
<i>ITEM</i>	Поле, такое как Покупатель или <code>extract_concept</code> .
<i>DATE</i>	Поле даты, такое как <code>начальная_дата</code> , где значения приведены в таком формате, как ДД-МЕС-ГГГГ.
<i>TIME</i>	Поле времени, такое как <code>power_flux</code> , где значения приведены в таком формате, как ЧЧММСС.

В этом руководстве перечислены функции и их назначение в одном столбце, тип результата (целочисленный, строка и т.д.) - в другом, описания (где они доступны) - в третьем столбце. Например, ниже представлено описание функции rem.

Таблица 15. Описание функции rem.

Функция	Результат	Описание
INT1 rem INT2	Число	Возвращает остаток от деления <i>INT1</i> на <i>INT2</i> . Например, INT1 - (INT1 div INT2) * INT2.

Подробности соглашений об использовании, например, как перечислять элементы или задавать символы в функции, описываются далее. Дополнительную информацию смотрите в разделе “Типы данных CLEM” на стр. 83.

## Информационные функции

Информационные функции используются для получения сведений о значениях определенного поля. Обычно они используются для получения флаговых полей. Например, можно использовать функцию @BLANK для создания флагового поля, обозначающего записи с пустыми значениями для выбранного поля. Аналогично, вы можете проверить тип хранения для поля, используя любую из функций типа хранения, например, is\_string.

Таблица 16. Информационные функции CLEM.

Функция	Результат	Описание
@BLANK(FIELD)	Логический	Возвращает значение true для всех записей, значения которых пустые, в соответствии с правилами обработки пустых значений, заданными на вышележащем узле Тип или на исходном узле (вкладка Типы).
@NULL(ITEM)	Логический	Возвращает значение true для всех записей с не определенными значениями. Не определенные значения - это системные значения null, выводимые в IBM SPSS Modeler как \$null\$.
is_date(ITEM)	Логический	Возвращает true для всех записей, тип которых - дата.
is_datetime(ITEM)	Логический	Возвращает true для всех записей, тип которых - дата, время или отметка времени.
is_integer(ITEM)	Логический	Возвращает true для всех записей, тип которых - целое число.
is_number(ITEM)	Логический	Возвращает true для всех записей, тип которых - число.
is_real(ITEM)	Логический	Возвращает true для всех записей, тип которых - действительное число.
is_string(ITEM)	Логический	Возвращает true для всех записей, тип которых - строка.
is_time(ITEM)	Логический	Возвращает true для всех записей, тип которых - время.
is_timestamp(ITEM)	Логический	Возвращает true для всех записей, тип которых - отметка времени.

## Функции преобразования

Функции преобразования позволяют конструировать новые поля и преобразовывать тип хранения существующих полей. Например, можно образовывать новые строки, присоединяя их друг к другу, или рассматривая строки отдельно. Для присоединения двух строк друг к другу используйте оператор ><. Например, если у поля Site есть значение "BRAMLEY", "xx" >< Site возвратит значение "xxBRAMLEY". Результат оператора >< - всегда строка, даже если аргументы - это не строки. То есть, если у поля V1 значение 3, а у поля V2 значение 5, результатом V1 >< V2 будет "35" (строка, а не число).



Функции преобразования (и любые другие функции, требующие конкретного типа входных данных, такого как значение даты или времени) зависят от текущих форматов, заданных в диалоговом окне Опции потока. Например, если вы хотите преобразовать строковое поле со значениями *Jan 2003*, *Feb 2003* и так далее, выберите совпадающий формат даты **МЕС ГГГГ** как формат даты по умолчанию для потока.

Таблица 17. Функции преобразования CLEM

Функция	Результат	Описание
ITEM1 >> ITEM2	Строчное	Объединяет значения для двух полей и возвращает итоговую строку <i>ITEM1ITEM2</i> .
to_integer(ITEM)	Целое число	Преобразует тип хранения заданного поля в целое число.
to_real(ITEM)	Действительное число	Преобразует тип хранения заданного поля в действительное число.
to_number(ITEM)	Число	Преобразует тип хранения заданного поля в число.
to_string(ITEM)	Строчное	Преобразует тип хранения заданного поля в строку.
to_time(ITEM)	Время	Преобразует тип хранения заданного поля во время.
to_date(ITEM)	Дата	Преобразует тип хранения заданного поля в дату.
to_timestamp(ITEM)	Метка даты/времени	Преобразует тип хранения заданного поля в отметку времени.
to_datetime(ITEM)	Дата и время	Преобразует тип хранения заданного поля в значение даты, времени или отметки времени.
datetime_date(ITEM)	Дата	Возвращает значение даты для <i>числа</i> , <i>строки</i> или <i>отметки времени</i> . Обратите внимание на то, что это единственная функция, позволяющая преобразовать число (секунд) обратно в дату. Если ITEM - это строка, создает дату по результатам синтаксического анализа строки с текущим форматом даты. Формат даты, заданный в диалоговом окне свойств потока, должен быть правильным, тогда эта функция будет успешно выполнена. Если ITEM - это число, оно интерпретируется как число секунд от базовой даты (или эпохи). Доли дня урезаются. Если ITEM - это отметка времени, возвращается часть этой отметки, представляющая собой дату. Если ITEM - это дата, она возвращается без изменений.

## Функции сравнения

Функции сравнения используются для сравнения значений в полях друг с другом или с заданной строкой. Например, можно проверить строки на равенство при помощи =. Пример проверки эквивалентности строк следующий: `Class = "class 1"`.

Для целей численного сравнения *больше* означает более близкое положение к положительной бесконечности, а *меньше* - к отрицательной. То есть, все отрицательные числа меньше любого положительного числа.

Таблица 18. Функции сравнения CLEM

Функция	Результат	Описание
count_equal(ITEM1, LIST)	Целое число	Возвращает количество значений из списка полей, которые равны <i>ITEM1</i> , или null, если <i>ITEM1</i> - это null.
count_greater_than(ITEM1, LIST)	Целое число	Возвращает количество значений из списка полей, которые больше <i>ITEM1</i> , или null, если <i>ITEM1</i> - это null.
count_less_than(ITEM1, LIST)	Целое число	Возвращает количество значений из списка полей, которые меньше <i>ITEM1</i> , или null, если <i>ITEM1</i> - это null.
count_not_equal(ITEM1, LIST)	Целое число	Возвращает количество значений из списка полей, которые не равны <i>ITEM1</i> , или null, если <i>ITEM1</i> - это null.

Таблица 18. Функции сравнения CLEM (продолжение)

Функция	Результат	Описание
count_nulls(LIST)	Целое число	Возвращает количество значений null из списка полей.
count_non_nulls(LIST)	Целое число	Возвращает количество значений не null из списка полей.
date_before(DATE1, DATE2)	Логический	Используется для проверки порядка значений даты. Возвращает значение true, если DATE1 предшествует DATE2.
first_index(ITEM, LIST)	Целое число	Возвращает индекс первого поля, содержащего ITEM, из списка полей LIST или 0, если значение не найдено. Поддерживается только для типов строки, целого числа и действительного числа.
first_non_null(LIST)	Все	Возвращает первое отличное от null значение в предоставленном списке полей. Поддерживаются все типы хранения.
first_non_null_index(LIST)	Целое число	Возвращает индекс первого поля с отличным от null значением из заданного списка LIST или 0, если все значения равны null. Поддерживаются все типы хранения.
ITEM1 = ITEM2	Логический	Возвращает значение true для записей, в которых ITEM1 равен ITEM2.
ITEM1 /= ITEM2	Логический	Возвращает значение true, если две строки не идентичны, или 0, если они идентичны.
ITEM1 < ITEM2	Логический	Возвращает значение true для записей, в которых ITEM1 меньше ITEM2.
ITEM1 <= ITEM2	Логический	Возвращает значение true для записей, в которых ITEM1 меньше или равен ITEM2.
ITEM1 > ITEM2	Логический	Возвращает значение true для записей, в которых ITEM1 больше ITEM2.
ITEM1 >= ITEM2	Логический	Возвращает значение true для записей, в которых ITEM1 больше или равен ITEM2.
last_index(ITEM, LIST)	Целое число	Возвращает индекс последнего поля, содержащего ITEM, из списка полей LIST или 0, если значение не найдено. Поддерживается только для типов строки, целого числа и действительного числа.
last_non_null(LIST)	Все	Возвращает последнее отличное от null значение в предоставленном списке полей. Поддерживаются все типы хранения.
last_non_null_index(LIST)	Целое число	Возвращает индекс последнего поля с отличным от null значением из заданного списка LIST или 0, если все значения равны null. Поддерживаются все типы хранения.
max(ITEM1, ITEM2)	Все	Возвращает большее из двух значений - ITEM1 или ITEM2.
max_index(LIST)	Целое число	Возвращает индекс поля, содержащего максимальное значение из списка числовых полей, или 0, если все значения равны null. Например, если максимальное значение содержится в третьем из перечисленных полей, возвращается значение индекса 3. Если максимальное значение содержится в нескольких полях, возвращается индекс перечисленного первым (самого левого) поля.
max_n(LIST)	Число	Возвращает максимальное значение из списка числовых полей или null, если во всех полях значения null.
member(ITEM, LIST)	Логический	Возвращает true, если ITEM входит в заданный список LIST. В противном случае возвращает значение false. Можно задавать также список имен полей.
min(ITEM1, ITEM2)	Все	Возвращает меньшее из двух значений - ITEM1 или ITEM2.

Таблица 18. Функции сравнения CLEM (продолжение)

Функция	Результат	Описание
min_index(LIST)	Целое число	Возвращает индекс поля, содержащего минимальное значение из списка числовых полей, или 0, если все значения равны null. Например, если минимальное значение содержится в третьем из перечисленных полей, возвращается значение индекса 3. Если минимальное значение содержится в нескольких полях, возвращается индекс перечисленного первым (самого левого) поля.
min_n(LIST)	Число	Возвращает минимальное значение из списка числовых полей или null, если во всех полях значения null.
time_before(TIME1, TIME2)	Логический	Используется для проверки порядка значений времени. Возвращает значение true, если <i>TIME1</i> раньше <i>TIME2</i> .
value_at(INT, LIST)		Возвращает значение каждого перечисленного поля со сдвигом INT или NULL, если сдвиг оказывается вне диапазона допустимых значений (то есть, меньше 1 или больше число перечисленных полей). Поддерживаются все типы хранения.

## Логические функции

Выражения CLEM можно использовать для выполнения логических операций.

Таблица 19. Логические функции CLEM

Функция	Результат	Описание
COND1 and COND2	Логический	Эта операция логической конъюнкции возвращает значение true, если и <i>COND1</i> , и <i>COND2</i> - это true. Если <i>COND1</i> - это false, значение <i>COND2</i> не вычисляется; это делает возможным операции конъюнкции, в которых <i>COND1</i> сначала проверяет, что операция <i>COND2</i> допустимая. Например, <code>length(Label) &gt;=6</code> и <code>Label(6) = 'x'</code> .
COND1 or COND2	Логический	Эта операция логической (включительной) дизъюнкции возвращает значение true, если или <i>COND1</i> , или <i>COND2</i> равны true, или оба значения - true. Если <i>COND1</i> - это true, <i>COND2</i> не вычисляется.
not(COND)	Логический	Эта операция логического отрицания возвращает значение true, если <i>COND</i> - это false. В противном случае эта операция возвращает значение 0.
if COND then EXPR1 else EXPR2 endif	Все	Это операция условного вычисления. Если <i>COND</i> - это true, данная операция возвращает результат <i>EXPR1</i> . В противном случае возвращается результат вычисления <i>EXPR2</i> .
if COND1 then EXPR1 elseif COND2 then EXPR2 else EXPR_N endif	Все	Это операция разветвленного условного вычисления. Если <i>COND1</i> - это true, данная операция возвращает результат <i>EXPR1</i> . В противном случае, если <i>COND2</i> - это true, данная операция возвращает результат вычисления <i>EXPR2</i> . В противном случае возвращается результат вычисления <i>EXPR_N</i> .

## Числовые функции

CLEM содержит несколько общеизвестных числовых функций.

Таблица 20. Числовые функции CLEM.

Функция	Результат	Описание
-NUM	Число	Используется для изменения знака <i>NUM</i> . Возвращает соответствующее число с противоположным знаком.

Таблица 20. Числовые функции CLEM (продолжение).

Функция	Результат	Описание
NUM1 + NUM2	Число	Возвращает сумму NUM1 и NUM2.
NUM1 – NUM2	Число	Возвращает результат вычитания NUM2 из NUM1.
NUM1 * NUM2	Число	Возвращает произведение значений NUM1 и NUM2.
NUM1 / NUM2	Число	Возвращает результат деления NUM1 на NUM2.
INT1 div INT2	Число	Используется для выполнения целочисленного деления. Возвращает результат деления значения INT1 на INT2.
INT1 rem INT2	Число	Возвращает остаток от деления INT1 на INT2. Например, INT1 – (INT1 div INT2) * INT2.
INT1 mod INT2	Число	Эта функция объявлена устаревшей. Используйте вместо нее функцию rem.
BASE ** POWER	Число	Возвращает результат возведения значения BASE в степень POWER, где каждое из чисел произвольно (за исключением того, что BASE не может быть нулем, если POWER - это ноль любого типа, кроме целого 0). Если POWER - это целое число, вычисление производится последовательным умножением степеней BASE. То есть, если BASE - это целое число, результат также будет целочисленным. Если POWER - это целочисленный 0, результатом всегда будет 1 того же типа, что и BASE. В противном случае, если POWER - не целое число, результат вычисляется как $\exp(\text{POWER} * \log(\text{BASE}))$ .
abs(NUM)	Число	Возвращает абсолютное значение NUM, которое всегда того же типа, что и само число.
exp(NUM)	Действительное число	Возвращает значение $e$ в степени NUM, где $e$ - основание натурального логарифма.
fracof(NUM)	Действительное число	Возвращает дробную часть числа NUM, определяемую как NUM – intof(NUM).
intof(NUM)	Целое	Усекает аргумент до целочисленного значения. Возвращает целое число того же знака, что и NUM, с максимальным значением модуля, такого что $\text{abs}(\text{INT}) \leq \text{abs}(\text{NUM})$ .
log(NUM)	Действительное число	Возвращает значение натурального логарифма числа NUM (по основанию $e$ ), которое не должно быть нулем ни для какого типа.
log10(NUM)	Действительное число	Возвращает значение десятичного логарифма числа NUM, которое не должно быть нулем ни для какого типа. Эта функция определяется как $\log(\text{NUM}) / \log(10)$ .
negate(NUM)	Число	Используется для изменения знака NUM. Возвращает соответствующее число с противоположным знаком.
round(NUM)	Целое	Используется для округления NUM до целого значения с помощью функции intof(NUM+0,5), если NUM положительное, или intof(NUM-0,5), если NUM отрицательное.
sign(NUM)	Число	Используется для определения знака числа NUM. Этот оператор возвращает –1, 0 или 1, если NUM - целое число. Если NUM - действительное число, возвращается –1,0, 0,0 или 1,0, в зависимости от числа NUM - отрицательное оно, ноль или положительное.
sqrt(NUM)	Действительное число	Возвращает квадратный корень из NUM. NUM должно быть положительным числом.
sum_n(LIST)	Число	Возвращает сумму значений из списка числовых полей или значение null, если значения всех полей в списке - это null.
mean_n(LIST)	Число	Возвращает среднее значение из списка числовых полей или null, если во всех полях значения null.

Таблица 20. Числовые функции CLEM (продолжение).

Функция	Результат	Описание
sdev_n(LIST)	Число	Возвращает значение среднеквадратичного отклонения из списка числовых полей или null, если во всех полях значения null.

## Тригонометрические функции

У всех функций этого раздела аргумент или возвращаемое значение - это угол. В обоих случаях единицы измерения угла (радианы или градусы) управляются параметром соответствующей опции потока.

Таблица 21. Тригонометрические функции CLEM

Функция	Результат	Описание
arccos(NUM)	Действительное число	Вычисляет арккосинус заданного числа.
arccosh(NUM)	Действительное число	Вычисляет гиперболический арккосинус заданного числа.
arcsin(NUM)	Действительное число	Вычисляет арксинус заданного числа.
arcsinh(NUM)	Действительное число	Вычисляет гиперболический арксинус заданного числа.
arctan(NUM)	Действительное число	Вычисляет арктангенс заданного числа.
arctan2(NUM_Y, NUM_X)	Действительное число	Вычисляет арктангенс NUM_Y / NUM_X и использует знаки двух чисел для определения информации о квадранте. Результат - это действительное значение в диапазоне $-\pi < \text{ANGLE} \leq \pi$ (радиан), $180 < \text{ANGLE} \leq 180$ (градусов)
arctanh(NUM)	Действительное число	Вычисляет гиперболический арктангенс заданного числа.
cos(NUM)	Действительное число	Вычисляет косинус заданного угла.
cosh(NUM)	Действительное число	Вычисляет гиперболический косинус заданного угла.
pi	Действительное число	Эта константа - наилучшее приближение числа пи (действительное число).
sin(NUM)	Действительное число	Вычисляет синус заданного угла.
sinh(NUM)	Действительное число	Вычисляет гиперболический синус заданного угла.
tan(NUM)	Действительное число	Вычисляет тангенс заданного угла.
tanh(NUM)	Действительное число	Вычисляет гиперболический тангенс заданного угла.

## Вероятностные функции

Вероятностные функции возвращают вероятности на основании различных распределений, например, вероятность того, что значение из  $t$ -распределения Стьюдента будет меньше конкретного значения.

Таблица 22. Вероятностные функции CLEM

Функция	Результат	Описание
<code>cdf_chisq(NUM, DF)</code>	Действительное число	Возвращает вероятность того, что значение из распределения хи-квадрат с заданным числом степеней свободы будет меньше указанного числа.
<code>cdf_f(NUM, DF1, DF2)</code>	Действительное число	Возвращает вероятность того, что значение из <i>F</i> -распределения с числом степеней свободы <i>DF1</i> и <i>DF2</i> будет меньше указанного числа.
<code>cdf_normal(NUM, MEAN, STDDEV)</code>	Действительное число	Возвращает вероятность того, что значение из нормального распределения с заданным средним значением и среднеквадратичным отклонением будет меньше указанного числа.
<code>cdf_t(NUM, DF)</code>	Действительное число	Возвращает вероятность того, что значение из <i>t</i> -распределения Стьюдента с заданным числом степеней свободы будет меньше указанного числа.

## Побитовые целочисленные операции

Эти функции позволяют работать с целыми числами как с битовыми структурами, представляющими дополнительные значения в двоичной системе исчисления, где у позиции бита *N* вес  $2^{**N}$ . Биты нумеруются от нуля и выше. При этих операциях бит знака целого числа отнесен на неопределенную позицию слева. Таким образом, над каждым старшим битом у положительного целого числа будет находиться 0 бит, а у отрицательного - 1 бит.

Таблица 23. Побитовые целочисленные операции CLEM.

Функция	Результат	Описание
<code>~ INT1</code>	Целое	Создает побитовое дополнение целого числа <i>INT1</i> . То есть, в результате 1 находится на каждой битовой позиции, в которой у числа <i>INT1</i> был 0. Всегда верно равенство <code>~ INT = -(INT + 1)</code> .
<code>INT1    INT2</code>	Целое	Результат этой операции - побитовое "ИЛИ включительное" для двух целых чисел, <i>INT1</i> и <i>INT2</i> . То есть, в результате будет 1 на каждой битовой позиции, для которой 1 есть или в числе <i>INT1</i> , или в <i>INT2</i> , или в обоих числах.
<code>INT1   /&amp; INT2</code>	Целое	Результат этой операции - побитовое "ИЛИ исключительное" для двух целых чисел, <i>INT1</i> и <i>INT2</i> . То есть, в результате будет 1 на каждой битовой позиции, для которой 1 есть или в числе <i>INT1</i> , или в <i>INT2</i> , но не в обоих числах одновременно.
<code>INT1 &amp;&amp; INT2</code>	Целое	Результат - это побитовое "И" целых чисел <i>INT1</i> и <i>INT2</i> . То есть, в результате будет 1 на каждой битовой позиции, для которой 1 есть и в числе <i>INT1</i> , и в <i>INT2</i> .
<code>INT1 &amp;&amp;~ INT2</code>	Целое	Результат - это побитовое "И" числа <i>INT1</i> и побитового дополнения числа <i>INT2</i> . То есть, в результате будет 1 на каждой битовой позиции, для которой в числе <i>INT1</i> есть 1, а в числе <i>INT2</i> - 0. Это то же самое, что и <code>INT1 &amp;&amp; (~INT2)</code> , и может быть полезно для очистки битов набора <i>INT1</i> в <i>INT2</i> .
<code>INT &lt;&lt; N</code>	Целое	Создает битовую структуру числа <i>INT1</i> , сдвинутую влево на <i>N</i> позиций. При отрицательном значении <i>N</i> возникает сдвиг вправо.
<code>INT &gt;&gt; N</code>	Целое	Создает битовую структуру числа <i>INT1</i> , сдвинутую вправо на <i>N</i> позиций. При отрицательном значении <i>N</i> возникает сдвиг влево.

Таблица 23. Побитовые целочисленные операции CLEM (продолжение).

Функция	Результат	Описание
<code>INT1 &amp;&amp;=_0 INT2</code>	Логический	Эквивалентно логическому выражению <code>INT1 &amp;&amp; INT2 != 0</code> , но более эффективно.
<code>INT1 &amp;&amp;/_0 INT2</code>	Логический	Эквивалентно логическому выражению <code>INT1 &amp;&amp; INT2 == 0</code> , но более эффективно.
<code>integer_bitcount(INT)</code>	Целое	Считает количество битов 1 или 0 в представлении двоичного дополнения числа <code>INT</code> . Если <code>INT</code> не отрицательное, <code>N</code> - это количество бит 1. Если <code>INT</code> отрицательное, это количество бит 0. Из-за знакового расширения существует бесконечное число бит 0 в неотрицательном числе или бит 1 в отрицательном числе. Всегда выполнено равенство <code>integer_bitcount(INT) = integer_bitcount(-(INT+1))</code> .
<code>integer_leastbit(INT)</code>	Целое	Возвращает положение <code>N</code> младшего бита, заданного для целого числа <code>INT</code> . <code>N</code> - эта наибольшая степень двойки, на которую нацело делится <code>INT</code> .
<code>integer_length(INT)</code>	Целое	Возвращает длину в битах числа <code>INT</code> как целого в представлении двоичного дополнения. То есть, <code>N</code> - это наименьшее целое, такое что <code>INT &lt; (1 &lt;&lt; N)</code> если <code>INT &gt;= 0</code> <code>INT &gt;= (-1 &lt;&lt; N)</code> если <code>INT &lt; 0</code> . Если <code>INT</code> неотрицательное, для представления <code>INT</code> как целого числа без знака требуется по меньшей мере <code>N</code> бит. Как вариант, для представления <code>INT</code> как целого со знаком требуется как минимум <code>N+1</code> бит независимо от его знака.
<code>testbit(INT, N)</code>	Логический	Проверяет бит в положении <code>N</code> для целого числа <code>INT</code> и возвращает состояние бита <code>N</code> как логическое значение, равное <code>true</code> для 1 и <code>false</code> для 0.

## Случайные функции

Следующие функции используются для случайного выбора элементов или для генератора псевдослучайных чисел.

Таблица 24. Случайные функции CLEM.

Функция	Результат	Описание
<code>oneof(LIST)</code>	Любое	Возвращает случайно выбранный элемент <code>LIST</code> . Элементы списка должны быть введены как <code>[ITEM1, ITEM2, ..., ITEM_N]</code> . Обратите внимание на то, что можно задавать также список имен полей.
<code>random(NUM)</code>	Число	Возвращает равномерно распределенное случайное число того же типа ( <code>INT</code> или <code>REAL</code> ) от 1 до <code>NUM</code> . Если используются целые числа, возвращаться будут также только целые числа. Если используются действительные (десятичные) числа, возвращаются действительные числа (десятичная точность определяется опциями потока). Максимальное случайное число, возвращаемое этой функцией, может быть равно <code>NUM</code> .
<code>random0(NUM)</code>	Число	У этой функции те же свойства, что у функции <code>random(NUM)</code> , но возвращаемые числа начинаются с нуля. Максимальное случайное число, возвращаемое этой функцией, никогда не будет равно <code>NUM</code> .

## Функции для работы с текстовыми значениями

В CLEM можно выполнять следующие операции со строками:

- Сравнивать строки
- Создавать строки

- Обращаться к символам

В CLEM строка - это любой набор символов между знаками согласования в виде двойных кавычек ("строковые кавычки"). Символом (CHAR) может быть любой одиночный алфавитно-цифровой символ. Они объявляются в выражениях CLEM при помощи обратных одинарных кавычек в форме `*символ*` , например, `z`, `A` или `2`. Не объявленные символы и отрицательные индексы для строки приведут к неопределенному поведению.

*Примечание.* Сравнения между строками, использующими и не использующими SQL pushback, могут генерировать различные результаты, если существуют заключительные пробелы.

Таблица 25. Строковые функции CLEM.

Функция	Результат	Описание
allbutfirst(N, STRING)	Строка	Возвращает строку, совпадающую со строкой <i>STRING</i> с удаленными <i>N</i> первыми символами.
allbutlast(N, STRING)	Строка	Возвращает строку, совпадающую со строкой <i>STRING</i> с удаленными <i>N</i> последними символами.
alphabefore(STRING1, STRING2)	Логический	Используется для проверки алфавитно-цифрового упорядочивания строк. Возвращает значение true, если <i>STRING1</i> следует перед <i>STRING2</i> .
endstring(LENGTH, STRING)	Строка	Извлекает последние <i>N</i> символов из заданной строки. Если длина строки меньше или равна заданному значению, она не изменяется.
hasendstring(STRING, SUBSTRING)	Целое	Это та же функция, что и <i>isendstring</i> (SUBSTRING, STRING).
hasmidstring(STRING, SUBSTRING)	Целое	Это та же функция, что и <i>ismidstring</i> (SUBSTRING, STRING) (встроенная подстрока).
hasstartstring(STRING, SUBSTRING)	Целое	Это та же функция, что и <i>isstartstring</i> (SUBSTRING, STRING).
hassubstring(STRING, N, SUBSTRING)	Целое	Это та же функция, что и <i>issubstring</i> (SUBSTRING, N, STRING), где у <i>N</i> значение по умолчанию 1.
count_substring(STRING, SUBSTRING)	Целое	Возвращает число, сколько раз заданная подстрока встречается в данной строке. Например, <i>count_substring</i> ("foooo.txt", "oo") возвращает 3.
hassubstring(STRING, SUBSTRING)	Целое	Это та же функция, что и <i>issubstring</i> (SUBSTRING, 1, STRING), где у <i>N</i> значение по умолчанию 1.
isalphacode(CHAR)	Логический	Возвращает значение true, если <i>CHAR</i> - это символ в заданной строке (часто - имя поля), у которого код символа - это буква. В противном случае эта функция возвращает значение 0. Например, <i>isalphacode</i> ( <i>produce_num</i> (1)).
isendstring(SUBSTRING, STRING)	Целое	Если строка <i>STRING</i> заканчивается подстрокой <i>SUBSTRING</i> , эта функция возвращает целую подстроку <i>SUBSTRING</i> в <i>STRING</i> . В противном случае эта функция возвращает значение 0.
islowercode(CHAR)	Логический	Возвращает значение true, если <i>CHAR</i> - это символ строчной буквы для заданной строки (часто - имя поля). В противном случае эта функция возвращает значение 0. Например, <i>islowercode</i> (`) и <i>islowercode</i> ( <i>country_name</i> (2)) - это два допустимых выражения.



Таблица 25. Строковые функции CLEM (продолжение).

Функция	Результат	Описание
<code>ismidstring(SUBSTRING, STRING)</code>	Целое	Если <i>SUBSTRING</i> - это подстрока строки <i>STRING</i> , но она не начинается с первого символа <i>STRING</i> или не заканчивается последним, эта функция возвращает индекс, с которого эта подстрока начинается. В противном случае эта функция возвращает значение 0.
<code>isnumbercode(CHAR)</code>	Логический	Возвращает значение true, если <i>CHAR</i> - это символ в заданной строке (часто - имя поля), у которого код символа - это цифра. В противном случае эта функция возвращает значение 0. Например, <code>isnumbercode(product_id(2))</code> .
<code>isstartstring(SUBSTRING, STRING)</code>	Целое	Если строка <i>STRING</i> начинается с подстроки <i>SUBSTRING</i> , эта функция возвращает индекс 1. В противном случае эта функция возвращает значение 0.
<code>issubstring(SUBSTRING, N, STRING)</code>	Целое	Ищет в строке <i>STRING</i> , начиная с ее <i>N</i> -ого символа, подстроку, совпадающую с <i>SUBSTRING</i> . Если такая подстрока находится, эта функция возвращает целочисленный индекс, с которого начинается совпадающая подстрока. В противном случае эта функция возвращает значение 0. Если значение <i>N</i> не задано, используется значение функции по умолчанию (1).
<code>issubstring(SUBSTRING, STRING)</code>	Целое	Ищет в строке <i>STRING</i> , начиная с ее <i>N</i> -ого символа, подстроку, совпадающую с <i>SUBSTRING</i> . Если такая подстрока находится, эта функция возвращает целочисленный индекс, с которого начинается совпадающая подстрока. В противном случае эта функция возвращает значение 0. Если значение <i>N</i> не задано, используется значение функции по умолчанию (1).
<code>issubstring_count(SUBSTRING, N, STRING):</code>	Целое	Возвращает индекс <i>N</i> -ого появления подстроки <i>SUBSTRING</i> в заданной строке <i>STRING</i> . Если подстрока <i>SUBSTRING</i> встречается в строке меньше <i>N</i> раз, возвращается 0.
<code>issubstring_lim(SUBSTRING, N, STARTLIM, ENDLIM, STRING)</code>	Целое	Это та же функция, что и <code>issubstring</code> , но совпадения ограничены началом на подстроке <i>STARTLIM</i> или до нее и окончанием на подстроке <i>ENDLIM</i> или до нее. Ограничения <i>STARTLIM</i> или <i>ENDLIM</i> можно отключить, использовав значение false для любого из аргументов, например, <code>issubstring_lim(SUBSTRING, N, false, false, STRING)</code> - это то же самое, что и <code>issubstring</code> .
<code>isuppercode(CHAR)</code>	Логический	Возвращает значение true, если <i>CHAR</i> - это буква верхнего регистра. В противном случае эта функция возвращает значение 0. Например, <code>isuppercode(`)`)</code> и <code>isuppercode(country_name(2))</code> - это два допустимых выражения.
<code>last(CHAR)</code>	Строка	Возвращает последний символ <i>CHAR</i> строки <i>STRING</i> (которая должна быть длиной по крайней мере в один символ).
<code>length(STRING)</code>	Целое	Возвращает длину строки <i>STRING</i> , то есть, количество символов в ней.

Таблица 25. Строковые функции CLEM (продолжение).

