



System i

Справочник по индексированию IBM Content Manager OnDemand для общего сервера i5/OS

Версия 6 Выпуск 1

SC89-1059-04





System i

Справочник по индексированию IBM Content Manager OnDemand для общего сервера i5/OS

Версия 6 Выпуск 1

SC89-1059-04

Это издание относится к IBM Content Manager OnDemand i5/OS версии 6 выпуска 1 и ко всем последующим выпускам и модификациям, если в новых изданиях не будет оговорено другое. Это издание заменяет SC27-1160-03.

© Copyright International Business Machines Corporation 2001, 2008. Все права защищены.

Содержание

О публикации IBM Справочник по индексированию Content Manager OnDemand для общего сервера i5/OS (SC27-1160)	vii
Для кого предназначена эта книга	vii
Структура книги	vii

Часть 1. Индексатор OS/400 1

Глава 1. Использование индексатора OS/400	3
Принципы индексирования	3
Параметры индексирования	4
Применение BREAK=YES в зависимости от BREAK=NO в параметрах индексатора	6
Управление максимальным числом страниц в группе	7
Определение многоключевых индексов	7
Пример	8
Определение операционных полей	10
Пример	11
Назначение значений индексов по умолчанию	14
Определение полей поиска текста	15
Работа с буферными файлами SCS с приложениями AFP	16
Применение маски при определении полей приложения	16
Применение логических элементов тегов (TLE)	17

Часть 2. Индексатор PDF 19

Глава 2. Обзор	21
Что такое индексатор PDF?	21
Как OnDemand использует информацию об индексировании	23
Обработка входных файлов PDF с помощью графического индексатора	24
Индексирование входных данных вручную	27
Принципы индексирования	27
Система координат	28
Параметры индексирования	28
Как создать параметры индексирования	30
Глава 3. Системные особенности	33
Системные ограничения	33
Требования к входным данным	33
Особенности поддержки национальных языков	34
Глава 4. Справочник параметров	35
Параметр COORDINATES	35
Синтаксис	35
Опции и значения	35
FIELD	35
Синтаксис поля триггера	35
Синтаксис поля константы	38
Связанные параметры	38
FONTLIB	39
Синтаксис	39
Опции и значения	39
INDEX	39

Синтаксис	39
Опции и значения	39
Примеры	40
Связанные параметры.	40
INDEXDD	40
Синтаксис	40
Опции и значения	40
INDEXSTARTBY	41
Синтаксис	41
Опции и значения	42
INPUTDD	42
Синтаксис	42
Опции и значения	42
MSGDD	42
Синтаксис	42
Опции и значения	42
OUTPUTDD	43
Синтаксис	43
Опции и значения	43
PARMDD	43
Синтаксис	43
Опции и значения	43
TEMPDIR	43
Синтаксис	44
Опции и значения	44
параметр TRACEDD	44
TRIGGER	44
Синтаксис	44
Опции и значения	44
Примеры	45
Связанные параметры.	46
Глава 5. Справочник сообщений	47
Глава 6. Справочник по программе ARSPDOC1	49
Назначение	49
Синтаксис	49
Описание.	49
Параметры	50
Расположение IFS	50
Глава 7. Справочник по программе ARSPDUMP	51
Назначение	51
Синтаксис	51
Описание.	51
Параметры	51
Примеры	52
Расположение IFS	52
Глава 8. Трассировщик	53

Часть 3. Стандартный индексатор 55

Глава 9. Обзор	57
Загрузка данных	57
Обработка данных AFP	58

Глава 10. Задание файла параметров	59
CODEPAGE:	59
Синтаксис	59
Опции и значения	59
Пример	60
Параметр COMMENT:	60
Синтаксис	60
Опции и значения	60
Пример	60
Параметр GROUP_FIELD_NAME:	60
Синтаксис	60
Опции и значения	61
Пример	61
Параметр GROUP_FIELD_VALUE:	61
Синтаксис	61
Опции и значения	61
Пример	61
Параметр GROUP_FILENAME:	62
Синтаксис	63
Опции и значения	63
Пример	63
Параметр GROUP_LENGTH:	63
Синтаксис	63
Опции и значения	63
Пример	63
Параметр GROUP_OFFSET:	63
Синтаксис	64
Опции и значения	64
Пример	64
Глава 11. Примеры файлов параметров	65
Глава 12. Дополнительные разделы по индексированию	67
Программа постпроцессора	67
Файлы индексов (.ind) в IFS.	67
Рекомендуемый порядок определения триггеров и полей	67
Индекс	69

I
I
I
I

О публикации IBM Справочник по индексированию Content Manager OnDemand для общего сервера i5/OS (SC27-1160)

| В этой книге содержится информация о методах индексирования, подготовке
| данных индекса и работе с инструментами индексации отчетов, которые
| планируется хранить и востребовать с помощью IBM Content Manager
| OnDemand для общего сервера i5/OS версии 6 выпуска 1 (OnDemand).

Для кого предназначена эта книга

Эта книга в основном предназначена для администраторов и других работников организации, которые отвечают за подготовку данных для хранения с помощью функции OnDemand.

Структура книги

Данная книга состоит из следующих частей. Каждая часть содержит описание одного из инструментов индексирования, предусмотренных в функции OnDemand:

- В части Часть 1, “Индексатор OS/400”, на стр. 1 описывается применение графического инструмента администрирующего клиента для определения критериев индексирования, которые индексатор OS/400 использует для поиска и создания данных индекса для ваших буферных файлов.
- В части Часть 2, “Индексатор PDF”, на стр. 19 описывается применение индексатора PDF функции OnDemand для создания данных индексирования файлов Adobe PDF.
- В части Часть 3, “Стандартный индексатор”, на стр. 55 описывается применение стандартного индексатора функции OnDemand для создания данных индексирования файлов других типов.

Часть 1. Индексатор OS/400

В этой части описывается индексатор OS/400. Индексатор OS/400 можно использовать для определения параметров индексирования для SCS, расширенной SCS, Advanced Function Presentation (AFP) и буферных файлов Line, которые требуется хранить в системе.

Индексатор OS/400 является первичным индексатором, применяемым системой i5/OS.

Глава 1. Использование индеклятора OS/400

Индексатор OS/400 является самым универсальным индексатором функции OnDemand для буферных файлов i5/OS. Индексатор OS/400 для SCS, расширенной SCS, Advanced Function Presentation (AFP) и буферных файлов Line вызывается командой ADDRPTOND. Для определения критериев индексирования, которые индексатор OS/400 применяет для нахождения и создания данных индексирования для ваших буферных файлов, используется графическая утилита индексирования административного клиента OnDemand.

Графическую утилиту можно вызвать двумя способами:

- Кнопкой Выбор образца данных в мастере отчетов или
- Выбором образца данных и нажатием кнопки Изменить в панели Информация индеклятора при создании определения приложения OnDemand

OnDemand использует индексатор OS/400 по умолчанию для SCS, расширенной SCS, AFP и буферных файлов Line. Дополнительная информация о мастере отчетов содержится в разделе Мастер отчетов во введении в *IBM Content Manager OnDemand для общего сервера i5/OS: Руководство по администрированию*. Дополнительная информация об определении приложения без использования мастера отчетов приведена в разделе Добавление приложения в главе Примеры в *IBM Content Manager OnDemand для общего сервера i5/OS: Руководство по администрированию*.

Принципы индексирования

Параметры индексирования содержат информацию, которая позволяет функции OnDemand определить ключевые элементы в потоке данных на печать и создать элементы индекса, указывающие на эти элементы. Вы можете задавать параметры индексирования, которые функция OnDemand применяет для разбиения потока данных на отдельные элементы, называемые группами. Группа - это набор из одной или нескольких страниц. Для данного набора задаются ограничения: например, выписка с банковского счета, страховой полис, счет за телефон или другой логический сегмент файла отчета. Группа также может представлять конкретное число страниц в отчете. Например, отчет на 10 000 страницах можно разбить на группы по 100 страниц. Функция OnDemand создает индексы для каждой группы. Группы определяются при изменении значения индекса (например, номера счета) или по достижению максимального числа страниц для группы.

Данные индекса состоят из имени атрибута (например, Имя_клиента) и значения атрибута (например, Фрэнк Бус) с тегом, который определяет расположение этих данных на печатной странице. Например, тег номер_счета с указателем 1,21,16 показывает функции OnDemand, что значения номер_счета можно найти в столбце 21 конкретных входных отчетов. Функция OnDemand набирает 16 байтов информации, начиная со столбца 21, и добавляет их к списку значений атрибутов, найденных во входных данных. Функция OnDemand создает файл индекса при индексации файлов отчетов. Файл индекса содержит элементы индекса, которые включают сдвиг и длину группы. Функция OnDemand вычисляет элемент индекса для каждой группы, найденной во входном файле. Затем она записывает полученные из входного файла значения атрибута в файл индекса.

Параметры индексирования

Параметры индексирования могут содержать параметры индексирования, преобразования и набора ресурсов, опции и значения. Для большинства отчетов функции OnDemand требуются три параметра индексирования для получения или создания данных индексирования:

- **TRIGGER** - функция OnDemand использует триггеры для определения расположения данных. По триггеру функция OnDemand ищет конкретную информацию в конкретном месте файла отчета. После нахождения записи в потоке данных, которая содержит информацию, определенную в триггере, функция OnDemand может начать просмотр информации об индексе.
 - Она сравнивает данные в файле отчета с набором знаков, заданных в триггере, по каждому байту.
 - Можно задать максимум восемь триггеров
 - Все постоянные триггеры группы должны совпасть, чтобы функция OnDemand могла создать информацию об индексе. Однако в потоке данных могут встречаться и плавающие триггеры. То есть данные индексирования, основанные на плавающем триггере, могут быть взяты из любой записи в файле отчета.
- **FIELD** - этот параметр показывает расположение, сдвиг и длину данных, которые функция OnDemand использует для создания значений индекса.
 - По умолчанию определения полей основываются на TRIGGER1, но могут основываться на любом из восьми параметров TRIGGER.
 - Можно задать максимум 32 поля.
 - Поле может определять все или часть фактического значения индекса, хранящегося в базе данных.
- **INDEX** - в этом параметре определяется имя атрибута, идентифицируется поле (или поля), на котором основывается индекс, и указывается тип индекса, создаваемого функцией OnDemand. Для индексов групп, которые функция OnDemand хранит в базе данных, имена атрибутов должны соответствовать именам полей базы данных групп приложения.
 - Функция OnDemand может создавать индексы для одной страницы, группы страниц и первого и последнего отсортированных значений на странице или группе страниц. Значения индексов групп функция OnDemand хранит в базе данных. С помощью индексов групп пользователи могут отыскивать элементы. Индексы страниц хранятся вместе с документом (например, счетом). После получения документа со страничными индексами по ним вы можете переходить на конкретные страницы.
Функция OnDemand может создавать такую информацию на страничном уровне только при преобразовании входных данных в AFP. Создание информации на страничном уровне задается параметрами CONVERT=YES и INDEXOBJ=ALL и созданием поля индекса с опцией TYPE=PAGE.
 - Для формирования индекса параметры полей можно объединять.
 - Можно задать максимум 32 параметра индексов.

Функция OnDemand создает новую группу и получает новые значения индекса, когда одно или несколько значений постоянных индексов группы изменяются или достигается максимальное число страниц в группе.

```

|  -----1-----2-----3-----4-----5-----6-----7-----8-----9
| 01                                     Страница 0001
| 1
| 2                               Jack Straw
| 3                               4 Buxanchange Way
| 4                               Wichitaw KS 99999-9999
| 5
| 6                               Дата выписки со счета: 06/15/07
| 7                               Номер счета: 1234-5678-9876-0000
| 8
| 9                               Баланс: 2984,17 доллара

```

Рисунок 1. Индексирование отчета

Следующие параметры индексирования могут быть использованы для создания данных индексирования для отчета, приведенного в рис. 1. Определения TRIGGER показывают функции OnDemand, как найти начало группы во входных данных. Для нахождения начала группы (выписки со счета) в данном примере файла функции OnDemand требуются два определения TRIGGER. Пример:

- Параметр TRIGGER1 отыскивает 1 в первом байте каждой входной записи.
- Параметр TRIGGER2 ищет строку Страница 0001 в столбце 72 той же записи.

Вместе эти триггеры однозначно определяют начало выписки отчета в этой записи.

Определения параметров FIELD показывают расположение значений индекса в выписках со счета. Поля основываются на расположении записей триггеров. Пример:

- Параметр FIELD1 задает значения индекса имени клиента, начиная со столбца 40 второй записи после записи, определенной параметром TRIGGER1.
- Параметр FIELD2 задает значения индекса данных выписки со счета, начиная со столбца 56 шестой записи после записи, определенной параметром TRIGGER1.
- Параметр FIELD3 задает значения индекса номера счета, начиная со столбца 56 седьмой записи после записи, определенной параметром TRIGGER1.

Определение параметра INDEX указывает имя атрибута поля индекса. Индексы основываются на одном или нескольких определениях полей. Пример:

- Параметр INDEX1 задает имя атрибута имя_клиента для значений, полученных с помощью параметра FIELD1.
- Параметр INDEX2 задает имя атрибута дата_выписки для значений, полученных с помощью параметра FIELD2.
- Параметр INDEX3 задает имя атрибута номер_счета для значений, полученных с помощью параметра FIELD3.

В следующей таблице приводятся максимальные значения для определенных атрибутов индексирования:

Атрибут индексирования	Макс. значение
Максимальное число строк на странице буферного файла	512
Максимальное число записей индексов в группе (документе)	5000
Максимальное число значений индекса в группе (документе)	9999
Максимальное число страниц в группе (документе)	9999
Максимальное число триггеров на странице	512
Максимальное число значений индекса на странице	1024

Атрибут индексирования	Макс. значение
Максимальное число полей на странице	1024

Применение BREAK=YES в зависимости от BREAK=NO в параметрах индекса

Группа - это набор страниц, которые логически связаны между собой. Например, все страницы в одной выписке с банковского счета могут составлять группу. Группа - это отдельный документ или *сегмент*, как он принят в Архиве буферных файлов. Разрыв группы - это процедура закрывания текущей группы и начала новой группы. В случае Архива буферных файлов эта процедура называется сегментацией. Для некоторого данного индекса группы задание параметра BREAK определяет, начать ли индексу OS/400 новый документ, когда значение этого индекса изменяется.