Функция	Результат	Описание
locchar(CHAR, N, STRING)	Целое	Используется для определения положения символов в символических полях. Эта функция ищет в строке <i>STRING</i> символ <i>CHAR</i> , начиная поиск с <i>N</i> -ого символа строки <i>STRING</i> . Эта функция возвращает значение, указывающее на положение (начиная с <i>N</i> ), где найден символ. Если символ не найден, эта функция возвращает значение 0. Если у функции недопустимое смещение ( <i>N</i> ) (например, сдвиг, выходящий за пределы длины строки), функция возвращает значение \$null\$.  Например, функция locchar(`n`, 2, web_page) ищет в поле с названием <i>web_page</i> символ `n`, начиная со второго символа в значении поля. <i>Примечание:</i> Не забудьте использовать одинарные обратные кавычки для выделения заданного символа.
locchar_back(CHAR, N, STRING)	Целое	Аналогично функции locchar, за тем исключением, что поиск выполняется в обратном направлении, начиная с <i>N</i> -го символа. Например, функция locchar_back(`n`, 9, web_page) ищет в поле <i>web_page</i> , начиная с девятого символа и двигаясь назад к началу строки. Если у функции недопустимое смещение (например, сдвиг, выходящий за пределы длины строки), функция возвращает значение \$null\$. Лучше всего использовать функцию locchar_back вместе с функцией length(<поле>), чтобы динамически использовать длину текущего значения поля. Например, locchar_back(`n`, (length(web_page)), web_page).
lowertoupper(CHAR) lowertoupper (STRING)	CHAR или строка	Входные данные могут быть или строкой, или символом, которые используются в этой функции, чтобы вернуть новый элемент того же типа, в котором все символы нижнего регистра преобразованы в верхний регистр. Например, lowertoupper(`a`), lowertoupper("My string") и lowertoupper(field_name(2)) - это допустимые выражения.
matches	Логический	Возвращает значение true, если строка соответствует заданному шаблону. Шаблон должен быть строковым литералом; он не может быть именем поля, содержащего шаблон. Знак вопроса (?) можно включить в шаблон для совпадения ровно одного символа; звездочка (*) указывает на совпадение нулевого или любого числа символов. Чтобы использовать для поиска совпадений знак вопроса или звездочку в литерале (вместо применения их в качестве подстановочных знаков), можно использовать символ обратной дробной черты (\) как знак перехода.
replace(SUBSTRING, NEWSUBSTRING, STRING)	Строка	В заданной строке <i>STRING</i> заменяет все экземпляры подстроки <i>SUBSTRING</i> на <i>NEWSUBSTRING</i> .

Таблица 25. Строковые функции CLEM (продолжение).

Функция	Результат	Описание
replicate(COUNT, STRING)	Строка	Возвращает строку, состоящую из скопированной заданное число раз исходной строки.
stripchar(CHAR, STRING)	Строка	Позволяет удалить заданные символы из строки или поля. Например, эту функцию можно использовать для удаления из данных дополнительных символов, таких как обозначения валют, для получения более простого числа или названия. Например, при использовании синтаксиса stripchar(``, 'Cost') возвращается новое поле с удаленным из всех значений знаком доллара. <i>Примечание:</i> Не забудьте использовать одинарные обратные кавычки для выделения заданного символа.
skipchar(CHAR, N, STRING)	Целое	Ищет в строке <i>STRING</i> любой символ, кроме <i>CHAR</i> , начиная с <i>N</i> -го символа. Эта функция возвращает целочисленную подстроку, обозначающую точку, где такое значение найдено, или 0, если все символы после <i>N</i> -ого - это <i>CHAR</i> . Если у функции недопустимое смещение (например, сдвиг, выходящий за пределы длины строки), функция возвращает значение \$null\$. Функция locchar часто используется совместно с функциями skipchar для определения значения <i>N</i> (точки, с которой начинается поиск в строке). Например, skipchar(``, (locchar(``, 1, "MyString")), "MyString").
skipchar_back(CHAR, N, STRING)	Целое	Аналогично функции skipchar, за тем исключением, что поиск выполняется в <b>обратном</b> направлении, начиная с <i>N</i> -го символа.
startstring(LENGTH, STRING)	Строка	Извлекает первые <i>N</i> символов из заданной строки. Если длина строки меньше или равна заданному значению, она не изменяется.
strmember(CHAR, STRING)	Целое	Эквивалентно функции locchar(CHAR, 1, STRING). Эта функция возвращает целочисленную подстроку, обозначающую точку, в которой впервые встречается <i>CHAR</i> , или 0. Если у функции недопустимое смещение (например, сдвиг, выходящий за пределы длины строки), функция возвращает значение \$null\$.
subscrs(N, STRING)	CHAR	Возвращает <i>N</i> -ый символ <i>CHAR</i> из входной строки <i>STRING</i> . В краткой форме эту функцию можно записать как STRING(N). Например, lowertoupper("name"(1)) - это допустимое выражение.
substring(N, LEN, STRING)	Строка	Возвращает подстроку <i>SUBSTRING</i> , состоящую из символов <i>LEN</i> строки <i>STRING</i> , начиная с символа с индексом <i>N</i> .
substring_between(N1, N2, STRING)	Строка	Возвращает подстроку строки <i>STRING</i> , которая начинается с индекса <i>N1</i> и заканчивается индексом <i>N2</i> .
trim(STRING)	Строка	Удаляет начальные и конечные пробельные символы из заданной строки.

Таблица 25. Строковые функции CLEM (продолжение).

Функция	Результат	Описание
trim_start(String)	Строка	Удаляет начальные пробельные символы из заданной строки.
trimend(String)	Строка	Удаляет конечные пробельные символы из заданной строки.
unicode_char(NUM)	CHAR	Возвращает символ Unicode со значением NUM.
unicode_value(CHAR)	NUM	Возвращает значение Unicode для CHAR
uppertolower(CHAR) uppertolower (STRING)	CHAR или строка	Входные данные могут быть или строкой, или символом, которые используются в этой функции, чтобы вернуть новый элемент того же типа, в котором все символы верхнего регистра преобразованы в нижний регистр. <i>Примечание:</i> Не забудьте задать строки с двойными кавычками и символы с одинарными обратными кавычками. Простые имена полей должны быть заданы без кавычек.

## Функции SoundEx

SoundEx - это способ найти строки, для которых известно произношение, но неизвестно точное правописание. Этот способ разработан в 1918 году, он ищет слова со сходными звуками на основании фонетических предположений, как произносятся определенные буквы. Например, его можно использовать для поиска имен в базе данных, когда написание и произношение для аналогичных имен может отличаться. Основной алгоритм SoundEx описан во многих источниках; несмотря на известные ограничения (например, начальные буквы ph и f не будут соответствовать друг другу при схожем звучании), он поддерживается в каком-то виде большинством баз данных.

Таблица 26. Функции SoundEx CLEM

Функция	Результат	Описание
soundex(String)	Целое число	Возвращает четырехсимвольный код SoundEx для заданной строки STRING.
soundex_difference(String1, String2)	Целое число	Возвращает целое число от 0 до 4, обозначающее количество символов, которые одинаковы в кодировке SoundEx для двух строк, где 0 обозначает отсутствие сходства, а 4 - сильное сходство или идентичность двух строк.

## Функции даты и времени

CLEM включает в себя семейство функций для работы с полями типа datetime для строковых переменных, представляющих даты и время. Используемые форматы даты и времени свои для каждого потока и задаются в диалоговом окне свойств потока. Функции даты и времени проводят синтаксический анализ строк даты и времени в соответствии с текущим выбранным форматом.

При указании в дате года с использованием только двух цифр (то есть, когда век не задан) IBM SPSS Modeler по умолчанию использует век, указанный в диалоговом окне свойств потока.

Таблица 27. Функции даты и времени CLEM.

Функция	Результат	Описание
@TODAY	Строка	При выборе опции <b>Дни/минуты разворачивания</b> в диалоговом окне свойств потока эта функция возвращает текущую дату как строку в текущем формате даты. Если используется формат даты с двумя цифрами и не выбрана опция <b>Дни/минуты разворачивания</b> , эта функция возвращает \$null\$ для текущего сервера.
to_time(ITEM)	Время	Преобразует тип хранения заданного поля во время.
to_date(ITEM)	Дата	Преобразует тип хранения заданного поля в дату.
to_timestamp(ITEM)	Отметка времени	Преобразует тип хранения заданного поля в отметку времени.
to_datetime(ITEM)	Datetime	Преобразует тип хранения заданного поля в значение даты, времени или отметки времени.
datetime_date(ITEM)	Дата	Возвращает значение даты для <i>числа</i> , <i>строки</i> или <i>отметки времени</i> . Обратите внимание на то, что это единственная функция, позволяющая преобразовать число (секунд) обратно в дату. Если ITEM - это строка, создает дату по результатам синтаксического анализа строки с текущим форматом даты. Формат даты, заданный в диалоговом окне свойств потока, должен быть правильным, тогда эта функция будет успешно выполнена. Если ITEM - это число, оно интерпретируется как число секунд от базовой даты (или эпохи). Доли дня урезаются. Если ITEM - это отметка времени, возвращается часть этой отметки, представляющая собой дату. Если ITEM - это дата, она возвращается без изменений.
date_before(DATE1, DATE2)	Логический	Возвращает значение true, если DATE1 представляет дату или отметку времени, предшествующую DATE2. В противном случае эта функция возвращает значение 0.
date_days_difference(DATE1, DATE2)	Целое	Возвращает в виде целого числа время в количестве дней от даты или отметки времени, представленных DATE1, до даты или отметки времени, представленных DATE2. Если DATE2 предшествует DATE1, эта функция возвращает отрицательное число.
date_in_days(DATE)	Целое	Возвращает в виде целого числа время в количестве дней от базовой даты до даты или отметки времени, представленных DATE. Если DATE предшествует базовой дате, эта функция возвращает отрицательное число. Для правильной работы вычислений необходимо включить в поле допустимую дату. Например, не нужно задавать в качестве даты 29 февраля 2001. Так как 2001 год не високосный, этой даты не существует.
date_in_months(DATE)	Действительное число	Возвращает в виде действительного числа время в количестве месяцев от базовой даты до даты или отметки времени, представленных DATE. Это приблизительное значение, основанное на значении 30,4375 дней в месяце. Если DATE предшествует базовой дате, эта функция возвращает отрицательное число. Для правильной работы вычислений необходимо включить в поле допустимую дату. Например, не нужно задавать в качестве даты 29 февраля 2001. Так как 2001 год не високосный, этой даты не существует.

Таблица 27. Функции даты и времени CLEM (продолжение).

Функция	Результат	Описание
date_in_weeks (DATE)	Действительное число	Возвращает в виде действительного числа время в количестве недель от базовой даты до даты или отметки времени, представленных DATE. Предполагается, что неделя равна 7,0 дней. Если DATE предшествует базовой дате, эта функция возвращает отрицательное число. Для правильной работы вычислений необходимо включить в поле допустимую дату. Например, не нужно задавать в качестве даты 29 февраля 2001. Так как 2001 год не високосный, этой даты не существует.
date_in_years (DATE)	Действительное число	Возвращает в виде действительного числа время в количестве лет от базовой даты до даты или отметки времени, представленных DATE. Это приблизительное значение, основанное на значении 365,25 дней в году. Если DATE предшествует базовой дате, эта функция возвращает отрицательное число. Для правильной работы вычислений необходимо включить в поле допустимую дату. Например, не нужно задавать в качестве даты 29 февраля 2001. Так как 2001 год не високосный, этой даты не существует.
date_months_difference (DATE1, DATE2)	Действительное число	Возвращает в виде действительного числа время в количестве месяцев от даты или отметки времени, представленных DATE1, до даты или отметки времени, представленных DATE2. Это приблизительное значение, основанное на значении 30,4375 дней в месяце. Если DATE2 предшествует DATE1, эта функция возвращает отрицательное число.
datetime_date (YEAR, MONTH, DAY)	Дата	Создает значение даты для данных YEAR, MONTH и DAY. Аргументы должны быть целыми числами.
datetime_day (DATE)	Целое	Возвращает день месяца из данного DATE или отметки времени. Результат - это целое число в диапазоне от 1 до 31.
datetime_day_name (DAY)	Строка	Возвращает название дня недели для данного номера дня недели DAY. Аргумент должен быть целым числом в диапазоне от 1 (воскресенье) до 7 (суббота).
datetime_hour (TIME)	Целое	Возвращает значение часа из значения TIME или отметки времени. Результат - это целое число в диапазоне от 0 до 23.
datetime_in_seconds (TIME)	Действительное число	Возвращает часть секунд, хранимых в TIME.
datetime_in_seconds (DATE), datetime_in_seconds (DATETIME)	Действительное число	Возвращает аккумулярованное число, преобразованное в секунды, от разности между текущим значением DATE или DATETIME и базовой датой (1900-01-01).
datetime_minute (TIME)	Целое	Возвращает значение минут из значения TIME или отметки времени. Результат - это целое число в диапазоне от 0 до 59.
datetime_month (DATE)	Целое	Возвращает значение месяца из значения DATE или отметки времени. Результат - это целое число в диапазоне от 1 до 12.
datetime_month_name (MONTH)	Строка	Возвращает полное название месяца для данного номера месяца MONTH. Аргумент должен быть целым числом в диапазоне от 1 до 12.
datetime_now	Отметка времени	Возвращает текущее время как отметку времени.
datetime_second (TIME)	Целое	Возвращает значение секунд из значения TIME или отметки времени. Результат - это целое число в диапазоне от 0 до 59.
datetime_day_short_name (DAY)	Строка	Возвращает сокращенное название дня недели для данного номера дня недели DAY. Аргумент должен быть целым числом в диапазоне от 1 (воскресенье) до 7 (суббота).

Таблица 27. Функции даты и времени CLEM (продолжение).

Функция	Результат	Описание
datetime_month_short_name (MONTH)	Строка	Возвращает сокращенное название месяца для данного номера месяца <i>MONTH</i> . Аргумент должен быть целым числом в диапазоне от 1 до 12.
datetime_time(HOUR, MINUTE, SECOND)	Время	Возвращает значение времени для заданных значений <i>HOUR</i> , <i>MINUTE</i> и <i>SECOND</i> . Аргументы должны быть целыми числами.
datetime_time(ITEM)	Время	Возвращает значение времени для данного <i>ITEM</i> .
datetime_timestamp(YEAR, MONTH, DAY, HOUR, MINUTE, SECOND)	Отметка времени	Возвращает значение отметки времени для данных значений <i>YEAR</i> , <i>MONTH</i> , <i>DAY</i> , <i>HOUR</i> , <i>MINUTE</i> и <i>SECOND</i> .
datetime_timestamp(DATE, TIME)	Отметка времени	Возвращает значение отметки времени для данных значений <i>DATE</i> и <i>TIME</i> .
datetime_timestamp (NUMBER)	Отметка времени	Возвращает значение отметки времени при данном значении числа секунд.
datetime_weekday(DATE)	Целое	Возвращает день недели из данного <i>DATE</i> или отметки времени.
datetime_year(DATE)	Целое	Возвращает значение года из значения <i>DATE</i> или отметки времени. Результат - это целое число, например, 2002.
date_weeks_difference (DATE1, DATE2)	Действительное число	Возвращает в виде действительного числа время в количестве недель от даты или отметки времени, представленных <i>DATE1</i> , до даты или отметки времени, представленных <i>DATE2</i> . Вычисления основаны на значении 7,0 дней в неделю. Если <i>DATE2</i> предшествует <i>DATE1</i> , эта функция возвращает отрицательное число.
date_years_difference (DATE1, DATE2)	Действительное число	Возвращает в виде действительного числа время в количестве лет от даты или отметки времени, представленных <i>DATE1</i> , до даты или отметки времени, представленных <i>DATE2</i> . Это приблизительное значение, основанное на значении 365,25 дней в году. Если <i>DATE2</i> предшествует <i>DATE1</i> , эта функция возвращает отрицательное число.
time_before(TIME1, TIME2)	Логический	Возвращает значение true, если <i>TIME1</i> представляет время или отметку времени, предшествующие <i>TIME2</i> . В противном случае эта функция возвращает значение 0.
time_hours_difference (TIME1, TIME2)	Действительное число	Возвращает разность времени в часах (действительное число) между значениями времени или отметками времени, представленными <i>TIME1</i> и <i>TIME2</i> . При выборе опции <b>Дни/минуты разворачивания</b> в диалоговом окне свойств потока большее значение <i>TIME1</i> берется для ссылки на предыдущий день. Если опция разворачивания не выбрана, большее значение <i>TIME1</i> определяет отрицательное значение возвращаемой величины.
time_in_hours(TIME)	Действительное число	Возвращает значение времени в часах (действительное число), представленное <i>TIME</i> . Например, при формате времени ЧЧММ выражение <code>time_in_hours('0130')</code> дает результат 1,5. <i>TIME</i> может представлять время или отметку времени.
time_in_mins(TIME)	Действительное число	Возвращает значение времени в минутах (действительное число), представленное <i>TIME</i> . <i>TIME</i> может представлять время или отметку времени.
time_in_secs(TIME)	Целое	Возвращает значение времени в секундах (целое число), представленное <i>TIME</i> . <i>TIME</i> может представлять время или отметку времени.

Таблица 27. Функции даты и времени CLEM (продолжение).

Функция	Результат	Описание
time_mins_difference(TIME1, TIME2)	Действительное число	Возвращает разность времени в минутах (действительное число) между значениями времени или отметками времени, представленными <i>TIME1</i> и <i>TIME2</i> . При выборе опции <b>Дни/минуты разворачивания</b> в диалоговом окне свойств потока большее значение <i>TIME1</i> берется для ссылки на предыдущий день (или на предыдущий час, если в текущем формате заданы только минуты и секунды). Если опция разворачивания не выбрана, большее значение <i>TIME1</i> определяет отрицательное значение возвращаемой величины.
time_secs_difference (TIME1, TIME2)	Целое	Возвращает разность времени в секундах (целое число) между значениями времени или отметками времени, представленными <i>TIME1</i> и <i>TIME2</i> . При выборе опции <b>Дни/минуты разворачивания</b> в диалоговом окне свойств потока большее значение <i>TIME1</i> берется для ссылки на предыдущий день (или на предыдущий час, если в текущем формате заданы только минуты и секунды). Если опция разворачивания не выбрана, большее значение <i>TIME1</i> определяет отрицательное значение возвращаемой величины.

## Преобразование значений даты и времени

Обратите внимание на то, что функции преобразования (и любые другие функции, требующие конкретного типа входных данных, таких как значение даты и времени) зависят от конкретного формата, заданного в диалоговом окне Опции потока. Например, если есть поле с именем *DATE*, сохраненное как строка со значениями *Янв 2003*, *Фев 2003* и так далее, можно преобразовать его в тип хранения даты следующим образом:

```
to_date(DATE)
```

Чтобы это преобразование сработало, выберите соответствующий формат даты **МЕС ГГГГ** как формат даты по умолчанию для потока.

Пример преобразования значений строки в даты с использование узла Заполнитель смотрите в потоке *broadband\_create\_models.str*, установленном в папке *\Demos* в подпапке *streams*.

**Даты, хранящиеся как числа.** Обратите внимание на то, что *DATE* в предыдущем примере - это имя поля, а *to\_date* - это функция CLEM. Если у вас есть даты, хранимые как числа, можно преобразовать их при помощи функции *datetime\_date*, где число интерпретируется как число секунд от базовой даты (или эпохи).  
*datetime\_date(DATE)*

Преобразуя дату в число секунд (и обратно), можно выполнить такие вычисления, как определение текущей даты плюс/минус фиксированное число дней, например:

```
datetime_date((date_in_days(DATE)-7)*60*60*24)
```

## Функции последовательности

Для некоторых операций важна последовательность событий. Эта прикладная программа позволяет вам работать со следующими последовательностями записей:

- Последовательности и временные ряды
- Функции последовательности
- Индексация записей
- Значения усреднения, суммирования и сравнения
- Мониторинг изменений - вычисление разностей
- @SINCE



- Значения сдвига
- Дополнительные возможности последовательностей

Для многих прикладных программ каждую запись, переданную в потоке, можно рассматривать как индивидуальное наблюдение, независимое от всех остальных. В этих случаях порядок записей обычно не важен.

Однако для некоторых классов задач последовательность записей очень важна. Обычно это ситуации, связанные с использованием временных рядов, в которых последовательность записей представляет собой упорядоченную последовательность событий или наблюдений. Каждая запись представляет собой снимок в конкретный момент времени; однако наиболее богатая информация может содержаться не в мгновенных значениях, а в том, как они изменяются и ведут себя во времени.

Конечно, соответствующим параметром может быть не время, а что-то иное. Например, записи могут представлять собой результаты анализа, выполняемого на некоторых расстояниях по прямой, но применяются те же принципы обработки.

Функции последовательности и специальные функции сразу распознаются по следующим характеристикам:

- Все они сопровождаются префиксом @.
- Имена этих функций пишутся заглавными буквами.

Функции последовательности могут относиться к записи, обрабатываемой в настоящее время на узле, к записям, уже прошедшим через узел и даже, в одном случае, к записям, которая еще только должны пройти через узел. Функции последовательности могут свободно использоваться с другими компонентами выражений CLEM, хотя есть некоторые ограничения на их аргументы.

### Примеры

Может оказаться полезной информация о том, сколько времени прошло после какого-то события или после выполнения некоторого условия. Для этого используйте функцию @SINCE, например:

```
@SINCE(Income > Outgoings)
```

Эта функция возвращает сдвиг последней записи, где у этого условия было значение true, то есть, количество записей до той записи, когда это условие выполнялось. Если это условие никогда не выполнялось, функция @SINCE возвращает @INDEX + 1.

Иногда может потребоваться сослаться на значение текущей записи в выражении, использованное функцией @SINCE. Это можно сделать при помощи функции @THIS, которая задает, что имя поля всегда применяется к текущей записи. Чтобы найти сдвиг последней записи, у которой значение в поле Concentration более чем вдвое превосходит это значение в текущей записи, можно использовать следующее выражение:

```
@SINCE(Concentration > 2 * @THIS(Concentration))
```

В некоторых случаях используемое в функции @SINCE условие выполняется для текущей записи по определению, например:

```
@SINCE(ID == @THIS(ID))
```

Поэтому функция @SINCE не обрабатывает данное условие для текущей записи. Используйте аналогичную функцию @SINCE0, если вы хотите обработать условие для текущей записи, а также для предыдущих; если значение этого условия равно true в текущей записи, @SINCE0 возвращает 0.

Таблица 28. Функции последовательности CLEM.

Функция	Результат	Описание
MEAN(FIELD)	Действительное число	Возвращает среднее значение для заданных FIELD или FIELDS.

Таблица 28. Функции последовательности CLEM (продолжение).

Функция	Результат	Описание
@MEAN(FIELD, EXPR)	Действительное число	Возвращает среднее значение в поле <i>FIELD</i> по последним записям <i>EXPR</i> , полученным текущим узлом, включая текущую запись. <i>FIELD</i> должно быть именем числового поля. <i>EXPR</i> может быть любым выражением, дающим при вычислении целое число, большее 0. Если <i>EXPR</i> пропущено или превосходит число полученных до настоящего момента записей, будет возвращаться среднее по всем полученным до сих пор записям.
@MEAN(FIELD, EXPR, INT)	Действительное число	Возвращает среднее значение в поле <i>FIELD</i> по последним записям <i>EXPR</i> , полученным текущим узлом, включая текущую запись. <i>FIELD</i> должно быть именем числового поля. <i>EXPR</i> может быть любым выражением, дающим при вычислении целое число, большее 0. Если <i>EXPR</i> пропущено или превосходит число полученных до настоящего момента записей, будет возвращаться среднее по всем полученным до сих пор записям. <i>INT</i> задает максимальное количество предшествующих значений, которые просматриваются. Это гораздо более эффективно, чем просто использование двух аргументов.
@DIFF1(FIELD)	Действительное число	Возвращает дифференциал первого порядка для <i>FIELD1</i> . Таким образом, функция с одним аргументом просто возвращает разность между текущим и предыдущим значением в поле. Возвращает 0, если соответствующие предыдущие записи не существуют.
@DIFF1(FIELD1, FIELD2)	Действительное число	Функция с двумя аргументами возвращает первый дифференциал <i>FIELD1</i> по отношению к <i>FIELD2</i> . Возвращает 0, если соответствующие предыдущие записи не существуют.
@DIFF2(FIELD)	Действительное число	Возвращает второй дифференциал <i>FIELD1</i> . Таким образом, функция с одним аргументом просто возвращает разность между текущим и предыдущим значением в поле. Возвращает 0, если соответствующие предыдущие записи не существуют.
@DIFF2(FIELD1, FIELD2)	Действительное число	Функция с двумя аргументами возвращает первый дифференциал <i>FIELD1</i> по отношению к <i>FIELD2</i> . Возвращает 0, если соответствующие предыдущие записи не существуют.
@INDEX	Целое	Возвращает индекс текущей записи. Индексы выделяются записям при их поступлении на текущий узел. Первой записи дается индекс 1, а для каждой последующей записи значение индекса увеличивается на 1.
@LAST_NON_BLANK(FIELD)	Любое	Возвращает последнее непустое значение для <i>FIELD</i> , как это определено на вышележащем узле Тип или узле источника. Если в прочитанных до сих пор записях для <i>FIELD</i> нет непустых значений, возвращается \$null\$. Обратите внимание на то, что пустые значения, которые называют также пользовательскими значениями отсутствия, могут определяться отдельно для каждого поля.
@MAX(FIELD)	Число	Возвращает максимальное значение для заданного поля <i>FIELD</i> .
@MAX(FIELD, EXPR)	Число	Возвращает максимальное значение в поле <i>FIELD</i> по последним записям <i>EXPR</i> , полученным текущим узлом, включая текущую запись. <i>FIELD</i> должно быть именем числового поля. <i>EXPR</i> может быть любым выражением, результат вычисления которого - целое число больше нуля.

Таблица 28. Функции последовательности CLEM (продолжение).

Функция	Результат	Описание
@MAX(FIELD, EXPR, INT)	Число	Возвращает максимальное значение в поле <i>FIELD</i> по последним записям <i>EXPR</i> , полученным текущим узлом, включая текущую запись. <i>FIELD</i> должно быть именем числового поля. <i>EXPR</i> может быть любым выражением, дающим при вычислении целое число, большее 0. Если <i>EXPR</i> пропущено или превосходит число полученных до настоящего момента записей, будет возвращаться максимальное значение по всем полученным до сих пор записям. <i>INT</i> задает максимальное количество предшествующих значений, которые просматриваются. Это гораздо более эффективно, чем просто использование двух аргументов.
@MIN(FIELD)	Число	Возвращает минимальное значение для заданного поля <i>FIELD</i> .
@MIN(FIELD, EXPR)	Число	Возвращает минимальное значение в поле <i>FIELD</i> по последним записям <i>EXPR</i> , полученным текущим узлом, включая текущую запись. <i>FIELD</i> должно быть именем числового поля. <i>EXPR</i> может быть любым выражением, результат вычисления которого - целое число больше нуля.
@MIN(FIELD, EXPR, INT)	Число	Возвращает минимальное значение в поле <i>FIELD</i> по последним записям <i>EXPR</i> , полученным текущим узлом, включая текущую запись. <i>FIELD</i> должно быть именем числового поля. <i>EXPR</i> может быть любым выражением, дающим при вычислении целое число, большее 0. Если <i>EXPR</i> пропущено или превосходит число полученных до настоящего момента записей, будет возвращаться минимальное значение по всем полученным до сих пор записям. <i>INT</i> задает максимальное количество предшествующих значений, которые просматриваются. Это гораздо более эффективно, чем просто использование двух аргументов.
@OFFSET(FIELD, EXPR)	Любое	Возвращает значение <i>FIELD</i> в записи, сдвинутой от текущей на значение <i>EXPR</i> . Положительное смещение относится к записи, которая уже прошла, в отрицательное задает "предвиденную" запись, которая еще не поступила. Например, функция @OFFSET(Status, 1) возвращает значение поля Status в предыдущей записи, а @OFFSET(Status, -4) "заглядывает вперед" на четыре записи в последовательности (то есть на запись, которая еще не прошла через узел), чтобы получить значение. <i>Обратите внимание на то, что отрицательное смещение нужно задавать в виде константы.</i> Только для положительных смещений, <i>EXPR</i> может быть также произвольным выражением CLEM, которое вычисляется для текущей записи, чтобы получить смещение. В таком случае версия этой функции с тремя аргументами должна повысить производительность (смотрите следующую функцию). Если выражение возвращает что-то отличное от неотрицательного целого числа, это приводит к ошибке, то есть вычислять сдвиги для будущих записей запрещено. <i>Примечание:</i> Автореферентная функция @OFFSET не может использовать предварительный просмотр литералов. Например, на узле Заполнитель нельзя заменить значение поля field1 при помощи такого выражения как @OFFSET(field1, -2).

Таблица 28. Функции последовательности CLEM (продолжение).

Функция	Результат	Описание
@OFFSET(FIELD, EXPR, INT)	Любое	<p>Выполняет ту же операцию, что и функция @OFFSET, с добавлением третьего аргумента, <i>INT</i>, который задает максимальное количества просматриваемых предыдущих значений. В тех случаях, когда смещение вычисляется выражением, этот третий аргумент должен повысить производительность.</p> <p>Например, при использовании выражения @OFFSET(Foo, Month, 12) системе известно, что нужно сохранять только последние 12 значений Foo; в противном случае она должна была хранить все значения. Если значение смещения - это константа (в том числе отрицательная для предварительных смещений, когда смещение должно быть константой), третий аргумент не требуется и будет использоваться версия этой функции с двумя аргументами. Смотрите также примечание об автореферентных функциях в описанной выше версии функции с двумя аргументами.</p>
@SDEV(FIELD)	Действительное число	Возвращает среднее квадратичное отклонение значений для заданных <i>FIELD</i> или <i>FIELDS</i> .
@SDEV(FIELD, EXPR)	Действительное число	Возвращает среднее квадратичное отклонение значений для поля <i>FIELD</i> по последним записям <i>EXPR</i> , полученным текущим узлом, включая текущую запись. <i>FIELD</i> должно быть именем числового поля. <i>EXPR</i> может быть любым выражением, дающим при вычислении целое число, большее 0. Если <i>EXPR</i> пропущено или превосходит число полученных до настоящего момента записей, будет возвращаться среднее квадратичное отклонение по всем полученным до сих пор записям.
@SDEV(FIELD, EXPR, INT)	Действительное число	Возвращает среднее квадратичное отклонение значений для поля <i>FIELD</i> по последним записям <i>EXPR</i> , полученным текущим узлом, включая текущую запись. <i>FIELD</i> должно быть именем числового поля. <i>EXPR</i> может быть любым выражением, дающим при вычислении целое число, большее 0. Если <i>EXPR</i> пропущено или превосходит число полученных до настоящего момента записей, будет возвращаться среднее квадратичное отклонение по всем полученным до сих пор записям. <i>INT</i> задает максимальное количество предшествующих значений, которые просматриваются. Это гораздо более эффективно, чем просто использование двух аргументов.
@SINCE(EXPR)	Любое	Возвращает количество записей, переданных с момента, когда значением <i>EXPR</i> , произвольного выражения CLEM, было true.
@SINCE(EXPR, INT)	Любое	Добавление второго аргумента <i>INT</i> задает максимальное количество предшествующих записей, которые просматриваются. Если у <i>EXPR</i> никогда не было значения true, <i>INT</i> равен @INDEX+1.
@SINCE0(EXPR)	Любое	Рассматривает в том числе и текущую запись, чем отличается от @SINCE; @SINCE0 возвращает 0, если <i>EXPR</i> равно true для текущей записи.
@SINCE0(EXPR, INT)	Любое	Добавление второго аргумента <i>INT</i> задает максимальное количество предшествующих записей, которые просматриваются.

Таблица 28. Функции последовательности CLEM (продолжение).

Функция	Результат	Описание
@SUM(FIELD)	Число	Возвращает сумму значений для заданных <i>FIELD</i> или <i>FIELDS</i> .
@SUM(FIELD, EXPR)	Число	Возвращает сумму значений в поле <i>FIELD</i> по последним записям <i>EXPR</i> , полученным текущим узлом, включая текущую запись. <i>FIELD</i> должно быть именем числового поля. <i>EXPR</i> может быть любым выражением, дающим при вычислении целое число, большее 0. Если <i>EXPR</i> пропущено или превосходит число полученных до настоящего момента записей, будет возвращаться сумма значений по всем полученным до сих пор записям.
@SUM(FIELD, EXPR, INT)	Число	Возвращает сумму значений в поле <i>FIELD</i> по последним записям <i>EXPR</i> , полученным текущим узлом, включая текущую запись. <i>FIELD</i> должно быть именем числового поля. <i>EXPR</i> может быть любым выражением, дающим при вычислении целое число, большее 0. Если <i>EXPR</i> пропущено или превосходит число полученных до настоящего момента записей, будет возвращаться сумма значений по всем полученным до сих пор записям. <i>INT</i> задает максимальное количество предшествующих значений, которые просматриваются. Это гораздо более эффективно, чем просто использование двух аргументов.
@THIS(FIELD)	Любое	Возвращает значение в поле с именем <i>FIELD</i> в текущей записи. Используется только в выражениях @SINCE.

## Глобальные функции

В основном функции @MEAN, @SUM, @MIN, @MAX и @SDEV работают со всеми прочитанными до сих пор записями, в том числе с текущей. Однако в некоторых случаях полезна возможность выяснения, как значения в текущей записи сравниваются со значениями, представленными во всем наборе данных. Используя узел Задать глобальные значения для генерирования значений по всему набору данных, можно получить доступ к этим значениям в выражении CLEM с использованием глобальных функций.

Например,

@GLOBAL\_MAX(Age)

возвращает самое большое значение поля Age в наборе данных, а выражение  
 $(Value - @GLOBAL\_MEAN(Value)) / @GLOBAL\_SDEV(Value)$

выражает разность между Значением записи и глобальным средним в виде числа среднеквадратичных отклонений. Использовать глобальные значения можно только после того, как они вычислены на узле Задать глобальные значения. Все текущие глобальные значения можно отменить, нажав кнопку **Очистить глобальные значения** на вкладке Глобальное в диалоговом окне свойств потока.

Таблица 29. Глобальные функции CLEM.

Функция	Результат	Описание
@GLOBAL_MAX(FIELD)	Число	Возвращает максимальное значение для <i>FIELD</i> по полному набору данных, ранее сгенерированных на узле задания глобальных значений. <i>FIELD</i> должно быть именем числового поля. Если соответствующее глобальное значение не было задано, возникнет ошибка.

Таблица 29. Глобальные функции CLEM (продолжение).

Функция	Результат	Описание
@GLOBAL_MIN(FIELD)	Число	Возвращает минимальное значение для <i>FIELD</i> по полному набору данных, ранее сгенерированных на узле задания глобальных значений. <i>FIELD</i> должно быть именем числового поля. Если соответствующее глобальное значение не было задано, возникнет ошибка.
@GLOBAL_SDEV(FIELD)	Число	Возвращает среднеквадратичное отклонение значений для <i>FIELD</i> по полному набору данных, ранее сгенерированных на узле задания глобальных значений. <i>FIELD</i> должно быть именем числового поля. Если соответствующее глобальное значение не было задано, возникнет ошибка.
@GLOBAL_MEAN(FIELD)	Число	Возвращает среднее значение для <i>FIELD</i> по полному набору данных, ранее сгенерированных на узле задания глобальных значений. <i>FIELD</i> должно быть именем числового поля. Если соответствующее глобальное значение не было задано, возникнет ошибка.
@GLOBAL_SUM(FIELD)	Число	Возвращает сумму значений для <i>FIELD</i> по полному набору данных, ранее сгенерированных на узле задания глобальных значений. <i>FIELD</i> должно быть именем числового поля. Если соответствующее глобальное значение не было задано, возникнет ошибка.

## Функции для обработки пустых значений и значений Null

При использовании CLEM можно указать, что определенные значения в поле рассматриваются как "пустые" (или отсутствующие) значения. С пустыми значениями работают следующие функции.

Таблица 30. Функции пустых и null-значений CLEM.

Функция	Результат	Описание
@BLANK(FIELD)	Логический	Возвращает значение true для всех записей, значения которых пустые, в соответствии с правилами обработки пустых значений, заданными на вышележащем узле Тип или на исходном узле (вкладка Типы).
@LAST_NON_BLANK(FIELD)	Любой	Возвращает последнее непустое значение для <i>FIELD</i> , как это определено на вышележащем узле Тип или узле источника. Если в прочитанных до сих пор записях для <i>FIELD</i> нет непустых значений, возвращается \$null\$. Обратите внимание на то, что пустые значения, которые называют также пользовательскими значениями отсутствия, могут определяться отдельно для каждого поля.
@NULL(FIELD)	Логический	Возвращает true, если значение поля <i>FIELD</i> - это системное значение отсутствия \$null\$. Возвращает false для всех других значений, в частности, для определенных пользователем пустых значений. Если вы хотите проверить оба варианта, используйте @BLANK(FIELD) и @NULL(FIELD).
undef	Любой	Обычно используется в CLEM для ввода значения \$null\$, например, для заполнения пустых значений значениями null на узле Заполнитель.

Пустые поля могут быть "заполнены" на узле Заполнитель. И на узле заполнитель, и на узле извлечения (только в множественном режиме) специальная функция CLEM @FIELD относится к текущему изучаемому полю или полям.

## Специальные поля

Специальные функции используются для обозначения конкретных исследуемых полей или для создания списка полей в качестве входных. Например, при извлечении нескольких полей сразу необходимо использовать @FIELD для обозначения действия "выполнить это извлечение для выбранных полей". Использование выражение  $\log(@FIELD)$  извлекает новое поле логарифма для каждого выбранного поля.

Таблица 31. Специальные поля CLEM.

Функция	Результат	Описание
@FIELD	Любое	Выполняет действие для всех полей, определенных в контексте выражения.
@TARGET	Любое	Когда выражение CLEM используется в функции пользовательского анализа, @TARGET представляет поле назначения или "правильное значение" для анализируемой пары назначение/предсказано. Эта функция обычно используется на узле Анализ.
@PREDICTED	Любое	Когда выражение CLEM используется в функции пользовательского анализа, @PREDICTED представляет предсказанное значение для анализируемой пары назначение/предсказано. Эта функция обычно используется на узле Анализ.
@PARTITION_FIELD	Любое	Подставляет имя текущего поля раздела.
@TRAINING_PARTITION	Любое	Возвращает значение текущего раздела обучения. Например, для выбора записей обучения с помощью узла Выбор используйте выражение CLEM: @PARTITION_FIELD = @TRAINING_PARTITION. Это обеспечивает, что узел Выбор будет работать всегда, независимо от того, какие значения используются для представления каждого раздела в данных.
@TESTING_PARTITION	Любое	Возвращает значение текущего испытательного раздела.
@VALIDATION_PARTITION	Любое	Возвращает значение текущего проверочного раздела.
@FIELDS_BETWEEN(start, end)	Любое	Возвращает список имен полей между заданными начальным и конечным полями (включительно) на основании естественного (то есть по вставке) порядка полей в данных.

Таблица 31. Специальные поля CLEM (продолжение).

Функция	Результат	Описание
@FIELDS_MATCHING(pattern)	<i>Любое</i>	Возвращает список имен полей, соответствующих заданному шаблону. Знак вопроса (?) можно включить в шаблон для совпадения ровно одного символа; звездочка (*) указывает на совпадение нулевого или любого числа символов. Чтобы использовать для поиска совпадений знак вопроса или звездочку в литерале (вместо применения их в качестве подстановочных знаков), можно использовать символ обратной дробной черты (\) как знак перехода.
@MULTI_RESPONSE_SET	<i>Любое</i>	Возвращает список полей в названном наборе нескольких ответов.



---

## Глава 9. Использование IBM SPSS Modeler совместно с репозиторием

---

### О программе IBM SPSS Collaboration and Deployment Services Repository

IBM SPSS Modeler может использоваться совместно с репозиторием IBM SPSS Collaboration and Deployment Services, давая возможность управлять жизненным циклом моделей исследования данных и соответствующих прогнозирующих объектов и использовать эти объекты в прикладных программах предприятия, инструментах и решениях. Таким образом могут совместно использоваться такие объекты IBM SPSS Modeler, как потоки, узлы, выводы потоков, сценарии, проекта и модели. Объекты записываются в центральный репозиторий, откуда к ним открывается совместный доступ для других прикладных программ и где они отслеживаются при помощи расширенных возможностей версий, метаданных и поиска.

*Примечание:* Для получения доступа к репозиторию IBM SPSS Collaboration and Deployment Services требуется отдельная лицензия. Дополнительную информацию смотрите в разделе <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>

Чтобы использовать IBM SPSS Modeler совместно с репозиторием, нужно установить адаптер на хосте репозитория. Без этого адаптера при попытке доступа к объектам репозитория из некоторых узлов или моделей IBM SPSS Modeler может появиться следующее сообщение:

Возможно, этот репозиторий надо изменить для поддержки нового узла, модели и типов вывода.

Указания по установке адаптера смотрите в руководстве *Установка внедрения IBM SPSS Modeler*, доступном на IBM SPSS Modeler Deployment DVD. Подробности о том, как получить доступ к объектам репозитория IBM SPSS Modeler из IBM SPSS Collaboration and Deployment Services Deployment Manager, есть в *Руководстве по внедрению IBM SPSS Modeler*.

В приведенных ниже разделах содержатся сведения о получении доступа к репозиторию из IBM SPSS Modeler.

#### Расширенная поддержка версий и поиска

Репозиторий предлагает всесторонние возможности по работе с версиями и поиску объектов. Например, пусть вы создали поток и записали его в репозиторий, где к нему возможен совместный доступ исследователей из других отделов. Если в дальнейшем вы изменяете поток в IBM SPSS Modeler, вы можете добавить измененную версию в репозиторий, не перезаписывая предыдущую версию. Все версии остаются доступными, и поиск среди них можно выполнять по имени, метке, используемым полям и другим атрибутам. Например, можно искать все версии модели, на вход которых подается чистая выручка, или все модели, созданные конкретным автором. (В традиционной файловой системе для этого пришлось бы сохранять версии под разными именами файлов, и тогда программы не знали бы связей между версиями.)

#### Единая регистрация

Благодаря единой регистрации пользователи могут соединяться с репозиторием, не вводя каждый раз подробности об имени пользователя и пароле. Существующие подробности о регистрации пользователя в локальной сети достаточны для аутентификации в IBM SPSS Collaboration and Deployment Services. Эта возможность зависит от следующего:

- Использование единой регистрации должно быть сконфигурировано в IBM SPSS Collaboration and Deployment Services.
- Пользователь должен быть зарегистрирован на хосте, совместимом с провайдером.

Дополнительную информацию смотрите в разделе “Соединение с репозиторием” на стр. 117.

---

## Сохранение и внедрение объектов репозитория

Потоки, созданные в IBM SPSS Modeler, можно **сохранять** в точности таком виде в репозитории как файлы с расширением *.str*. Таким образом один и тот же поток доступен разным пользователям на всем предприятии. Дополнительную информацию смотрите в разделе “Запись объектов в репозиторий” на стр. 118.

Кроме того, поток можно **внедрить** в репозиторий. Внедренный поток сохраняется как файл с дополнительными метаданными. Внедренный поток может использоваться всеми возможностями уровня предприятия IBM SPSS Collaboration and Deployment Services, например, автоматизированными оценками и обновлением модели. Например, чтобы учитывать новые доступные данные, модель может автоматически обновляться по регулярному расписанию. Другой вариант - для анализа претендента чемпиона можно внедрить набор потоков, в котором они сравниваются и среди них выявляется наиболее эффективная прогнозная модель.

Есть два способа внедрить поток: как поток (файл с расширением *.str*) или как сценарий (файл с расширением *.scn*). Если внедрять поток как поток, то его сможет использовать прикладная программа тонкого клиента IBM SPSS Modeler Advantage. Дополнительную информацию смотрите в разделе “Открытие потока в IBM SPSS Modeler Advantage” на стр. 133. Если внедрять поток как сценарий, то его смогут использовать прогнозирующие прикладные программы версии 5, предшественник IBM SPSS Modeler Advantage.

Дополнительную информацию смотрите в разделе “Опции внедрения потока” на стр. 128.

Требования к потокам, внедряемым как сценарии

- Чтобы обеспечить согласованный доступ к данным предприятия, потоки, внедренные как сценарии, должны быть доступны через компонент представления предприятия IBM SPSS Collaboration and Deployment Services. Это значит, что в IBM SPSS Modeler должен быть как минимум один исходный узел представления предприятия в каждой ветви потока, предназначенной для оценки и моделирования.
- Чтобы использовать узел представления предприятия, нужно установить IBM SPSS Collaboration and Deployment Services, сконфигурировать и обеспечить доступ к ним с вашей площадки после того, как уже были заданы определения для представления предприятия, представлений прикладных программ и провайдера данных (Data Provider Definition, DPD). Дополнительную информацию получите у своего администратора или найдите на сайте корпорации по адресу <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>.
- DPD задается для конкретного источника данных ODBC. Чтобы использовать DPD из IBM SPSS Modeler, нужно, чтобы источник данных ODBC был определен на хосте сервера IBM SPSS Modeler, имеющего то же имя и соединяющегося с тем же складом данных, что указанный в DPD.
- Кроме того, IBM SPSS Collaboration and Deployment Services Enterprise View Driver должен быть установлен на каждом компьютере, используемом для модификации или запуска потока. В Windows просто установите драйвер на компьютер, где установлены SPSS Modeler или сервер SPSS Modeler, и никакого дополнительного конфигурирования драйвера не потребуется. В UNIX нужно добавить в сценарий запуска ссылку на сценарий *pev.sh*. Обратитесь к администратору для получения дополнительных сведений по установке IBM SPSS Collaboration and Deployment Services Enterprise View Driver.

Другие опции внедрения

Хотя IBM SPSS Collaboration and Deployment Services предлагает самый широкий выбор возможностей для управления содержимым предприятия, доступен также ряд других механизмов внедрения или экспорта потоков, в том числе:

- Экспорт потока и модели для дальнейшего использования совместно с IBM SPSS Modeler Solution Publisher Runtime.

- Экспорт одной или нескольких моделей в PMML, формат на основе XML для кодирования информации о модели. Дополнительную информацию смотрите в разделе “Импорт и экспорт моделей в виде PMML” на стр. 134.

---

## Соединение с репозиторием

1. Чтобы соединиться с репозиторием, в главном меню IBM SPSS Modeler выберите:

**Инструменты > Репозиторий > Опции...**

2. В поле **URL репозитория** введите или выберите путь каталога или URL установки репозитория, к которому вы хотите обратиться. Вы можете подключиться одновременно только к одному репозиторию. Настройка зависит от площадки или установки. Подробности регистрационных данных узнайте у администратора вашей системы.

*Примечание:* Для получения доступа к репозиторию IBM SPSS Collaboration and Deployment Services требуется отдельная лицензия. Дополнительную информацию смотрите в разделе <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>

**Задать параметры авторизации.** Оставьте этот переключатель выключенным, чтобы была доступна возможность **единой регистрации**, которая попытается зарегистрировать вас при помощи информации об имени пользователя и пароле на локальном компьютере. Если единая регистрация невозможна или вы включили этот переключатель, запретив единую регистрацию (например, чтобы зарегистрироваться под учетной записью администратора), откроется новый экран для ввода регистрационных данных.

## Ввод регистрационных данных для репозитория

В зависимости от настройки могут понадобиться следующие поля в диалоговом окне Репозиторий:  
Регистрационные данные:

**ID пользователя и пароль.** Задайте допустимое имя пользователя и пароль для регистрации. При необходимости обращайтесь к вашему администратору для получения более полной информации.

**Провайдер.** Выберите провайдера защиты для аутентификации. В репозитории может быть сконфигурировано использование различных провайдеров защиты; при необходимости обращайтесь к вашему администратору для получения более полной информации.

**Запомнить репозиторий и имя пользователя.** Сохраняет текущие настройки как значения по умолчанию, чтобы не вводить их снова при каждом соединении.

---

## Просмотр содержимого репозитория

Содержимое репозитория можно просматривать так же, как в Проводнике Windows; кроме того, можно просматривать *версии* записанных объектов.

1. Чтобы открыть окно IBM SPSS Collaboration and Deployment Services Repository, в меню SPSS Modeler выберите:

**Инструменты > Репозиторий > Показать...**

1. При необходимости задайте параметры соединения с репозиторием. Дополнительную информацию смотрите в разделе “Соединение с репозиторием”. Порт, пароль и другие подробности соединения узнайте у местного администратора системы.

Вначале окно просмотра содержит дерево иерархии папок. Щелкните по имени папки, чтобы вывести ее содержимое.

Объекты, которые соответствуют текущему выбору или критериям поиска, показаны на правой панели, и в нижней части правой панели показаны подробности о текущей версии. Выводимые атрибуты применяются к последней версии.

---

## Запись объектов в репозиторий

Потоки, узлы, модели, палитры моделей, проекты и объекты вывода можно записать в репозиторий, где они будут доступны другим пользователям и прикладным программам.

*Примечание:* Для получения доступа к репозиторию IBM SPSS Collaboration and Deployment Services требуется отдельная лицензия. Дополнительную информацию смотрите в разделе <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>

Кроме того, вывод потока можно публиковать в репозитории в формате, поддерживающем просмотр другими пользователями через интернет в IBM SPSS Collaboration and Deployment Services Deployment Portal.

## Задание свойств объекта

При записи объекта открывается диалоговое окно Репозиторий: Записать, в котором можно задать ряд свойств объекта. Варианты:

- Выбрать, под каким именем и в какую папку записывается объект в репозитории
- Добавить информацию об объекте, такую как метка версии и другие свойства, по которым поддерживается поиск объектов
- Назначить объекту одну или несколько тем классификации
- Задать для объекта опции защиты

Свойства, которые можно задать, описаны в приведенных ниже разделах.

## Выбор положения для записи объектов

В диалоговом окне Репозиторий: записать введите следующее.

**Сохранить в.** Содержит текущую папку -- положение, куда будет записан объект. Щелкните дважды по имени папки в списке, чтобы сделать эту папку текущей папкой. Нажмите кнопку Переход на один уровень вверх, чтобы перейти в родительскую папку. Нажмите кнопку Новая папка, чтобы создать папку на текущем уровне.

**Имя файла.** Имя, под которым будет записан объект.

**Записать.** Записывает объект в текущее положение.

## Добавление информации о записанных объектах

Все поля на вкладке Информация в диалоговом окне Репозиторий: записать - это необязательные поля.

**Автор.** Имя пользователя, создавшего объект в репозитории. По умолчанию содержит имя пользователя, использованного для соединения с репозиторием, но здесь это имя можно изменить.

**Метка версии.** Выберите метку из списка, чтобы пометить версию объекта, или нажмите кнопку **Добавить**, чтобы создать новую метку. Не используйте в метке символ "[". Убедитесь, что все переключатели выключены, если не хотите поставить метку на эту версию объекта. Здесь можно Дополнительную информацию смотрите в разделе "Просмотр и редактирование свойств объекта" на стр. 126.

**Описание.** Описание объекта. Пользователи могут искать объекты по описанию (смотрите примечание).

**Ключевые слова.** Одно или несколько ключевых слов, связанных с объектом и пригодных для поиска (смотрите примечание).

**Срок действия.** Дата, после которой объект больше не виден обычным пользователям, хотя владельцу, а также администратору репозитория объект виден по-прежнему. Чтобы задать дату истечения срока действия, выберите опцию **Дата** и введите дату или выберите ее при помощи кнопки календаря.

**Записать.** Записывает объект в текущее положение.

*Примечание:* Информация в полях **Описание** и **Ключевые слова** обрабатывается не так, как информация, введенная в SPSS Modeler на вкладке аннотаций объекта. Поиск в репозитории по описанию или ключевым словам не вернет информацию со вкладки аннотаций. Дополнительную информацию смотрите в разделе “Поиск объектов в репозитории” на стр. 123.

## Назначение тем хранимому объекту

Темы - это иерархическая система классификации содержимого, записанного в репозитории. При записи объектов и при их поиске можно выбирать среди доступных тем. Список доступных тем задается пользователями репозитория, обладающими соответствующими привилегиями (подробности смотрите в *Руководстве пользователя менеджера внедрения*).

Чтобы назначить тему объекту, на вкладке Темы диалогового окна Репозиторий: записать:

1. Нажмите кнопку **Добавить**.
2. Щелкните по имени темы в списке доступных тем.
3. Щелкните по **ОК**.

Чтобы удалить назначение темы:

4. Выберите тему в списке назначенных тем.
5. Щелкните **Удалить**.

## Задание опций защиты для хранимых объектов

Вы можете задать или изменить ряд опций защиты для хранимого объекта на вкладке Защита в диалоговом окне Репозиторий: записать. Для одного или нескольких **принципалов** (то есть пользователей или групп пользователей) можно:

- Назначить права доступа к объекту
- Изменить права доступа к объекту
- Лишить прав доступа к объекту

**Принципал.** Имя пользователя или группы пользователей репозитория, у которых есть права доступа для этого объекта.

**Разрешения.** Права доступа этого пользователя или группы к объекту.

**Добавить.** Вы можете добавить одного или несколько пользователей или групп в список тех, кому предоставлены права доступа к этому объекту. Дополнительную информацию смотрите в разделе “Добавление пользователя в список разрешений”.

**Изменить.** Вы можете изменить права доступа выбранного пользователя или группы к этому объекту. По умолчанию предоставляются права доступа для чтения. Эта опция дает возможность предоставить дополнительные права доступа, а именно разрешения Владельца, Записи, Удаления и Изменения.

**Удалить.** Удаляет выбранного пользователя или группу из списка разрешений для этого объекта.

**Добавление пользователя в список разрешений:** Указанные ниже поля доступны, если нажать кнопку **Добавить** на вкладке Защита в диалоговом окне Репозиторий: записать:

**Выбрать провайдера.** Выберите провайдера защиты для аутентификации. В репозитории может быть сконфигурировано использование различных провайдеров защиты; при необходимости обращайтесь к вашему администратору для получения более полной информации.

**Найти.** Введите имя пользователя или группы пользователей репозитория, которых вы хотите добавить, и нажмите кнопку **Поиск**, чтобы это имя появилось в списке пользователей. Чтобы добавить несколько имен пользователя одновременно, оставьте это поле пустым и просто нажмите кнопку **Поиск**, чтобы вывести список всех имен пользователей репозитория.

**Список пользователей.** Выберите одно или несколько имен пользователей из списка и нажмите кнопку **ОК**, чтобы добавить их в список разрешений.

**Изменение прав доступа к объекту:** Указанные ниже поля доступны, если нажать кнопку **Изменить** на вкладке **Защита** в диалоговом окне **Репозиторий: записать**:

**Владелец.** Выберите эту опцию, чтобы предоставить этому пользователю или группе права владельца этого объекта. Владелец имеет полный контроль над объектом, включая права доступа для удаления и изменения.

**Чтение.** По умолчанию тот пользователь или группа, которые не владеют объектом, получают права доступа только для чтения объекта. Включите нужные переключатели, чтобы добавить этому пользователю или группе права доступа для **Записи**, **Удаления** или **Изменения**.

## Запись потоков

Поток можно записать в репозиторий как файл *.str*, и он будет доступен другим пользователям.

*Примечание:* Информацию о том, как внедрить поток и пользоваться преимуществами дополнительных возможностей репозитория смотрите в разделе “Внедрение потоков” на стр. 127.

Чтобы записать текущий поток:

1. Выберите в главном меню:  
**Файл > Записать > Записать как поток...**
2. При необходимости задайте параметры соединения с репозиторием. Дополнительную информацию смотрите в разделе “Соединение с репозиторием” на стр. 117. Порт, пароль и другие подробности соединения узнайте у местного администратора системы.
3. В репозитории: запишите диалоговое окно, выберите папку для записи объекта, задайте любую другую информацию, которую нужно записать, и нажмите кнопку **Записать**. Дополнительную информацию смотрите в разделе “Задание свойств объекта” на стр. 118.

## Запись проектов

В репозиторий можно записать весь проект IBM SPSS Modeler как файл *.cpi*, и он будет доступна другим пользователям.

Поскольку файл проекта - это контейнер для других объектов IBM SPSS Modeler, нужно указать, чтобы IBM SPSS Modeler записал объекты проекта в репозиторий. Для этого служит настройка в диалоговом окне **Свойства проекта**. Дополнительную информацию смотрите в разделе “Задание свойств проекта” на стр. 140.

После того, как вы сконфигурировали запись объектов проекта в репозиторий, при каждом добавлении в проект нового объекта IBM SPSS Modeler автоматически предлагает записать объект.

При завершении сеанса в IBM SPSS Modeler нужно записать файл новой версии проекта, чтобы сохранить добавленные объекты. Файл проекта автоматически выбирает для хранения и получения последние версии объектов. Если вы не добавили новых объектов в проект за время сеанса IBM SPSS Modeler, перезаписывать файл проекта не требуется. Однако нужно записать новые версии объектов проекта (потоков, вывода и так далее), в которые были внесены изменения.

Чтобы записать проект

1. Выберите проект на вкладке **CRISP-DM** или **Классы** на панели менеджеров в IBM SPSS Modeler и выберите в главном меню:

### Файл > Проект > Записать проект...

2. При необходимости задайте параметры соединения с репозиторием. Дополнительную информацию смотрите в разделе “Соединение с репозиторием” на стр. 117. Порт, пароль и другие подробности соединения узнайте у местного администратора системы.
3. В репозитории: запишите диалоговое окно, выберите папку для записи объекта, задайте любую другую информацию, которую нужно записать, и нажмите кнопку **Записать**. Дополнительную информацию смотрите в разделе “Задание свойств объекта” на стр. 118.

## Запись узлов

Узел можно записать в репозиторий как определение отдельного узла из текущего потока в виде файла *.nod*, и он будет доступен другим пользователям.

Чтобы записать узел:

1. Щелкните правой кнопкой по узлу на холсте потока и выберите **Сохранить узел**.
2. При необходимости задайте параметры соединения с репозиторием. Дополнительную информацию смотрите в разделе “Соединение с репозиторием” на стр. 117. Порт, пароль и другие подробности соединения узнайте у местного администратора системы.
3. В репозитории: запишите диалоговое окно, выберите папку для записи объекта, задайте любую другую информацию, которую нужно записать, и нажмите кнопку **Записать**. Дополнительную информацию смотрите в разделе “Задание свойств объекта” на стр. 118.

## Запись объектов вывода

Объект вывода из текущего потока можно записать в репозиторий как файл *.soi*, и он будет доступен другим пользователям.

Чтобы записать объект вывода:

1. Щелкните по объекту на вкладке Вывод на панели менеджеров в SPSS Modeler и выберите в главном меню:  
**Файл > Вывод > Записать вывод...**
2. Другой вариант - щелкнуть правой кнопкой по объекту на вкладке Вывод и выбрать **Записать**.
3. При необходимости задайте параметры соединения с репозиторием. Дополнительную информацию смотрите в разделе “Соединение с репозиторием” на стр. 117. Порт, пароль и другие подробности соединения узнайте у местного администратора системы.
4. В репозитории: запишите диалоговое окно, выберите папку для записи объекта, задайте любую другую информацию, которую нужно записать, и нажмите кнопку **Записать**. Дополнительную информацию смотрите в разделе “Задание свойств объекта” на стр. 118.

## Запись моделей и палитр моделей

Модель можно записать в репозиторий как файл *.gm*, и она будет доступна другим пользователям. Кроме того, в репозиторий можно записать все содержимое палитры моделей, как файл *.gen*.

Запись модели:

1. Щелкните по объекту на палитре моделей в SPSS Modeler и выберите в главном меню:  
**Файл > Модели > Записать модель...**
2. Другой вариант - щелкнуть правой кнопкой по объекту на палитре моделей и выбрать **Записать модель**.
3. Далее выполните действия из раздела “Выполнение процедуры записи”.

Запись палитры моделей:

1. Щелкните правой кнопкой по фоновому участку палитры моделей.
2. Во всплывающем меню выберите **Записать палитру**.

3. Далее выполните действия из раздела "Выполнение процедуры записи".

Выполнение процедуры записи:

1. При необходимости задайте параметры соединения с репозиторием. Дополнительную информацию смотрите в разделе "Соединение с репозиторием" на стр. 117. Порт, пароль и другие подробности соединения узнайте у местного администратора системы.
2. В репозитории: запишите диалоговое окно, выберите папку для записи объекта, задайте любую другую информацию, которую нужно записать, и нажмите кнопку **Записать**. Дополнительную информацию смотрите в разделе "Задание свойств объекта" на стр. 118.

---

## Получение объектов из репозитория

Вы можете получить из репозитория записанные в нем потоки, модели, палитры моделей, узлы, проекты и объекты вывода.

*Примечание:* Помимо пунктов меню, описанных здесь, для получения потоков, объектов вывода, моделей и палитр моделей можно использовать щелчок правой кнопкой в нужной вкладке панели менеджеров в верхней правой части окна SPSS Modeler.

1. Чтобы получить поток, в главном меню IBM SPSS Modeler выберите:  
**Файл > Получить поток...**
2. Чтобы получить модель, палитру моделей, проект или объект вывода, в главном меню IBM SPSS Modeler выберите:  
**Файл > Модели > Получить модель...**  
*или*  
**Файл > Модели > Получить палитру моделей...**  
*или*  
**Файл > Проекты > Получить проект...**  
*или*  
**Файл > Вывод > Получить вывод...**
3. Другой вариант - щелкнуть правой кнопкой на панели менеджеров или панели проекта и во всплывающем меню выбрать **Получить**.
4. Чтобы получить узел, в главном меню IBM SPSS Modeler выберите:  
**Вставить > Узел (или надузел) из репозитория...**
  - а. При необходимости задайте параметры соединения с репозиторием. Дополнительную информацию смотрите в разделе "Соединение с репозиторием" на стр. 117. Порт, пароль и другие подробности соединения узнайте у местного администратора системы.
5. В диалоговом окне Репозиторий: Получить найдите нужный объект, выберите его и нажмите кнопку **Получить**. Дополнительную информацию смотрите в разделе .

## Выбор получаемого объекта

В диалоговом окне Репозиторий: Получить или найти доступны следующие поля:

**Папка.** Содержит иерархию папок для текущей папки. Для перехода к другой папке просто выберите ее в дереве или используйте список объектов, показанный ниже этого поля.