Если задан BREAK=YES, то индексатор OS/400 начнет новую группу, когда значение поля, на котором основан индекс, изменится. Например, когда изменяется номер счета. Задание BREAK=NO применяется, если определяются два или более полей, и индексатор OS/400 должен начать новую группу, только когда изменится значение второго поля. Параметр BREAK=YES задается только для индекса, который основан на поле, которое вы хотите, чтобы индексатор OS/400 использовал для управления разрывом группы. Параметр BREAK=NO задается для всех остальных индексов в группе.

В продолжение примера с выпиской с банковского счета рассмотрим хранение выписок со счетов. Каждая выписка начинается с изменения номера счета. Были определены индексы для номера счета, имени клиента и даты выписки. Скорее всего, вы захотите для номера счета задать параметр BREAK=YES, для имени клиента - BREAK=NO и для даты выписки - BREAK=NO. В результате разрыв группы будет делаться только при изменении номера счета. Соответствующие параметры индекса в определении приложения могут выглядеть следующим образом:

```
INDEX1=X'C1838396A495A3D5A494828599',FIELD1,(TYPE=GROUP,BREAK=YES) /* Номер_счета */
INDEX2=X'C3A4A2A396948599D5819485',FIELD2,(TYPE=GROUP,BREAK=NO) /* Имя_клиента */
INDEX3=X'E2A381A385948595A3C481A385',FIELD3,(TYPE=GROUP,BREAK=NO) /* Дата_выписки */
```

Мастер отчетов клиента-администратора функции OnDemand предназначен упростить процедуру определения групп приложений, приложений и папок. Мастер предполагает, что любое изменение в индексе должно быть связано с разрывом группы. Поэтому он устанавливает параметр BREAK=YES для всех полей индекса. Если требования к данным таковы, что не все индексы должны быть связаны с разрывом группы, то в параметрах индекса приложения необходимо изменить BREAK=YES на BREAK=NO. Эти изменения можно сделать двумя способами: с помощью клиента-администратора выберите опцию Обновление для приложения, вызовите вкладку Индексатор и вручную отредактируйте параметры индекса, внося соответствующие изменения, или с помощью графического индекса откройте свойства индекса и затем нажмите переключатель **Нет**.

Если же индекс основан на плавающем триггере, то параметр BREAK можно устанавливать только на **NO**. Кроме того, если выбрана опция **Разрешить несколько значений**, то параметр BREAK автоматически устанавливается на **NO**.

Если данные уже заархивированы и все индексы имеют параметр BREAK=YES, это изменение еще вносить можно. Изменить параметр BREAK=YES на BREAK=NO можно любое время. В случае изменения параметров индекса необходимо убедиться в правильности архивирования отчетов. Заархивированные отчеты разархивировать не требуется; однако изменение будет действовать только на отчеты, которые архивируются после внесения изменения.

Управление максимальным числом страниц в группе

Для каждой индексируемой группы можно задать максимальное число страниц в группе. Для определения максимального числа страниц в группе функция OnDemand может применять параметр индекса GROUPMAXPAGES. Пуст требуется проиндексировать отчет, содержащий тысячи страниц информации. Если использование параметра BREAK=YES не приводит к созданию достаточно небольших групп страниц (или сегментов) отчета, то можно воспользоваться параметром, например GROUPMAXPAGES=100, чтобы проинструктировать функцию OnDemand закрывать текущую группу и начинать новую, как только число страниц в текущей группе достигает 100. То есть, если значение параметра GROUPMAXPAGES достигается до изменений индекса группы, функция OnDemand создает новую группу принудительно. Если значение параметра GROUPMAXPAGES не задано, функция OnDemand не вставляет разрыв, пока значение одного из полей INDEX с установкой BREAK=YES не изменится.

Определение многоключевых индексов

Когда значение индекса встречается несколько раз в пределах одного документа, могут использоваться многоключевые индексы. Например, накладные могут иметь номер накладной, номер клиента и имя клиента в первых трех полях индекса, которые в данной накладной встречаются по одному разу. В этом случае номер продукта можно определить как многоключевой индекс, поскольку в одной накладной может быть несколько номеров продуктов. С помощью многоключевого индекса конечный пользователь может осуществить поиск по номеру продукта и найти любую накладную с данным номером продукта независимо от того, в каком месте списка заказанных продуктов находится этот номер. Без этой опции будет проиндексирован только первый номер продукта на странице.

Для многоключевого индексирования к каждому оператору INDEX, для которого в одном документе есть несколько значений, необходимо добавить ключевое слово ALLOWMULTIPLEVALUES=YES. Например:

```
INDEX2=X'97969596',FIELD2,(TYPE=GROUP,BREAK=NO,ALLOWMULTIPLEVALUES=YES)
```

Новое ключевое слово будет добавлено к определению приложения функции OnDemand. Для изменения параметров индекса приложения откройте вкладку Сведения об индексе, затем выберите пункт Клавиатура и нажмите кнопку Изменить. Это новое ключевое слово ALLOWMULTIPLEVALUES доступно только для BREAK=NO. Кроме того, в отличие от правила использования нескольких ключей Архива буферных файлов OnDemand, определение индекса как многоключевого не требует определения всех последующих полей индексов тоже как многоключевых. В среде общего сервера, как видно из примера, можно определить некоторый индекс как многоключевой, а следующий - как не многоключевой. Однако поле,

используемое для многоключевого индекса, должно находиться в той же или в следующей строке, которая содержит плавающий триггер, используемый для поиска этого поля.

Пример

В следующем примере показывается, как определить многоключевой индекс с помощью мастера отчетов и графического индексатора. Примером архивируемого отчета является накладная AFP. В качестве индексов используются следующие части информации:

- Номер клиента
- Номер накладной
- Дата заполнения накладной
- Номер изделия (это будет многоключевой индекс)
- Сумма платежа

Обычно триггеры и поля определяются от левого верха до правого низа отчета. Это упрощает процедуру создания параметров индексатора.

рис. 2 показывает страницу примера отчета.

OnDemand - [04/16/03]

File Edit View Search Notes Options Window Help

400 CPU Parkway
Vegetation, NJ 55000

Super Sun Seeds
A Growth Company

Office: 555-499-2367
Fax: 555-415-9794

THE LAST LEAF
340 DESPERADO COURT
LONGVIEW
CA 12345-6789

CAMBIUM LAYER LIMITED
2222 SAPLING CIRCLE
BARKERSVILLE
BC 47365-7290

-- Sold To --

Customer Number: 154 Invoice Number: 31354 Invoice Date: 3/16/03 Payment Date: 4/16/03

SHIP Via: TREE TRUCK Shipped Date: 3/16/03 Terms: NET 15 Salesman: MARY PINETREE

4	PK	03698741	STRING GRAPEFRUIT	2.01	8.04
300	EA	11000146	AZALIA, GIANT ROSE SEEDS	.55	165.00
6	CT	11005010	EARLY DWARF DANISH SEEDS	3.01	18.06
24	DZ	11005013	MINCOR NANTES CARROT SEED	.87	20.88
24	PK	11005020	FRENCH PICKLING SEEDS	2.39	57.36
12	CT	12382910	SUCCATASH SEEDS	.38	4.56
55	CT	13145340	SOUR GRAPE SEEDS	.15	8.25
14	BZ	32165478	BLACK EYED BANANA	3.01	42.14
600	DZ	44646510	PLUMP RED PLUMS	.49	294.00
40	DZ	45613712	CRANAPPLE BERRY SEEDS	1.28	51.20

Ready bmarshall - iSeries (7.1.1.0) Page 1 of 6 100% 24

Рисунок 2. Пример отчета с многоключевым индексированием

Сначала запустите администрирующего клиента функции OnDemand и зарегистрируйтесь на сервере вашего экземпляра. Затем нажмите кнопку Мастер отчетов панели инструментов. Выберите тип данных, например выберите AFP. Затем выберите образец входного файла. Графический индексатор покажет буферный файл.

Пример отчета содержит данные AFP и графический индекатор отображает только текст без ресурсов AFP (таких как специальные шрифты, штриховые коды, графика и аппликации).

Определите первый триггер. В дате отгрузки в качестве триггера 1 выберите символ / (прямая косая черта). Этот триггер будет использоваться для поиска номера клиента, номера накладной и даты отгрузки.

Определите второй триггер. В цене в качестве триггера 2 выберите символ . (точка). Этот триггер должен быть определен как плавающий и будет использоваться для поиска номеров изделий.

Определите третий триггер. В сумме платежа в качестве триггера 3 выберите символ / (прямая косая черта). Этот триггер будет использоваться для поиска суммы платежа.

Теперь определите поля и индексы. С помощью мастера отчетов поля и индексы определяются на одном этапе. Если же используется графический индекатор из определения приложения, то поля и индексы определяются на разных этапах.

Первое поле и индекс определите для номера клиента. Номер клиента находится триггером 1. На странице Атрибуты полей базы данных поле номера клиента определяется как данные типа строка.

Второе поле и индекс определите для номера накладной. Номер накладной находится триггером 1. На странице Атрибуты полей базы данных номер накладной определяется как данные типа строка.

Третье поле и индекс определите для даты заполнения накладной. Дата заполнения накладной находится триггером 1. На странице Атрибуты полей базы данных дата заполнения накладной определяется как данные типа данные и выбирается как поле сегмента.

Четвертое поле и индекс определите для номера изделия. Номер изделия находится триггером 2. На странице Атрибуты полей базы данных номер изделия определяется как данные типа строка.

Параметр *Маска* служит для задания шаблона для данных поля, чтобы они использовались как индекс. В нашем примере поле должно состоять из восьми численных знаков (каждый символ # представляет один численный знак). Было бы удобно, если бы один триггер (точка) мог представлять строку, не содержащую номер изделия.

После определения всех полей необходимо вернуться назад и отметить индекс номера изделия как многоключевой (как описано выше).

Пятое поле и индекс определите для суммы платежа. Сумма платежа находится триггером 3. На странице Атрибуты полей базы данных сумма платежа определяется как данные типа строка.

На этом определение полей и индексов завершено.

Теперь необходимо вернуться назад и определить номер изделия, который является индексом 4, как многоключевой. На панели инструментов нажмите кнопку Переключение выбора Триггер, Индекс, Параметры полей.

Администрирующий клиент открывает окно диалога Выбор.

Щелкните на индексе 4. Затем нажмите кнопку Свойства для вызова окна диалога Обновление индекса.

Включите переключатель Разрешить несколько значений. **Примечание:** Для этого требуется версия 7.1.0.8 администрирующего клиента OnDemand или более поздняя.

Для сохранения индекса номера изделия как многоключевого нажмите кнопку ОК.

Закройте окно диалога Выбор.

Для проверки, как система будет индексировать документ, на панели инструментов нажмите кнопку Переключение между Просмотр и Добавление параметров.

Определенные триггеры будут выделены красным. Определенные поля будут выделены синим.

Теперь можно закрыть окно графического индексатора и завершить процедуру, воспользовавшись мастером отчетов для определения группы приложений, приложения и папки.

Параметры индексатора, которые были созданы для данного примера отчета, показаны на рис. 3.

```
TRIGGER1=*,55,X'61',(TYPE=GROUP) /* / */
TRIGGER2=*,64,X'4B',(TYPE=FLOAT) /* . */
TRIGGER3=*,31,X'61',(TYPE=FLOAT) /* / */
FIELD1=0,15,6,(TRIGGER=1,BASE=0)
FIELD2=0,33,6,(TRIGGER=1,BASE=0)
FIELD3=0,50,8,(TRIGGER=1,BASE=0)
FIELD4=0,19,8,(TRIGGER=2,BASE=0,MASK='#####')
FIELD5=0,69,12,(TRIGGER=3,BASE=0)
INDEX1=X'83A4A2A39596',FIELD1,(TYPE=GROUP,BREAK=YES) /* ном_зак */
INDEX2=X'8995A59596',FIELD2,(TYPE=GROUP,BREAK=YES) /* ном_накл */
INDEX3=X'8995A58481A385',FIELD3,(TYPE=GROUP,BREAK=YES) /* дата_накл */
INDEX4=X'89A3859495A494',FIELD4,(TYPE=GROUP,BREAK=NO,ALLOWMULTIPLEVALUES=YES) /* ном_изд */
INDEX5=X'A396A3819384A485',FIELD5,(TYPE=GROUP,BREAK=NO) /* сумма_плат*/
```

Рисунок 3. Параметры индексатора с многоключевыми индексами

После загрузки примера отчета можно запустить OnDemand Client, открыть новую папку и выполнить поиск документов.

Определение операционных полей

Отчет об операциях содержит страницы записей с одним или несколькими столбцами отсортированных данных. Например, каждая страница гроссбуха содержит до 80 записей операций. Каждая запись имеет уникальное значение, такое как номер операции. Записи в отчете сортируются по номеру операции.

Вместо сохранения каждого номера операции в базе данных (возможно, сотни тысяч строк) можно разбить отчет на группы страниц (скажем, по 100 страниц в группе), получить первый и последний номера операций в каждой группе и сохранить эти значения в базе данных. После этого для получения конкретной группы страниц отчета с конкретным номером операции пользователь задает

номер операции. Функция OnDemand сравнивает номер операции с первыми и последними значениями, хранящимися в базе данных, и выдает нужную группу.

Для определения отчета об операциях, который содержит один или несколько столбцов отсортированных данных, как в примере, используется операционное поле. Операционное поле позволяет функции OnDemand проиндексировать группу страниц, используя первое значение индекса на первой странице и последнее значение индекса на последней странице.

Определить операционное поле проще всего с помощью мастера отчетов и графического индексатора.

Параметр индексатора для операционного поля будет выглядеть аналогично следующему:

```
FIELD1=*,*,10,(OFFSET=(3:12),MASK='#####',ORDER=BYCOL)
```

Параметр индексатора для индекса, созданный на основе операционного поля, будет выглядеть аналогично следующему:

```
INDEX1=X'D3968195',FIELD1,(TYPE=GROUPRANGE,BREAK=NO)
```

Эти параметры индексатора мастер отчетов добавит в определение приложения OnDemand. Чтобы увидеть их, зайдите на вкладку Сведения об индексаторе, затем выберите пункт Клавиатура и нажмите кнопку Изменить для просмотра списка Параметры индексатора приложения.