**Кнопка Переход на один уровень вверх.** Переходит от текущей папки на один уровень выше в иерархии.

**Кнопка Создать папку.** Создает новую папку на текущем уровне иерархии.

**Имя файла.** Имя файла выбранного объекта в репозитории. Для получения объекта нажмите кнопку **Получить**.



**Тип файла.** Тип объекта, выбранного для получения. Видны только объекты этого типа и папки. Чтобы показать и получить объекты другого типа, выберите нужный тип в списке.

**Открыть как заблокированный.** По умолчанию при получении объекта он блокируется в репозитории, чтобы его не изменили другие пользователи. Если вы не хотите блокировать объект при получении, выключите этот переключатель.

**Описание, ключевые слова.** Дополнительные подробности об объекте, если они были заданы при записи этого объекта. Дополнительную информацию смотрите в разделе “Добавление информации о записанных объектах” на стр. 118.

**Версия.** Нажмите эту кнопку, чтобы получить версию объекта, отличную от последней. Будет показана подробная информация по всем версиям объекта, чтобы можно было выбрать нужную версию.

## Выбор версии объекта

Чтобы выбрать конкретную версию объекта в репозитории, в диалоговом окне Репозиторий: Выбрать версию:

1. (Необязательно) отсортируйте список по версиям, меткам, размеру, дате создания или создавшему пользователю двойным щелчком по заголовку соответствующего столбца.
2. Выберите нужную версию объекта.
3. Нажмите кнопку Продолжить.

---

## Поиск объектов в репозитории

Объекты можно искать по имени, папке, типу, метке, дате и другим критериям.

Поиск по имени

Чтобы найти объекты по имени:

1. Выберите в главном меню IBM SPSS Modeler:

**Инструменты > Репозиторий > Показать...**

- a. При необходимости задайте параметры соединения с репозиторием. Дополнительную информацию смотрите в разделе “Соединение с репозиторием” на стр. 117. Порт, пароль и другие подробности соединения узнайте у местного администратора системы.
2. Щелкните по вкладке **Поиск**.
  3. В поле **Поиск объектов с именем** введите имя искомого объекта.

При поиске объектов по имени можно использовать звездочку (\*) как символ подстановки, соответствующий любой строке символов, и знак вопроса (?) - как соответствующий одному произвольному символу. Например, *\*кластер\** соответствует всем объектам, в имени которых (в любом месте) содержится строка *кластер*. Строке поиска *\*\_m0?\** соответствует *кластер\_M01.str* и *кластер\_M02.str*, но не *кластер\_M01a.str*. Регистр символов при поиске не учитывается (*кластер* соответствует строке *Кластер* и *КЛАСТЕР*).

*Примечание:* Если объектов много, поиск может занять несколько минут.

Поиск по другим критериям

Поиск можно выполнить с учетом заголовка, метки, дат, автора, ключевых слов, проиндексированного содержимого или описания. Будут найдены только объекты, удовлетворяющие *всем* заданным критериям поиска. Например, можно найти все потоки, содержащие одну или несколько моделей кластеризации, причем еще с конкретной меткой и датой изменения после заданной даты.

**Типы объектов.** Поиск можно ограничить одним типом объектов, таким как модели, потоки, выводы, узлы, надузлы, проекты, палитры моделей, сценарии и другие типы.

- **Модели.** Модели можно искать по категории (классификация, аппроксимация, кластеризация и др.) или по алгоритму моделирования, такому как алгоритм Коонена.

Кроме того, можно учитывать использованные поля - например, искать модели, использующие поле с именем *доход* на входе или на выходе (как поле назначения).

- **Потоки.** Для уточнения поиска потоков можно указать используемые поля или тип модели (ее категорию или алгоритм), содержащейся в потоке.

**Темы.** Можно искать модели, связанные с определенными темами из списка, который задается пользователями репозитория, обладающими соответствующими привилегиями (подробности смотрите в *Руководстве пользователя менеджера внедрения*). Чтобы использовать список, включите этот переключатель, нажмите появившуюся кнопку **Добавить темы**, выберите одну или несколько тем в списке и нажмите кнопку **ОК**.

**Метка.** Уточняет поиск при помощи меток версий объекта.

**Даты.** Можно указать дату создания или изменения и искать только объекты до, после и в пределах диапазона дат.

**Автор.** Ограничивает поиск объектами, которые были созданы заданным пользователем.

**Ключевые слова.** Поиск заданных ключевых слов. В IBM SPSS Modeler ключевые слова задаются на вкладке **Аннотация для потока, модели или объекта вывода**.

**Описание.** Поиск заданных терминов в поле описания. В IBM SPSS Modeler описание задается на вкладке **Аннотация для потока, модели или объекта вывода**. Можно задать несколько искомых фраз через точку с запятой, например, *доход*; тип обрезки; величина иска. (Обратите внимание на то, что пробелы в искомой фразе учитываются. Например, различаются тип обрезки с одним пробелом и тип обрезки с двумя пробелами.)

---

## Изменение объектов репозитория

Существующие в репозитории объекты можно удалять непосредственно из SPSS Modeler. Варианты:

- Создать, переименовать или удалить папки
- Заблокировать или разблокировать объекты
- Удалить объекты

## Создание, переименование и удаление папок

1. Чтобы выполнить операции над папками в репозитории, выберите в главном меню SPSS Modeler:

**Инструменты > Репозиторий > Показать...**

а. При необходимости задайте параметры соединения с репозиторием. Дополнительную информацию смотрите в разделе “Соединение с репозиторием” на стр. 117. Порт, пароль и другие подробности соединения узнайте у местного администратора системы.

2. Убедитесь, что активна вкладка **Папки**.
3. Чтобы создать новую папку, щелкните правой кнопкой по родительской папке и выберите **Новая папка**.
4. Чтобы переименовать папку, щелкните по ней правой кнопкой и выберите **Переименовать папку**.
5. Чтобы удалить папку, щелкните по ней правой кнопкой и выберите **Удалить папку**.

## Блокировка и разблокировка объектов репозитория

Объект можно заблокировать, и другие пользователи не смогут вносить изменения ни в одну из его существующих версий или создавать новые версии. Значок заблокированного объекта помечается

маленьким символом замка.



Рисунок 15. Заблокированный объект

Чтобы заблокировать объект

1. В окне просмотра репозитория щелкните правой кнопкой по нужному объекту.
2. Выберите **Заблокировать**.

Чтобы разблокировать объект

1. В окне просмотра репозитория щелкните правой кнопкой по нужному объекту.
2. Нажмите кнопку **Разблокировать**.

## Удаление объектов репозитория

Прежде чем удалить объект из репозитория, сначала решите, нужно ли удалить все версии объекта или только одну конкретную версию.

Чтобы удалить все версии объекта

1. В окне просмотра репозитория щелкните правой кнопкой по нужному объекту.
2. Выберите **Удалить объекты**.

Чтобы удалить последнюю версию объекта

1. В окне просмотра репозитория щелкните правой кнопкой по нужному объекту.
2. Щелкните **Удалить**.

Чтобы удалить предыдущую версию объекта

1. В окне просмотра репозитория щелкните правой кнопкой по нужному объекту.
2. Выберите **Удалить версии**.
3. Выберите одну или несколько удаляемых версий и нажмите кнопку **ОК**.

---

## Управление свойствами объектов репозитория

Из SPSS Modeler можно управлять различными свойствами объектов. Варианты:

- Просмотреть свойства папки
- Просмотреть и отредактировать свойства объекта
- Создать, поставить и снять метки версий для объекта

### Просмотр свойств папки

Чтобы просмотреть свойства любой папки в окне репозитория, щелкните правой кнопкой по нужной папке. Выберите **Свойства папки**.

Вкладка **Общее**

Эта вкладка содержит имя папки, дату создания и дату изменения.

Вкладка **Разрешения**

На этой вкладке задаются права чтения и записи для папки. Перечислены все пользователи и группы, имеющие доступ к родительской папке. Разрешения образуют иерархию. Например, если у кого-то нет права на чтение, ему нельзя предоставить право на запись. Если у кого-то нет права на запись, ему нельзя предоставить право на удаление.

**Пользователи и группы.** Содержит пользователей и группы репозитория, у которых есть хотя бы право на чтение этой папки. Включите переключатели **Запись** и **Удаление**, чтобы добавить такие права доступа к этой папке отдельному пользователю или группе. Щелкните по значку **Добавить пользователей или группы** у правого края вкладки **Разрешения**, чтобы предоставить доступ дополнительным пользователям и группам. Список доступных пользователей и групп управляется администратором.

**Каскадировать разрешения.** Выберите опцию, чтобы управлять применением изменений, внесенных в свойства текущей папки, к ее дочерним папкам, если такие есть.

- **Каскадировать все разрешения.** Каскадирует параметры разрешений из текущей папки на все дочерние папки и папки-потомки. Это быстрый способ задать разрешения для нескольких папок сразу. Задайте нужные разрешения для родительской папки, а потом каскадируйте их, сколько нужно.
- **Каскадировать только изменения.** Каскадирует только изменения, внесенные после последнего применения изменений. Например, если была добавлена новая группа и нужно предоставить ей доступ ко всем папкам в ветви **Продажи**, предоставьте этой группе доступ к корневой папке **Продажи** и каскадируйте это изменение на все подпапки. Остальные разрешения для существующих подпапок останутся, какие были.
- **Не каскадировать.** Любые изменения в свойствах текущей папки применяются только к ней, не каскадируясь на дочерние папки.

## Просмотр и редактирование свойств объекта

Свойства объекта можно просматривать и редактировать в диалоговом окне **Свойства объекта**. Хотя некоторые свойства изменить нельзя, всегда можно изменить объект, добавив новую версию.

1. В окне репозитория щелкните правой кнопкой по нужному объекту.
2. Выберите **Свойства объекта**.

Вкладка **Общие**

**Имя.** Имя, под которым объект виден в репозитории.

**Дата создания.** Дата создания объекта (не путать с версией).

**Последнее изменение.** Дата последнего изменения версии.

**Автор.** Регистрационное имя пользователя.

**Описание.** По умолчанию содержит то описание, которое задано на вкладке **Аннотация объекта** в **SPSS Modeler**.

**Связанные темы.** Если нужно, модели и связанные с ними объекты можно сгруппировать в репозитории по темам. Список доступных тем задается пользователями репозитория, обладающими соответствующими привилегиями (подробности смотрите в *Руководстве пользователя менеджера внедрения*).

**Ключевые слова.** Ключевые слова задаются на вкладке **Аннотация** для потока, модели или объекта вывода. Ключевые слова должны быть разделены пробелами, максимум допустимо 255 символов. (Если ключевые слова содержат пробелы, заключайте такие ключевые слова в кавычки.)

Вкладка **Версии**

Объекты, записанные в репозиторий, могут иметь по несколько версий. На вкладке **Версии** выводится информация о каждой версии.

Следующие свойства можно задать или изменить для отдельной версии записанного объекта:

**Версия.** Уникальный идентификатор для версии, сгенерированный на основе времени сохранения версии.

**Метка.** Текущая метка для версии (если есть). В отличие от идентификатора версии, метки можно переставлять с одной версии на другую.

Кроме того, для каждой версии выводится размер файла, дата создания и автор.

**Редактировать метки.** Щелкните по значку **Редактировать метки** вверху справа на вкладке Версии, чтобы задать, поставить или снять метки записанных объектов. Дополнительную информацию смотрите в разделе “Работа с метками версий объекта”.

Вкладка Разрешения

На вкладке Разрешения можно задать права чтения и записи объекта. Перечислены все пользователи и группы, имеющие доступ к текущему объекту. Разрешения образуют иерархию. Например, если у кого-то нет права на чтение, ему нельзя предоставить право на запись. Если у кого-то нет права на запись, ему нельзя предоставить право на удаление.

**Пользователи и группы.** Содержит пользователей и группы репозитория, у которых есть хотя бы право на чтение этого объекта. Включите переключатели Запись и Удаление, чтобы добавить такие права доступа к этому объекту отдельному пользователю или группе. Щелкните по значку **Добавить пользователей или группы** у правого края вкладки Разрешения, чтобы предоставить доступ дополнительным пользователям и группам. Список доступных пользователей и групп управляется администратором.

## Работа с метками версий объекта

В диалоговом окне Редактировать метки версий можно:

- Поставить метки на выбранный объект
- Снять метки с выделенного объекта
- Задать новую метку и поставить ее на объект

Чтобы поставить метки на объект

1. Выберите одну или несколько меток в списке **Доступные метки**.
2. Щелкните по кнопке со стрелкой вправо, чтобы перенести выбранные метки в список **Поставленные метки**.
3. Щелкните по **ОК**.

Чтобы снять метки с объекта

1. Выберите одну или несколько меток в списке **Поставленные метки**.
2. Щелкните по кнопке со стрелкой влево, чтобы перенести выбранные метки в список **Доступные метки**.
3. Щелкните по **ОК**.

Чтобы задать новую метку и поставить ее на объект

1. Введите имя метки в поле **Новая метка**.
2. Щелкните по кнопке со стрелкой вправо, чтобы перенести новую метку в список **Поставленные метки**.
3. Щелкните по **ОК**.

---

## Внедрение потоков

Чтобы разрешить использование потока совместно с прикладной программой тонкого клиента IBM SPSS Modeler Advantage, ее надо внедрить в репозитории как поток (файл *.str*).

Как бы ни был внедрен поток, как поток (файл *.str*) или как сценарий (файл *.scn*), объект может использовать все возможности уровня предприятия IBM SPSS Collaboration and Deployment Services. Дополнительную информацию смотрите в разделе “Сохранение и внедрение объектов репозитория” на стр. 116.

Чтобы внедрить текущий поток (метод меню Файл)

1. Выберите в главном меню:  
**Файл > Записать > Внедрить**
2. Выберите тип внедрения и, при необходимости, другие задайте другие опции в этом диалоговом окне.
3. Выберите **Внедрить как поток**, чтобы внедрить поток для использования совместно с IBM SPSS Modeler Advantage или IBM SPSS Collaboration and Deployment Services. Выберите **Внедрить как сценарий**, чтобы внедрить поток как сценарий для использования совместно с IBM SPSS Collaboration and Deployment Services или прогнозирующими прикладными программами версии 5.
4. Нажмите кнопку **Записать**. Для получения дополнительной информации нажмите кнопку **Справка**.
5. Далее выполните действия из раздела “Выполнение процесса внедрения”.

Чтобы внедрить текущий поток (метод меню Инструменты)

1. Выберите в главном меню:  
**Инструменты > Свойства потока > Внедрение**
2. Выберите тип внедрения и, при необходимости, задайте другие опции на вкладке Внедрения; нажмите кнопку **Записать**. Дополнительную информацию смотрите в разделе “Опции внедрения потока”.

Выполнение процесса внедрения

1. При необходимости задайте параметры соединения с репозиторием. Дополнительную информацию смотрите в разделе “Соединение с репозиторием” на стр. 117. Порт, пароль и другие подробности соединения узнайте у местного администратора системы.
2. В репозитории: запишите диалоговое окно, выберите папку для записи объекта, задайте любую другую информацию, которую нужно записать, и нажмите кнопку **Записать**. Дополнительную информацию смотрите в разделе “Задание свойств объекта” на стр. 118.

## Опции внедрения потока

На вкладке Внедрение в диалоговом окне опций потока можно задать опции для внедрения потока. Поток можно внедрить либо как поток, либо как сценарий.

Если поток внедряется как поток, его можно открывать и редактировать в прикладной программе тонкого клиента IBM SPSS Modeler Advantage. Поток записывается в репозиторий как файл с расширением *.str*.

При внедрении потока как сценария поток записывается в репозиторий как файл с расширением *.scn*. Поток сможет использоваться прогнозирующими прикладными программами версии 5.

Как бы ни внедрялся поток, как поток или как сценарий, становятся доступны все преимущества дополнительных возможностей IBM SPSS Collaboration and Deployment Services, такие как множественный доступ, автоматизация оценок, обновление модели и анализ претендента чемпиона.

Кроме того, на вкладке Внедрение доступен предварительный просмотр описания потока, которое создается для потока IBM SPSS Modeler. Дополнительную информацию смотрите в разделе “Описания потока” на стр. 52.

*Примечание:* Для надежного доступа к данным предприятия у потока, внедряемого как сценарий, должен быть доступ к данным источника через IBM SPSS Collaboration and Deployment Services Enterprise View, так что такой поток должен содержать как минимум один исходный узел представления предприятия в каждой ветви потока, предназначенной для оценки и моделирования.

**Тип внедрения.** Выберите способ внедрения потока. Для всех потоков перед внедрением сначала нужно указать узел оценки; дополнительные требования и опции зависят от типа внедрения.

- **<нет>**. Поток не будет внедряться в репозиторий. Никакие опции не доступны, кроме предварительного просмотра описания потока.
- **Только оценка.** Поток внедряется в репозиторий при нажатии кнопки **Записать**. Данные можно оценить по узлу, прописанному в поле **Узел оценки**.
- **Обновление модели.** То же, что для типа внедрения Только оценка, но, кроме того, модель можно обновлять в репозитории, обслуживая объекты, указанные в полях **Узел моделирования** и **Слепок модели**.  
*Примечание:* Автоматическое обновление в IBM SPSS Collaboration and Deployment Services по умолчанию не поддерживается, так что, если вы хотите использовать эту возможность при выполнении потока из репозитория, выберите этот тип внедрения. Дополнительную информацию смотрите в разделе “Обновление модели” на стр. 131.

**Узел оценок.** Выберите узел диаграммы, вывода или экспорта, чтобы указать ветвь для оценки данных. Поток может содержать сколь угодно много допустимых ветвей, моделей и конечных узлов, но для оценок внедряемого потока нужно указать одну и только одну ветвь оценок. Это требование относится к внедрению любого потока.

**Параметры оценок.** Можно задать параметры, которые можно модифицировать при выполнении ветви оценок. Дополнительную информацию смотрите в разделе “Параметры оценки и моделирования” на стр. 130.

**Узел моделирования.** Задает узел моделирования, используемый для обновления модели в репозитории или генерирования ее заново. Нужен узел моделирования того же типа, какой задан как **Слепок модели**.

**Параметры построения модели.** Можно задать параметры, которые можно модифицировать при выполнении узла моделирования. Дополнительную информацию смотрите в разделе “Параметры оценки и моделирования” на стр. 130.

**Слепок модели.** При обновлении модели задает тот слепок модели, который будет обновлен или сгенерирован заново при каждом обновлении потока в репозитории (обычно в ходе планового задания). Модель должна располагаться на ветви оценок. На ветви оценок может быть сколь угодно много моделей, но только одну можно задать для данной цели. Обратите внимание на то, что вначале после создания потока эта модель может быть пустышкой, которая будет наполняться смыслом после обновлений и генерирования заново по мере того, как станут доступны новые данные.

**Внедрить как поток.** Выберите эту опцию, если хотите использовать поток совместно с IBM SPSS Modeler Advantage или IBM SPSS Collaboration and Deployment Services (и посмотрите приведенное ниже примечание).

**Внедрить как сценарий.** Выберите эту опцию, если хотите использовать поток совместно с IBM SPSS Collaboration and Deployment Services или прогнозирующими прикладными программами версии 5 (и посмотрите приведенное ниже примечание).

**Проверить.** Нажмите эту кнопку, чтобы проверить, что поток допустим для внедрения. Для всех потоков перед внедрением сначала нужно указать узел оценки. Поток, внедряемый как сценарий, должен содержать допустимый исходный узел представления предприятия. Если эти условия не соблюдены, выводятся сообщения об ошибках.

**Записать.** Внедряет поток, если он допустим. В противном случае выводится сообщение об ошибке. Нажмите кнопку **Исправить**, чтобы исправить ошибку, и повторите попытку.

**Предварительный просмотр описания потока.** Здесь можно просмотреть содержимое описания потока, созданное IBM SPSS Modeler для данного потока. Дополнительную информацию смотрите в разделе “Описание потока” на стр. 52.

*Примечание:* (Внедрение как потока или как сценария) в одной ветви моделирования может быть несколько узлов представления предприятия. В большинстве таких случаев рекомендуется использовать единое соединение данных для всех узлов представления предприятия в пределах ветви, а для анализа претендента чемпиона это обязательное требование.

- Если поддержка претендента чемпиона не требуется, использовать в пределах одной ветви разные соединения представления предприятия можно, но эти соединения должны отличаться только определениями провайдера данных (data provider definition, DPD).
- Эти ограничения действуют только в пределах данной ветви. У ветвей оценки и построения моделей могут быть сколь угодно разные соединения представления предприятия.

## Параметры оценки и моделирования

При внедрении потока в IBM SPSS Collaboration and Deployment Services можно выбрать параметры, поддерживающие просмотр и редактирование при каждом обновлении или оценке модели. Например, можно задать максимальное и минимальное значение или еще какое-то значение, которое будет изменяться при каждом выполнении задания.

1. Чтобы параметр был видимым и, следовательно, стал доступен для просмотра или редактирования после внедрения потока, выберите нужный параметр в списке в диалоговом окне Оценка параметров.

Список доступных параметров определяется на вкладке Параметры в диалоговом окне свойств потока. Дополнительную информацию смотрите в разделе “Задание параметров потока и сеанса” на стр. 48.

## Ветвь оценки

Если вы внедряете поток, нужно одну из ветвей назначить **ветвью оценки** (то есть ветвью, содержащей узел оценки). Ветвь, назначенная ветвью оценки, выделяется на холсте потока; кроме того, выделяется связь модели со слепком в ветви оценки. Эта визуализация особенно полезна в сложных потоках с многочисленными ветвями, где выделить взглядом ветвь оценки не так легко.

*Примечание:* В качестве ветви оценки может назначаться только ветвь потока.

Если в потоке уже задана ветвь оценки, новое назначение переопределяет старое. Цвет для выделения ветви оценки можно задать в опции Пользовательский цвет. Дополнительную информацию смотрите в разделе “Задание опций вывода” на стр. 148.

Показать или скрыть выделение ветви оценки можно кнопкой Показать или скрыть разметку потока на панели инструментов.



Рисунок 16. Кнопка Показать или скрыть разметку потока на панели инструментов

## Выбор ветви оценки для внедрения

Ветвь оценки можно назначить либо во всплывающем меню конечного узла, либо в меню Инструменты. Если действовать через всплывающее меню, узел оценки автоматически назначается на вкладке Внедрение окна свойств диалога.

Чтобы назначить ветвь в качестве ветви оценки (всплывающее меню)

1. Соедините слепок модели с конечным узлом (это узел обработки или вывода ниже по потоку от слепка).
2. Щелкните правой кнопкой по конечному узлу.
3. В меню выберите **Использовать как ветвь оценки**.

Чтобы назначить ветвь в качестве ветви оценки (меню Инструменты)



1. Соедините слепок модели с конечным узлом (это узел обработки или вывода ниже по потоку от слепка).
2. В главном меню выберите:  
**Инструменты > Свойства потока > Внедрение**
3. В списке **Тип внедрения** выберите **Только оценка** или **Обновление модели**, как нужно. Дополнительную информацию смотрите в разделе “Опции внедрения потока” на стр. 128.
4. Выберите поле **Узел оценки** и выберите конечный узел в списке.
5. Нажмите кнопку **ОК**.

## Обновление модели

Обновление модели - это процесс, в котором существующая в потоке модель строится еще раз с использованием новых данных. Сам поток в репозитории не изменяется. В частности, тип алгоритма и настройки, связанные с потоком, остаются теми же; однако модель заново обучается по новым данным и обновляется, если оказывается, что новая версия работает лучше старой.

Обновление можно задать только для одного слепка модели в потоке, называемого **моделью обновления**. Если выбрать опцию **Обновление модели** на вкладке Внедрение в окне свойств потока (смотрите раздел “Опции внедрения потока” на стр. 128), то выбранный слепок модели становится моделью обновления. Кроме того, назначить модель в качестве модели обновления можно во всплывающем меню слепка модели. Это возможно, только если слепок в этот момент принадлежит ветви оценки.

Если для некоторого слепка выключить статус “модели обновления”, это будет равносильно заданию типа внедрения потока Только оценка; в этом случае соответствующим образом изменится вкладка Внедрение диалогового окна свойств потока. Этот статус можно включать и выключать при помощи опции **Использовать как модель обновления** во всплывающем меню слепка на текущей ветви оценки.

При удалении ссылки модели некоторого слепка на ветви оценки удаляется и статус слепка как “модели обновления”. Удаление связи модели можно откатить при помощи меню Правка или панели инструментов; при этом также восстанавливается статус слепка как “модели обновления”.

## Как выбирается Обновить модель

Наряду с ветвью оценок, в потоке выделяется связь с моделью обновления. Слепок модели, выбранный как модель обновления, а, следовательно, и выделенная связь зависят от числа слепков в потоке.

Одна модель в потоке

Если при выборе ветви потока в качестве ветви оценок с ней соединен лишь один слепок модели, этот слепок становится моделью обновления для потока.

Несколько моделей в потоке

Если к потоку присоединено несколько слепков, модель обновления выбирается следующим образом.

Если один из слепков моделей был определен на вкладке Внедрение в диалоговом окне Свойства потока и находится в потоке, то этот слепок становится моделью обновления.

Если ни один из слепков моделей не был определен на вкладке Внедрение или один был определен, но не находится в ветви оценок, то моделью обновления становится слепок, ближайший к конечному узлу.

Если затем вы снимете выбор со всех связей с моделями, выбранных как связи обновления, останется выделена только ветвь оценки, без связей. Будет задан тип внедрения Только оценка.

*Примечание:* Одну из связей можно перевести в состояние Замена, но только одну. В этом случае, если выбирается ветвь оценок, то в качестве модели обновления выбирается слепок модели со связью обновления и ближайший к конечному узлу.

Ни одной модели в потоке

Если в потоке моделей нет или есть только модели без связей моделей, будет задан тип внедрения Только оценка.

### **Проверка ветви оценок на наличие ошибок**

При назначении ветви оценок она проверяется на наличие ошибок, таких как отсутствие узла представление предприятия в потоке, внедряемом как сценарий.

Если найдена ошибка, ветвь оценки выделяется цветом ошибки ветви оценки и выводится сообщение об ошибке. Вы можете задать цвет ошибки при помощи опции Пользовательский цвет. Дополнительную информацию смотрите в разделе “Задание опций вывода” на стр. 148.

Если найдена ошибка, выполните следующее:

1. Исправьте ошибку согласно содержанию сообщения об ошибке.
2. Сделайте одно из следующего:
  - Щелкните правой кнопкой по конечному узлу и во всплывающем меню выберите **Проверить сценарий**.
  - В главном меню выберите:  
**Инструменты > Свойства потока > Внедрение**  
и выберите **Проверка**.
3. При необходимости повторите процесс, пока не будут устранены все ошибки.

---

## Глава 10. Экспорт во внешние прикладные программы

---

### Об экспорте во внешние прикладные программы

IBM SPSS Modeler предлагает ряд механизмов, чтобы экспортировать весь процесс исследования данных во внешние прикладные программы, чтобы проделанную работу по подготовке данных и построению моделей можно было использовать и вне IBM SPSS Modeler.

В предыдущем разделе показано, как внедрять потоки в репозиторий IBM SPSS Collaboration and Deployment Services, чтобы получить такие возможности, как доступ нескольких пользователей, расписание заданий и другие. Аналогичным образом потоки IBM SPSS Modeler можно также использовать совместно со следующим:

- IBM SPSS Modeler Advantage
- Прикладные программы, способные импортировать и экспортировать файлы в формате PMML

Дополнительную информацию об использовании потоков совместно с IBM SPSS Modeler Advantage смотрите в разделе “Открытие потока в IBM SPSS Modeler Advantage”.

Информацию об экспорте и импорте моделей как файлов PMML, чтобы совместно использовать модели с любыми прикладными программами, которые поддерживают этот формат, смотрите в разделе “Импорт и экспорт моделей в виде PMML” на стр. 134.

---

### Открытие потока в IBM SPSS Modeler Advantage

Потоки IBM SPSS Modeler можно использовать вместе с прикладной программой тонкого клиента IBM SPSS Modeler Advantage. Можно полностью создать прикладную программу тонкого клиента внутри IBM SPSS Modeler Advantage, но можно также за основу рабочего потока прикладной программы взять поток, созданный в IBM SPSS Modeler.

Чтобы открыть поток в IBM SPSS Modeler Advantage:

1. Внедрите поток в репозиторий IBM SPSS Collaboration and Deployment Services, обязательно выбрав вариант **Внедрить как поток**. Дополнительную информацию смотрите в разделе “Внедрение потоков” на стр. 127.
2. Нажмите кнопку Открыть на панели инструментов IBM SPSS Modeler Advantage или выберите в главном меню:

#### Файл > Открыть в IBM SPSS Modeler Advantage

1. При необходимости задайте параметры соединения с репозиторием. Дополнительную информацию смотрите в разделе “Соединение с репозиторием” на стр. 117. Порт, пароль и другие подробности соединения узнайте у местного администратора системы.

*Примечание:* На сервере репозитория также должна быть установлено программное обеспечение IBM SPSS Modeler Advantage.

1. В репозитории: запишите диалоговое окно, выберите папку для записи объекта, задайте любую другую информацию, которую нужно записать, и нажмите кнопку **Записать**. Дополнительную информацию смотрите в разделе “Задание свойств объекта” на стр. 118.

Запустится IBM SPSS Modeler Advantage с уже открытым потоком. В IBM SPSS Modeler поток закроется.

---

## Импорт и экспорт моделей в виде PMML

Язык разметки предсказательных моделей (predictive model markup language, PMML) - это формат XML, описывающий исследование данных и статистические модели, в том числе входные данные моделей, а также преобразования, используемые для подготовки данных к исследованию, и параметры, определяющие сами модели. IBM SPSS Modeler может экспортировать и импортировать PMML, делая возможным совместное использование моделей несколькими поддерживающими этот формат прикладными программами, например, IBM SPSS Statistics.

Более подробную информацию о PMML смотрите на сайте группы исследования данных (<http://www.dmg.org>).

Экспортировать модель

Экспорт PMML поддерживается для большинства типов моделей, сгенерированных в IBM SPSS Modeler. Дополнительную информацию смотрите в разделе “Типы моделей, поддерживающие PMML”.

1. Щелкните правой кнопкой мыши по слепку модели на палитре моделей. (Можно также дважды щелкнуть по слепку модели на холсте и выбрать меню Файл).
2. В меню щелкните по **Экспортировать PMML**.
3. В диалоговом окне Экспорт (или Сохранить) укажите каталог назначения и уникальное имя для модели.

*Примечание:* Опции для экспорта PMML можно изменить в диалоговом окне Опции пользователя. В главном меню выберите:

**Инструменты > Опции > Опции пользователя**

и перейдите на вкладку PMML.

Дополнительную информацию смотрите в разделе “Задание опций экспорта PMML” на стр. 149.

Импортировать модель, сохраненную как PMML

Модели, экспортированные в виде PMML из IBM SPSS Modeler или других прикладных программ, можно импортировать на палитру моделей. Дополнительную информацию смотрите в разделе “Типы моделей, поддерживающие PMML”.

1. На палитре моделей щелкните по ней правой кнопкой мыши и выберите из меню **Импортировать PMML**.
2. Выберите файл для импорта и нужным образом задайте опции для меток переменных.
3. Нажмите кнопку **Открыть**.

**Использовать метки переменных, если они есть в модели.** Для переменных в словаре данных PMML может задавать и имена, и метки переменных (например, Referrer ID для *RefID*). Выберите эту опцию для использования меток переменных, если они присутствуют в исходном экспортированном PMML.

Если выбрана опция меток переменных, но в PMML нет меток переменных, имена переменных используются обычным способом.

## Типы моделей, поддерживающие PMML

Экспорт PMML

**IBM SPSS Modeler модели.** Следующие созданные в IBM SPSS Modeler модели можно экспортировать как PMML 4.0:

- Дерево C&R
- QUEST

- CHAID
- линейная регрессия
- Нейросеть
- C5.0
- логистическая регрессия
- Обобщенная линейная регрессия
- SVM
- Априорный анализ
- Carma
- К-средние
- Коонена
- TwoStep
- GLMM (поддерживается только для моделей GLMM с только фиксированными эффектами)
- Список решений
- Модель Кокса
- Последовательность (скоринг для моделей PMML последовательностей не поддерживается)
- StatisticsМодель

**Внутренние модели баз данных.** Если модели сгенерированы с помощью внутренних алгоритмов баз данных, экспорт PMML доступен только для моделей IBM InfoSphere Warehouse. Модели, созданные при помощи Analysis Services от Microsoft или Oracle Data Miner, экспортировать нельзя. Обратите внимание также на то, что модели IBM, экспортированные как PMML, нельзя импортировать обратно в IBM SPSS Modeler.

#### Импорт PMML

IBM SPSS Modeler может импортировать и оценивать модели PMML, сгенерированные текущими версиями всех продуктов IBM SPSS Statistics, в том числе модели, экспортированные из IBM SPSS Modeler, а также модель или ее преобразованную в PMML форму, сгенерированную IBM SPSS Statistics 17.0 или более новой версии. Существенно, что это относится к любому PMML, который может оценить механизм скоринга, со следующими исключениями:

- Нельзя импортировать модели Априори, CARMA, обнаружения аномалий и последовательности.
- После импорта в IBM SPSS Modeler модели PMML нельзя просматривать, хотя их можно использовать при скоринге. (Обратите внимание на то, что в первую очередь это относится к моделям, экспортированным из IBM SPSS Modeler. Чтобы избежать этого ограничения, экспортируйте модель как файл сгенерированной модели [*\*.gm*], а не PMML).
- Модели IBM InfoSphere Warehouse, экспортированные как PMML, нельзя импортировать.
- При импорте происходит ограниченная проверка, но при попытке проведения скоринга модели - полная. Таким образом, возможны ситуации, когда импорт прошел успешно, но скоринг завершился неудачно или привел к неправильным результатам.



---

## Глава 11. Проекты и отчеты

---

### Введение в проекты

**Проект** - это группа файлов, связанных с некоторой задачей исследования данных. Проекты содержат потоки данных, диаграммы, сгенерированные модели, отчеты и любые другие файлы, созданные в IBM SPSS Modeler. На первый взгляд может показаться, что проекты IBM SPSS Modeler - просто способ организации вывода, но в действительности их возможности куда больше. Пользуясь проектами, можно:

- Аннотировать объекты в файле проекта.
- Использовать методологию CRISP-DM как путеводитель по исследованию данных. Проекты также содержат справочную систему CRISP-DM со всеми подробностями и примерами из практики к исследованию данных по CRISP-DM.
- Добавлять в проект объекты, созданные вне IBM SPSS Modeler, например, слайдовые презентации PowerPoint, рассказывающие о целях вашего исследования данных, или "белые книги" по алгоритмам, которые вы планируете использовать.
- Создавать как всесторонние, так и простые отчеты об изменениях на основе ваших аннотаций. Эти отчеты можно генерировать в формате HTML для удобства публикации во внутренней сети вашей организации.

*Примечание:* Если панель проекта в окне IBM SPSS Modeler не видна, выберите **Проект** в меню Вид.

Есть два способа просмотреть добавленные в проект объекты: **Представление классов** и **Представление CRISP-DM**. Все, что вы добавляете в проект, добавляется в оба представления, между которыми можно переключаться, выбирая, что удобнее в данный момент.

### Представление CRISP-DM

Проекты IBM SPSS Modeler поддерживают межотраслевой стандарт CRISP-DM (Cross-Industry Standard Process for Data Mining) и предлагают проверенный на практике непроприетарный способ организовать различные участки работы по исследованию данных. CRISP-DM использует шесть этапов для описания процесса от начала (сбор бизнес-требований) до конца (внедрение результатов). Хотя некоторые из этапов обычно не требуют работы в IBM SPSS Modeler, панель проектов содержит все шесть этапов, чтобы у вас был центральный пункт хранения и отслеживания всех материалов, связанных с проектом. Например, этап понимания бизнеса обычно включает в себя сбор требований и совещания с коллегами, чтобы определить цели исследования, а не работу с данными в IBM SPSS Modeler. Вы можете сохранить свои заметки о таких совещаниях на панели проектов в папке *Понимание бизнеса*, чтобы использовать их как справку и включить в отчеты.

Кроме того, представление CRISP-DM на панели проектов снабжено своей справочной системой, которая поможет вам выполнить весь жизненный цикл исследования данных. В IBM SPSS Modeler эта справка доступна как **Справка CRISP-DM** в меню Справка.

*Примечание:* Если панель проекта в окне не видна, выберите **Проект** в меню Вид.

### Задание этапа по умолчанию для проекта

Объекты, добавляемые в проект, добавляются в этап CRISP-DM по умолчанию. Это значит, что вам будет нужно вручную сгруппировать объекты согласно тому этапу исследования данных, на котором вы их использовали. Есть смысл задать в качестве папки по умолчанию тот этап, с которым вы сейчас работаете.

#### Чтобы выбрать, какой этап будет этапом по умолчанию:

1. В представлении CRISP-DM щелкните правой кнопкой по папке для этапа, задаваемого как этап по умолчанию.
2. В меню выберите **Задать как параметры по умолчанию**.

Папка по умолчанию выводится жирным шрифтом.

## Представление классов

В представлении Классы ваша работа IBM SPSS Modeler разбита на категории по типу создаваемых объектов. Сохраненные объекты можно добавлять в любые из следующих категорий:

- Потоки
- Узлы
- Модели
- Таблицы, диаграммы и отчеты
- Прочее (файлы, созданные вне IBM SPSS Modeler, такие как слайдовая презентация или "белые книги", относящиеся к вашей работе по исследованию данных)

При добавлении объектов в представление Классы эти объекты также добавляются в папку этапа по умолчанию в представлении CRISP-DM.

*Примечание:* Если панель проекта в окне не видна, выберите **Проект** в меню Вид.

---

## Построение проекта

Проект - это файл, содержащий ссылки на все файлы, которые вы связали с проектом. Это значит, что элементы проекта сохраняются и сами по себе, и как ссылки в файле проекта (.cprj). Поскольку проект - это реляционная структура, обратите внимание на следующее:

- Чтобы добавить в проект новый элемент, его сначала нужно сохранить отдельно. Если элемент несохраненный, при попытке добавить его в проект выводится приглашение сначала сохранить элемент.
- Объекты, которые обновляются отдельно, такие как потоки, обновляются также и в файле проекта.
- Если вручную переместить или удалить объекты (такие как потоки, узлы и объекты вывода) из файловой системы, ссылки в файле проекта станут недопустимыми.

## Создание нового проекта

Новые проекты удобно создавать в окне IBM SPSS Modeler. Если ни какой проект не открыт, можете сразу начинать строить новый; существующий проект можно закрыть и начать с нуля.

В главном меню выберите:

**Файл > Project > Новый проект...**

## Добавление в проект

После того, как вы создали или открыли проект, вы можете, пользуясь рядом методов, добавлять в проект объекты, такие как потоки данных, узлы и отчеты.

Добавление объектов с панели менеджеров

На панели менеджеров в верхнем правом углу окна IBM SPSS Modeler можно добавлять потоки или объекты вывода.

1. Выберите объект, такой как таблица или поток, на одной из вкладок панели менеджеров.
2. Щелкните по нему правой кнопкой и выберите **Добавить в проект**.  
Если объект уже был сохранен, он будет автоматически добавлен в соответствующую папку объектов (в представлении классов) или в папку этапа по умолчанию (в представлении CRISP-DM).
3. Другой вариант - перетащить объекты с панели менеджеров на панель проекта.



*Примечание:* Возможно, вам предложат сначала сохранить объект. При сохранении убедитесь, что в диалоговом окне Сохранить включен переключатель **Добавить файл в проект**. Тогда после сохранения объект будет добавлен в проект автоматически.

Добавление узлов с холста

Можно добавлять и отдельные узлы с холста потока, пользуясь диалоговым окном Сохранить.

1. Выберите узел на холсте.
2. Щелкните по нему правой кнопкой и выберите **Сохранить узел**. Другой вариант - выберите в главном меню:  
**Правка > Узел > Сохранить узел...**
3. В диалоговом окне Сохранить включите переключатель **Добавить файл в проект**.
4. Создайте имя для узла и нажмите кнопку **Сохранить**.

Файл будет сохранен и добавлен в проект. Узлы добавляются в папку *Узлы* в представлении классов и в папку этапа по умолчанию в представлении CRISP-DM.

Добавление внешних файлов

В проект можно добавлять широкий спектр объектов, созданных вне IBM SPSS Modeler. Это полезно, когда весь процесс исследования данных управляется из IBM SPSS Modeler. Например, в проекте можно хранить ссылки на данные, заметки, презентации и диаграммы. В представлении CRISP-DM внешние файлы можно добавлять в любую папку. В представлении классов внешние файлы можно сохранить только в папке *Другие*.

**Чтобы добавить внешние файлы в проект:**

1. Перетащите файлы с рабочего стола в проект.  
*или*
2. Щелкните правой кнопкой по папке назначения в представлении CRISP-DM или в представлении классов.
3. В меню выберите **Добавить в папку**.
4. Выберите файл в диалоговом окне и нажмите кнопку **Открыть**.

Ссылка на выбранный объект будет добавлена в проекты IBM SPSS Modeler.

## Передача проектов в IBM SPSS Collaboration and Deployment Services Repository

Можно передать весь проект, включая все файлы компонентов, в IBM SPSS Collaboration and Deployment Services Repository за одно действие. Все объекты, которые уже находятся в положении назначения, не перемещаются. Эта возможность работает и в обратную сторону: проекты можно целиком перемещать из IBM SPSS Collaboration and Deployment Services Repository в локальную файловую систему.

*Примечание:* Для получения доступа к репозиторию IBM SPSS Collaboration and Deployment Services требуется отдельная лицензия. Дополнительную информацию смотрите в разделе <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>

Перенос проекта

Убедитесь, что переносимый проект открыт на панели проектов.

**Чтобы перенести проект:**

1. Щелкните правой кнопкой по папке проекта и выберите **Перенести проект**.

2. Если появится приглашение, зарегистрируйтесь в IBM SPSS Collaboration and Deployment Services Repository.
3. Укажите новое положение для проекта и нажмите кнопку **ОК**.

## Задание свойств проекта

В диалоговом окне свойств проекта можно настроить содержимое проекта и документацию по проекту. Чтобы вызвать свойства проекта:

1. Щелкните правой кнопкой по объекту или папке на панели проекта и выберите **Свойства проекта**.
2. Щелкните по вкладке **Проект**, чтобы задать основную информацию о проекте.

**Создан.** Содержит дату создания проекта (редактировать нельзя).

**Сводная информация.** Можно ввести сводку по проекту исследования данных, которая войдет в отчет по проекту.

**Содержимое.** Содержит тип и число компонентов, на которые ссылается файл проекта (редактировать нельзя).

**Сохранить несохраненный объект как.** Задает, нужно ли сохранять несохраненные объекты в локальной файловой системе или записывать их в репозиторий. Дополнительную информацию смотрите в разделе “О программе IBM SPSS Collaboration and Deployment Services Repository” на стр. 115.

**Изменять ссылки на объекты при загрузке проекта.** Выберите эту опцию, чтобы обновить ссылки проекта на его компоненты. *Примечание:* Добавленные в проект файлы не сохраняются в самом файле проекта. Вместо этого в проекте сохраняется ссылка на файлы. Это значит, что при перемещении или удалении файла объект удаляется из проекта.

## Аннотирование проекта

На панели проектов поддерживается ряд способов аннотировать вашу работу по исследованию данных. Аннотации на уровне проектов часто используются для отслеживания целей и решений "в целом", а аннотации к папке или узлу - для детализации. На вкладке Аннотации вам предоставляется место для изложения подробностей на уровне проекта, таких как исключение данных с неустраняемыми пропусками или многообещающие гипотезы, сформировавшиеся при осмотре данных.

**Чтобы аннотировать проект:**

1. Выберите папку проекта в представлении CRISP-DM или в представлении классов.
2. Щелкните правой кнопкой по папке и выберите **Свойства проекта**.
3. Щелкните по вкладке **Аннотации**.
4. Введите ключевые слова и текст, чтобы описать проект.

### Свойства папки и аннотации

Аннотировать можно отдельные папки проектов (в представлении CRISP-DM и в представлении классов). В представлении CRISP-DM это весьма эффективный способ документировать, каковы цели организации на каждом этапе исследования данных. Например, при использовании утилиты аннотаций для папки *Как устроен бизнес*, вы можете включить такие документы, как "Бизнес-задача этого исследования - уменьшить панические настроения среди высокоценных клиентов". Затем этот текст может быть автоматически включен в отчет по проекту, если выбрать опцию **Включить в отчет**.

**Чтобы аннотировать папку:**

1. Выберите папку на панели проектов.
2. Щелкните правой кнопкой по папке и выберите **Свойства папки**.

В представлении CRISP-DM к аннотациям папки относятся сводки задач каждого этапа, а также указания по выполнению соответствующих задач исследования данных. Любую из этих аннотаций можно удалить или отредактировать.

**Имя.** Эта область содержит имя выбранного поля.

**Контекстная справка по инструменту.** Создайте пользовательскую подсказку, которая будет выводиться, если остановить указатель мыши на папке проекта. В представлении CRISP-DM это полезно, например, для того, чтобы предложить краткий обзор целей каждого этапа или отметить состояние этапа, такое как "Выполняется" или "Выполнен".

**Поле аннотации.** Это поле служит для более пространственных аннотаций, которые могут быть собраны в отчет по проекту. В представлении CRISP-DM в этой аннотации описывается каждый этап исследования данных, но вы можете изменить ее смысл в своем проекте.

**Включить в отчет.** Чтобы эта аннотация вошла в отчеты, выберите **Включить в отчет**.

## Свойства объекта

Вы можете просмотреть свойства объекта и выбрать отдельные объекты для включения в отчет по проекту. Чтобы вызвать свойства объекта:

1. Щелкните правой кнопкой по объекту на панели проектов.
2. Выберите в меню **Свойства объекта**.

**Имя.** Эта область содержит имя сохраненного объекта.

**Путь.** Эта область содержит положение сохраненного объекта.

**Включить в отчет.** Выберите эту опцию, чтобы включить подробности объекта в генерируемый отчет.

## Закрытие проекта

Когда вы выходите из IBM SPSS Modeler или открываете новый проект, существующий файл проекта (.cpr) закрывается.

Некоторые связанные с проектом файлы (потoki, узлы или диаграммы) могут остаться открытыми. Если нужно оставить эти файлы открытыми, ответьте Нет на сообщение **... Хотите сохранить и закрыть эти файлы?**

Если вы измените и сохраните связанные с проектом файлы после его закрытия, измененные версии будут включены в проект при следующем его открытии. Чтобы избежать этого, удалите сохраняемый файл из проекта или сохраните его под новым именем.

---

## Генерирование отчета

Одна из наиболее полезных возможностей проекта - генерирование отчетов на основе элементов проекта и аннотаций. Это критически важный компонент эффективного исследования данных, как постоянно отмечалось при обсуждении методологии CRISP-DM. Отчет можно сгенерировать непосредственно в ряде типов файлов, а также в окне для немедленного просмотра на экране. Из окна отчет можно напечатать, сохранить или открыть в браузере. Сохраненные ответы можно распространить среди других пользователей организации.

Нередко в процессе исследования данных отчеты несколько раз генерируются из файлов проектов и распространяются среди участников проекта. Отчет вбирает в себя главное из объектов, на которые ссылается файл проекта, а также все созданные аннотации. Отчеты можно создавать либо на основе представления классов, либо на основе представления CRISP-DM.

### Чтобы сгенерировать отчет:

1. Выберите папку проекта в представлении CRISP-DM или в представлении классов.
2. Щелкните правой кнопкой по папке и выберите **Отчет по проекту**.
3. Задайте опции отчета и нажмите кнопку **Сгенерировать отчет**.

При помощи опций в диалоговом окне отчета поддерживается ряд способов сгенерировать отчет нужного типа:

**Имя вывода.** Задайте имя окна вывода, если вы решили отправить вывод отчета на экран. Можно задать пользовательское имя или разрешить IBM SPSS Modeler автоматически назвать это окно без вашего участия.

**Вывод на экран.** Выберите эту опцию, чтобы сгенерировать и вывести отчет в окне вывода. Обратите внимание на то, что из окна вывода тоже можно экспортировать отчет в различные типы файлов.

**Вывод в файл.** Включите эту опцию, чтобы сгенерировать и сохранить отчет как файл того типа, который задан в списке Тип файла.

**Имя файла.** Задайте имя файла для генерируемого отчета. По умолчанию файлы сохраняются в каталог IBM SPSS Modeler *\bin*. Чтобы сохранить в другом положении, нажмите кнопку просмотра (...).

**Тип файла.** Доступны следующие типы файлов:

- **Документ HTML.** Отчет сохраняется как отдельный файл HTML. Если ваш отчет содержит диаграммы, они будут сохранены как файлы PNG и ссылки на них будут помещены в файл HTML. При публикации отчета в Интернете убедитесь, что закачиваете и файл HTML, и все изображения, на которые он ссылается.
- **Текстовый документ.** Отчет сохраняется как отдельный текстовый файл. Если отчет содержит диаграммы, в итоге в отчет попадут только путь и имена файлов.
- **Документ Microsoft Word.** Отчет сохраняется как отдельный документ со всеми диаграммами, встроенными непосредственно в документ.
- **Документ Microsoft Excel.** Отчет сохраняется как отдельная электронная таблица со всеми диаграммами, встроенными непосредственно в эту электронную таблицу.
- **Документ Microsoft PowerPoint.** Каждый этап выводится как отдельный слайд. Все диаграммы непосредственно встраиваются в слайды PowerPoint.
- **Объект вывода.** При открытии в IBM SPSS Modeler этот файл (.oui) эквивалентен опции **Вывод на экран** в группе **Формат отчета**.

*Примечание:* Чтобы экспортировать в тот или иной файл Microsoft Office, нужно установить соответствующую прикладную программу.

**Заголовок.** Задайте заголовок отчета.

**Структура отчета.** Выберите **CRISP-DM** или **Классы**. В представлении CRISP-DM генерируется отчет о состоянии с обзорными сведениями и подробностями по каждому этапу исследования данных. Представление классов основано на объектах и больше подходит для внутреннего отслеживания данных и потоков.

**Автор.** Выводится имя пользователя по умолчанию, но его можно изменить.

**Отчет включает.** Выберите метод включения объектов в отчет. Выберите **все папки и объекты**, чтобы взять все элементы, добавленные в файл отчета. Кроме того, можно брать элементы с учетом того, выбрана ли опция **Включить в отчет** в свойствах объекта. Другой вариант - чтобы проверить элементы, не вошедшие в отчет, можно взять только элементы, помеченные как исключаемые (те, у которых не выбрана опция **Включить в отчет**).

**Выбрать.** При помощи этой опции можно создавать отчеты об обновлениях проекта, беря только **последние элементы**. Другой вариант - можно отслеживать более старые и, возможно, не решенные вопросы, задав параметры для **старых элементов**. Выберите **все элементы**, чтобы игнорировать время как параметр при составлении отчета.

**Упорядочить по.** Можно выбрать сочетание следующих характеристик, чтобы упорядочивать по ним объекты в папке:

- **Тип.** Группировать объекты по типу.
- **Имя.** организовать объекты по алфавиту.
- **Дата добавления.** Сортировать объекты по дате добавления в проект.

## Сохранение и экспорт сгенерированных отчетов

Отчет, сгенерированный на экран, открывается в новом окне вывода. Все диаграммы в отчете показаны как изображения, размещенные в тексте.

Терминология отчета

В отчете приводится общее число узлов в каждом потоке. Эти числа даны под следующими заголовками, терминология которых взята из IBM SPSS Modeler, а не из CRISP-DM:

- **Узлы чтения данных.** Узлы источников.
- **Узлы записи данных.** Узлы экспорта.
- **Построители моделей.** Узлы моделирования.
- **Применители моделей.** Сгенерированные модели, так называемые слепки.
- **Построители вывода.** Узлы диаграмм или вывода.
- **Другое.** Остальные узлы, относящиеся к проекту. Например, узлы, доступные на вкладках Операции с полями и Операции на палитре узлов.

**Чтобы сохранить отчет:**

1. В меню Файл выберите **Сохранить**.
2. Укажите имя файла.  
Отчет сохраняется как отдельный объект вывода.

**Чтобы экспортировать отчет:**

3. В меню Файл выберите **Экспорт** и тип файла, в который нужно экспортировать информацию.
4. Укажите имя файла.

Отчет сохраняется как указанный вами формат.

Вы можете экспортировать информацию в следующие типы файлов:

- HTML
- Текст
- Microsoft Word
- Microsoft Excel
- Microsoft PowerPoint

*Примечание:* Чтобы экспортировать в тот или иной файл Microsoft Office, нужно установить соответствующую прикладную программу.

Кнопки в верхней части окна служат для следующего:

- Печать отчета.
- Просмотр отчета как HTML во внешнем браузере.



---

## Глава 12. Настройка IBM SPSS Modeler

---

### Настройка опций IBM SPSS Modeler

Доступен ряд операций для настройки IBM SPSS Modeler под ваши потребности. Прежде всего настройка включает в себя задание опций пользователя, таких как выделение памяти, каталоги по умолчанию и использование звука и цвета. Кроме того, можно настроить палитру узлов в нижней части окна IBM SPSS Modeler.

---

### Настройка опций IBM SPSS Modeler

Есть несколько способов задать свои опции для IBM SPSS Modeler:

- Задать опции системы, такие как использование памяти и национальная версия, выбрав **Опции системы** в меню **Инструменты > Опции**.
- Задать опции пользователя, такие как экранные шрифты и цвета, выбрав **Опции пользователя** в меню **Инструменты > Опции**.
- Указать положение прикладных программ, работающих совместно с IBM SPSS Modeler, выбрав **Вспомогательные программы** в меню **Инструменты > Опции**.
- Указать каталоги по умолчанию, используемые в IBM SPSS Modeler, выбрав **Задать каталог** или **Задать каталог сервера** в меню **Файл**.

Кроме того, вы можете задать опции, применимые только к вашим потокам - ко всем или только некоторым. Дополнительную информацию смотрите в разделе “Задание опций для потоков” на стр. 41.

### Системные опции

Вы можете задать предпочтительный язык или национальную версию IBM SPSS Modeler, выбрав **Опции системы** в меню **Инструменты > Опции**. Здесь же можно задать максимальное использование памяти для IBM SPSS Modeler. Обратите внимание на то, что изменения, внесенные в этом диалоговом окне, вступят в силу только после перезапуска IBM SPSS Modeler.

**Максимальная память.** Выберите, чтобы задать предельное число мегабайт для использования памяти IBM SPSS Modeler. На некоторых платформах IBM SPSS Modeler ограничивает размер своего процесса, чтобы ограничить нагрузку на компьютер с ограниченными ресурсами или тяжелой нагрузкой. Если вы имеете дело с большими объемами данных, такое самоограничение может привести к ошибке “Нехватка памяти”. Вы можете уменьшить нагрузку на память, задав новое пороговое значение.

**Использовать национальную версию системы.** По умолчанию эта опция выбрана и для нее задано значение Английский (США). Снимите это выделение и задайте другой язык из списка доступных языков и национальных версий.

### Работа с памятью

Помимо параметра **Максимальный размер памяти**, задаваемого в диалоговом окне Опции системы, есть ряд способов оптимизировать использование памяти:

- Настройте опцию **Максимальное число элементов для номинальных полей** в диалоговом окне свойства потока. Эта опция задает максимальное число элементов для номинальных полей, после которого тип измерений поля становится *Без типа*. Дополнительную информацию смотрите в разделе “Задание общих опций потоков” на стр. 42.
- Чтобы IBM SPSS Modeler освободил память, щелкните в нижнем правом углу окна, где выводится память, используемая IBM SPSS Modeler, и выделенный объем (*xx* Мбайт / *xx* Мбайт). После щелчка по этой области она потемнеет, а затем размеры выделения памяти уменьшатся. Когда область вернется к обычному цвету, IBM SPSS Modeler освободит максимально возможный объем памяти.

## Задание каталогов по умолчанию

Вы можете задать каталог по умолчанию, используемый для проводника по файлам и вывода, выбрав **Задать каталог** или **Задать каталог сервера** в меню **Файл**.

- **Задать каталог.** При помощи этой опции задается рабочий каталог. По умолчанию рабочий каталог учитывает путь установки версии IBM SPSS Modeler или путь командной строки, использованный для запуска IBM SPSS Modeler. В локальном режиме рабочий каталог - это путь, используемый всеми операциями на стороне клиента и файлами вывода (если они заданы с относительным путем).
- **Задать каталог сервера.** Опция **Задать каталог сервера** доступна в меню **Файл**, если есть соединение с удаленным сервером. При помощи этой опции задается каталог по умолчанию для всех файлов сервера и файлов данных, заданных для ввода или вывода. По умолчанию каталог сервера - это *\$CLEO/data*, где *\$CLEO* - каталог, в котором установлена версия сервера IBM SPSS Modeler. При помощи командной строки можно переопределить это значение по умолчанию, задав флаг `-server_directory` в аргументе командной строки `modelerclient`.

## Задание опций пользователя

Общие опции для IBM SPSS Modeler можно задать, выбрав команду **Опции пользователя** в меню **Инструменты > Опции**. Эти опции применяются ко всем потокам, используемым в IBM SPSS Modeler.

Следующие опции можно задать, щелкнув по соответствующей вкладке:

- Опции уведомления, например, сообщения о перезаписи модели или об ошибке.
- Опции вывода, например, цвета диаграмм и фона.
- Опции цветовой разметки синтаксиса.
- Опции экспорта PMML, которые используются при экспорте моделей в формате PMML (Predictive Model Markup Language).
- Информация о пользователе или об авторе, такая как фамилия, инициалы, электронный адрес. Эта информация будет выводиться на вкладке **Аннотации** для узлов и других объектов, которые вы создадите.

Чтобы задать опции для потока, такие как десятичные разделители, форматы даты и времени, оптимизация, макет потока и сценарии потока, откройте диалоговое окно **Свойства потока**, доступное из меню **Файл** и из меню **Инструменты**.

## Задание опций уведомления

На вкладке **Уведомление** диалогового окна **Опции пользователя** можно задать различные опции, касающиеся вывода и типа окон предупреждений и подтверждений в IBM SPSS Modeler. Кроме того, можно задать поведение вкладок **Вывод** и **Модели** на панели менеджеров при генерировании нового вывода и моделей.

**Показать диалоговое окно обратной связи выполнения потока.** Включите этот переключатель, чтобы вывести диалоговое окно с индикатором хода выполнения, если поток выполняется уже больше трех секунд. Диалоговое окно также содержит подробности о выходных объектах, созданных потоком.

- **Закрыть диалоговое окно по завершении.** По умолчанию после выполнения потока диалоговое окно закрывается. Выключите этот переключатель, если хотите видеть это диалоговое окно и после выполнения потока.

**Предупреждать, если узел перезаписывает файл.** Включите этот переключатель, чтобы при помощи сообщения об ошибке предупреждать, когда операции с узлами ведут к перезаписи существующего файла.

**Предупреждать, если узел перезаписывает таблицу базы данных.** Включите этот переключатель, чтобы при помощи сообщения об ошибке предупреждать, когда операции с узлами ведут к перезаписи существующей таблицы базы данных.

Звуковые уведомления



В этом списке можно задать уведомление звуками о событиях и ошибках. Доступен ряд звуков. Выбранный звук можно воспроизвести кнопкой Воспроизвести (кнопка с динамиком). Чтобы найти и выбрать звук, нажмите кнопку просмотра (...).

*Примечание:* Файлы .wav, которые служат для создания звуков в IBM SPSS Modeler, хранятся в подкаталоге /media/sounds каталога установки.

- **Отключить все звуковые сигналы.** Включите этот переключатель, чтобы отключить звуковое уведомление для всех событий.

Визуальные уведомления

Опции в этой группе задают поведение вкладок Вывод и Модели на панели менеджеров в верхней правой части дисплея при генерировании новых элементов. Чтобы задать поведение соответствующих вкладок, выберите в списке **Новая модель** или **Новый вывод**.

Для варианта **Новая модель** доступны следующие опции:

**Добавить модель в поток.** Если этот переключатель включен (состояние по умолчанию), то после построения модели она добавляется и в поток, и на вкладку моделей. В потоке модель показана как ссылка на узел моделирования, из которого эта модель была создана. Если этот переключатель выключен, модель добавляется только на вкладку Модели.

**Заменить предыдущую модель.** Если этот переключатель включен (состояние по умолчанию), то существующая модель перезаписывается в этом потоке на вкладке моделей и на холсте потока. Если этот переключатель выключен, то модель добавляется к существующим моделям на вкладке и на холсте. Обратите внимание на то, что эта настройка пересиливается настройкой замены моделей в ссылке на модель.

Приведенные ниже опции доступны при выборе **Нового вывода**:

**Предупреждать, если вывод превышает [n].** Выберите, показывать ли предупреждение, когда число элементов на вкладке Вывод превышает заранее заданное количество. По умолчанию это количество - 20, но его можно изменить при необходимости.

Во всех случаях доступны следующие опции:

**Выбрать вкладку.** Выбрать, надо ли переключиться на вкладку Вывод или на вкладку Модели, когда соответствующий объект генерируется при выполнении потока.

- Выберите **Всегда**, чтобы переключаться на соответствующую вкладку на панели менеджеров.
- Выберите **Если генерируется текущим потоком**, чтобы переключаться на соответствующую вкладку, только если объекты генерируются тем потоком, который в настоящее время виден на холсте.
- Выберите **Никогда**, чтобы запретить программе переключиться на соответствующую вкладку и уведомить о сгенерированном выводе или модели.

**Выделить вкладку.** Выберите, выделить ли вкладку Вывод или Модели на панели менеджеров, после того как сгенерирован новый вывод или модель.

- Выберите **Если не выбрана**, чтобы, генерируя новый объект на панели менеджеров, выделить соответствующую вкладку, если она не выбрана.
- Выберите **Никогда**, чтобы запретить программе выделить соответствующую вкладку и уведомить о сгенерированном объекте.

**Прокрутить палитру, чтобы вывести на экран (только для новой модели).** Выберите, нужно ли автоматически прокрутить вкладку Модели на панели менеджеров, чтобы стала видна новейшая модель.

- Выберите **Всегда**, чтобы разрешить прокрутку.