Пример

В следующем примере показывается, как определить отчет об операциях с помощью мастера отчетов и графического индексатора. В качестве примера отчета для архивирования возьмем Отчет о непогашенных ссудах. Каждая страница отчета содержит записи ссуд. Каждая запись содержит уникальное значение, а именно, номер ссуды. Записи в отчете отсортированы по номеру ссуды. В качестве индексов мы используем следующие части информации:

- Дата записи
- Начальный номер на странице
- Номер ссуды (будет операционным полем)

Обычно триггеры и поля определяются от левого верха до правого низа отчета. Это упрощает процедуру создания параметров индексатора.

Пример страницы отчета показан на рис. 4 на стр. 12.

ОТЧЕТ	D33313001	НАЦИОНАЛЬНЫЙ БАНК ONDEMAND	ДАТА	01-15-00
БАНК	001		ВРЕМЯ	16:03:46
С	01/01/99		РЕЖИМ	9
ПО	12/31/99	ОТЧЕТ О НЕПОГАШЕННЫХ ССУДАХ	СТР	0001

НОМЕР ССУДЫ	ИМЯ КЛИЕНТА	СУММА ССУДЫ	ЗАДЕРЖКА 30 ДНЕЙ	ЗАДЕРЖКА 60 ДНЕЙ	ЗАДЕРЖКА 90 ДНЕЙ
0100000000	AARON, ROBERT	\$10000000.00	\$ 50.00	\$ 50.00	\$.00
0100000001	ABBOTT, DAVID	\$ 11000.00	\$ 100.00	\$ 200.00	\$.00
0100000002	ABBOTT, DAVID	\$ 12000.00	\$ 140.00	\$.00	\$.00
0100000003	ABBOTT, DAVID	\$ 13000.00	\$ 150.00	\$.00	\$.00
0100000005	ROBINS, STEVEN	\$ 500.00	\$ 50.00	\$.00	\$.00
0100000006	ARNOLD, SAMUEL	\$ 1000.00	\$ 75.00	\$ 150.00	\$ 225.00
0100000007	PETERS, PAUL	\$ 650.00	\$ 50.00	\$.00	\$.00
0100000008	ROBERTS, ABRAHAM	\$ 9000.00	\$ 120.00	\$.00	\$.00
0100000009	SMITH, RANDOLPH	\$ 8000.00	\$ 115.00	\$.00	\$.00
0100000010	KLINE, PETER	\$ 8500.00	\$ 110.00	\$.00	\$.00

Рисунок 4. Пример отчета с операционными полями

Сначала запустите администрирующего клиента функции OnDemand и зарегистрируйтесь на сервере вашего экземпляра. Затем нажмите кнопку Мастер отчетов панели инструментов. Выберите тип данных, например выберите SCS. Затем выберите образец входного файла. Графический индекса́тор покажет буферный файл.

Определите первый триггер. Для триггера 1 выберите слово ОТЧЕТ. Этот триггер будет использоваться для определения начала документа и поиска полей Дата отчета и Начальный номер на странице.

В этом примере требуется только триггер 1. Теперь определите поля и индексы. С помощью мастера отчетов поля и индексы определяются на одном этапе. Если же используется графический индекса́тор из определения приложения, то поля и индексы определяются на разных этапах.

Первые поле и индекс определите для даты отчета. Дата отчета находится триггером 1. На странице Атрибуты полей базы данных дата отчета определяется как данные типа данные и выбирается как поле сегмента.

Вторые поле и индекс определите для начального номера на странице. Начальный номер на странице находится триггером 1. На странице Атрибуты полей базы данных начальный номер на странице определяется как данные целочисленного типа.

После определения всех полей необходимо изменить начальный номер на странице, чтобы новая группа не создавалась при каждом изменении номера страницы.

Вторые поле и индекс определите для номера ссуды. Номер ссуды находится с помощью маски. Параметр Маска служит для задания шаблона для данных операционного поля, чтобы они использовались как индекс. В нашем примере поле должно состоять из десяти численных знаков (каждый символ # представляет один численный знак). Операционное поле не применяет триггер для поиска данных; оно использует маску для определения структуры данных и принимает все данные на странице, которые совпадают с маской.

Страница Атрибуты полей базы данных имеет специальные параметры для поддержки операционного поля. Конечный пользователь примера отчета увидит имена полей папок. Имена полей базы данных OnDemand скрыты от конечных пользователей.

Конечный пользователь заносит критерии поиска (номер ссуды) в поле, которое идентифицируется Полем запроса папки. В списке документов будут присутствовать два номера ссуды. Это начальный и конечный ссуд в группе отчета, которая содержит искомый номер ссуды.

Номер ссуды определяется как данные типа строка.

Теперь необходимо вернуться назад и определить то, чтобы изменение значения начального номера на странице, которым является индекс 2, не приводило к созданию начала новой группы документа. На панели инструментов нажмите кнопку Переключение выбора Триггер, Индекс, Параметры полей.

Администрирующий клиент открывает окно диалога Выбор.

Щелкните на индексе 2. Затем нажмите кнопку Свойства для вызова окна диалога Обновление индекса.

Для параметра Разрыв выберите опцию Нет. Для сохранения начального номера на странице как индекса без разрыва нажмите кнопку ОК. Теперь изменение начального номера на странице не будет приводить к созданию новой группы документа.

Закройте окно диалога Выбор.

Для проверки, как система будет индексировать документ, на панели инструментов нажмите кнопку Переключение между Просмотр и Добавление параметров.

Определенные триггеры будут выделены красным. Определенные поля будут выделены синим. Определенные операционные поля будут выделены зеленым.

Теперь можно закрыть окно графического индексатора и завершить процедуру, воспользовавшись мастером отчетов для определения группы приложений, приложения и папки.

Параметры индексатора, которые были созданы для данного примера отчета, показаны на рис. 5.

```
TRIGGER1=*,2,X'D9C5D7D6D9E3',(TYPE=GROUP) /* ОТЧЕТ */
FIELD1=0,83,8,(TRIGGER=1,BASE=0)
FIELD2=3,87,4,(TRIGGER=1,BASE=0)
FIELD3=*,*,10,(OFFSET=(3:12),MASK='#####',ORDER=BYROW)
INDEX1=X'998481A385',FIELD1,(TYPE=GROUP,BREAK=YES) /* дата_отч */
INDEX2=X'A297818785',FIELD2,(TYPE=GROUP,BREAK=NO) /* нач_ном_стр */
INDEX3=X'D396819540D5A494828599',FIELD3,(TYPE=GROUPRANGE,BREAK=NO) /* Номер ссуды */
```

Рисунок 5. Параметры индексатора с операционными полями

После архивирования примера отчета можно запустить OnDemand Client, открыть новую папку и выполнить поиск документов.

Назначение значений индексов по умолчанию

Вы можете создать определение приложения OnDemand с полем индекса, которое не всегда присутствует на печатной странице. Если при индексации ни одно значение поля не находится (то есть, если находятся только пробелы или расположение поля отсутствует на конкретной печатной странице), тогда ключевое слово DEFAULT позволяет определить значение, используемое по умолчанию. Ключевое слово DEFAULT можно поместить в строку параметра индекса FIELD параметров индекса для конкретного определения приложения.

Ключевое слово DEFAULT можно задать двумя способами. Первый способ позволяет задать некоторое фактическое значение (в алфавитно-цифровом или шестнадцатеричном формате). Второй способ позволяет использовать значение по умолчанию, заданное на вкладке Загрузка информации определения приложения OnDemand и распространения индекса (описывается далее).

Примеры первого способа:

```
DEFAULT='ваше_значение' (например, DEFAULT='ABC')
```

или

```
DEFAULT=x'ваше_шестн_значение' (например, DEFAULT=x'C1C2C3')
```

Примеры второго способа:

```
DEFAULT='_*USELOADDEFAULTORPROPAGATION'
```

или

```
DEFAULT=x'6D5CE4E2C5D3D6C1C4C4C5C6C1E4D3E3D6D9D7D9D6D7C1C7C1E3C9D6D55C6D' (В этом случае заданное шестнадцатеричное значение является шестнадцатеричным представлением строки символов *_USELOADDEFAULTORPROPAGATION*_.)
```

Второй способ (использование *_USELOADDEFAULTORPROPAGATION*_ или его шестнадцатеричного представления) предусматривает процесс загрузки с присвоением значения по умолчанию с вкладки Загрузка информации определения приложения или для распространения индекса. Для того чтобы значение по умолчанию присваивалось в процессе загрузки с вкладки Загрузка информации, необходимо задать это значение с помощью OnDemand Administrator Client. Если значение по умолчанию не будет задано, начнется распространение. Распространение представляет собой процедуру переноса найденного предыдущего значения. Это удобно, но может дать непредвиденные результаты. Например, если в поле был номер клиента, то это значение номера клиента будет перенесено из предыдущего документа, если оно не было задано для текущего документа. Такой результат может оказаться нежелательным. Будьте внимательны при применении второго способа, так как может начаться распространение индекса.

Определение полей поиска текста

Функция поиска текста служит для поиска документов, содержащих заданное слово или фразу, которая не была определена в качестве поля индекса для этих документов. Сначала для поиска документа используются заданные значения полей индекса. После этого документ, соответствующий критериям полей индекса, подвергается поиску заданного слова или фразы. Например, если другими полями индекса являются дата и номер счета, то заданное слово или фраза будет искаться только в документах с этими датой и номером счета. Если документ содержит это заданное слово или фразу, он добавляется в список документов.

Примечания:

1. Для каждой папки можно определить только одно поле поиска текста.
2. Единственным доступным поисковым оператором для поля поиска текста является EQUAL.
3. Поле поиска текста не поддерживает подстановочные символы и маски.
4. Регистр заданного слова или фразы не учитывается. Например, фраза *клиент абв* в поиске является тем же самым, что и *клиент абв*, *Клиент Абв* и *КЛИЕНТ АБВ*.

Функция поиска текста полностью выполняется сервером System i5. Производительность будет зависеть от размера и количества документов, по которым осуществляется поиск, и от быстродействия нагруженной системы. Для ограничения количества документов пользователям следует задавать критерии для некоторых или всех полей индекса.

Для создания поля поиска текста в определении папки OnDemand выполните следующее:

1. С помощью мастера отчетов создайте группу приложений, приложение и папку. (В мастере отчетов не предусмотрено создание поля поиска текста. Однако это делается выполнением всего лишь нескольких действий вне мастера отчетов.
2. Скопируйте папку.
3. Измените имя новой папки.
4. На вкладке **Определение полей** добавьте поле с именем **Поиск полного текста** и выберите тип поля **Поиск текста**. Для добавления поля нажмите кнопку **Добавить**.
5. Нажмите **ОК** для сохранения новой папки.

Вы можете удалить ранее созданную мастером отчетов папку и всегда пользоваться новой папкой, которая содержит поле **Поиск текста**. После архивирования нескольких документов в группе приложений можно попробовать применить функцию поиска текста.

Для усиления функции поиска текста вы можете задать ряд опций в пределах OnDemand Client:

- В меню **Опции** выберите опцию **Показывать искомую строку**. При выборе этой опции введенная для поиска текстовая строка будет выделяться в открытом документе.
- При установке опции **Автопросмотр на Первый документ** или **Отдельный документ** документ будет показываться с автоматически выделенной искомой текстовой строкой. Опция **Отдельный документ** автоматически покажет документ только в случае, если только один документ удовлетворяет

критериям поиска. Опция Первый документ автоматически покажет первый документ из списка документов независимо от того, сколько документов отвечают критериям поиска.

Когда вы будете готовы к тому, чтобы проверить ваше поле поиска текста, откройте папку, содержащую поле поиска текста и выполните поиск текста. Искомая строка текста может содержать одно или несколько слов. Откройте один из документов в списке. В открытом документе искомая строка должна быть выделена. Кнопка Найти следующей панели инструментов позволяет найти следующее совпадение для искомой строки в документе. В пределах этой папки также можно выполнять стандартные поиски, то есть вам не нужно будет каждый раз задавать искомый в документах текст.

Для использования функции поиска текста в документах AFP или расширенной SCS необходимо установить приложение Portable Application Solutions Environment (PASE; один из продуктов i5/OS). Если PASE не установлено, то при попытке поиска текста в документах AFP или расширенной SCS вы получите сообщение об ошибке 161 в протоколе системы OnDemand. Для применения функции поиска текста в документах SCS или Line PASE не требуется.

Работа с буферными файлами SCS с приложениями AFP

Для работы с буферными файлами SCS, имеющими приложение AFP в соответствующем файле принтера рекомендуется просто изменить параметр DEVTYPE файла принтера, использовавшегося для создания оригинального буферного файла, на *AFPDS. Система i5/OS будет помещать данные в буфер как *AFPDS, которые для функции OnDemand наиболее эффективны для получения (загрузки) этого типа буферных данных. Однако это изменение будет требовать печати оригинального рабочего буферного файла на принтере AFPDS. В большинстве случаев, если вы действительно выполняете печать с аппликацией, это не представляет проблемы. Однако если вы печатаете на построчном принтере на типографских бланках, этот способ не работает.

Если по какой-то причине вы не можете изменить параметр DEVTYPE оригинального файла принтера на *AFPDS, функция OnDemand может выполнить преобразование в AFP автоматически с обеспечением точного соответствия вида и печати буферного файла. (Этот способ требует больше времени, чем использование параметра DEVTYPE файла принтера системой i5/OS.) Для этого преобразования в определении приложения OnDemand просто задайте тип данных и параметр индекатора DOCTYPE как AFP, а не SCS. Когда функция OnDemand встретит буферный файл *SCS с аппликацией и при этом определение приложения и параметр индекатора DOCTYPE, тип данных для которых задан как AFP, функция OnDemand преобразует данные *SCS в данные *AFPDS и сохранит этот вновь созданный буферный файл *AFPDS. Перепечатки без функции OnDemand потребуют принтер с поддержкой AFP, но только из-за аппликации. В случае задания типа данных как AFP в определениях приложений OnDemand для какого-то другого типа буферного файла (не AFP) загрузку данных выполнить не удастся.

Применение маски при определении полей приложения

Маска задает шаблон символов, с которым программа индексации сопоставляет данные в некотором конкретном поле. С индексатором OS/400 маска может применяться к полю, основанному на триггере, или к операционному полю. Если данные соответствуют маске, индексатор выбирает

это поле. Если данные не соответствуют маске, то данное поле так, как будто триггер или операционное поле не было найдено.