- Выберите **Если генерируется текущим потоком**, чтобы прокрутить только для объектов, сгенерированных тем потоком, который в настоящее время виден на холсте.
- Выберите **Никогда**, чтобы запретить программе автоматически прокрутить вкладку Модели.

**Открыть окно (только для нового вывода).** Выберите, нужно ли автоматически открыть окно, после того как сгенерирован вывод.

- Выберите **Всегда**, чтобы разрешить новое окно вывода.
- Выберите **Если генерируется текущим потоком**, чтобы открыть новое окно для вывода, сгенерированного тем потоком, который в настоящее время виден на холсте.
- Выберите **Никогда**, чтобы запретить программе автоматически открыть новое окно для сгенерированного вывода.

Чтобы вернуться на этой вкладке к значениям системы по умолчанию, выберите **Значения по умолчанию**.

## Задание опций вывода

На вкладке Дисплей диалогового окна Опции пользователя можно задать опции вывода шрифтов и цветов в IBM SPSS Modeler.

**Выводить при запуске окно приветствия.** Включите этот переключатель, чтобы выводить диалоговое окно приветствия при запуске. Из диалогового окна приветствия можно запустить обучающую программу по прикладным программам, открыть демонстрационный поток или существующий поток или проект или создать новый поток.

**Показать разметку потоков и надузлов.** Если этот переключатель включен, то разметка (если есть) на потоках и надузлах выводится по умолчанию. Разметка включает в себя комментарии к потоку, ссылки на модели и выделение ветвей оценок.

**Стандартные шрифты и цвета (действует после перезагрузки).** В этой секции можно задать оформление экрана IBM SPSS Modeler, цветовую схему и размер шрифтов. Выбранные здесь опции действуют только после закрытия и перезапуска IBM SPSS Modeler.

- **Внешний вид.** Дает возможность выбрать одну из стандартных цветовых схем и оформление экрана.  
Доступные варианты:  
**Стандарт SPSS** (по умолчанию), общее оформление для продуктов IBM SPSS.  
**Классика SPSS**, оформление, привычное пользователям прошлых версий IBM SPSS Modeler.  
**Windows**, оформление Windows, полезное для повышения контраста на холсте потока и палитрах.
- **Размер шрифта по умолчанию для узлов.** Укажите размер шрифта на палитрах узлов и для узлов на холсте потока.

*Примечание:* Размер значков узлов можно задать на панели Макет на вкладке Опции в диалоговом окне свойств потока. В главном меню выберите

**Инструменты > Свойства потока > Опции > Макет.**

**Пользовательские цвета.** Эта таблица содержит текущие выбранные цвета для различных элементов экрана. Для каждого элемента в таблице можно изменить текущий цвет, щелкнув дважды по соответствующей строке в столбце **Цвет** и выбрав цвет из списка. Чтобы задать пользовательский цвет, прокрутите экран вниз до конца списка и щелкните по пункту **Цвет....**

**Порядок цветов категории диаграмм.** Эта таблица содержит текущие выбранные цвета для новых диаграмм. Порядок цветов отражает порядок, в котором они будут использоваться на диаграмме. Например, если номинальное поле, которое используется для наложения цветов, содержит четыре уникальных значения, то для него будут использоваться только первые четыре цвета из списка. Для каждого элемента в таблице можно изменить текущий цвет, щелкнув дважды по соответствующей строке в столбце **Цвет** и выбрав цвет

из списка. Чтобы задать пользовательский цвет, прокрутите экран вниз до конца списка и щелкните по пункту **Цвет....** Внесенные здесь изменения не влияют на диаграммы, созданные ранее.

Чтобы вернуться на этой вкладке к значениям системы по умолчанию, выберите **Значения по умолчанию**.

## Задание опций вывода синтаксиса

На странице Синтаксис диалогового окна Пользовательские опции можно задать опции для атрибутов шрифта и цветов вывода в сценариях, которые вы создаете в IBM SPSS Modeler.

**Выделение синтаксиса.** Эта таблица содержит текущие выбранные цвета для различных элементов синтаксиса, в том числе для шрифта и для окна, в котором он выводится. Для каждого элемента в таблице можно изменить цвет, щелкнув по соответствующему выпадающему списку в строке и выбрав цвет из списка. Кроме того, для элементов шрифта можно выбрать выделение жирным шрифтом и курсивом.

**Просмотр.** В этой таблице выводится пример вывода синтаксиса с использованием атрибутов цветов и шрифта, которые вы выбрали в таблице **Выделение синтаксиса**. Этот просмотр изменяется, когда вы изменяете свой выбор.

Чтобы вернуться на этой вкладке к значениям системы по умолчанию, выберите **Значения по умолчанию**.

## Задание опций экспорта PMML

На вкладке PMML можно управлять тем, как IBM SPSS Modeler экспортирует модели в формат PMML (Predictive Model Markup Language). Дополнительную информацию смотрите в разделе “Импорт и экспорт моделей в виде PMML” на стр. 134.

**Экспорт PMML.** Здесь можно сконфигурировать варианты PMML, оптимальные для открытия в прикладной программе назначения.

- Выберите **С расширениями**, чтобы разрешить расширения PMML для особых случаев, где нет эквивалента в стандартном PMML. Заметим, что в большинстве случаев результат не будет отличаться от стандартного PMML.
- Выберите **Как стандартный PMML....**, чтобы экспортировать PMML, максимально приближенный к стандартному PMML.

**Опции стандартного PMML.** Если выбрана опция **Как стандартный PMML....**, можно выбрать один из двух допустимых способов экспортировать модели линейной и логистической регрессии:

- Как **<общие регрессионные> модели PMML**
- Как **<регрессионные> модели PMML**

Дополнительную информацию о PMML смотрите на сайте группы Data Mining Group по адресу <http://www.dmg.org>.

## Задание информации о пользователе

**Информация о пользователе/авторе.** Введенная здесь информация будет выводиться на вкладке Аннотации для узлов и других объектов, которые вы создадите.

---

## Настройка палитры узлов

Потоки строятся из узлов. Все узлы, доступные при построении потока, приведены на палитре узлов в нижней части окна IBM SPSS Modeler. Дополнительную информацию смотрите в разделе “Палитра узлов” на стр. 16.

Палитру узлов можно реорганизовать двумя способами:

- Настроить менеджер палитр. Дополнительную информацию смотрите в разделе “Настройка менеджера палитры” на стр. 150.

- Изменить вид вкладок палитр, содержащих подпалитры, на палитре узлов. Дополнительную информацию смотрите в разделе “Создание подпалитры” на стр. 151.

## Настройка менеджера палитры

Менеджер палитры можно настроить для вашего удобства при использовании IBM SPSS Modeler. Например, если вы часто анализируете данные временных рядов из некоторой базы данных, есть смысл обеспечить на специальной вкладке палитры доступ к таким узлам, как источник базы данных, интервал, временные ряды и диаграммы временного графика. Такого рода настройку удобно выполнять в менеджере палитры, создавая на палитре узлов пользовательские вкладки.

В менеджере палитры можно выполнить ряд задач:

- Управлять видимостью вкладок на палитре узлов под холстом потока.
- Изменять порядок вкладок на палитре узлов.
- Создавать и редактировать свои вкладки, а на них, в свою очередь, - подпалитры.
- Редактировать выбор узла по умолчанию на вкладке избранных узлов.

Чтобы открыть менеджер палитры, в меню Инструменты выберите **Управление палитрами**.

**Имя палитры.** Содержит имена всех вкладок палитры, показанных и не показанных на палитре узлов. Сюда входят и все вкладки палитры, созданные вами. Дополнительную информацию смотрите в разделе “Создание вкладки на палитре”.

**Число узлов.** Число узлов на каждой вкладке палитры. Если узлов много, будет удобнее разбить такую вкладку на подпалитры. Дополнительную информацию смотрите в разделе “Создание подпалитры” на стр. 151.

**Показан?.** Выберите это поле, чтобы показать эту вкладку на палитре узлов. Дополнительную информацию смотрите в разделе “Видимые вкладки на палитре узлов” на стр. 151.

**Подпалитры.** Чтобы выбрать подпалитры, показываемые на вкладке, выделите нужное **Имя палитры** и нажмите эту кнопку; откроется диалоговое окно подпалитр. Дополнительную информацию смотрите в разделе “Создание подпалитры” на стр. 151.

**Восстановить значения по умолчанию.** Нажмите эту кнопку, чтобы полностью отказаться от всех изменений и добавлений, сделанных на палитрах и подпалитрах, и вернуться к настройке палитры по умолчанию.

## Создание вкладки на палитре

Чтобы создать пользовательскую вкладку на палитре

1. Через меню Инструменты откройте менеджер палитры.
2. Справа от столбца *Показана?* нажмите кнопку Добавить палитру; откроется диалоговое окно Создать или редактировать палитру.
3. Введите уникальное **Имя палитры**.
4. В области **Доступные узлы** выберите узлы для добавления на вкладку палитры.
5. Нажмите кнопку со стрелкой вправо под названием Добавить узел, чтобы переместить выделенный узел в область **Выбранные узлы**. Повторяйте, чтобы добавить все нужные узлы.  
Добавив все нужные узлы, вы можете изменить порядок их вывода на вкладке палитры:
6. Кнопками с простыми стрелками перемещайте узел вверх и вниз на одну строку.
7. Кнопками со стрелкой и линией перемещайте узел в начало или конец списка.
8. Чтобы удалить узел с палитры, выделите узел и нажмите кнопку Удалить справа от области **Выделенные узлы**.

## Видимые вкладки на палитре узлов

Возможно, в IBM SPSS Modeler доступны варианты, которые не используете никогда; вкладки с такими узлами можно скрыть, задав это в менеджере палитры.

Чтобы выбрать, какие вкладки видны на палитре узлов:

1. Через меню Инструменты откройте менеджер палитры.
2. Пользуясь переключателями в столбце *Видима?*, включите или выключите видимость каждой вкладки палитры.

Чтобы навсегда удалить узел с палитры, выделите узел и нажмите кнопку Удалить справа от столбца *Видима?*. После удаления восстановить вкладку невозможно.

*Примечание:* Нельзя удалить с палитры вкладки по умолчанию, поставляемые с IBM SPSS Modeler, кроме вкладки Избранное.

Изменение порядка на палитре узлов

Выделив вкладки, которые нужно показать на палитре узлов, вы можете изменить для них порядок вывода:

1. Кнопками с простыми стрелками перемещайте вкладку вверх и вниз на одну строку. Сдвиг вверх в списке означает сдвиг влево на палитре узлов и наоборот.
2. Кнопками со стрелкой и линией перемещайте вкладку палитры в начало или конец списка. Вкладка в начале списка выводится на левом краю палитры узлов.

## Видимые подпалитры на вкладке палитры

Аналогично тому, как можно задавать, какие вкладки видимы на палитре узлов, для вкладки можно задавать, какие подпалитры на ней доступны.

Чтобы выбрать подпалитры, которые должны быть видимы на вкладке палитры:

1. Через меню Инструменты откройте менеджер палитры.
2. Выделите нужную палитру.
3. Нажмите кнопку Подпалитры; откроется диалоговое окно Подпалитры.
4. Пользуясь переключателями в столбце *Видима?*, включите или выключите видимость каждой подпалитры на вкладке палитры. Подпалитра **Все** видима всегда и не может быть удалена.
5. Чтобы навсегда удалить подпалитру со вкладки палитры, выделите подпалитру и нажмите кнопку Удалить справа от столбца *Видима?*.

*Примечание:* Нельзя удалить подпалитры по умолчанию, поставляемые с вкладкой палитры Моделирование.

Изменение порядка на вкладке палитры

Выделив подпалитры, которые нужно показать на вкладке палитры, вы можете изменить для них порядок вывода:

1. Кнопками с простыми стрелками перемещайте подпалитру вверх и вниз на одну строку.
2. Кнопками со стрелкой и линией перемещайте подпалитру в начало или конец списка.

Созданные вами подпалитры будут показаны на палитре узлов при выборе родительской вкладки палитры. Дополнительную информацию смотрите в разделе “Изменение представления вкладки палитры” на стр. 152.

## Создание подпалитры

Поскольку любой существующий узел можно добавить на пользовательские вкладки палитры, может оказаться, что вы выбрали столько узлов, что для их просмотра требуется прокрутка. Чтобы предотвратить необходимость в прокрутке, можно создавать подпалитры и на них помещать узлы, которые вы выбрали

для вкладки палитры. Например, если вы создали вкладку палитры с узлами, часто используемыми при создании ваших потоков, можете разбить ее на подпалитры с узлами источника, операций над полями, моделирования и вывода.

*Примечание:* Узлы подпалитры выбираются только из узлов, которые вы добавили на родительскую вкладку палитры.

Чтобы создать подпалитру:

1. Через меню Инструменты откройте менеджер палитры.
2. Выберите палитру, на которую добавить подпалитры.
3. Нажмите кнопку Подпалитры; откроется диалоговое окно Подпалитры.
4. Справа от столбца *Показана?* нажмите кнопку Добавить подпалитру; откроется диалоговое окно Создать или редактировать подпалитру.
5. Введите уникальное **Имя подпалитры**.
6. В области **Доступные узлы** выберите узлы для добавления на подпалитру.
7. Нажмите кнопку со стрелкой вправо под названием Добавить узел, чтобы переместить выделенный узел в область **Выбранные узлы**.
8. Добавив нужные узлы, нажмите кнопку **ОК**, чтобы вернуться в диалоговое окно подпалитр.

Созданные вами подпалитры будут показаны на палитре узлов при выборе родительской вкладки палитры. Дополнительную информацию смотрите в разделе “Изменение представления вкладки палитры”.

## Изменение представления вкладки палитры

Число узлов, доступных в IBM SPSS Modeler, весьма велико, и на небольших экранах палитру узлов иногда приходится прокручивать налево и направо, чтобы увидеть нужные узлы; это особенно заметно в отношении вкладки палитры моделирования. Чтобы сократить требуемую прокрутку, можно вывести только узлы из некоторой подпалитры (если заданы подпалитры). Дополнительную информацию смотрите в разделе “Создание подпалитры” на стр. 151.

Чтобы показать или скрыть узлы на вкладке палитры, выберите вкладку палитры, а затем в меню слева выберите показ всех узлов или только конкретной подпалитры.

---

## Управление узлами SEMI

SEMI объявлен устаревшим и заменен на CLEF, где предлагается более гибкий и удобный комплект возможностей. Дополнительную информацию смотрите в *Руководстве разработчика IBM SPSS Modeler 16 CLEF* в составе этого выпуска.

---

## Глава 13. Вопросы производительности для потоков и узлов

Проектируя потоки, вы можете добиться максимальной производительности, упорядочив узлы в наиболее эффективной конфигурации, разрешив кэши узлов, где нужно, и уделив внимание другим особенностям, описанным в этом разделе.

Помимо учета рассматриваемых здесь особенностей, дополнительный и более существенный выигрыш в производительности обычно достигается благодаря эффективному использованию базы данных, особенно путем оптимизации SQL.

---

### Порядок узлов

Даже если вы не пользуетесь оптимизацией SQL, порядок узлов в потоке может влиять на производительность. Общая цель в том, чтобы уменьшить объем данных, обрабатываемых на узлах ниже по потоку; следовательно, если есть узлы, уменьшающие объем данных, желательно разместить их ближе к началу потока. сервер IBM SPSS Modeler может применить некоторые правила переупорядочивания автоматически во время компиляции и вынести вперед определенные узлы, если можно убедиться, что это безопасно. (По умолчанию эта возможность включена. Узнайте у своего администратора системы, включена ли эта возможность в вашей установке.)

При использовании оптимизации SQL цель - максимальная доступность и эффективность. Поскольку оптимизация останавливается, когда поток содержит операцию, не выполняемую в базе данных, то рекомендуется сгруппировать оптимизируемые операторы SQL в начале потока. При такой стратегии больший объем обработки остается внутри базы данных, после чего меньше данных переносится в IBM SPSS Modeler.

Приведенные ниже операции можно выполнять в большинстве баз данных. Постарайтесь сгруппировать их в *начале* потока:

- Объединить по ключу (join)
- Выбрать
- Агрегировать
- Сортировка
- Выборка
- Добавлять в конец
- Операции DISTINCT в режиме *include*, в котором выбираются все поля
- Операции заполнения
- Основные вычисления со стандартной арифметикой и обработкой строк (зависит от базы данных)
- Задание флага

Приведенные ниже операции невозможно выполнять в большинстве баз данных. Их рекомендуется поместить *после* операций из приведенного выше списка:

- Операции над данными не из базы данных, а, например, из плоского файла
- Объединение по порядку
- Баланс
- Операции DISTINCT в режиме *discard* или операции, в которых как особый тип выбран некоторый поднабор полей
- Все операции, для которых требуется доступ к данным из записей помимо обрабатываемой записи

- Отклонения полей состояния и количества
- Операции узла хронологии
- Операции с функциями временных рядов ("функции, начинающиеся на @")
- Режимы проверки типы *Warn* и *Abort*
- Построение моделей, их применение и анализ

*Примечание:* Такие модели, как деревья решений, наборы правил, линейная регрессия и модели, сгенерированные по факторам, могут генерировать SQL, и, следовательно, их можно загнать в базу данных.

- Вывод данных в любое место, кроме той базы данных, в которой обрабатываются данные

## Кэши узлов

Чтобы оптимизировать выполнение потока, для любого неконечного узла можно задать **кэш**. После того, как для узла задан кэш, при следующем выполнении потока данных этот кэш будет заполнен прошедшими через этот узел данными. С этого момента данные будут считываться из этого кэша (который хранится на диске во временном каталоге), а не из источника данных.

Кэширование особенно полезно после операции, которая требует много времени, такой как сортировка, объединение или суммирование. Например, пусть у вас есть узел источника, сконфигурированный для чтения данных о продажах из базы данных, и узел суммирования, который вычисляет сводку продаж по положению. Имеет смысл задать кэш для узла суммирования, а не для узла источника, чтобы кэшировать сводные данные, а не весь набор данных.

*Примечание:* Кэширование на узлах источника, то есть простое копирование исходных данных, прочитанных в IBM SPSS Modeler, в большинстве случаев не повышает производительность.

Узлы, для которых включено кэширование, помечаются значком документа в верхнем правом углу. Если в кэше узла есть данные, этот значок документа зеленый.

Чтобы включить кэш

1. На холсте потока щелкните правой кнопкой по узлу и в меню выберите **Кэш**.
2. В подменю кэширования выберите **Включить**.
3. Выключить кэш можно, щелкнув правой кнопкой по узлу и выбрав **Выключить** в подменю кэширования.

Кэширование узлов в базе данных

Для потоков, выполняемых в базе данных, возможно кэширование промежуточных данных во временную таблицу в базе данных, а не в файловой системе. В сочетании с оптимизацией SQL возможен значительный выигрыш в производительности. Например, можно кэшировать и повторно использовать вывод потока, объединяющего несколько таблиц для создания производной таблицы исследования данных. Можно дополнительно повысить производительность, автоматически сгенерировав SQL для всех узлов ниже по потоку.

При использовании кэширования в базе данных для строковых значений длиннее 255 символов нужно либо убедиться, что выше по потоку есть узел типа и значения поля считаны, либо задать длину строки в параметре `default_sql_string_length` в файле `options.cfg`. Тогда для соответствующего столбца во временной таблице будет задана ширина, достаточная для таких строковых значений.

Чтобы воспользоваться преимуществами кэширования в базе данных, нужно разрешить и оптимизацию SQL, и кэширование в базе данных. Обратите внимание на то, что параметрами оптимизации сервера переопределяются параметры клиента. Дополнительную информацию смотрите в разделе “Задание опций оптимизации для потоков” на стр. 44.



Если кэширование в базе данных разрешено, то просто щелкните правой кнопкой по любому неконечному узлу, чтобы кэшировать данные в этой точке, и кэш автоматически будет создан непосредственно в базе данных при следующем выполнении потока. Если же кэширование в базе данных или оптимизация SQL не разрешены, кэш будет записан в файловую систему.

*Примечание:* Временные таблицы для кэширования поддерживаются в следующих базах данных: DB2, Netezza, Oracle, SQL Server и Teradata. В остальных базах данных при кэшировании используются обычные таблицы. Код SQL можно оптимизировать под конкретную базу данных - за помощью обратитесь в службу поддержки.

---

## Производительность: узлы обработки

**Сортировка.** Чтобы сортировать данные, узел сортировки сначала должен прочитать весь входной набор данных. До некоторого предела данные считываются в оперативную память, сверх этого данные скидываются на диск. При сортировке используется сочетание алгоритмов: в оперативную память считывается максимальный объем данных, который сортируется эффективным алгоритмом гибридной ускоренной сортировки. Если все данные уместились в памяти, сортировка на этом выполнена. В противном случае применяется алгоритм объединения-сортировки. Отсортированные данные записываются в файл, затем в память считывается следующий чанк данных, сортируется и записывается на диск. Так продолжается, пока не будут прочитаны все данные; затем начинается объединение чанков. При объединении может потребоваться многократное прохождение данных, записанных на диск. При максимальной загрузке узел сортировки может держать на диске две полные копии набора данных: сортированный и несортированный.

Общее время выполнения алгоритма получается порядка  $N \cdot \log(N)$ , где  $N$  - число записей. Сортировка в памяти быстрее, чем объединение с диском, так что время выполнения может быть уменьшено выделением большей памяти для сортировки. Алгоритм выделяет себе долю физической оперативной памяти, которая задается в конфигурации сервер IBM SPSS Modeler опцией *Коэффициент использования памяти*. Чтобы увеличить объем памяти для сортировки, установите физическую оперативную память большего объема или увеличьте это значение. Обратите внимание на то, что когда доля использованной памяти превышает рабочий набор процесса и приходится прибегать к подкачке страниц с диска, производительность падает, так как у алгоритма сортировки в памяти структура обращений к памяти носит произвольный характер и может приводить к большому объему операций подкачки. Алгоритм сортировки используется рядом узлов помимо узла сортировки, и для них действуют те же правила касательно производительности.

**Категоризация.** Узел разделения на интервалы считывает весь набор входных данных, чтобы вычислить границы между интервалами, и только после этого распределяет записи по интервалам. На время вычисления границ набор данных кэшируется; затем он снова просматривается для распределения записей. Если метод категоризации - *постоянная ширина* или *среднее + стандартное отклонение*, то набор данных кэшируется на непосредственно на диск. У этих методов линейное время выполнения и требуется дисковое пространство для хранения всего набора данных. Если метод категоризации - *ранги* или *плитки*, набор данных сортируется при помощи описанного выше алгоритма сортировки, и отсортированные данные используются как кэш. Благодаря сортировке у этих методов время выполнения  $M \cdot N \cdot \log(N)$ , где  $M$  - число категоризируемых полей, а  $N$  - число записей; требуется место на диске, равное удвоенному размеру набора данных.

Производительность последующих прогонов повышается, если на основе сгенерированных интервалов сгенерировать узел вычислений. Операции вычислений намного быстрее, чем категоризация.

**Объединение по ключу (Join).** Узел объединения, когда метод объединения - *ключи* (аналогичен операции JOIN в базе данных), сортирует все входные наборы данных по ключевым полям. У этой части процедуры время выполнения  $M \cdot N \cdot \log(N)$ , где  $M$  - число входных наборов данных и  $N$  - число записей в самом большом входном наборе данных; требуется место на диске для записи всех входных наборов данных плюс второй копии самого большого набора данных. Время выполнения собственно объединения пропорционально размеру выходного набора данных, зависящему от частоты ключей соответствия. В худшем случае выходной набор данных - декартово произведение входных наборов, и тогда время выполнения

приближается к  $MM$ . Это редкий случай; у большинства объединений ключей соответствия намного меньше. Если один из наборов данных велик по сравнению с остальными или входные данные уже отсортированы по ключевому полю, то производительность этого узла можно существенно повысить, используя вкладку Оптимизация.

**Агрегировать.** Если не задана опция *Ключи непрерывны*, это узел считывает, не сохраняя, весь входной набор данных, и только после этого генерирует сводный вывод. В крайнем случае, когда размер сводных данных достигает предельно допустимого (заданного в конфигурации сервер IBM SPSS Modeler опцией *Коэффициент использования памяти*), остаток набора данных сортируется и обрабатывается так, как будто опция *Ключи непрерывны* задана. Когда эта опция задана, данные не сохраняются, поскольку сводные записи генерируются по мере чтения входных данных.

**Особый тип.** Узел особого типа записывает все поля ключей уникальности из входного набора данных; в тех случаях, когда все поля ключевые и все записи уникальны, записывается весь входной набор данных. По умолчанию узел особого типа сортирует данные по ключевым полям и затем отбирает или отбрасывает первую запись особого типа из каждой группы. Для небольших наборов данных с малым числом различных ключей или заранее отсортированных наборов данных можно выбрать опции, повышающие скорость и эффективность обработки.

**Тип.** В некоторых случаях узел типа кэширует входные данные при чтении значений; кэш используется при обработке ниже по потоку. Для кэша требуется дисковое пространство под весь набор данных, зато обработка становится быстрее.

**Оценка.** Узел оценки должен отсортировать входные данные, чтобы вычислить плитки. Сортировка повторяется для каждой оцениваемой модели, поскольку в каждом случае будут свои оценки и, следовательно, свой порядок записей. Время выполнения составляет  $M \cdot N \cdot \log(N)$ , где  $M$  - число моделей и  $N$  - число записей.

---

## Производительность: узлы моделирования

**Нейросеть и алгоритм Коонена.** Алгоритмы обучения нейросети (включая алгоритм Коонена) пропускают через сеть обучающие данные много раз. До некоторого предела данные считываются в оперативную память, сверх этого данные скидываются на диск. Доступ к обучающим данным с диска стоит дорого, поскольку используется произвольный доступ с большим числом дисковых операций. Можно запретить использование дискового хранения для этих алгоритмов, задав принудительное считывание данных в оперативную память, если включить опцию **Оптимизировать скорость** на вкладке модели в диалоговом окне узла. Обратите внимание на то, что если требуемый объем памяти превышает рабочий набор процесса сервера, для части данных будет использована подкачка страниц на диске с соответствующим ущербом для производительности.

Когда включена опция **Оптимизировать память**, алгоритму выделяется процент физической оперативной памяти согласно значению опции конфигурации сервер IBM SPSS Modeler *Предельный процент памяти моделирования*. Чтобы использовать больший объем памяти для обучения нейросетей, либо установите больший объем оперативной памяти, либо увеличьте значение этой опции, но имейте в виду, что слишком высокое значение этой опции вызовет использование дисковых страниц подкачки.

Время выполнения алгоритмов нейросетей зависит от требуемого уровня точности. Временем выполнения можно управлять, задав условие остановки в диалоговом окне узла.

**К-средних.** У алгоритма кластеризации  $k$ -средних те же опции для управления использованием памяти, что и у алгоритмов нейросетей. Однако производительность работы с диском выше, поскольку используется последовательный доступ к данным.

---

## Производительность: выражения CLEM

Функции последовательности CLEM (“функции на @”), которые пользуются предшествующими данными потока, должны сохранять достаточно данных для доступа к данным максимально глубокого прошлого. Для выполнения операций без ограничения использования предшествующих данных для поля должны сохраняться все значения. Операция без ограничения - это такая операция, в которой смещение задается не литеральным целым. Например, @OFFSET(Продажи, Месяц). В качестве значения смещения использовано имя поля *Месяц*, значение которого до выполнения не известно. Сервер должен сохранить все значения поля *Продажи*, чтобы обеспечить искомые результаты. В тех случаях, когда верхняя граница вам известна, задавайте ее как дополнительный аргумент, например, так: @OFFSET(Продажи, Месяц, 12). Такая операция указывает серверу хранить только 12 последних значений поля *Продажи*. Функции последовательности, как с ограничением, так и без, почти всегда замедляют генерирование SQL.



---

## Глава 14. Средства доступности в IBM SPSS Modeler

---

### Обзор средств доступности в IBM SPSS Modeler

IBM SPSS Modeler обеспечивает средства доступности для всех пользователей, а также средства, облегчающие работу специально для пользователей с недостатками зрения и других функций. В этом разделе описываются возможности и методы работы с использованием улучшений доступности, такими как программы чтения экрана и клавиши быстрого вызова.

---

### Типы поддержки средств доступности

Как для пользователей с недостатками зрения, так и для пользователей, работающих через клавиатуру, есть ряд альтернативных методов применения комплекта инструментов исследования данных. Так, можно, не прибегая к мыши, создавать потоки, задавать опции и просматривать вывод. Доступные клавиши быстрого вызова перечислены в приведенных ниже разделах. Кроме того, IBM SPSS Modeler предоставляет обширную поддержку для программ чтения экрана, таких как JAWS для Windows. Можно также оптимизировать цветовую схему, чтобы обеспечить больший контраст. Эти типы поддержки рассматриваются в приведенных ниже разделах.

### Специальные возможности для людей с ограничениями по зрению

В IBM SPSS Modeler можно задать целый ряд параметров для повышения эффективности работы с программным обеспечением.

#### Параметры вывода

Можно выбрать цвета для вывода диаграмм. Кроме того, можно выбрать цветовую схему Windows для самой программы. Возможно, вам удастся улучшить видимый контраст.

1. Чтобы задать параметры дисплея, выберите в меню Инструменты пункт **Опции пользователя**.
2. Щелкните по вкладке **Дисплей**. Опции на этой вкладке включают в себя цветовую схему программного обеспечения, цвета диаграммы и размеры шрифта для узлов.

#### Использование звуков для уведомления

Включая и выключая звуки, вы можете управлять уведомлением о тех или иных операциях программного обеспечения. Например, можно активировать звуки для событий, таких как создание и удаление узла или генерирование нового вывода или моделей.

1. Чтобы задать опции уведомления, выберите в меню Инструменты пункт **Опции пользователя**.
2. Щелкните по вкладке **Уведомления**.

#### Управление автоматическим открытием новых окон

На той же вкладке Уведомления диалогового окна Опции пользователя задается, открывать ли для нового вывода, такого как таблицы и диаграммы, отдельное окно. Возможно, вам будет проще выключить эту опцию и открывать окна вывода по мере необходимости.

1. Чтобы задать эти опции, в меню Инструменты выберите пункт **Опции пользователя**.
2. Щелкните по вкладке **Уведомления**.
3. В диалоговом окне в группе **Визуальные уведомления** выберите в списке **Новый вывод**.
4. Для опции **Открывать окно** выберите **Никогда**.

#### Размер узла

Узлы можно выводить на экран в одном из двух размеров, стандартном или уменьшенном. Вы можете перенастроить эти два размера, как вам нужно.

1. Чтобы задать опции размера узлов, выберите в меню Файл пункт **Свойства потока**.
2. Щелкните по вкладке **Макет**.
3. В списке **Размер значка** выберите **Стандартный**.

## Специальные возможности для слепых пользователей

Поддержка для слепых пользователей зависит в основном от типа используемой программы чтения экрана, например, JAWS для Windows. Есть ряд настроек, которые можно задать для оптимального использования программы чтения экрана совместно с IBM SPSS Modeler.

### Параметры дисплея

Обычно программы чтения экрана лучше справляются с текстом, имеющим высокий видимый контраст. Если вы уже настроили в Windows контрастную цветовую схему, можете использовать ее и для программы чтения экрана.