Маска может содержать следующие символы:

- @ Соответствует алфавитно-цифровым знакам
- # Соответствует численным знакам
- = Соответствует любому символу
- ~ Соответствует любому символу, кроме пробела
- ^ Соответствует любому символу, кроме пробела
- % Соответствует пробелу и численным знакам

Например, маска #####.## заставит индекатор выбрать поле с данными (слева направо), состоящими из четырехчисленных знаков с десятичной точкой и двумя последующими численными знаками.

Примером синтаксиса параметра индекатора для поля с маской является следующий параметр:

```
FIELD4=0,-24,7,(TRIGGER=3),BASE=TRIGGER,MASK='#####.##')
```

Примечание: При использовании поля группы на основе триггера для правильного определения поля может потребоваться добавление ключевого слова MASK вручную. В используемой вами версии администрирующего клиента OnDemand с графической утилитой индексирования маски полей групп на основе триггеров могут не поддерживаться. Поддержка масок полей на основе плавающих триггеров была добавлена в версию 7.1.0.6 администрирующего клиента.

Применение логических элементов тегов (TLE)

Логические элементы тегов (TLE) для идентификации данных индекса не требуют специальных переключателей или других специальных настроек. Графический индекатор OnDemand (который привлекается клиентом-администратором OnDemand при определении приложения) автоматически представляет данные TLE вверху каждой страницы для печати перед представлением самих данных, что позволяет использовать данные TLE точно так же, как сами данные для печати, для получения информации об индексе (такой как номер клиента или номер накладной).

Пример данных, представляемых в графическом индексаторе клиента-администратора OnDemand при работе с TLE в буферном файле AFPDS, показан далее. Четыре строки вверху сразу после строки начала группы представляют информацию TLE. Поток данных AFP *text* должен кодироваться в EBCDIC, а не в ASCII. То же относится к TLE.

```
*GROUP_START          113928
Номер накладной       113928
Дата заполнения       06/15/07
Номер клиента         44332
Сумма платежа        2,859.36 доллара
```

```
ABC COMPANY
101 Plagioclase Blvd.
Deva Station          VA 55564
```

```
528 555-1234
```

ДАТА ОТГРУЗКИ 04/07/73
Dewey Cheatham & Howe
P.O. Box 47899
Ridiculous TN 79832

НОМЕР КЛИЕНТА 44332

НОМЕР ЗАКАЗА - C3050279

17 ИГУАНЫ	3.23	0.11	77.34
93 РОЖКИ ДЛЯ ОБУВИ	18.95	13.13	127.83
55 ВИЛКИ	43.43	9.23	239.01
55 ЗАВЯЗКИ	97.00	43.83	4,721.64
93 ШНУРКИ	0.54	2.32	14.12
09 ЗУБОЧИСТКИ	53.00	19.91	102.43
			5282.37

Часть 2. Индексатор PDF

В этой части описывается индексатор PDF функции OnDemand. Индексатор PDF может применяться для задания параметров индексирования входных файлов Adobe PDF, которые вы хотите хранить в системе.

Глава 2. Обзор

Что такое индекатор PDF?

Индексатор OnDemand PDF - это программа, с помощью которой можно извлекать данные индексирования из файлов Adobe PDF или создавать эти данные для таких файлов. Данные индексирования позволяют расширить возможности хранения, получения и просмотра документов PDF в OnDemand. Индексатор PDF поддерживает потоки входных и выходных данных PDF версии 1.3. Дополнительную информацию о потоке данных PDF см. в *Portable Document Format Reference Manual*, опубликованном Adobe Systems Incorporated. Компания Adobe также предлагает электронную информацию вместе с продуктами Acrobat Exchange и Acrobat Distiller, включая электронные руководства по приложениям Adobe Capture, PDFWriter, Distiller и Exchange.

Вы определяете и сохраняете документы PDF на сервере с помощью стандартных OnDemand функций. Для этого необходимо определить приложение и группы приложений функции OnDemand. Как часть приложения необходимо определить параметры индексирования, применяемые индексатором PDF для обработки входных файлов. Вы можете автоматизировать индексирование и загрузку данных с помощью специальных параметров команд ADDRPTOND (используя *STMF для параметра INPUT) или STRMONOND (используя *DIR для параметра TYPE) или программы ARSLOAD API. Дополнительная информация о командах ADDRPTOND и STRMONOND содержится в приложении Справочник команд в книге *IBM Content Manager OnDemand для общего сервера i5/OS: Руководство по администрированию*. Дополнительная информация о программе ARSLOAD API и ее параметрах содержится в приложении Справочник API в книге *IBM Content Manager OnDemand для общего сервера i5/OS: Руководство по администрированию*.

После индексирования и сохранения входных файлов с помощью функции OnDemand вы используете клиентскую программу OnDemand для просмотра документов (или документа) PDF, созданных в процессе индексирования и загрузки. Из клиентской программы OnDemand также можно печатать страницы просматриваемого документа PDF.

рис. 6 на стр. 22 представляет пример процесса индексирования и загрузки входных файлов PDF.

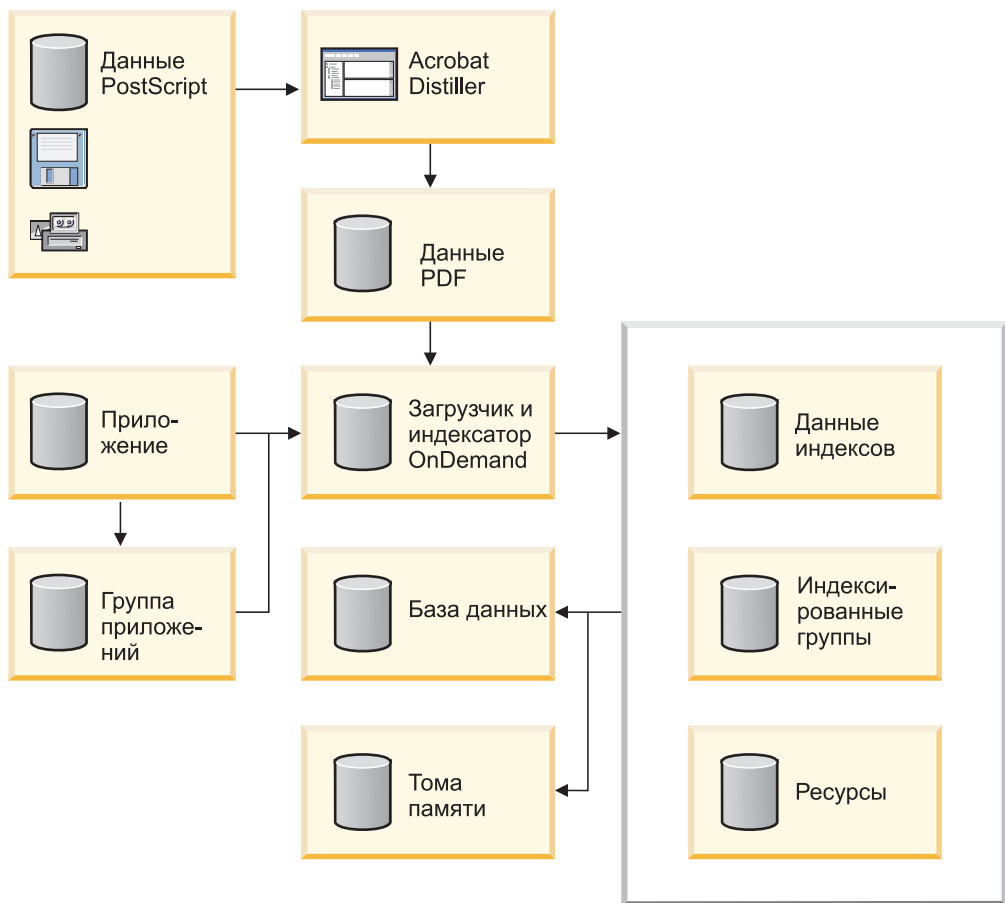


Рисунок 6. Обработка входных файлов PDF в OnDemand

Индексатор PDF обрабатывает входные файлы PDF. Файл PDF - это преобразованный файл PostScript, который обладает структурой и позволяет более эффективно работать с содержимым.

Функция OnDemand получает информацию об обработке из определений приложения и группы приложений, хранящихся в базе данных. Определение приложения содержит тип входных данных, применяемую программу индексирования для входных файлов, параметры индексирования и другую информацию о входных данных. Группа приложений определяет базу данных и характеристики управления памятью, где хранятся данные. Для создания приложения и задания параметров индексирования можно использовать административный клиент.

При обработке входного файла PDF и задании индексатора PDF на странице Информация об индексировании приложения функция OnDemand автоматически вызывает индексатор PDF. Индексатор PDF обрабатывает входной файл PDF, применяя параметры индексирования, которые определяют расположение и атрибуты индексируемых данных. Индексатор PDF получает данные индексирования из файла PDF и создает файл индекса и выходной файл. Выходной файл содержит группы индексированных страниц. Группа индексированных страниц может представлять весь входной файл или, чаще всего, одну или несколько страниц входного файла. Если входной файл содержит логические группы страниц, такие как выписки со счетов или страховые полисы, то индексатор PDF может создать проиндексированную группу для каждой выписки или полиса во входном файле. После этого

пользователи могут вызвать конкретную выписку с банковского счета или набор выписок, а не весь файл. После индексирования данных функция OnDemand сохраняет данные индексирования и проиндексированные группы на диске или в архивных томах хранения.

Как OnDemand использует информацию об индексировании

Каждый элемент данных, сохраняемый в OnDemand, индексируется одним или несколькими индексами на *уровне групп*. Группы определяются при изменении значения индекса (например, номера счета). При загрузке файла PDF в систему функция OnDemand вызывает индекатор PDF для обработки параметров индексирования и создает данные индексирования. Затем функция OnDemand загружает данные индексирования в базу данных, сохраняя значения атрибутов для групп, которые программа индексирования PDF извлекла из данных, в соответствующих полях базы данных. рис. 7 представляет процедуру создания индекса и загрузки данных.



Рисунок 7. Индексирование и загрузка данных

Для каждого отчета, который планируется хранить в OnDemand обычно создается приложение. При создании приложения определяются параметры индексирования, которые программа индексирования использует для обработки отчета и создания данных индексирования, которые загружаются в базу данных. Например, параметр INDEX включает имя атрибута и определяет параметр FIELD, который программа индексирования использует для нахождения значения атрибута во входных данных. При создании приложения его необходимо отнести к некоторой группе приложений. Имя атрибута, задаваемого в параметре INDEX, должно совпадать с именем поля базы данных группы приложений, в котором функция OnDemand будет хранить значения индексов.

Поля базы данных определяются при создании группы приложений. OnDemand создает столбец в таблице группы приложений для каждого определяемого поля данных. При индексировании отчета вы создаете данные индексирования, которые содержат имена полей индекса и значения индексов, полученные из отчета. Функция OnDemand сохраняет данные индексирования в полях базы данных.

Для поиска отчетов, хранящихся в OnDemand, пользователь открывает папку. Появляющиеся при открывании пользователем папки поисковые поля соответствуют полям базы данных в группе приложений (которые, в свою очередь, представляют имена атрибутов индексирования). Пользователь формирует запрос, вводя значения в одно или несколько поисковых полей.

OnDemand просматривает базу данных на элементы, которые содержат значения (значения атрибутов индекса), соответствующие значениям, введенным пользователем. Каждый элемент содержит информацию об индексе группы. OnDemand создает список элементов, соответствующих запросу. Когда пользователь выбирает элемент для просмотра, клиентская программа OnDemand вызывает выбранный элемент с диска или архивной памяти.

Обработка входных файлов PDF с помощью графического индексатора

В этом разделе описывается применение графического индексатора для создания информации об индексировании для входных файлов PDF.

Важная информация: Если вы планируете обрабатывать файлы ввода PDF с помощью мастера отчетов или графического индексатора, то необходимо предварительно установить программу Adobe Acrobat на компьютере, на котором будет запускаться административный клиент. Программу Adobe Acrobat необходимо приобрести у компании Adobe или другого поставщика программных продуктов.

Функция OnDemand содержит файл ARSPDF32.API, который позволяет просматривать документы PDF из клиентской программы. Если клиентская программа устанавливается после установки программы Adobe Acrobat, то программа установки скопирует файл API в каталог встраиваемых модулей Acrobat. Если клиентская программа была установлена до установки программы Adobe Acrobat, то необходимо скопировать файл API в каталог встраиваемых модулей Acrobat. При установке более новой версии приложения Acrobat необходимо скопировать файл API в новый каталог встраиваемых модулей Acrobat. В стандартном варианте файл API находится в каталоге \Program Files\IBM\OnDemand32\PDF. Стандартным каталогом встраиваемых модулей Acrobat является \Program Files\Adobe\Acrobat x.y\Acrobat\Plug_ins, где x.y - версия приложения Acrobat, например 4.0, 5.0 и т.д.

Начиная с версии 5.2, информацию об индексировании можно определять как среду визуализации. Процедура начинается с открывания образца входного файла графическим индексатором. (**Примечание:** При использовании графического индексатора PDF входным файлом может быть только локальный файл. (Графический индексатор PDF предназначен для работы с файлами PDF на рабочей станции, но не с буферными файлами PDF в очереди вывода на сервере System i5.) Графический индексатор можно запустить из мастера отчетов или выбором опции образца данных на странице Информация об индексировании приложения. После открывания входного файла в графическом индексаторе определяются триггеры, поля и индексы. Индексатор PDF использует триггеры, поля и индексы для поиска начала документа во входных данных и получения значений индексов из входных данных. После определения

триггеров, полей и индексов вы можете сохранить их в приложении, чтобы впоследствии функция OnDemand могла применить их для обработки входных файлов, загружаемых вами в систему.

Триггер, поле или индекс определяется очерчиванием рамкой текстовой строки с помощью мыши и определением свойств. Например, для определения триггера, который обозначает начало документа, вы можете нарисовать рамку вокруг текстовой строки Номер счета на первой странице ведомости во входном файле. Затем, в окне диалога Добавление триггера вы можете принять значение по умолчанию, такое как расположение текстовой строки на странице. При обработке входного файла индексатор PDF пытается найти указанную строку в заданном месте. При условии совпадения индексатор PDF знает, что найдено начало документа. Поля и индексы основываются на расположении триггера.

Открываемый графическим индексатором файл PDF должен содержать представительный образец типа входных данных, которые вы планируете загружать в систему. Например, образец входного файла должен содержать по крайней мере один документ. Хороший образец может содержать несколько документов, чтобы можно было проверить расположение триггеров, полей и индексов более чем на одном документе. Образец входного файла должен содержать информацию, которую необходимо определить как начало документа во входном файле. Также он должен содержать информацию, необходимую для определения индексов. При загрузке входного файла в систему индексатор PDF будет использовать созданную информацию об индексировании для поиска и извлечения значений индексов для каждого документа во входном файле.

Ниже приводится пример применения графического индексатора мастера отчетов для создания информации об индексировании для входного файла. Информация об индексировании содержит триггер, который однозначно определяет начало документа во входном файле, и поля и индексы для каждого документа.

1. Прежде всего запустите администрирующего клиента.
2. Войдите в систему сервера.
3. Запустите мастер отчетов, щелкнув на его значке в панели инструментов. Мастер отчетов откроет окно диалога Образец данных.
4. Выберите пункт Выбор образца данных; откроется окно диалога Открыть.
Примечание: При использовании графического индексатора PDF образцом данных может быть только локальный файл. Графический индексатор PDF предназначен для работы с файлами PDF на рабочей станции, но не с буферными файлами PDF в очереди вывода на сервере System i5.
5. Введите имя или полное имя файла в предусмотренное поле или используйте команду Найти или Обзор, чтобы найти файл.
6. Нажмите кнопку Открыть. Графический индексатор откроет этот входной файл в окне отчета.
7. Для просмотра основного раздела справки по окну отчета нажмите F1. Основной раздел содержит общее описание окна отчета и ссылки на другие разделы с информацией о добавлении триггеров, полей и индексов. В окне Опции и команды вызовите страницу Информация об индексаторе и откройте раздел Команды индексирования. (Для вызова информации о значках панели инструментов также можно воспользоваться инструментом

содержания справки. На странице Информация об индексаторе в подразделе Задачи выберите Добавление триггера (PDF).

8. Закройте открытые разделы справки и вернитесь в окно отчета.
9. Определение триггера.
 - Найдите текстовую строку, которая однозначно определяет начало документа. Например, Номер счета, Номер накладной, Имя клиента и т.д.
 - Мышью начертите рамку вокруг этой текстовой строки. Начните точно с внешнего верхнего левого угла строки. Для этого нажмите и держите кнопку мыши. Перемещайте мышь в направлении нижнего правого угла строки. По мере перемещения мыши графический индексатор чертит рамку пунктирной линией. Когда нужная текстовая строка будет обведена полностью, отпустите кнопку мыши. Графический индексатор выделит текст внутри рамки.
 - Щелкните на значке Определение триггера панели инструментов; откроется окно диалога Добавление триггера. Проверьте атрибуты триггера. Например, выбранная в окне отчета текстовая строка должна показываться в позиции Значение; для Триггера 1 Страницы для поиска должны быть заданы для каждой страницы. Вызовите справку по другим опциям и значениям, которые вы можете задать.
 - Определите триггер нажатием кнопки ОК.
 - Для проверки того, что триггер однозначно определяет начало документа, сначала введите окно отчета в режим просмотра. Щелкните на инструменте Выбор, чтобы открыть окно диалога Выбор. В списке Триггеры дважды щелкните на вашем триггере. Графический индексатор выделит тестовую строку в текущем документе. Дважды щелкните на этом триггере. Графический индексатор должен выделить текстовую строку на первой странице следующего документа. С помощью окна диалога Выбор пройдите первые страницы всех документов и вернитесь к первому документу во входном файле.
 - Введите окно отчета в режим добавления.
10. Определение поля и индекса.
 - Найдите текстовую строку для задания расположения поля. Эта текстовая строка должна содержать образец значения индекса. Например, если вы хотите получать значения номеров счетов из входного файла, то найдите, в каком месте страницы печатается номер счета.
 - Мышью начертите рамку вокруг этой текстовой строки. Начните точно с внешнего верхнего левого угла строки. Для этого нажмите и держите кнопку мыши. Перемещайте мышь в направлении нижнего правого угла строки. По мере перемещения мыши графический индексатор чертит рамку пунктирной линией. Когда нужная текстовая строка будет обведена полностью, отпустите кнопку мыши. Графический индексатор выделит текст внутри рамки.
 - Щелкните на значке Определение поля панели инструментов; откроется окно диалога Добавление поля.
 - На странице Информация о поле проверьте атрибуты поля индекса. Например, выбранная в окне отчета текстовая строка должна показываться в позиции Ссылочная строка; Триггер должен определять триггер, на котором основано это поле. Вызовите справку по опциям и значениям, которые вы можете задать.
 - На странице Атрибуты полей базы данных проверьте атрибуты поля поля базы данных. В поле Имя поля базы данных введите имя поля

группы приложений, в котором функция OnDemand должна хранить значение индекса. В поле имя поля папки введите, имя поля папки, которое будет присутствовать на поисковом экране клиента. Вызовите справку по другим опциям и значениям, которые вы можете задать.

- Определите поле и индекс нажатием кнопки ОК.
 - Для проверки расположений полей сначала введите окно отчета в режим просмотра. Поля должны быть очерчены синей рамкой. Затем щелкните на инструменте Выбор, чтобы открыть окно диалога Выбор. В списке полей дважды щелкните на поле 1. Графический индексатор выделит текстовую строку в текущем документе. Снова дважды щелкните на полей 1. Графический индексатор должен перейти на следующий документ и выделить ту же текстовую строку. С помощью окна диалога Выбор пройдите все документы и просмотрите поле. Вернитесь к первому документу во входном файле.
 - Введите окно отчета в режим добавления.
11. Щелкните на инструменте Просмотр параметров индексатора; появится окно диалога Просмотр параметров индексатора. Оно содержит списки параметров индексирования, которые индексатор PDF будет применять для обработки входных файлов, загружаемых в приложение. Как минимум, требуются один триггер, одно поле и один индекс. Дополнительная информация о параметрах индексирования приведена в разделе Глава 4, “Справочник параметров”, на стр. 35.
 12. После завершения определения всех триггеров, полей и индексов закройте окно отчета.
 13. Нажмите кнопку Да для сохранения параметров индексатора.
 14. В окне Образец данных нажмите кнопку Далее, чтобы продолжить работу с мастером отчетов.

Индексирование входных данных вручную

Примечание: Если вы предпочитаете создать параметры индексирования PDF вручную, а не с помощью графического индексатора PDF, то для этого вы можете использовать команды, описываемые в остальной части этой главы.

Принципы индексирования

Параметры индексирования включают в себя информацию, которая позволяет Индексатор PDF определить ключевые элементы в потоке данных на печать, *отметить* эти элементы и создать *элементы индекса*, указывающие на эти отмеченные элементы. Функция OnDemand использует эти метку и данные индексирования для эффективного, структурированного поиска и получения данных. Информация об индексировании позволяет Индексатор PDF разбить поток данных на отдельные элементы, называемые *группами*. Группа является набором из одной или нескольких страниц, таких как выписка с банковского счета, страховой полис, счет за телефон или другой логический сегмент отчета. Когда значение индекса изменяется (например, номер счета), индексатор PDF создает индексы для каждой группы.

Метка состоит из *имени атрибута*, например Имя клиента, и *значения атрибута*, например Эрл Хокинс. Метки также включают информацию о расположении значения атрибута на странице для индексатора PDF. Например, метка для сбора значений индексов имен клиентов указывает индексатору PDF начальное

и конечное положения значений индекса имени клиента на странице. Индексатор PDF создает данные и индексирования и сохраняет их в базовом файле индексов.

Система координат

Расположение текстовых строк индексатор PDF использует для определения начала группы, а значения индексов описываются как координаты x и y системы координат, наложенной на страницу. Для каждой текстовой строки определяются верхнее левое и нижнее правое положения на странице. Верхний левый угол и нижний правый угол образуют рамку строки. Это наименьший прямоугольник, который полностью заключает в себе текстовую строку. Начало находится в верхнем левом углу страницы. Координата x увеличивается вправо, а координата y - вниз по странице. Также указывается страница, на которой находится текстовая строка. Например, текстовая строка Имя клиента, которая начинается на 4 дюйма вправо и 1 дюйм вниз и заканчивается на 5,5 дюйма вправо и 1,5 дюйма вниз на первой странице входного файла, может быть описана следующим образом:

```
u1(4,1),lr(5.5,1.5),1,'Имя клиента'
```

Функция OnDemand имеет команду ARSPDUMP, которая помогает задать расположения текстовых строк на странице. Дополнительная информация о команде ARSPDUMP приведена в разделе Глава 7, “Справочник по программе ARSPDUMP”, на стр. 51.

Параметры индексирования

Параметры обработки могут содержать параметры индексирования и преобразования, опции и значения. Для создания данных индексирования : большинства отчетов индексатору PDF требуются по крайней мере три параметра.

- TRIGGER

Функция OnDemand использует триггеры для определения. По триггеру индексатор PDF ищет конкретную информацию в конкретном месте на странице. После нахождения текстовой строки во входном файле, которая содержит информацию, определенную в триггере, индексатор PDF может начать просмотр информации об индексировании.

- Индексатор PDF сравнивает слова во входном файле с текстовой строкой, заданной в триггере.
- Расположение значения строки триггера должно определяться с помощью системы координат x, y и сдвигов страниц.
- Можно задать максимум 16 триггеров.
- Для того чтобы индексатор PDF мог начать поиск информации об индексе, все триггеры должны совпасть.

- FIELD

Параметр поля определяет расположение данных, которые индексатор PDF использует для создания значений индекса.

- По умолчанию определения полей основываются на TRIGGER1, но могут основываться на любом из 16 параметров TRIGGER.
- Расположение поля должно определяться с помощью системы координат x, y и сдвигов страниц.
- Можно задать максимум 32 поля.
- Параметр поля также может определять все или часть фактического значения индекса, хранящегося в базе данных.

- INDEX

В параметре индекса определяются имя атрибута и поле или поля, на которых основан индекс. Мы настоятельно рекомендуем присваивать атрибуту то же имя, что и у поля базы данных группы приложений.

- Индексатор PDF создает индексы для группы из одной или нескольких страниц.
- Для формирования индекса параметры полей можно объединять.
- Можно задать максимум 32 параметра индексов.

Индексатор PDF создает новую группу и получает новые значения индекса, когда одно или несколько значений индексов изменяются.

рис. 8 представляет часть страницы образца входного файла. Текстовые строки, определяющие начало группы, и значения индексов заключены в прямоугольные рамки.

0,25	0,75	0,25	1,00	0,75	0,50
0,25	Стр. 001				
0,75					
0,25		Иван Кузнецов			
0,75		Ул. Ермака, 12			
0,25		Змеиногорск, 123456			
0,75					
0,25		Дата:	31/08/2003		
0,25		Номер счета:	0000-3727-1644-0099		
0,50					
0,25		Баланс:	\$1 096,54		

Рисунок 8. Данные индексирования для индексатора PDF

Параметры TRIGGER показывают индексатору PDF, как найти начало группы во входных данных. Для нахождения начала группы (выписки со счета) в данном образце файла индексатору PDF требуется один параметр TRIGGER. Параметры FIELD показывают расположение значений индекса в выписках со счета. Поля основываются на расположении записей триггеров. Параметры INDEX задают имена атрибутов полей индексов. Индексы основываются на

одном или нескольких параметрах полей. Параметры, которые могут использоваться для индексирования отчета, приведены в рис. 8 на стр. 29. Описание синтаксиса параметров приведено в разделе Глава 4, “Справочник параметров”, на стр. 35.

- Определите триггер для поиска текстовой строки на каждой странице входных данных, которая определяет начало группы (выписки со счета).
`TRIGGER1=u1(0,0),lr(.75,.25),*, 'Страница 001'`
- Определите поля для обозначения расположения данных индекса. Для данного образца отчета можно определить четыре поля:
 - Поле FIELD1 определяет расположение значений индекса Имя клиента.
`FIELD1=u1(1,1),lr(2,1.25),0`
 - Поле FIELD2 определяет расположение значений индекса Дата выписки.
`FIELD2=u1(2,2),lr(2.75,2.25),0`
 - Поле FIELD3 определяет расположение значений индекса Номер счета.
`FIELD3=u1(2,2.25),lr(3.25,2.5),0`
 - Поле FIELD4 определяет расположение значений индекса Баланс.
`FIELD4=u1(2,3),lr(2.75,3.25),0`
- Определите индексы для задания имени атрибута значения индекса и параметра поля, используемых для поиска значения индекса.
 - Параметр INDEX1 задает имя клиента для значений, получаемых с помощью параметра FIELD1.
`INDEX1='имя_клиента',FIELD1`
 - Параметр INDEX2 задает дату выписки для значений, получаемых с помощью параметра FIELD2.
`INDEX2='дата_вып',FIELD2`
 - Параметр INDEX3 задает номер счета для значений, получаемых с помощью параметра FIELD3.
`INDEX3='ном_счета',FIELD3`
 - Параметр INDEX4 задает баланс для значений, получаемых с помощью параметра FIELD4.
`INDEX4='баланс',FIELD4`

Как создать параметры индексирования

Создание параметров индексирования состоит из двух частей. Сначала образец входных данных обрабатывается для задания координат *x,y* текстовых строк, которые индексатор PDF использует для определения групп и поиска данных индексов. После этого с помощью административный клиент создаются параметры индексирования..

Функция OnDemand имеет команду ARSPDUMP, которая помогает определить расположение триггера и значения строк полей во входных данных. Команда ARSPDUMP обрабатывает одну или несколько страниц образца отчетных данных и создает выходной файл. Выходной файл содержит одну запись для каждой текстовой строки на странице. Каждая запись содержит координаты *x,y* рамки, заключающей в себе текстовую строку (верхний левый угол, нижний правый угол). Дополнительная информация о команде ARSPDUMP приведена в разделе Глава 7, “Справочник по программе ARSPDUMP”, на стр. 51.

Процедура реализуется следующим образом:

- Получите печатную копию образца отчета.