1. Чтобы задать параметры дисплея, выберите в меню Инструменты пункт **Опции пользователя**.
2. Щелкните по вкладке **Дисплей**.

### Использование звуков для уведомления

Включая и выключая звуки, вы можете управлять уведомлением о тех или иных операциях программного обеспечения. Например, можно активировать звуки для событий, таких как создание и удаление узла или генерирование нового вывода или моделей.

1. Чтобы задать опции уведомления, выберите в меню Инструменты пункт **Опции пользователя**.
2. Щелкните по вкладке **Уведомления**.

### Управление автоматическим открытием новых окон

На той же вкладке Уведомления диалогового окна Опции пользователя задается, открывать ли отдельные окна для нового вывода. Возможно, вам будет проще выключить эту опцию и открывать окна вывода по мере необходимости.

1. Чтобы задать эти опции, в меню Инструменты выберите пункт **Опции пользователя**.
2. Щелкните по вкладке **Уведомления**.
3. В диалоговом окне в группе **Визуальные уведомления** выберите в списке **Новый вывод**.
4. Для опции **Открывать окно** выберите **Никогда**.

## Средства доступности через клавиатуру

Функции продукта доступны с клавиатуры. На самом простом уровне можно активировать меню окна, нажав клавишу Alt плюс нужную клавишу (например, Alt+F, чтобы открыть меню Файл), или перебирать управляющие элементы диалогового окна, нажимая клавишу Tab. У каждого вида окон IBM SPSS Statistics есть свои особенности и полезные возможности для перемещения по диалоговым окнам.

В этом разделе описаны основные возможности действовать при помощи клавиатуры, от открытия потока до использования диалоговых окон узла для работы с выводом. Дополнительно приводятся списки сочетаний клавиш для упрощения навигации в программе.

### Клавиши быстрого вызова для навигации по главному окну

При исследовании данных большая часть работы выполняется в главном окне IBM SPSS Modeler. Основная область называется **холст потока** и служит для построения и запуска потоков данных. В нижней части окна расположены **палитры узлов**, содержащие все доступные узлы. Узлы, которые соответствуют тому или иному типу операций при исследовании данных, собраны в палитры, которые собраны во вкладки.

Например, узлы, которые поставляют данные для IBM SPSS Modeler, собраны на вкладке Источники, а узлы, которые вычисляют, фильтруют или приглашают вводить поля, собраны на вкладке Операции с полями.

В правой части окна находится несколько инструментов для управления потоками, выводом и проектами. Вверху справа расположены **менеджеры**; имеется три вкладки для управления потоками, выводом и сгенерированными моделями. Для доступа к этим объектам нужно выбрать нужную вкладку, а затем нужный объект в списке. В нижней половине справа расположена **панель проектов**, на которой можно организовывать работу по проектам. В этой области есть две вкладки, по которым доступно два разных представления проектов. В **Представлении классов** проекты сортируются по типу, а в **Представлении CRISP-DM** - по стадии исследования данных, такой как Подготовка данных или Моделирование. Эти особенности окна IBM SPSS Modeler рассматриваются в справочной системе и в Руководстве пользователя.

Следующая таблица содержит клавиши быстрого вызова для перемещения в главном окне IBM SPSS Modeler и построения потоков. Клавиши быстрого вызова для диалоговых окон и вывода приведены в дальнейших разделах. Обратите внимание на то, что эти клавиши быстрого вызова доступны только из главного окна.

Таблица 32. Клавиши быстрого вызова в главном окне

Клавиши	Функция
Ctrl+F5	Перемещает фокус на палитру узлов.
Ctrl+F6	Перемещает фокус на холст потока.
Ctrl+F7	Перемещает фокус на панель менеджеров.
Ctrl+F8	Перемещает фокус на панель проектов.

Таблица 33. Клавиши быстрого вызова для узлов и потоков

Клавиши	Функция
Ctrl+N	Создает новый чистый холст потока.
Ctrl+O	Открывает диалоговое окно Открыть, в котором можно выбрать существующий поток.
Ctrl+клавиша с цифрой	Перемещает фокус на вкладку соответствующего номера на панели или в окне. Например, если на панели или в окне есть вкладки, то Ctrl+1 перемещает на первую слева вкладку, Ctrl+2 на вторую и т.д.
Ctrl+Стрелка вниз	На палитре узлов перемещает фокус с выступа вкладки палитры на первый узел на этой вкладке.
Ctrl+Стрелка вверх	На палитре узлов перемещает фокус с узла на выступ его вкладки палитры.
клавиша Enter	Если на палитре узлов выделен узел (в том числе уточненная модель на палитре сгенерированных моделей), эта клавиша добавляет этот узел на холст потока. Если узел выделен уже холсте, то нажатие клавиши <b>Enter</b> открывает диалоговое окно для этого узла.
Ctrl+Enter	Если на палитре узлов выделен узел, добавляет этот узел на холст потока, не выделяя его, и перемещает фокус на первый узел на палитре.
Alt+Enter	Если на палитре узлов выделен узел, добавляет этот узел на холст потока и выделяет его, а фокус перемещает на первый узел на палитре.
Shift+Пробел	Если фокус находится на палитре на узле или на комментарии, то включает или выключает выделение этого узла или комментария.  Если одновременно выделены другие узлы или комментарии, с них выделение также снимется.

Таблица 33. Клавиши быстрого вызова для узлов и потоков (продолжение)

Клавиши	Функция
Ctrl+Shift+Пробел	Если фокус находится на узле или на комментарии в потоке или на палитре, то включает или выключает выделение этого узла или комментария.  При этом существующее выделение других узлов или комментариев не меняется.
Стрелка влево или вправо	Если фокус находится на холсте потока, то перемещает весь поток по экрану в горизонтальном направлении. Если фокус находится на вкладке палитры, то перебирает вкладки. Если фокус находится на узле палитры, то перебирает узлы на палитре.
Стрелка вверх или вниз	Если фокус находится на холсте потока, то перемещает весь поток по экрану в вертикальном направлении. Если фокус находится на узле палитры, то перебирает узлы на палитре. Если фокус находится на подпалитре, то перебирает подпалитры на данной вкладке палитры.
Alt+Стрелка влево или вправо	Перемещает выделенные узлы и комментарии по холсту потока в горизонтальном направлении согласно направлению стрелки на клавише.
Alt+Стрелка вверх или вниз	Перемещает выделенные узлы и комментарии по холсту потока в вертикальном направлении согласно направлению стрелки на клавише.
Ctrl+A	Выделяет все узлы в потоке.
Ctrl+Q	Если фокус находится на узле, выделяет этот узел и все узлы, расположенные ниже; снимает выделение с узлов, расположенных выше.
Ctrl+W	Если фокус находится на выделенном узле, то снимает выделение с этого узла и со всех выделенных узлов, расположенных ниже.
Ctrl+Alt+D	Дублирует выделенный узел.
Ctrl+Alt+L	Если в потоке выделен слепок модели, открывает диалоговое окно Вставка, в котором можно загрузить сохраненную модель из файла .pod в поток.
Ctrl+Alt+R	Для выделенного узла выводит на экран вкладку Аннотации, на которой можно переименовать узел.
Ctrl+Alt+U	Создает узел Ввод пользователя.
Ctrl+Alt+C	Включает или выключает кэш для узла.
Ctrl+Alt+F	Очищает кэш для узла.
клавиша Tab	На холсте потоков циклически перебирает все узлы источника и комментарии в текущем потоке. На палитре узлов перебирает ее узлы. На выделенной подпалитре перемещается на ее первый узел.
Shift+Клавиша Tab	Выполняет ту же операцию, что клавиша Tab, но в обратном порядке.
Ctrl+Tab	Если фокус находится на панели менеджеров или на панели проектов, перемещает фокус на холст потока. Если фокус находится на палитре узлов, перемещает фокус с узла на выступ вкладки или обратно.
Любая буквенная клавиша	Если фокус находится на узле в текущем потоке, перемещает фокус на следующий узел, имя которого начинается с нажатой клавиши.
F1	Открывает справочную систему на разделе, соответствующем положению фокуса.
F2	Начинает соединение с узла, выделенного на холсте. Чтобы завершить соединение, дойдите до нужного узла на холсте клавишей Tab и нажмите Shift+Пробел.
F3	Удаляет все соединения для выделенного узла на холсте.
F6	Перемещает фокус между панелью менеджеров, панелью проектов и палитрами узлов.



Таблица 33. Клавиши быстрого вызова для узлов и потоков (продолжение)

Клавиши	Функция
F10	Открывает меню Файл.
Shift+F10	Открывает всплывающее меню для узла или потока.
Delete	Удаляет выделенный узел с холста.
Esc	Закрывает всплывающее меню или диалоговое окно.
Ctrl+Alt+X	Раскрывает надузел.
Ctrl+Alt+Z	Увеличивает представление надузла.
Ctrl+Alt+Shift+Z	Уменьшает представление надузла.
Ctrl+E	Когда фокус находится на холсте потока, запускает текущий поток.

Кроме того, в IBM SPSS Modeler используется ряд стандартных сочетаний клавиш, таких как Ctrl+C для копирования. Дополнительную информацию смотрите в разделе “Использование клавиш быстрого вызова” на стр. 22.

### Клавиши быстрого вызова для диалоговых окон и таблиц

Ряд клавиш быстрого вызова и клавиш программы чтения экрана будут полезны при работе с диалоговыми окнами, таблицами и таблицами в диалоговых окнах. Ниже приведен полный список специальных клавиатурных сокращений и клавиш быстрого вызова программы чтения экрана.

Таблица 34. Клавиши быстрого вызова в диалоговых окнах и в построителе выражений

Клавиши	Функция
Alt+4	Служит для отказа от всех открытых диалоговых окон или окон вывода. Выводимую информацию можно получить со вкладки Вывод на панели менеджера.
Ctrl+End	Если фокус находится на любом элементе управления в построителе выражений, точка ввода перейдет в конец выражения.
Ctrl+1	В построителе выражений перемещает фокус на элемент управления редактирования выражения.
Ctrl+2	В построителе выражений перемещает фокус в список функций.
Ctrl+3	В построителе выражений перемещает фокус в список полей.

#### Клавиши быстрого вызова для таблиц

Клавиши быстрого вызова для таблиц работают для таблиц вывода и для табличных элементов управления в диалоговых окнах таких узлов, как Тип, Фильтр и Объединение. Обычно для перемещения между ячейками таблиц служит клавиша Tab, а для выхода из табличного элемента управления служит Ctrl+Tab. *Примечание:* Иногда программа чтения экрана не начинает читать содержимое ячейки сразу после перехода в эту ячейку. Чтобы перезапустить процесс чтения, нажмите один или два раза на клавишу со стрелкой.

Таблица 35. Клавиши быстрого вызова для таблиц

Клавиши	Функция
Ctrl+W	В таблице читает краткое описание выбранной строки (от английского слова go <b>W</b> ). Например: "Выбрана строка 2, значения - пол, флаг, м/ж..."
Ctrl+Alt+W	В таблице читает полное описание выбранной строки (от английского слова go <b>W</b> ). Например: "Выбрана строка 2, значения такие. Поле = пол, тип = флаг, пол = м/ж..."

Таблица 35. Клавиши быстрого вызова для таблиц (продолжение)

Клавиши	Функция
Ctrl+D	В таблице читает краткое описание выбранной области (от английского слова <b>Description</b> ). Например: "Выбран фрагмент одна строка на шесть столбцов."
Ctrl+Alt+D	В таблице читает полное описание выбранной области (от английского слова <b>Description</b> ). Например: "Выбран фрагмент одна строка на шесть столбцов. Выбранные столбцы - Поле, Тип, Пропущенные значения. Выбранная строка - 1."
Ctrl+T	В таблицах озвучивает краткое описание выбранных столбцов. Например: "Поля, Тип, Пропущенные значения."
Ctrl+Alt+T	В таблицах озвучивает полное описание выбранных столбцов. Например: "Выбранные столбцы - это Поля, Тип, Пропущенные значения."
Ctrl+R	В таблицах озвучивает число записей в таблице (от английского слова <b>Records</b> ).
Ctrl+Alt+R	В таблицах озвучивает число записей в таблице и имена столбцов (от английского слова <b>Records</b> ).
Ctrl+I	Если фокус на ячейке таблицы, читает информацию о ячейке или содержимое ячейки (от английского слова <b>Information</b> ).
Ctrl+Alt+I	Если фокус на ячейке таблицы, читает полное описание информации о ячейке - имя столбца и содержимое ячейки (от английского слова <b>Information</b> ).
Ctrl+G	В таблицах озвучивает краткую общую информацию о выбранном (от английского слова <b>General</b> ).
Ctrl+Alt+G	В таблицах озвучивает полную общую информацию о выбранном (от английского слова <b>General</b> ).
Ctrl+Q	В таблицах переключает режим краткого и полного описания ячеек таблицы (от английского слова <b>Quick</b> ). Нажмите Ctrl+Q, чтобы при навигации по ячейкам при помощи клавиш со стрелками зачитывались полные описания, например, "Пол=Женский". Нажмите Ctrl+Q еще раз, чтоб переключиться на краткие описания (только содержимое ячеек).

## Клавиши быстрого вызова для комментариев

При работе с экранными комментариями можно использовать следующие клавиши быстрого вызова.

Таблица 36. Клавиши быстрого вызова для комментариев

Клавиши	Функция
Alt+C	Показать/скрыть комментарии.
Alt+M	Вставляет новый комментарий, если в настоящее время вывод комментариев включен; если комментарии скрыты, показывает их.
клавиша Tab	На холсте потоков циклически перебирает все узлы источника и комментарии в текущем потоке.
клавиша Enter	Когда фокус находится на комментарии, начинает редактирование.
Alt+Enter или Ctrl+Tab	Завершает редактирование с сохранением внесенных изменений.
Esc	Отменяет редактирование. Внесенные во время редактирования изменения теряются.
Alt+Shift+стрелка вверх	Уменьшает высоту текстовой области на одну ячейку таблицы или на один пиксел, если привязка к сетке соответственно включена или выключена.

Таблица 36. Клавиши быстрого вызова для комментариев (продолжение)

Клавиши	Функция
Alt+Shift+стрелка вниз	Увеличивает высоту текстовой области на одну ячейку таблицы или на один пиксел, если привязка к сетке соответственно включена или выключена.
Alt+Shift+стрелка влево	Уменьшает ширину текстовой области на одну ячейку таблицы или на один пиксел, если привязка к сетке соответственно включена или выключена.
Alt+Shift+стрелка вправо	Увеличивает ширину текстовой области на одну ячейку таблицы или на один пиксел, если привязка к сетке соответственно включена или выключена.

## Клавиши быстрого вызова для средства просмотра кластеров и средства просмотра моделей

Доступны клавиши и сочетания клавиш для навигации по окнам средства просмотра кластеров и средства просмотра моделей

Таблица 37. Общие клавиши быстрого вызова - средство просмотра кластеров и средство просмотра моделей

Клавиши	Функция
Клавиша Tab	Перемещает фокус на следующий элемент управления экрана.
Shift+Клавиша Tab	Перемещает фокус на предыдущий элемент управления экрана.
Стрелка вниз	Если фокус находится на выпадающем списке, открывает список или перемещается на следующий пункт списка.  Если фокус находится на меню, перемещается на следующий пункт меню.  Если фокус находится на мини-изображении, перемещается на следующее мини-изображение в наборе (или на первое, если в настоящий момент фокус находится на последнем).
Стрелка вверх	Если открыт выпадающий список, перемещается на предыдущий пункт списка.  Если фокус находится на меню, перемещается на предыдущий пункт меню.  Если фокус находится на мини-изображении, перемещается на предыдущее мини-изображение в наборе (или на последнее, если в настоящий момент фокус находится на первом).
Enter	Закрывает открытый выпадающий список или выбирает текущий пункт в открытом меню.
F6	Переключает фокус между левой и правой панелями окна.
Стрелки влево и вправо	Если фокус находится на вкладке, перемещается на предыдущую или следующую вкладку.  Если фокус находится на меню, перемещается на предыдущее или следующее меню.
Alt+буква	Нажимает кнопку или меню, в имени которых подчеркнута эта буква.
Esc	Закрывает открытое меню или выпадающий список.

Только в средстве просмотра кластеров

Средство просмотра кластеров содержит представление Кластеры с таблицей кластеров по показателям.

Чтобы выбрать представление Кластеры вместо представления Сводка по модели:

1. Нажимайте клавишу Tab, пока не будет выбрана кнопка **Вид**.
2. Нажмите стрелку вниз два раза, чтобы выбрать **Кластеры**.  
Отсюда можно выбрать отдельную ячейку таблицы:
3. Нажимайте клавишу Tab, пока не дойдете до последнего значка на панели инструментов визуализации.



Рисунок 17. Значок Показать дерево визуализации

4. Нажмите клавишу Tab еще раз, затем клавишу Пробел, затем одну из клавиш со стрелками.

Теперь доступны следующие клавиши быстрого вызова:

Таблица 38. Клавиши быстрого вызова для средства просмотра кластеров

Клавиши	Функция
Клавиша со стрелкой	Перемещает фокус между отдельными ячейками таблицы. Когда фокус переходит в другую ячейку, распределение ячеек на правой панели изменяется.
Ctrl+, (запятая)	Выполняет или отменяет выбор столбца, на одной из ячеек которого находится фокус.  Чтобы добавить столбец к набору выбранных столбцов, переместитесь клавишами со стрелками в ячейку в нужном столбце и нажмите Ctrl+, еще раз.
Табуляция	Уводит фокус из таблицы на следующий элемент управления экрана.
Shift+Клавиша табулятора	Возвращает фокус из таблицы на предыдущий элемент управления экрана.
F2	Входит в режим редактирования (только в ячейках меток и описаний).
Enter	Сохраняет внесенные изменения и выходит из режима редактирования (только в ячейках меток и описаний).
Esc	Выходит из режима редактирования без сохранения изменений (только в ячейках меток и описаний).

## Пример использования клавиш быстрого вызова: построение потоков

Для лучшего понимания, как может выполнить построение потока пользователь, работающий только с клавиатурой или пользователь, работающий с программой чтения экрана, ниже приведен пример построения потока без использования мыши. В этом примере вы создадите поток, состоящий из узла файла переменных, узла вычислений и узла гистограмм. Выполните следующие действия:

1. **Запустите IBM SPSS Modeler.** При запуске IBM SPSS Modeler фокус находится на вкладке Избранное палитры узлов.
2. **Ctrl+Стрелка вниз.** Перемещает фокус с выступа вкладки на тело вкладки.
3. **Стрелка вправо.** Перемещает фокус на узел файла переменных.
4. **Клавиша Пробел.** Выбирает узел файла переменных.
5. **Ctrl+Enter.** Добавляет узел файла переменных на холст потока. Это сочетание клавиш оставляет выделение на узле файла переменных, так что следующий узел будет соединен с этим узлом.
6. **Клавиша Tab.** Возвращает фокус на палитру узлов.
7. **Стрелка вправо 4 раза.** Перемещает на узел вычислений.

8. **Клавиша Пробел.** Выбирает узел вычислений.
9. **Alt+Enter.** Добавляет узел вычислений на холст и переносит выделение на узел вычислений. Теперь этот узел готов к соединению со следующим добавленным узлом.
10. **Клавиша Tab.** Возвращает фокус на палитру узлов.
11. **Стрелка вправо 5 раз.** Перемещает фокус на узел гистограмм на палитре.
12. **Клавиша Пробел.** Выбирает узел гистограмм.
13. **Клавиша Enter.** Добавляет узел в поток и перемещает фокус на холст потока.

Вы можете перейти к следующему примеру сейчас или сохранить этот пример и продолжить работу с ним в другое время.

### Пример использования сочетаний клавиш: редактирование узлов

В этом примере вы будете использовать поток, созданный в более раннем примере. Поток состоит из узла файла переменных, узла вычислений и узла гистограмм. Вначале фокус на третьем узле потока, узле гистограмм.

1. **Ctrl+стрелка влево 2 раза.** Возвращает фокус на узел файла переменных.
2. **Вход.** Открывает диалоговое окно Файл переменных. Клавишей Tab переместитесь на поле Файл и введите имя и путь нужного текстового файла. Нажмите Ctrl+Tab, чтобы переместиться в нижнюю часть диалогового окна, клавишей Tab дойдите до кнопки ОК и нажмите клавишу Enter, чтобы закрыть диалоговое окно.
3. **Ctrl+Стрелка вправо** Передает фокус второму узлу, узлу вычислений.
4. **Вход.** Открывает диалоговое окно Узел вычислений. Клавишей Tab переместитесь на нужные поля и задайте условия вычислений. Нажмите Ctrl+Tab, чтобы переместиться на кнопку ОК, и клавишу Enter, чтобы закрыть диалоговое окно.
5. **Ctrl+Стрелка вправо** Передает фокус третьему узлу, узлу гистограмм.
6. **Вход.** Открывает диалоговое окно Узел гистограмм. Клавишей Tab переместитесь на нужные поля и задайте опции диаграмм. В выпадающих списках для открытия списка и выделения пункта нажимайте стрелку вниз, для выбора пункта списка клавишу Enter. Клавишей Tab переместитесь на кнопку ОК и нажмите клавишу Enter, чтобы закрыть диалоговое окно.

В этой точке можно добавить дополнительные узлы или выполнить текущий поток. При создании потоков следует иметь в виду следующие советы:

- При соединении узлов вручную используйте клавишу F2, чтобы создать отправную точку соединения, клавишу Tab, чтобы перейти на конечную точку, и Shift+Пробел, чтобы завершить создание соединения.
- Чтобы разорвать все соединения для выбранного узла на холсте, нажмите клавишу F3.
- После создания потока можете нажать сочетание Ctrl+E, выполняющее текущий поток.

Доступен полный список клавиш быстрого вызова. Дополнительную информацию смотрите в разделе “Клавиши быстрого вызова для навигации по главному окну” на стр. 160.

## Использование программы чтения экрана

На рынке доступен ряд программ чтения экрана. IBM SPSS Modeler конфигурируется для поддержки JAWS для Windows с использованием соединителя Java Access Bridge, который устанавливается совместно с IBM SPSS Modeler. Если JAWS установлена, то просто запускайте ее перед запуском IBM SPSS Modeler, и вы будете пользоваться этим продуктом через эту программу чтения экрана.

Уникальное наглядное представление исследования данных в IBM SPSS Modeler через диаграммы и графы рассчитано прежде всего на визуальное восприятие. Однако доступно и текстовое представление вывода и моделей, при помощи которого можно изучать результаты и принимать решения через программу чтения экрана.

*Примечание:* На 64-битных клиентских компьютерах некоторые компоненты технологий для людей с ограниченными возможностями не работают. Причина в том, что соединитель Java Access Bridge не рассчитан на работу в 64-битной среде.

## Использование файла словаря IBM SPSS Modeler

Совместно с JAWS можно использовать файл словаря IBM SPSS Modeler (*Awt.JDF*). Чтобы использовать этот файл:

1. Перейдите в подкаталог */accessibility* каталога установки IBM SPSS Modeler и скопируйте файл словаря (*Awt.JDF*).
2. Вставьте его в каталог с вашими сценариями JAWS.

Если у вас уже работают другие прикладные программы JAVA, то, возможно, файл с именем *Awt.JDF* уже существует на компьютере. В этом случае вам, возможно, не удастся использовать этот файл словаря без ручного редактирования файла словаря.

## Использование программы чтения экрана при работе с выводом в формате HTML

Возможны трудности при просмотре вывода в формате HTML в IBM SPSS Modeler с использованием программы чтения экрана. Это касается ряда типов вывода, включая:

- Вывод, просматриваемый на вкладке Дополнительно для узлов регрессии, логистической регрессии и Фактор/PCA
- Вывод узла отчетов

В каждом из этих окон или диалоговых окон есть инструмент на панели инструментов, служащий для запуска вывода в браузер по умолчанию, предоставляющем стандартную поддержку программам чтения экрана. В этом случае можно изучать выходную информацию с использованием программы чтения экрана.

## Средства доступности в окне интерактивного дерева

Стандартное представление модели дерева решений в окне интерактивного дерева может вызвать трудности для программ чтения экрана. Чтобы вызвать версию повышенной доступности, в меню интерактивных деревьев выбирайте следующее:

### Вид > Окно повышенной доступности

Будет выведено представление, аналогичное стандартной карте дерева, но правильно читаемое программой JAWS. Для перемещения вверх, вниз, вправо и влево служат обычные клавиши со стрелками. При навигации по окну повышенной доступности фокус движется по интерактивному дереву вместе с прокруткой дерева. Клавиша Пробел заменяет текущее выделение, сочетание клавиш Ctrl+Пробел расширяет текущее выделение.

---

## Советы по использованию

Есть несколько рекомендаций, как сделать среду IBM SPSS Modeler более доступной. Вот некоторые общие соображения при работе в IBM SPSS Modeler.

- **Выход из расширенных текстовых полей.** Для выхода из расширенных текстовых полей нажмите Ctrl+Tab. *Примечание:* Сочетание клавиш Ctrl+Tab также служит для выхода из табличных элементов управления.
- **Использование клавиши Tab вместо клавиш со стрелками.** При выборе опций в диалоговом окне для перемещения между переключателями используйте клавишу Tab. Клавиши со стрелками в этом контексте не работают.
- **Выпадающие списки.** В выпадающем списке в диалоговых окнах можно нажать клавишу Escape или Пробел, чтобы выбрать пункт и закрыть список. Кроме того, клавиша Escape закрывает выпадающий список, оставшийся открытым после перехода на другой элемент управления при помощи клавиши Tab.

- **Состояние выполнения.** Иногда при выполнении потока на большой базе данных JAWS сообщает о состоянии потока с задержкой. Чтобы почаще получать последний отчет о состоянии, нажимайте клавишу Ctrl.
- **Использование палитры узлов.** Иногда после перехода на вкладку на палитрах узлов JAWS начинает зачитывать группу полей вместо имени узла. В этом случае можно нажать Ctrl+Стрелка вправо, затем Ctrl+Стрелка влево, чтобы сбросить программу чтения экрана и услышать имя узла.
- **Чтение меню.** Иногда после открытия меню JAWS начинает чтение не с первого пункта. Если вы подозреваете, что это произошло, нажмите стрелку вниз, затем стрелку вверх, чтобы услышать первый пункт меню.
- **Каскадные меню.** JAWS не зачитывает первый уровень каскадного меню. Если при движении по меню вы слышите паузу, нажмите стрелку вправо, чтобы услышать пункты дочернего меню.

Кроме того, если у вас установлен продукт IBM SPSS Modeler Text Analytics, следующие рекомендации помогут вам сделать более доступным интерфейс интерактивного рабочего стола.

- **Вход в диалоговые окна.** Иногда после входа в диалоговое окно нужно нажать клавишу Tab, чтобы переместить фокус на первый элемент управления.
- **Выход из расширенных текстовых полей.** Для выхода из расширенных текстовых полей и перехода к на следующий элемент управления нажмите Ctrl+Tab. *Примечание:* Сочетание клавиш Ctrl+Tab также служит для выхода из табличных элементов управления.
- **Нажатие первой буквы для перехода на элемент в древесном списке.** При поиске элемента на панели категорий, панели извлеченных результатов или в дереве библиотек можно нажать первую букву искомого элемента, когда фокус находится на нужной панели. Будет выбрано следующее вхождение элемента, который начинается с нажатой буквы.
- **Выпадающие списки.** В выпадающем списке в диалоговых окнах можно нажать клавишу Пробел, чтобы выбрать пункт и закрыть список.

Дополнительные советы разбираются в следующих темах.

## Конфликты с другими программами

При тестировании IBM SPSS Modeler с программами чтения экрана, такими как JAWS, наши разработчики обнаружили, что использование сервера управления системой (Systems Management Server, SMS) в организации может мешать способности JAWS читать прикладные программы на основе Java, такие как IBM SPSS Modeler. Эта ситуация устраняется, если выключить SMS. Чтобы получить дополнительную информацию об SMS, посетите сайт Microsoft.

## JAWS и Java

Различные версии JAWS обеспечивают разные уровни поддержки для прикладных программ на основе Java. Хотя IBM SPSS Modeler будет работать со всеми современными версиями JAWS, у некоторых из них могут быть те или иные затруднения при использовании с системами на основе Java. Посетите сайт JAWS for Windows по адресу <http://www.FreedomScientific.com>.

## Использование диаграмм в IBM SPSS Modeler

Трудно при помощи программы чтения экрана воспринимать информацию, представленную визуально, например, на гистограммах, диаграммах оценок, мультиграфиках и диаграммах рассеяния. Однако заметим, что Web-диаграммы и распределения можно изучать по текстовой сводке, доступной из окна вывода.





---

## Глава 15. Поддержка Unicode

---

### Поддержка Unicode в IBM SPSS Modeler

IBM SPSS Modeler полностью поддерживает Unicode как для IBM SPSS Modeler, так и для сервер IBM SPSS Modeler. Благодаря этому возможен обмен данными с другими прикладными программами, поддерживающими Unicode, включая многоязычные базы данных, без потери информации, которая случается при преобразовании в национальную схему кодировки и обратно.

- IBM SPSS Modeler хранит данные Unicode во внутреннем формате Unicode и может без потерь читать и записывать многоязычные данные, которые хранятся как Unicode в базах данных.
- IBM SPSS Modeler может читать и записывать текстовые файлы в кодировке UTF-8. По умолчанию при импорте и экспорте текста предполагается кодировка национальной версии, но поддерживается и вариант UTF-8. Можно указать такую настройку в узлах импорта и экспорта файлов, а можно изменить кодировку по умолчанию в диалоговом окне свойств потока. Дополнительную информацию смотрите в разделе “Задание общих опций потоков” на стр. 42.
- Statistics, SAS и текстовые файлы данных, сохраненные в кодировке национальной версии, будут преобразованы в UTF-8 при импорте и обратно при экспорте. Если при записи в файл обнаруживаются символы Unicode, которые отсутствуют в наборе символов национальной версии, то вместо них подставляются заменяющие символы и выводится предупреждение. Такой ситуации можно ожидать только после импорта данных из источника с поддержкой Unicode (база данных или текстовый файл UTF-8), содержащего символы из другой национальной версии или из многонациональной версии или кодовых наборов.
- Образы IBM SPSS Modeler Solution Publisher имеют кодировку UTF-8 и обладают полной переносимостью между платформами и национальными версиями.

#### О Unicode

Цель стандарта Unicode - предложить единый способ кодировать многоязычные тексты для удобного совместного использования в разных странах, национальных версиях и прикладных программах. В стандарте Unicode, который в настоящее время имеет версию 4.0.1, определяется кодовый набор, включающий в себя как поднаборы все широко употребительные в современном мире кодовые наборы; каждому символу назначается уникальное имя и код. Символы и коды такие же, как в наборе Universal Character Set (UCS), определенном в стандарте ISO-10646. Дополнительные сведения смотрите на домашней странице Unicode.



---

## Уведомления

Эта информация относится к продуктам и сервису, предлагаемым по всему миру.

IBM может не предоставлять в других странах продукты, услуги и аппаратные средства, описанные в данном документе. За информацией о продуктах и услугах, предоставляемых в вашей стране, обращайтесь к местному представителю IBM. Ссылки на продукты, программы или услуги IBM не означают и не предполагают, что можно использовать только указанные продукты, программы или услуги IBM. Разрешается использовать любые функционально эквивалентные продукты, программы или услуги, если при этом не нарушаются права IBM на интеллектуальную собственность. Однако ответственность за оценку и проверку работы любого продукта, программы или сервиса, не произведенного корпорацией IBM, лежит на пользователе.