- Определите значения строк для поиска триггеров и полей.
- Укажите номер страницы, на которой присутствует каждое значение строки. Номером является *номер листа*, а не идентификатор страницы. Номером листа является порядковый номер страницы в файле, начиная с номера 1 (один) для первой страницы в файле. Идентификатором страницы является пользовательское обозначение каждой страницы (например, iv, 5 или 17-3).
- Обработайте одну или несколько страниц командой ARSPDUMP.
- В выходном файле найдите записи, которые содержат значения строк, и запишите координаты x,y.
- Используя координаты x,y, номер страницы и значение строки, создайте параметры TRIGGER и FIELD.

Параметры индексирования являются частью приложения OnDemand. административный клиент содержит окно редактирования для определения параметров индексирования для приложения.

Глава 3. Системные особенности

Системные ограничения

При использовании индексатора PDF для создания данных индексирования для файлов PostScript и PDF, создаваемых пользовательскими программами, необходимо учитывать следующее:

- Индексатор PDF может обрабатывать входные файлы PDF, содержащие приблизительно до 3000 страниц, независимо от размера файла. Это максимальное значение изменяется (увеличивается или уменьшается) в зависимости от типа и числа шрифтов и изображений, содержащихся в файле.
- IBM рекомендует, чтобы CCSID входного файла PDF был 1252 (WinAnsiEncoding). Применение другого CCSID может привести к непредсказуемым результатам.
- Индексатор PDF поддерживает языки. Однако IBM не предоставляет шрифты DBCS. Шрифты DBCS можно приобрести у компании Adobe. Индексатор PDF поддерживает все шрифты DBCS, кроме закодированных японских шрифтов.
- Входные данные, выделенные маркерами PostScript Passthrough, индексироваться не могут.
- Пакет Adobe Toolkit не проверяет ссылки или закладки, использующие другие страницы документа или другие документы. Правильность обработки ссылок или закладок зависит от разбивки ваших документов.
- При ссылке на шрифт во входном файле, который отсутствует в нем, и невозможности нахождения этого шрифта индексатором PDF запрошенный шрифт заменяется одним из базовых шрифтов Adobe Type 1, предоставляемых IBM. Если вы приобрели дополнительные шрифты и установили их в систему, то они могут быть встроены в процессе индексирования, если они используются во входном файле и определены параметром FONTLIB. Дополнительная информация приведена в разделе "FONTLIB" на стр. 39.

Требования к входным данным

Индексатор PDF обрабатывает входные данные PDF. Данные PostScript, создаваемые приложениями, перед применением индексатора PDF должны обрабатываться приложением Acrobat Distiller. Электронная документация, прилагающаяся к продукту Acrobat Distiller, содержит описания способов создания данных PDF.

Для автоматизации процесса индексирования и загрузки данных на сервер OnDemand имя входного файла должно определять группу приложений и приложение для загрузки. При присвоении имени входному файлу соблюдайте следующее соглашение:

```
MVS.JOBNAME.DATASET.FORM.YYDDD.HHMMSS.PDF
```

Важная информация: Для запуска процесса загрузки требуется расширение имени файла .PDF.

По умолчанию для определения группы приложений для загрузки программа ARSLOAD использует часть FORM имени файла. Однако для задания другой части имени файла (MVS, JOBNAME или DATASET) для этой цели можно воспользоваться параметром **-G**. Например, `arsload -G JOBNAME`.

Если группа приложений содержит несколько приложений, то необходимо указать приложение для загрузки; в противном случае загрузка не будет выполнена. Для указания части имени файла (MVS, JOBNAME, DATASET или FORM), определяющей приложение для загрузки, можно запустить программу ARSLOAD с параметром **-A**. Например, `arsload -A DATASET`.

Регистр идентификатора PDF не учитывается. В именах группы приложений и приложения регистр учитывается, и они могут содержать специальные символы, такие как символ пробела.

Особенности поддержки национальных языков

Индексатор PDF поддерживает языки DBCS. Однако IBM не предоставляет шрифты DBCS. Шрифты DBCS можно приобрести у компании Adobe. Индексатор PDF поддерживает все шрифты DBCS, кроме закодированных японских шрифтов.

Значения данных в параметрах TRIGGER и FIELD должны кодироваться с применением той же кодовой страницы, что и документ. Например, если документ кодируется с применением кодовой страницы 1252, то к значениям данных в параметрах TRIGGER и FIELD также должна применяться кодовая страница 1252. Примерами значений данных, которые вы можете задавать, являются значения строк TRIGGER и стандартное и постоянное значения FIELD.

При загрузке данных с помощью Индексатор PDF локаль должна быть задана соответственно кодовой странице документов. Например, если кодовой страницей документов является 954, установите переменную среды локали на `ja_JP`, или определите другую локаль, которая правильно идентифицирует прописные и строчные символы на кодовой странице 954.

Дополнительную информацию о поддержке национальных языков в OnDemand можно найти в Руководстве по планированию и установке *IBM Content Manager OnDemand для общего сервера i5/OS*.

Глава 4. Справочник параметров

Этот справочник параметров предполагает, что для обработки входных файлов используется программа ARSLOAD. При этом индексатор PDF игнорирует любые значения, которые могли быть заданы для параметров INPUTDD, MSGDD, OUTPUTDD и PARMDD. При запуске программы ARSPDOC1 из командной строки или из пользовательской программы необходимо задать значения для параметров INPUTDD, OUTPUTDD и PARMDD и убедиться, что стандартные значения для параметров INDEXDD и MSGDD верны.

Параметр COORDINATES

Задаёт метрические значения координат x, y в параметрах FIELD и TRIGGER.

Применяется?

Нет

Стандартное значение

IN

Синтаксис

COORDINATES=*метрическое значение*

Опции и значения

Метрическое значение может измеряться в:

IN

Метрические значения координат задаются в дюймах (по умолчанию).

CM

Метрические значения координат задаются в сантиметрах.

MM

Метрические значения координат задаются в миллиметрах.

FIELD

Задаёт расположение данных индекса и может определить стандартное и постоянное значения индекса. Должно быть задано по крайней мере одно поле. Можно задать до 32 полей. Можно определять два типа полей: *поле триггера*, которое основано на расположении значения строки триггера и *поле константы*, которое задаёт фактическое значение индекса, которое хранится в базе данных.

Применяется?

Да

Стандартное значение

<нет

Синтаксис поля триггера

FIELD $n=ul(x,y),lr(x,y),страница[(TRIGGER=n,BASE=\{0 | TRIGGER\},$
MASK='маска_поля',DEFAULT='значение']

Опции и значения

n

Идентификатор параметра поля. При добавлении параметра поля используется следующее доступное число, начиная с 1 (единицы).

ul(x,y)

Координаты верхнего левого угла рамки строки поля. Рамка строки поля - это наименьший прямоугольник, который полностью заключает в себе значение строки поля (одно или несколько слов на странице). Индексатор PDF должен отыскать это значение строки поля в рамке строки поля. Поддерживаются значения от 0 (нуль) до 45, ширина и длина страницы в дюймах.

lr(x,y)

Координаты нижнего правого угла рамки строки поля. Рамка строки поля - это наименьший прямоугольник, который полностью заключает в себе значение строки поля (одно или несколько слов на странице). Индексатор PDF должен отыскать это значение строки поля в рамке строки поля. Поддерживаются значения от 0 (нуль) до 45, ширина и длина страницы в дюймах.

страница

Номер листа, с которого индексатор PDF начинает поиск поля относительно триггера или 0 (нуль), если триггер находится на той же странице. При задании $BASE=0$ значение *страницы* может быть равно от -16 до 16 . Если задается $BASE=TRIGGER$, то значение *страницы* должно быть равно 0 (нулю), которое является номером листа относительно расположения значения строки триггера.

TRIGGER=*n*

Задаёт параметр триггера для поиска поля. Это произвольное ключевое слово, но стандартным является значение TRIGGER1. Вместо *n* подставьте номер соответствующего параметра TRIGGER.

BASE={0|TRIGGER}

Указывает, должен ли индексатор PDF использовать координаты верхнего левого угла рамки строки триггера для поиска поля. Задайте 0 (нуль) или TRIGGER. В случае $BASE=0$ индексатор PDF к координатам рамки строки поля прибавит нуль. В случае $BASE=TRIGGER$ индексатор PDF прибавит координаты верхнего левого угла расположения рамки строки триггера к координатам рамки строки поля. Это произвольное ключевое слово, но стандартным значением является $BASE=0$.

Если данные поля всегда начинаются в конкретном месте на странице, то параметр должен быть $BASE=0$. Параметр $BASE=TRIGGER$ должен использоваться, если поле не всегда располагается в одном и том же месте на странице, но всегда находится на определенном расстоянии от триггера. Это удобно, когда число строк на странице изменяется и приводит к изменению расположения значений полей. Например, пусть заданы следующие параметры:

```
TRIGGER2=ul(4,4),lr(5,8),1,'Итого'  
FIELD2=ul(1,0),lr(2,1),0,(TRIGGER=2,BASE=TRIGGER)
```

Значение строки триггера можно найти в прямоугольнике один на четыре дюйма. Индексатор PDF всегда найдет поле внутри однодюймовой рамки на расстоянии одного дюйма вправо от расположения значения строки триггера. Если индексатор PDF найдет значение строки триггера в рамке $ul(4,4),lr(5,5)$, он попытается найти поле в рамке $ul(5,4),lr(6,5)$. Если индексатор PDF найдет значение строки триггера в рамке $ul(4,6),lr(5,7)$, он попытается найти поле в рамке $ul(5,6),lr(6,7)$.

Примечание: Начиная с версии 5.2, поле, основанное на расположении триггера (BASE=TRIGGER), может задаваться в любом месте на странице, содержащей триггер. До этого поле, основанное на расположении триггера, должно было задаваться справа и ниже верхней левой точки триггера. В новой версии значения x или y могут быть отрицательными, пока получающиеся абсолютные координаты поля прямоугольника строки поля еще находятся в пределах $0 \leq x \leq 45$ и $0 \leq y \leq 45$. Координаты $ul(x,y)$ и $lr(x,y)$ параметра FIELD представляют собой относительные смещения от координат $ul(x,y)$ триггера. Например, пусть прямоугольник строки поля находится в $ul(1,1)$, $lr(2,2)$, которое является абсолютным расположением на странице. Если прямоугольник строки триггера находится в $ul(5,5)$, $lr(7,7)$, то координатами поля будут $ul(-4,-4)$, $lr(-3,-3)$.

MASK='маска_поля'

Комбинация символов, с которой индексатор PDF сравнивает данные в поле. При задании поля с маской параметр INDEX, основанный на этом поле, не может быть связан ни с каким другим полем. Маска может включать следующие символы:

@ Соответствует алфавитным знакам. Пример:

MASK='@@@@@@@@@@@@@'

Заставляет индексатор PDF искать 15-значное буквенное поле, например имя.

Соответствует численным знакам. Пример:

MASK='#####'

Заставляет индексатор PDF искать 10-значное числовое поле, например номер счета.

~ Соответствует любому символу, кроме пробела.

^ Соответствует любому символу, кроме пробела.

% Соответствует пробелу и численным знакам.

= Соответствует любому символу.

Примечание: Строка, задаваемая для маски, может содержать любой символ. Например, пусть задано следующее:

```
TRIGGER2=*,25,'СЧЕТ'  
FIELD2=0,38,11,(TRIGGER=2,BASE=0,MASK='@000-###-#')
```

Индексатор PDF выберет поле, только если данные в столбцах поля содержат одиннадцатисимвольную строку, состоящую из любой буквы, трех нулей, символа черточки, любых четырех чисел, символа черточки и любого числа.

DEFAULT='значение'

Задаёт стандартное значение индекса, когда в пределах координат рамки строки поля нет слов.

Например, предположим, что некоторая прикладная программа создает ведомости, содержащие поле контроля. Поле может содержать значения ГОДЕН и НЕГОДЕН. Однако если для какой-то ведомости нет результатов контроля, программа не должна формировать значение. В этом случае

рамка строки поля должна быть пустой. Для сохранения стандартного значения в базе данных для актов без результатов контроля определим поле следующим образом:

```
FIELD3=u1(8,1),lr(8.5,1.25),1,(DEFAULT='БЕЗ КОНТРОЛЯ')
```

Индексатор PDF присвоит индексу, связанному с полем FIELD3, значение БЕЗ КОНТРОЛЯ, если рамка строки поля пустая.

Примеры

Следующий параметр поля заставляет индексатор PDF искать поле с координатами, заданными для рамки строки поля. Это поле основано на TRIGGER1 и расположено на той же странице, что и TRIGGER1. Зададим BASE=0, поскольку рамка строки поля всегда присутствует в одном месте на странице.

```
TRIGGER1=u1(0,0),lr(.75,.25),*, 'Страница 0001'  
FIELD1=u1(1,1),lr(3.25,1.25),0,(TRIGGER=1,BASE=0)
```

Синтаксис поля константы

FIELD*n*='константа'

Опции и значения

n

Идентификатор параметра поля. При добавлении параметра поля используется следующее доступное число, начиная с 1 (единицы).

'константа'

Буквенное (постоянное) значение строки поля. Это значение индекса, хранящееся в базе данных. Значение константы может быть длиной от 1 до 250 байтов. Индексатор PDF не проверяет тип или содержимое константы.

Примеры

Следующий параметр поля заставляет индексатор PDF сохранять одну и ту же текстовую строку во всех создаваемых им значениях INDEX1.

```
FIELD1='0000000000'  
INDEX1='счет',FIELD1
```

Следующие параметры полей заставляют индексатор PDF объединять значение константы со значением индекса, извлекаемым из данных. Индексатор PDF объединяет значение константы, определенное в параметре FIELD1 со всеми значениями индексов, найденными с помощью параметра FIELD2. Объединенное значение строки сохраняется в базе данных. В этом примере поле номера счета в данных имеет длину 14 байтов. Однако номер счета в базе данных имеет длину 19 байтов. Для добавления пятибайтового префикса ко всем номерам счетов, полученным из данных, используем поле константы (—).

```
FIELD1='0000-'  
FIELD2=u1(2,2),lr(2.5,2.25),0,(TRIGGER=1,BASE=0)  
INDEX1='ном_счета',FIELD1,FIELD2
```

Связанные параметры

Параметр INDEX на странице 39.

Параметр TRIGGER на странице 44.

FONTLIB

Задает каталог или каталоги, в которых хранятся шрифты. Укажите любое достоверное полное имя. Индексатор PDF ищет шрифты по списку путей. При ссылке на шрифт во входном файле, который отсутствует в нем, индексатор PDF пытается найти этот шрифт в каталоге или каталогах, перечисленных в параметре FONTLIB. Если шрифт находится, индексатор PDF добавляет его в выходной файл. При ссылке на шрифт во входном файле и невозможности нахождения этого шрифта Индексатор PDF запрошенный шрифт заменяется одним из базовых шрифтов Adobe Type 1, предоставляемых IBM. Если вы приобрели дополнительные шрифты и установили их в систему, то они могут быть встроены в процессе индексирования, если они используются во входном файле и присутствуют в одном из каталогов, заданных параметром FONTLIB.

Применяется?

Нет

Стандартное значение

/QIBM/ProdData/OnDemand/Adobe/fonts

Синтаксис

FONTLIB=*список_путей*

Опции и значения

Список_путей представляет собой строку из одного или нескольких достоверных полных имен, разделенных двоеточием. Пример:

/QIBM/ProdData/OnDemand/Adobe/fonts:/mycustom/fonts

Индексатор PDF просматривает пути в заданном порядке. Полные имена разделяются двоеточием (:).

INDEX

Задает имя индекса и поле или поля, на которых этот индекс основан. Должен быть задан по крайней мере один параметр индекса. Можно задать до 32 параметров индексов. IBM рекомендует, чтобы при создании параметров индексов имя индекса было таким же, что и имя поля базы данных группы приложений.

Применяется?

Да

Стандартное значение

<нет

Синтаксис

INDEX*n*'*с именем*',FIELD*n*[,...FIELD*n*]

Опции и значения

n

Идентификатор параметра индекса. При добавлении параметра индекса используется следующее доступное число, начиная с 1 (единицы).

'*ИМЯ*'

Определяет имя индекса, связанное с фактическим значением индекса. Например, пусть параметр INDEX1 содержит номера счетов. Строка

ном_счета является значимым именем индекса. Значением индекса INDEX1 является фактический номер счета, например 000123456789.

Именем индекса может быть строка длиной от 1 до 250 байтов. Мы настоятельно рекомендуем присваивать индексу то же имя, что и у поля базы данных группы приложений.

FIELD*nn*

Имя параметра поля или имена параметров полей, которые индекатор PDF использует для поиска индекса. Можно задать до 32 параметров полей.

Имена параметров полей разделяются запятой. Общая длина всех заданных параметров полей не может превышать 250 байтов.

Примеры

Следующий параметр индекса заставляет индекатор PDF создать индексы групп для значений индексов дат (индекатор PDF поддерживает только четырехуровневые индексы). При изменении значения индекса индекатор PDF закрывает текущую группу и начинает новую.

```
INDEX1='дата_акта',FIELD1
```

Следующие параметры индексов заставляют индекатор PDF создать индексы групп для значений имен клиентов и номеров счетов. При изменении значения имени клиента или номера счета индекатор PDF закрывает текущую группу и начинает новую.

```
INDEX1='имя',FIELD1  
INDEX2='ном_счета',FIELD2
```

Связанные параметры

Параметр FIELD на странице 35.

INDEXDD

Определяет имя или полное имя файла объектов индексирования. Индекатор PDF заносит информацию об индексировании в файл объектов индексирования. Если имя файла задано без пути, то индекатор PDF поместит файл объектов индексирования в текущий каталог. Если параметр INDEXDD не задан, то индекатор PDF запишет информацию об индексировании в файл INDEX.

Применяется?

Нет

Примечание: При обработке входных файлов программой ARSLOAD индекатор PDF игнорирует значение, которое могло быть задано для параметра INDEXDD. Если же входные файлы обрабатываются программой ARSPDOC1, то необходимо проверить значение параметра INDEXDD.

Стандартное значение

INDEX

Синтаксис

```
INDEXDD=имя_файла
```

Опции и значения

Имя_файла представляет достоверное имя файла или полное имя файла.

INDEXSTARTBY

Указывает номер страницы, по которой индекатор PDF должен находить первую группу (документ) во входном файле. Первая группа определяется по наличию всех триггеров и полей. Например, пусть заданы следующие параметры:

```
TRIGGER1=u1(4.72,1.28),lr(5.36,1.45),*, 'СЧЕТ'  
TRIGGER2=u1(6.11,1.43),lr(6.79,1.59),1, 'СВЕДЕНИЯ'  
INDEX1='Счет',FIELD1,FIELD2  
FIELD1=u1(6.11,1.29),lr(6.63,1.45),2  
FIELD2=u1(6.69,1.29),lr(7.04,1.45),2  
INDEX2='Итого',FIELD3  
FIELD3=u1(6.11,1.43),lr(6.79,1.59),2  
INDEXSTARTBY=3
```

Слово СЧЕТ должно находиться на странице в месте, описанном TRIGGER1. Слово СВЕДЕНИЯ должно находиться на следующей странице в месте, заданном TRIGGER2. Кроме того, должно быть найдено одно или несколько слов для полей FIELD1, FIELD2 и FIELD3 в местах, указанных FIELD1, FIELD2 и FIELD3, которые располагаются через две страницы после страницы с TRIGGER1.

В этом примере первая группа должна начинаться на странице один, странице два или странице три. Если TRIGGER1 находится на странице один, то TRIGGER2 должен находиться на странице два, а FIELD1, FIELD2 и FIELD3 - на странице три.

Индексатор PDF останавливает обработку, если он не находит первую группу по заданному номеру страницы. Этот параметр произволен, но по умолчанию индекатор PDF должен находить первую группу на первой странице входного файла. Этот параметр удобен, если входной файл содержит начальные страницы. Например, если входной файл содержит две начальные страницы, то вы можете указать номер страницы, на единицу больший числа начальных страниц (INDEXSTARTBY=3), чтобы индекатор PDF прекратил обработку только в том случае, если он не найдет первую группу по третьей странице входных данных.

Примечание: При применении INDEXSTARTBY для пропуска начальных страниц индекатор PDF не копирует непроиндексированные страницы в выходной файл и не сохраняет их в OnDemand. Например, при задании INDEXSTARTBY=3 и нахождении первой группы на странице три страницы один и два не будут скопированы в выходной файл и не будут сохранены в OnDemand. При задании INDEXSTARTBY=3 и нахождении первой группы на странице два страница один не будет скопирована в выходной файл и не будет сохранена в OnDemand.

Применяется?

Нет

Стандартное значение

1

Синтаксис

INDEXSTARTBY=значение

Опции и значения

Значение представляет номер страницы, по которой индекатор PDF должен находить первую группу (документ) во входном файле.

INPUTDD

Определяет имя или полное имя входного файла PDF для обработки индексатором PDF.

Применяется?

Нет

Примечание: При обработке входных файлов программой ARSLOAD индексатор PDF игнорирует значение, которое могло быть задано для параметра INPUTDD. Если же входные файлы обрабатываются программой ARSPDOC1, то необходимо задать значение параметра INPUTDD.

Стандартное значение

<нет

Синтаксис

INPUTDD=*ИМЯ*

Опции и значения

Имя представляет имя файла или полное имя входного файла. Если имя файла задано без пути, то индексатор PDF ищет указанный файл в текущем каталоге.

MSGDD

Задаёт имя или полное имя входного файла для записи сообщений об ошибках индексатором PDF. Если параметр MSGDD не задан, то индексатор PDF выводит сообщения на дисплей (в интерактивном режиме) или заносит в протокол задания (в пакетном режиме).

Применяется?

Нет

Примечание: При обработке входных файлов программой ARSLOAD индексатор PDF игнорирует значение, которое могло быть задано для параметра MSGDD. Если же входные файлы обрабатываются программой ARSPDOC1, то необходимо проверить значение параметра MSGDD.

Стандартное значение

на дисплее (интерактивный режим) или в протоколе задания (пакетный режим), которые иногда относятся к stderr (стандартной ошибке)

Синтаксис

MSGDD=*ИМЯ*

Опции и значения

Имя представляет имя файла или полное имя файла для записи сообщений об ошибках индексатора PDF. Если имя файла задано без пути, то индексатор PDF поместит файл ошибок в текущий каталог.

OUTPUTDD

Определяет имя или полное имя выходного файла.

Применяется?

Нет

Примечание: При обработке входных файлов программой ARSLOAD индексатор PDF игнорирует значение, которое могло быть задано для параметра OUTPUTDD. Если же входные файлы обрабатываются программой ARSPDOC1, то необходимо задать значение параметра OUTPUTDD.

Стандартное значение

<нет

Синтаксис

OUTPUTDD=ИМЯ

Опции и значения

Имя представляет имя файла или полное имя выходного файла. Если имя файла задано без пути, то индексатор PDF поместит выходной файл в текущий каталог.

PARMDD

Определяет имя или полное имя файла, содержащего параметры индексирования, применяемые для обработки входных данных.

Применяется?

Нет

Примечание: При обработке входных файлов программой ARSLOAD индексатор PDF игнорирует значение, которое могло быть задано для параметра PARMDD. Если же входные файлы обрабатываются программой ARSPDOC1, то необходимо задать значение параметра PARMDD.

Стандартное значение

<нет

Синтаксис

PARMDD=ИМЯ

Опции и значения

Имя представляет имя файла или полное имя файла, содержащего параметры индексирования. Если имя файла задано без пути, то индексатор PDF ищет файл в текущем каталоге.

TEMPDIR

Определяет имя каталога для индексатора PDF в качестве временного рабочего пространства.

Применяется?

Нет

Стандартное значение
/arstmp

Синтаксис

TEMPDIR=*каталог*

Опции и значения

Каталог представляет имя доступного каталога.

параметр TRACEDD

| Параметр TRACEDD добавлен в версии 5.3. Дополнительная информация
| приведена в разделе Глава 8, "Трассировщик", на стр. 53.

TRIGGER

| Задаёт расположения и значения строк для однозначного обозначения начала
| группы и расположения и значения строк полей, используемых для определения
| индексов. Необходимо задать по крайней мере один триггер; всего можно
| определить до 16 триггеров. Индексатор PDF не поддерживает триггеры
| плавающего типа (TYPE=FLOAT).

Применяется?

Да

Стандартное значение

<нет

Синтаксис

TRIGGER*n=ul(x,y),lr(x,y),страница,'значение'*

Опции и значения

n

Идентификатор параметра триггера. При добавлении параметра триггера используется следующее доступное число, начиная с 1 (единицы).

ul(x,y)

Координаты верхнего левого угла рамки строки триггера. Рамка строки триггера - это наименьший прямоугольник, который полностью заключает в себе значение строки триггера (одно или несколько слов на странице). Индексатор PDF должен отыскать это значение строки триггера в рамке строки триггера. Поддерживаются значения от 0 (нуль) до 45, ширина и длина страницы в дюймах.

lr(x,y)

Координаты нижнего правого угла рамки строки триггера. Рамка строки триггера - это наименьший прямоугольник, который полностью заключает в себе значение строки триггера (одно или несколько слов на странице). Индексатор PDF должен отыскать это значение строки триггера в рамке строки триггера. Поддерживаются значения от 0 (нуль) до 45, ширина и длина страницы в дюймах.

страница

Номер страницы входного файла, на которой должно находиться значение строки триггера.

- Для TRIGGER1 значением *страницы* должна быть звездочка (*), которая указывает, что значение строки триггера может находиться на любой странице входного файла. Индексатор PDF начинает поиск с первой страницы входного файла. Индексатор PDF продолжает поиск, пока значение строки триггера не будет найдено, не будет достигнуто значение INDEXSTARTBY или до последней страницы входного файла. Если индексатор PDF достигает значения INDEXSTARTBY или последней страницы и не находит значение строки триггера, то имеет место ошибка, и индексирование прекращается.
- Для всех других триггеров значение *страницы* может быть равно от 0 (нуля) до 16 относительно TRIGGER1. Например, значение страницы 0 (нуль) означает, что триггер находится на той же странице, что и TRIGGER1; значение 1 (единица) означает, что триггер находится на следующей странице относительно страницы, содержащей TRIGGER1, и т.д. Для TRIGGER2 по TRIGGER16 значение строки триггера может находиться максимум на 16 странице от TRIGGER1.

'значение'

Фактическое значение строки, с которым индексатор PDF сравнивает входные данные. В значении строки регистр учитывается. Значение может состоять из одного или нескольких слов, присутствующих на странице.

Примеры

TRIGGER1

Следующий параметр TRIGGER1 заставляет индексатор PDF искать заданную позицию на каждой странице входных данных для указанной строки. Необходимо определить TRIGGER1; значением страницы для TRIGGER1 должна быть звездочка.

```
TRIGGER1=u1(0,0),lr(.75,.25),*,'Страница 0001'
```

Триггеры группы

Следующий параметр триггера заставляет индексатор PDF искать соответствие значению строки Номер счета в пределах координат, определяемых рамкой строки триггера. Триггер может быть найден на той же странице, что и TRIGGER1.

```
TRIGGER2=u1(1,2.25),lr(2,2.5),0,'Номер счета'
```

Следующий параметр триггера заставляет индексатор PDF искать соответствие значению строки Итого в пределах координат, определяемых рамкой строки триггера. В этом примере задана рамка строки триггера один на четыре дюйма, поскольку вертикальное положение триггера на странице может изменяться. Предположим, например, что страница содержит номера счетов и балансы с общим итогом для всех счетов в списке. В списке может быть один или несколько счетов. Положение итоговой суммы изменяется в зависимости от числа счетов в списке. Параметр поля основан на триггере, чтобы индексатор PDF мог найти поле независимо от фактического расположения значения строки триггера. Полем является однодюймовая рамка, которая всегда начинается на расстоянии одного дюйма от правой границы триггера. После нахождения значения строки триггера индексатор PDF прибавляет верхние левые координаты рамки строки триггера к координатам, заданным для этого поля. Триггер может быть найден на следующей странице от TRIGGER1.

```
TRIGGER2=u1(4,4),lr(5,8),1,'Итого'  
FIELD2=u1(1,0),lr(2,1),0,(TRIGGER=2,BASE=TRIGGER)
```

Связанные параметры

Параметр FIELD на странице 35.

Глава 5. Справочник сообщений

Индексатор PDF создает список сообщений в конце каждого цикла индексирования. Код возврата 0 (нуль) означает, что обработка завершена без ошибок.