IBM может располагать патентами или рассматриваемыми заявками на патенты, относящимися к предмету данного документа. Предъявление данного документа не предоставляет какую-либо лицензию на эти патенты. Вы можете послать письменный запрос о лицензии по адресу:

IBM Director of Licensing  
IBM Corporation  
North Castle Drive  
Armonk, NY 10504-1785  
U.S.A.

По поводу лицензий, связанных с использованием наборов двухбайтных символов (DBCS), обращайтесь в отдел интеллектуальной собственности IBM в вашей стране или направьте запрос в письменной форме по адресу:

Intellectual Property Licensing  
Legal and Intellectual Property Law  
IBM Japan Ltd.  
1623-14, Shimotsuruma, Yamato-shi  
Kanagawa 242-8502 Japan

Следующий абзац не применяется в Великобритании или в любой другой стране, где подобные заявления противоречат местным законам: INTERNATIONAL BUSINESS MACHINES CORPORATION ПРЕДСТАВЛЯЕТ ДАННУЮ ПУБЛИКАЦИЮ "КАК ЕСТЬ", БЕЗ КАКИХ-ЛИБО ГАРАНТИЙ, КАК ЯВНЫХ, ТАК И ПОДРАЗУМЕВАЕМЫХ, ВКЛЮЧАЯ, НО НЕ ОГРАНИЧИВАЯСЬ ТАКОВЫМИ, ПРЕДПОЛАГАЕМЫЕ ГАРАНТИИ СОБЛЮДЕНИЯ ЧЬИХ-ЛИБО АВТОРСКИХ ПРАВ, ВОЗМОЖНОСТИ КОММЕРЧЕСКОГО ИСПОЛЬЗОВАНИЯ ИЛИ ПРИГОДНОСТИ ДЛЯ КАКИХ-ЛИБО ЦЕЛЕЙ И СООТВЕТСТВИЯ ОПРЕДЕЛЕННОЙ ЦЕЛИ. В некоторых штатах при определенных соглашениях не допускается отказ от выраженных или подразумеваемых гарантий, поэтому данное заявление может к вам не относиться.

Эта информация может содержать технические неточности и типографские ошибки. В представленную здесь информацию периодически вносятся изменения; эти изменения будут включаться в новые издания данной публикации. Фирма IBM может в любое время без уведомления вносить изменения и усовершенствования в продукты и программы, описанные в этой публикации.

Любые приводимые здесь ссылки на web-сайты, не относящиеся к компании IBM, даются исключительно для удобства и ни в коей мере не служат целям поддержки или рекламы этих web-сайтов. Материалы этих Web-сайтов не являются частью данного продукта IBM, и вы можете использовать их только на собственную ответственность.

Любую предоставленную вами информацию IBM может использовать или распространять любым способом, какой сочтет нужным, не беря на себя никаких обязательств по отношению к вам.

Если обладателю лицензии на данную программу понадобятся сведения о возможности: (i) обмена данными между независимо разработанными программами и другими программами (включая данную) и (ii) совместного использования таких данных, он может обратиться по адресу:

IBM Software Group  
ATTN: Licensing  
200 W. Madison St.  
Chicago, IL; 60606  
U.S.A.

Такая информация может быть доступна при соответствующих условиях и соглашениях, включая в некоторых случаях взимание платы.

Описанную в данном документе лицензионную программу и все прилагаемые к ней лицензированные материалы IBM предоставляет на основе положений Соглашения между IBM и Заказчиком, Международного Соглашения о Лицензиях на Программы IBM или любого эквивалентного соглашения между IBM и заказчиком.

Любые данные о выполнении, содержащиеся здесь, были определены в контролируемой среде. Поэтому результаты, полученные в других операционных средах, могут существенно отличаться. Некоторые измерения могли быть сделаны на системах в стадии разработки, и поэтому нет гарантии, что соответствующие показатели останутся теми же на общедоступных системах. Более того, некоторые показатели могли быть оценены путем экстраполяции. Реальные результаты могут отличаться. Пользователи этого документа должны проверить приводимые данные в их конкретной среде.

Информация о продуктах, не принадлежащих компании IBM, была получена от поставщиков этих продуктов, из их опубликованных сообщений или других общедоступных источников. Компания IBM не тестировала эти продукты и не может подтвердить правильность их работы, совместимость и другие утверждения, касающиеся продуктов, не принадлежащих компании IBM. Вопросы о возможностях этих продуктов следует направлять их поставщикам.

Все заявления, касающиеся будущих направлений деятельности или намерений корпорации IBM, подвержены изменению или отмене без предупреждения и являются не более чем выражением целей или намерений.

Эти сведения содержат примеры данных и отчетов, используемых в повседневных деловых операциях. Чтобы проиллюстрировать их настолько полно, насколько это возможно, данные примеры включают имена индивидуумов, названия компаний, брендов и продуктов. Все эти имена и названия являются вымышленными, и любое совпадение с названиями и адресами, используемыми реально действующими компаниями, является чисто случайными.

При просмотре данного электронного информационного документа фотографии и цветные иллюстрации могут не показываться.

---

## Товарные знаки

IBM, логотип IBM, и [ibm.com](http://ibm.com) являются товарными знаками или зарегистрированными товарными знаками компании International Business Machines Corp., зарегистрированными во многих странах мира. Прочие наименования продуктов и услуг могут быть товарными знаками, принадлежащими IBM или другим компаниям. Текущий список товарных знаков IBM можно найти в Интернете “Copyright and trademark information” по адресу: [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Intel, логотип Intel, Intel Inside, логотип Intel Inside, Intel Centrino, логотип Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium и Pentium являются товарными знаками или зарегистрированными товарными знаками компании Intel или ее дочерних компаний в Соединенных Штатах и других странах.

Linux является зарегистрированным товарным знаком Linus Torvalds в Соединенных Штатах и других странах.

Microsoft, Windows, Windows NT и логотип Windows являются товарными знаками корпорации Microsoft в Соединенных Штатах и других странах.

UNIX является зарегистрированным товарным знаком The Open Group в Соединенных Штатах и других странах.

Java и все основанные на Java товарные знаки и логотипы - товарные знаки или зарегистрированные товарные знаки Oracle и/или его филиалов.

Другие названия продуктов и услуг могут являться товарными знаками IBM или других компаний.



# Индекс

## Спец. символы

-password  
сервер IBM SPSS Modeler 12

## С

CLEM 77  
введение 24, 69  
выражения 71, 83  
построение выражений 77  
примеры 69  
проверка выражений 79  
типы данных 83, 84, 85  
функции 77  
язык 83  
COP 13  
count\_less\_than function 75  
count\_substring function 97  
CRISP-DM 19, 137  
представление проектов 137

## D

DTD 134

## F

f-распределение  
вероятностные функции 95  
factor 168  
first\_index function 76

## I

IBM InfoSphere Warehouse (ISW)  
экспорт PMML 134  
IBM SPSS Collaboration and Deployment  
Services 116  
IBM SPSS Modeler 1, 15  
выполнение из командной строки 11  
документация 3  
Начинаем работу 11  
обзор 11, 145  
параметры 145  
советы и клавиши быстрого  
вызова 63  
специальные возможности 159  
IBM SPSS Modeler Advantage 116, 127  
ID пользователя  
сервер IBM SPSS Modeler 12  
integer\_bitcount function 96

## J

Java 169  
JAWS 159, 167, 169

## L

last\_non\_null function 76

## M

min\_index function 76  
model nuggets 54

## P

PMML  
импорт моделей 134  
опции экспорта 149  
экспорт моделей 134

## S

substring\_between function 97

## T

t-распределение  
вероятностные функции 95  
ToolTips  
аннотирование узлов 58  
trim\_start function 97

## A

автоматизация 69  
анализ на основе дерева  
типичные применения 25  
анализ претендента чемпиона 116, 128  
аннотации  
папка 140  
преобразование в комментарии 58  
проект 140  
аннотирование  
потoki 54, 58  
узлы 54, 58  
атрибут 25

## Б

база данных  
функции 77  
блокировка объектов репозитория locking  
IBM SPSS Collaboration and Deployment  
Services 124  
блокировка узлов 41

## В

введение 83  
IBM SPSS Modeler 11, 145  
вероятностные функции 95

ветви, моделирование и оценки 54, 130,  
132  
визуальное программирование 15  
включить узлы 37  
внедрение 116  
внедрение сценариев 128  
во время выполнения, приглашения 49  
вставка 19  
вход в сервер IBM SPSS Modeler 12  
выбор узла на палитре 151  
вывод 17  
вывод в формате HTML  
программа чтения экрана 168  
вывод всех комментариев для потока 57  
выключить узлы 37, 38  
выполнение потоков 53  
выражения 83  
выражения CLEM  
производительность 157  
вырезать 19  
выходные узлы 35

## Г

генерирование SQL  
запись в журнал 45  
предварительный просмотр 45  
главное окно 15  
глобальные переменные  
в выражениях CLEM 78  
глобальные функции 111  
горячие клавиши 22  
градусы  
единицы измерения 43

## Д

данные  
предварительный просмотр 40  
данные с шумом 26  
даты  
преобразование 106  
управление 106  
действительные числа 83, 84  
дерева решений  
специальные возможности 168  
десятичные знаки  
форматы вывода 43  
десятичный разделитель  
форматы вывода чисел 42  
диаграммы  
добавление к проектам 138  
сохранение вывода 60  
диалоговое окно запуска 148  
диалоговое окно приветствия 148  
дни перехода через нуль 43  
добавление  
в проект 138  
добавление соединений с сервером IBM  
SPSS Modeler 13

документация 3

## Е

единая регистрация 12  
единая регистрация в репозитории IBM  
SPSS Collaboration and Deployment  
Services 115, 117

## З

загрузка  
состояния 61  
узлы 61  
задание имен для узлов и потоков 58  
замена моделей 146  
записи 25  
значения отсутствия 66  
запятая 42  
значения 71  
добавление в выражения CLEM 78  
просмотр из аудита данных 78  
значения даты и времени 74  
значения из одних пробелов 66, 73  
значения отсутствия 66, 73  
значки  
задание опций 22, 46

## И

изменение размеров 21  
имена потоков 58  
имена узлов 58  
импорт  
PMML 134  
имя домена (Windows)  
сервер IBM SPSS Modeler 12  
имя хоста  
сервер IBM SPSS Modeler 12, 13  
информационные функции 90  
исследование данных 25  
примеры прикладных программ 34  
стратегия 27  
исходные узлы 35  
обновление 42  
отображение данных 62

## К

каталог  
по умолчанию 146  
каталог temp 14  
классы 19, 137, 138  
клиент  
каталог по умолчанию 146  
ключевые слова  
аннотирование узлов 58  
кодировка 42, 171  
кодировка UTF-8 42, 171  
кодировка потока по умолчанию 42  
кодировка текста 42  
командная строка  
запуск IBM SPSS Modeler 11  
комментарии  
вывод всех в потоке 57

комментарии (*продолжение*)  
к узлам и потокам 54  
сочетания клавиш 164, 165  
конечные узлы 35  
конкатенация строк 90  
координатор процессов 13  
копия 19  
кэш  
включение 145  
задание кэша 38  
очистка 39, 42  
сохранение 39

## Л

линейная регрессия  
экспортировать в формате PMML 149  
логистическая регрессия 168  
экспортировать в формате PMML 149  
логические функции 93

## М

масштаб 19  
масштабирование потоков для  
просмотра 22  
машинное обучение 25  
менеджеры 17  
метки  
вывод на экран 42  
значение 134  
слоев (контрольной) 134  
метки версий объекта репозитория IBM  
SPSS Collaboration and Deployment  
Services 127  
метки объекта репозитория IBM SPSS  
Collaboration and Deployment  
Services 127  
модели 54  
добавление к проектам 138  
замена 146  
обновление 131  
сохранение в репозиторий IBM SPSS  
Collaboration and Deployment  
Services 121  
экспорт 149  
модели PMML  
линейная регрессия 149  
логистическая регрессия 149  
моделирование  
ветвь 54  
модель процесса CRISP-DM 27, 28  
мышь  
использование в IBM SPSS  
Modeler 22, 36

## Н

наблюдение 25  
наборы 42  
наборы дихотомических множественных  
ответов  
в выражениях CLEM 76  
наборы категориальных множественных  
ответов  
в выражениях CLEM 76

наборы множественных ответов  
в выражениях CLEM 76, 78  
наборы правил  
оценка 42  
навигация  
сочетания клавиш 160  
настройка вкладки палитры 152  
национальная версия  
параметры 145  
неотображаемые поля 61  
несколько сеансов IBM SPSS Modeler 14  
новые возможности 7  
номер порта  
сервер IBM SPSS Modeler 12, 13  
нормальное распределение  
вероятностные функции 95

## О

обнаружение знаний 25  
обновить  
исходные узлы 42  
обновление моделей 131  
обновление модели 128  
обработка пустых значений  
Функции CLEM 112  
объекты  
свойства 141  
объекты вывода  
сохранение в репозиторий IBM SPSS  
Collaboration and Deployment  
Services 121  
обязательные поля 63  
окно интерактивного дерева  
специальные возможности 168  
оператор and 93  
оператор not 93  
оператор or 93  
оператор больше 91  
оператор меньше 91  
оператор неравенства 91  
оператор равенства 91  
операторы  
присоединение строк 90  
операции  
в выражениях CLEM 77  
описания потока 52, 53  
опции  
вывода 148  
пользователя 146  
синтаксис 149  
опции внедрения  
сценарии 128  
опции пользователя 146  
остановить выполнение 19  
откат 19  
открытие  
вывод 61  
модели 61  
потоки 61  
проекты 138  
состояния 61  
узлы 61  
отображение данных 63  
отображение полей 61  
отчеты  
добавление к проектам 138



отчеты (продолжение)  
  задание свойств 141  
  создание 141  
  сохранение вывода 60

## П

палитра моделей 121  
палитра сгенерированных моделей 17  
палитры 15  
  настройка 150  
память  
  управление 145  
панель инструментов 19  
папки репозитория IBM SPSS Collaboration and Deployment Services 124, 125  
параметры 145  
  IBM SPSS Modeler 145  
  PMML 149  
  в выражениях CLEM 78  
  использование в сценариях 130  
  построение моделей 130  
  поток 49  
  приглашения во время выполнения 49  
  свойства потока 41, 42, 43, 44, 45, 46, 47  
  сеанс 49  
  скоринг 130  
  тип 49  
Параметры stream 49  
параметры сеанса 49  
переименование  
  потоки 52  
  узлы 58  
переменные 25  
печать 24  
  потоки 22, 38  
по умолчанию  
  этап проекта 137  
побитовые функции 96  
поддержка Unicode 171  
подпалитра  
  видимые на вкладке палитры 151  
  создание 151  
  удаление с вкладки палитры 151  
поиск  
  узлов в потоке 51  
поиск объектов в репозитории IBM SPSS Collaboration and Deployment Services 123  
поиск соединений в COP 13  
получение объектов из репозитория IBM SPSS Collaboration and Deployment Services 122  
пользовательские функции(UDF) 77  
поля 25, 83, 85  
  в выражениях CLEM 78  
  просмотр значений 78  
поля времени  
  преобразование 106  
поля шаблона 63  
Построитель выражений 163  
  доступ 77  
  использование 77  
  обзор 77  
поток 15  
потоки 11

потоки (продолжение)  
  аннотирование 54, 58  
  выключение узлов 37  
  добавление к проектам 138  
  добавление комментариев 54  
  добавление узлов 36, 38  
  загрузка 61  
  запуск 53  
  масштабирование для просмотра 22  
  обход узла 37  
  опции внедрения 128  
  параметры 41, 42, 43, 44, 45, 46  
  переименование 52, 58  
  просмотр числа выполнений 47  
  соединение узлов 36  
  создание 35  
  сохранение 59  
  сохранение в репозиторий IBM SPSS Collaboration and Deployment Services 120  
  файлы резервных копий 59  
потоков данных  
  создание 35  
предварительный просмотр  
  данные узла 40  
представление предприятия для IBM SPSS Collaboration and Deployment Services 116, 128  
предупреждения 47  
  задание опций 146  
прикладные программы исследования  
  данных 26  
применения 25  
примеры  
  обзор 4  
  Руководство по прикладным программам 3  
примеры прикладных программ 3  
пробелы  
  удаление из строк 73, 97  
проверка выражений CLEM 79  
прогнозирующие прикладные программы 128  
программы чтения экрана 160, 167, 168  
  пример 166, 167  
проекты 19, 137  
  аннотирование 140  
  в репозиторий IBM SPSS Collaboration and Deployment Services 139  
  генерирование отчетов 141  
  добавление объектов 138  
  задание папки по умолчанию 137  
  задание свойств 140  
  закрытие 141  
  Представление CRISP-DM 137  
  Представление классов 138  
  Свойства folder 140  
  свойства объекта 141  
  создание 138  
  создание нового 138  
  сохранение в репозиторий IBM SPSS Collaboration and Deployment Services 120  
производительность  
  выражения CLEM 157  
  узлов моделирования 156  
  узлов обработки 155

прокрутка  
  задание опций 46  
пропущенные значения  
  в записях 66  
  выражения CLEM 67  
пустое пространство  
  удаление из строк 73, 97  
пустые значения 73

## Р

радианы  
  единицы измерения 43  
разблокировка объектов репозитория  
  unlocking IBM SPSS Collaboration and Deployment Services 124  
распределение хи-квадрат  
  вероятностные функции 95  
регрессия 168  
репозиторий IBM SPSS Collaboration and Deployment Services 115, 116  
  блокировка и разблокировка объектов 124  
  единая регистрация 117  
  извлечение объектов 122  
  папки 124, 125  
  передача проектов 139  
  просмотр 117  
  свойства объекта 126  
  соединение с 117  
  сохранение объектов 118  
  удаление объектов и версий 125  
Репозиторий IBM SPSS Collaboration and Deployment Services  
  поиск 123

## С

свойства  
  для потоков данных 41  
  папка проекта 140  
  этапы отчета 141  
свойства объекта репозитория IBM SPSS Collaboration and Deployment Services 126  
свойства потока  
  Analytic Server 46  
сворачивание 21  
сервер  
  добавление соединений 13  
  каталог по умолчанию 146  
  подключиться к 12  
  поиск серверов в COP 13  
сервер IBM SPSS Modeler 1  
  -password 12  
  ID пользователя 12  
  имя домена (Windows) 12  
  имя хоста 12, 13  
  номер порта 12, 13  
символ группировки  
  форматы вывода чисел 42  
символ обратной дробной черты в выражениях CLEM 84  
символов 83, 84  
система  
  параметры 145

скоринг  
     ветвь 54, 130, 132  
 слепки 54  
     определение 17  
 советы  
     общее использование 63  
     по средствам доступности 168  
 соглашения 89  
 соединения  
     в сервер IBM SPSS Modeler 12, 13  
     кластер сервера 13  
     с IBM SPSS Collaboration and  
     Deployment Services 117  
 создание пользовательской вкладки 150  
     создание подпалитры 151  
 сообщения  
     вывод сгенерированных SQL 45  
 сообщения об ошибках 47  
 соответствие нормативу 508 159  
 состояния  
     загрузка 61  
     сохранение 60  
 сохранение  
     нескольких объектов 60  
     объекты вывода 60  
     потоки 59  
     состояния 60  
     узлы 59  
 сохранение объектов в репозиторий IBM  
 SPSS Collaboration and Deployment  
 Services 118  
 сочетания клавиш 160, 163, 164, 165  
     клавиатура 22, 160, 163, 164, 165  
     общее использование 63  
 специальные возможности 159, 169  
     пример 166, 167  
     советы в IBM SPSS Modeler 168  
 специальные символы  
     удаление из строк 73  
 специальные функции 113  
 списки 83, 84  
 средняя кнопка мыши  
     как обойтись другими средствами 22,  
     36  
 средства доступности  
     возможности в IBM SPSS Modeler 159  
 старшинство 86  
 старшинство операторов 86  
 степенная (экспоненциальная)  
     функция 93  
 строки 83, 84  
     замена 73  
     операции в выражениях CLEM 73  
     совпадение 73  
 существенные поля 61, 63  
 сценарии 24, 69, 127  
     определение 116  
     опции внедрения 128  
 считыватели экрана 163

## Т

таблицы 163  
     добавление к проектам 138  
     сохранение вывода 60  
 текстовые файлы данных  
     кодировка 171

тип внедрения 128  
 типичные применения 25  
 типы данных 71  
     в параметрах 49  
 точка 42  
 тригонометрические функции 95

## У

уведомление  
     задание опций 146  
 узел k-средних  
     большие наборы 42  
     производительность 156  
 узел Аудит данных  
     использование в исследовании  
     данных 26  
 узел аудита данных  
     использование в исследовательской  
     работе 25  
 узел заполнения  
     значения отсутствия 67  
 узел качества  
     значения отсутствия 66  
 узел Коонена  
     большие наборы 42  
     производительность 156  
 узел нейросети  
     большие наборы 42  
     производительность 156  
 узел особого типа  
     производительность 155  
 узел отбора показателей  
     значения отсутствия 66  
 узел Оценка  
     производительность 155  
 узел построения правил  
     загрузка 61  
 узел представления предприятия 128  
 узел разделения на интервалы  
     производительность 155  
 узел Слияние  
     производительность 155  
 Узел сортировки  
     производительность 155  
 узел суммирования  
     производительность 155  
 узел типа  
     значения отсутствия 67  
     производительность 155  
 узел файла кэша  
     загрузка 61  
 узлы 11  
     аннотирование 54, 58  
     блокировка 41  
     введение 35  
     видимые на палитре 151  
     включение 37  
     выключение 37, 38  
     выключение узлов в потоке 37  
     добавление 36, 38  
     добавление к проектам 138  
     добавление комментариев 54  
     дублирование 38  
     загрузка 61  
     задание опций 38  
     настройка вкладки палитры 152

узлы (*продолжение*)  
     обход в потоке 37  
     поиск 51  
     порядок 153  
     предварительный просмотр  
     данных 40  
     производительность 155, 156  
     редактирование 38  
     соединение в потоке 36  
     создание пользовательской  
     вкладки 150  
     создание пользовательской  
     подпалитры 151  
     сохранение 59  
     сохранение в репозиторий IBM SPSS  
     Collaboration and Deployment  
     Services 121  
     удаление 36  
     удаление с палитры 151  
     удаление соединений 38  
     число выполнений 47  
 узлы моделирования 35  
     настройка вкладки палитры  
     моделирования 152  
     производительность 156  
 узлы обработки 35  
     производительность 155  
 условия 71  
 утилита отображения данных 61, 62

## Ф

файл словаря 167  
 файлы PowerPoint 138  
 Файлы SAS  
     кодировка 171  
 файлы вывода  
     сохранение 60  
 файлы журналов  
     вывод сгенерированных SQL 45  
 файлы резервных копий потоков  
     восстановление 59  
 файлы статистики  
     кодировка 171  
 формат вывода валюты 43  
 форматы времени 43, 85, 86  
 форматы вывода  
     валюта 43  
     десятичные знаки 43  
     символ группировки 43  
     числа 43  
     экспоненциальная запись 43  
 форматы вывода чисел 43  
 форматы дат 43, 85, 86  
 функции 85, 86, 89, 90, 106  
     @BLANK 67  
     @FIELD 77, 113  
     @GLOBAL\_MAX 111  
     @GLOBAL\_MEAN 111  
     @GLOBAL\_MIN 111  
     @GLOBAL\_SDEV 111  
     @GLOBAL\_SUM 111  
     @PARTITION 113  
     @PREDICTED 77, 113  
     @TARGET 77, 113  
 база данных 77  
 в выражениях CLEM 77

функции (*продолжение*)  
 обработка пропущенных значений 67  
 пользовательские функции(UDF) 77  
 примеры 69  
 функции CLEM  
 вероятность 95  
 ДатаВремя 102  
 информация 90  
 логические 93  
 побитовые 96  
 последовательность 106  
 пустые значения и значения null 112  
 случайный 97  
 сравнение 91  
 тригонометрические 95  
 числовые 93  
 Функции CLEM  
 глобальные 111  
 преобразование 90  
 пропущенные значения 67  
 специальные функции 113  
 список доступных 88  
 текст 97  
 Функции date 85, 86  
 date\_before 91  
 Функции if, then, else 93  
 Функции time 85, 86  
 time\_before 91  
 функции time и date 85  
 функции времени  
 time\_before 102  
 time\_hours\_difference 102  
 time\_in\_hours 102  
 time\_in\_mins 102  
 time\_in\_secs 102  
 time\_mins\_difference 102  
 time\_secs\_difference 102  
 функции даты  
 date\_before 102  
 date\_days\_difference 102  
 date\_in\_days 102  
 date\_in\_months 102  
 date\_in\_weeks 102  
 date\_in\_years 102  
 date\_months\_difference 102  
 date\_weeks\_difference 102  
 date\_years\_difference 102  
 Функция @TODAY 102  
 функции даты и времени 86  
 datetime\_date 102  
 datetime\_day 102  
 datetime\_day\_name 102  
 datetime\_day\_short\_name 102  
 datetime\_hour 102  
 datetime\_in\_seconds 102  
 datetime\_minute 102  
 datetime\_month 102  
 datetime\_month\_name 102  
 datetime\_month\_short\_name 102  
 datetime\_now datetime\_second 102  
 datetime\_time 102  
 datetime\_timestamp 102  
 datetime\_weekday 102  
 datetime\_year 102  
 функции для работы с текстовыми значениями 97  
 функции последовательности 106  
 функции преобразования 90  
 функции распределения 95  
 функции сравнения 91  
 Функция @BLANK 67, 90, 112  
 Функция @DIFF 106  
 Функция @FIELD 67, 113  
 Функция @FIELDS\_BETWEEN 67, 75, 113  
 Функция @FIELDS\_MATCHING 67, 75, 113  
 Функция @INDEX 106  
 Функция @LAST\_NON\_BLANK 106, 112  
 Функция @MAX 106  
 Функция @MEAN 106  
 Функция @MIN 106  
 функция @MULTI\_RESPONSE\_SET 113  
 Функция @MULTI\_RESPONSE\_SET 76  
 Функция @NULL 67, 90, 112  
 Функция @OFFSET 106  
 вопросы производительности 157  
 функция @PARTITION\_FIELD 113  
 Функция @PREDICTED 113  
 Функция @SDEV 106  
 Функция @SINCE 106  
 Функция @SUM 106  
 Функция @TARGET 113  
 Функция @TESTING\_PARTITION 113  
 Функция @THIS 106  
 Функция @TODAY 102  
 Функция @TRAINING\_PARTITION 113  
 функция  
 @VALIDATION\_PARTITION 113  
 Функция abs 93  
 функция allbutfirst 97  
 Функция allbutlast 97  
 Функция alphabefore 97  
 Функция arccos 95  
 Функция arccosh 95  
 Функция arcsin 95  
 Функция arcsinh 95  
 Функция arctan 95  
 Функция arctan2 95  
 Функция arctanh 95  
 Функция cdf\_chisq 95  
 Функция cdf\_f 95  
 Функция cdf\_normal 95  
 Функция cdf\_t 95  
 Функция cos 95  
 Функция cosh 95  
 Функция count\_equal 75, 91  
 Функция count\_greater\_than 75, 91  
 Функция count\_less\_than 91  
 Функция count\_non\_nulls 91  
 Функция count\_not\_equal 75, 91  
 Функция count\_nulls 67, 75, 91  
 Функция date\_before 91  
 Функция datetime\_date 90  
 Функция DIFF 106  
 Функция div 93  
 Функция endstring 97  
 Функция exponential 93  
 Функция first\_index 91  
 Функция first\_non\_null 76, 91  
 Функция first\_non\_null\_index 76, 91  
 Функция fracof 93  
 функция hasendstring 97  
 Функция hasmidstring 97  
 Функция hasstartstring 97  
 Функция hassubstring 97  
 Функция INDEX 106  
 Функция integer\_leastbit 96  
 Функция integer\_length 96  
 Функция intof 93  
 Функция is\_date 90  
 Функция is\_datetime 90  
 Функция is\_integer 90  
 Функция is\_number 90  
 Функция is\_real 90  
 Функция is\_string 90  
 Функция is\_time 90  
 Функция is\_timestamp 90  
 Функция isalphanumeric 97  
 функция isendstring 97  
 Функция islowercode 97  
 Функция ismidstring 97  
 функция isnumbercode 97  
 Функция isstartstring 97  
 Функция issubstring 97  
 Функция issubstring\_count 97  
 Функция issubstring\_lim 97  
 функция isuppercode 97  
 Функция last\_index 76, 91  
 Функция LAST\_NON\_BLANK 106  
 Функция last\_non\_null 91  
 Функция last\_non\_null\_index 76, 91  
 Функция length 97  
 функция locchar 97  
 Функция locchar\_back 97  
 Функция log 93  
 Функция log10 93  
 Функция lowertoupper 97  
 Функция matches 97  
 Функция max 91  
 Функция MAX 106  
 Функция max\_index 76, 91  
 Функция max\_n 75, 91  
 Функция MEAN 106  
 Функция mean\_n 75, 93  
 Функция member 91  
 Функция min 91  
 Функция MIN 106  
 Функция min\_index 91  
 Функция min\_n 75, 91  
 Функция mod 93  
 Функция negate 93  
 Функция OFFSET 106  
 Функция oneof 97  
 Функция pi 95  
 Функция random 97  
 Функция random0 97  
 Функция rem 93  
 Функция replace 97  
 Функция replicate 97  
 Функция round 93  
 Функция SDEV 106  
 Функция sdev\_n 75, 93  
 Функция sign 93  
 Функция sin 95  
 Функция SINCE 106  
 Функция sinh 95  
 Функция skipchar 97  
 Функция skipchar\_back 97  
 Функция soundex 102  
 Функция soundex\_difference 102

Функция sqrt 93  
функция startstring 97  
Функция stripchar 97  
Функция strmember 97  
функция subscrs 97  
Функция substring 97  
Функция SUM 106  
Функция sum\_n 75, 93  
Функция tan 95  
Функция tanh 95  
Функция testbit 96  
Функция THIS 106  
Функция time\_before 91  
Функция to\_date 90, 102  
Функция to\_dateline 102  
Функция to\_datetime 90  
Функция to\_integer 90  
Функция to\_number 90  
Функция to\_real 90  
Функция to\_string 90  
Функция to\_time 90, 102  
Функция to\_timestamp 90, 102  
Функция trim 97  
Функция trimend 97  
функция undef 112  
Функция unicode\_char 97  
Функция unicode\_value 97  
функция uppertolower 97  
Функция value\_at 76, 91

## Я

язык  
параметры 145

## Х

холст 15  
холст потока  
параметры 46

## Ц

цвета  
настройка 148  
цвета сценария  
настройка 149  
целые числа 83

## Ч

числа 74, 83, 84  
число выполнений, просмотр 47  
числовые функции 93

## Ш

шаблоны 62  
шрифты 148, 149

## Э

экспоненциальная запись  
формат вывода 43  
экспорт  
PMML 134  
описания потока 53





Напечатано в Дании