Индексатор PDF обнаруживает многие состояния ошибок, которые могут быть логически сгруппированы в несколько категорий:

- **Информационные сообщения**

В процессе обработки файла Индексатор PDF выдает информационные сообщения, которые позволяют пользователю определить правильность задания параметров обработки. Эти сообщения могут быть полезны для создания контрольного журнала.

- **Предупреждающие сообщения**

Индексатор PDF выдает предупреждающее сообщение и код возврата 4 (четыре), когда корректность документа сомнительна.

- **Сообщения об ошибках**

Индексатор PDF выдает сообщение об ошибке и код возврата 8 (восемь) или 16 (шестнадцать) и завершает обработку текущего входного файла. Большинство состояний ошибок, обнаруживаемых Индексатором PDF, попадают в эту категорию. Завершение работы может быть различным. При некоторых серьезных ошибках Индексатор PDF может завершиться неудачно из-за сбойного сегмента. Это обычно наблюдается при отказах системных служб. В некоторых случаях Индексатор PDF завершает работу с соответствующими сообщениями об ошибках, занесенными в стандартную ошибку или в файл. Когда Индексатор PDF вызывается программой ARSLOAD, сообщения об ошибках автоматически заносятся в системный протокол. При запуске команды ARSPDOC1 можно указать имя или полное имя файла для сообщений об обработке с помощью параметра **MSGDD**.

- **Пакет Adobe Toolkit**

- **Внутренняя ошибка**

Индексатор PDF выдает сообщение об ошибке и код возврата 16 (шестнадцать) и завершает обработку текущего входного файла.

Информацию о сообщениях, формируемых Индексатором PDF, а также описания ошибок и действий, предпринимаемых в ответ на сообщения, можно найти в документе *IBM DB2 Content Manager OnDemand: Сообщения и коды*, SC27-1379. Сообщения, выдаваемые Индексатором PDF, перечислены в разделе *Общий сервер публикации о сообщениях*.

Глава 6. Справочник по программе ARSPDOC1

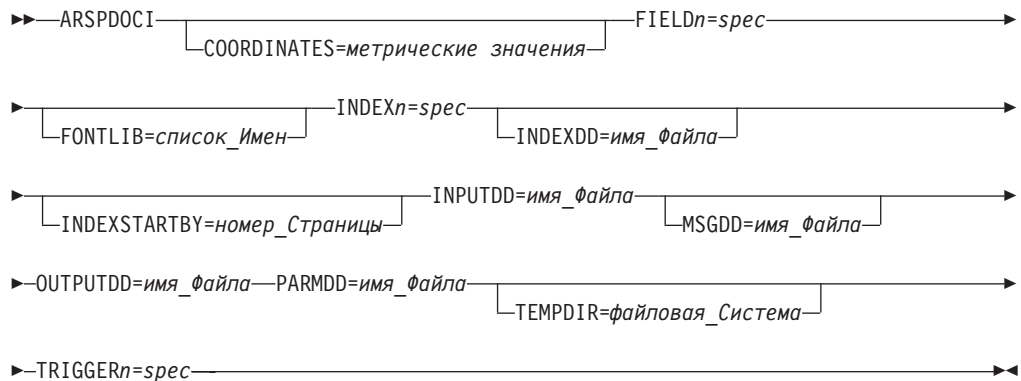
Назначение

Создание данных индексирования для файлов PDF.

Программа ARSPDOC1 использует определенные расположения текстовых строк на странице документа PDF для создания файла индексов текста, а также индексированного по сдвигу байтов документа PDF. Программу ARSPDUMP можно использовать для создания списка расположений текстовых строк в документе. Дополнительная информация приведена в разделе Глава 7, “Справочник по программе ARSPDUMP”, на стр. 51.

Синтаксис

Примечание: Следующий синтаксис должен использоваться только при запуске программы ARSPDOC1 из командной строки или при вызове ее из пользовательской программы.



Описание

Программа ARSPDOC1 может применяться для индексирования файла PDF. Программа ARSLOAD автоматически вызывает программу ARSPDOC1, если тип входных данных PDF и индексатор тоже для PDF. Если для индексирования файла PDF вы не хотите использовать программу ARSLOAD для обработки файла, то вы можете запустить программу ARSPDOC1 из командной строки или вызвать ее из программы.

Программе ARSPDOC1 нужны два входных файла: документ PDF и файл параметров.

При ссылке на шрифт во входном файле, который отсутствует в нем, и невозможности нахождения этого шрифта Индексатор PDF запрошенный шрифт заменяется одним из базовых шрифтов Adobe Type 1, предоставляемых IBM. Если вы приобрели дополнительные шрифты и установили их в систему, то они могут быть встроены в процессе индексирования, если они используются во входном файле и определены параметром FONTLIB. Дополнительная информация приведена в разделе “FONTLIB” на стр. 39.

Параметры

Информация о задаваемых параметрах при запуске программы ARSPDOC1 из командной строки или из пользовательской программы приведена в разделе Глава 4, “Справочник параметров”, на стр. 35.

Расположение IFS

/usr/bin/arspdoci
Выполняемая программа.

Глава 7. Справочник по программе ARSPDUMP

Назначение

Создание расположений текстовых строк на странице.

Программа ARSPDUMP создает список расположений текстовых строк на странице в файле PDF. Результатом работы программы ARSPDUMP являются список текстовых строк на странице и координаты каждой строки. Полученную программой ARSPDUMP информацию можно использовать для создания файла параметров, требуемого программой ARSPDOC1 для индексирования файлов PDF. Дополнительная информация приведена в разделе Глава 6, “Справочник по программе ARSPDOC1”, на стр. 49.

Синтаксис

```
ARSPDUMP -f входной_Файл [-F файл_Шрифтов] [-h]
-o выходной_Файл -r номер_Листа [-t временный_Каталог]
```

Описание

Программа ARSPDUMP определяет расположения текстовых строк на странице в файле PDF.

Результатом работы программы ARSPDUMP являются список текстовых строк на странице и координаты каждой строки.

Если в файле PDF применяется шрифт, не включенный в документ, то программа ARSPDUMP пытается найти этот шрифт на основании информации, заданной в параметре **-F**. Если программа ARSPDUMP не находит шрифт, то она использует заменяющий шрифт Adobe Type 1.

Параметры

-f входной_Файл

Имя файла или полное имя файла PDF для обработки.

-F каталог_Шрифтов

Задаёт каталоги, в которых хранятся шрифты. Укажите любое достоверное полное имя. Полные имена разделяются двоеточием (:). Программа ARSPDUMP просматривает пути в заданном порядке. Если флаг и имя каталога шрифтов не заданы, то программа ARSPDUMP пытается найти шрифты в каталоге /QIBM/ProdData/OnDemand/Adobe/fonts.

-h Задаёт список параметров и их описания для программы ARSPDUMP.

-o выходной_Файл

Имя файла или полное имя файла для записи выходных сообщений программой ARSPDUMP. Если флаг и имя файла не заданы, то

программа ARSPDUMP выводит сообщения на дисплей (в интерактивном режиме) или заносит в протокол задания (в пакетном режиме).

-p номер_Листа

Номер страницы в файле PDF для обработки программой ARSPDUMP. Это страница, которая содержит текстовые строки, которые вы ходите использовать для определения триггеров и полей. Номером листа является порядковый номер страницы в файле, начиная с номера 1 (один) для первой страницы в файле. Он не является идентификатором страницы, качестве которого используется пользовательское обозначение каждой страницы (например, iv, 5 или 17-3).

-t временный_Каталог

Определяет каталог для программы ARSPDUMP в качестве временного рабочего пространства. Укажите любое достоверное имя каталога. Если флаг и имя такого каталога не заданы, то программа ARSPDUMP в качестве временного рабочего пространства использует каталог /arstmp.

Примеры

В следующем примере показывается, как вызвать программу ARSPDUMP из приложения QSHHELL для создания строк и расположений текста на странице номер три документа sample.pdf для выходного документа sample.out:

```
arspdump -f sample.pdf -o sample.out -p 3
```

Информация о запуске программы ARSPDUMP из приложения QSHHELL содержится в *Руководстве по администрированию IBM Content Manager OnDemand для общего сервера i5/OS*.

Расположение IFS

/usr/bin/arspdump

Выполняемая программа.

Глава 8. Трассировщик

Начиная с версии Version 5.3, расширенная трассировка для Индексатор PDF недоступна. Средство трассировки может помочь пользователям при отладке неполадок, таких как системные сбои во время индексирования и загрузки документов PDF.

Для трассировки или отладки неполадки с помощью Индексатор PDF требуется следующее:

- Файл параметров, который определяет поля, триггеры, индексы и другую информацию для индексирования
- Входной файл PDF для обработки

Файл параметров и входной файл PDF могут обрабатываться запуском индексатора PDF из командной строки. Пример:

```
arspdoci parmdd=filen.parms inputdd=filen.pdf outputdd=filen.out indexdd=filen.ind  
tracedd=filen.trace
```

где:

arspdoci - имя программы-индексатора PDF для командной строки

parmdd= задает имя входного файла с параметрами индексирования

inputdd= задает имя входного файла PDF для обработки

outputdd= задает имя выходного файла с проиндексированными документами PDF, созданными Индексатор PDF

indexdd= задает имя выходного файла с информацией об индексировании для загрузки в базу данных

tracedd= задает имя выходного файла с информацией о трассировке

Примечание: Информация о задаваемых параметрах при выполнении программы ARSPDOC1 содержится в разделе Глава 6, “Справочник по программе ARSPDOC1”, на стр. 49.

После выполнения Индексатор PDF с трассировкой выходной файл, заданный параметром tracedd=, будет содержать подробную информацию о произведенной обработке и местах сбоев Индексатор PDF при обработке. Информация о трассировке содержит сведения о неудавшемся поиске триггера, неудавшемся поиске поля, испорченных данных PDF, затруднении с извлечением страницы PDF из документа и даже о нехватке памяти или дискового пространства для завершения требуемых операций. рис. 9 на стр. 54 иллюстрирует пример информации о трассировке, создаваемой Индексатор PDF.

```

COORDINATES=IN
ARSPDOCI completed code get_keyword <-----
ARSPDOCI completed code get_keyword 003 ----->
TRIGGER=UL(7.00,0.25),LR(7.70,0.57),*, 'Page: '
ARSPDOCI completed code get_keyword <-----
ARSPDOCI completed code get_keyword 003 ----->
ARSPDOCI completed code parse_trigger <-----
ARSPDOCI completed code parse_quoted_parm <-----
ARSPDOCI completed code parse_quoted_parm 001 ----->
ARSPDOCI completed code parse_trigger 001 ----->
FIELD1=UL(7.00,0.48),LR(7.90,0.77),0,(TRIGGER=1,BASE=0)
ARSPDOCI completed code get_keyword <-----
ARSPDOCI completed code get_keyword 003 ----->
ARSPDOCI completed code parse_field <-----
ARSPDOCI completed code parse_subfields <-----
ARSPDOCI completed code get_keyword <-----
ARSPDOCI completed code get_keyword 003 ----->
ARSPDOCI completed code get_keyword <-----
ARSPDOCI completed code get_keyword 003 ----->
ARSPDOCI completed code parse_subfields 001 ----->
ARSPDOCI completed code parse_field 001 ----->
FIELD2=UL(6.11,1.39),LR(7.15,1.57),0,(TRIGGER=1,BASE=0)
ARSPDOCI completed code get_keyword <-----
ARSPDOCI completed code get_keyword 003 ----->
ARSPDOCI completed code parse_field <-----
ARSPDOCI completed code parse_subfields <-----
ARSPDOCI completed code get_keyword <-----

.
.
.

ARSPDOCI completed code get_keyword <-----
ARSPDOCI completed code get_keyword 003 ----->
ARSPDOCI completed code arsparm_final_sanity_check <-----
ARSPDOCI completed code arsparm_final_sanity_check 001 ----->
ARSPDOCI completed code ArspProcessOpt <-----
ARSPDOCI completed code ArspOpenIndex <-----
ARSPDOCI completed code ArspOpenIndex 001 ----->
Adobe PDF Library version -732512488.-1
Editing is : -1
Number of input pages = 130
ARSPDOCI completed code ArspProcessOpt:Calling ArspSearchDocPages()
ARSPDOCI completed code ArspSearchDocPages <-----
ARSPDOCI completed code ArspSearchDocPages: ArspCreateWordFinder()
ARSPDOCI completed code ArspSearchDocPages: PDWordFinderAcquireWordList()
ARSPDOCI completed code ArspSearchDocPages: PDDocAcquirePage()
ARSPDOCI completed code ArspSearchDocPages: ArspSearchPage()
ARSPDOCI completed code ArspSearchDocPages: PDPPageRelease()
ARSPDOCI completed code ArspSearchDocPages: PDWordFinderReleaseWordList()
Trigger(s) not found by page 1
ARSPDOCI completed code ArspSearchDocPages 004 ----->
ARSPDOCI completed code ArspProcessOpt:Calling ArspCloseIndex()
ARSPDOCI completed code ArspCloseIndex <-----
ARSPDOCI completed code ArspCloseIndex 001 ----->
ARSPDOCI completed code ArspProcessOpt:Calling PDDocClose()
ARSPDOCI completed code ArspProcessOpt 002 ----->
ARSPDOCI completed code 1
ARSPDOCI completed code ArspFreeParms ()

```

Рисунок 9. Информация о трассировке для Индексатор PDF

Часть 3. Стандартный индексатор

В этой части описывается стандартный индексатор функции OnDemand. Стандартный индексатор может применяться для задания данных индексирования для других типов файлов, которые вы хотите хранить в системе. (Входных файлов, которые не содержат буферных данных PDF, SCS, расширенной SCS, Advanced Function Presentation (AFP) или Line.)

Глава 9. Обзор

Функция OnDemand предусматривает стандартный индексатор для возможности определения информации об индексировании для входных данных, которые вы не можете или не хотите индексировать индексатором OS/400 или индексатором PDF. Предположим, например, что вы хотите загрузить файлы в систему, которые были созданы текстовым процессором. Файлы могут храниться в системе в том же формате, в каком они были созданы. Эти файлы могут вызываться из системы и просматриваться с помощью текстового процессора. Однако поскольку эти документы не содержат буферных данных PDF, SCS, расширенной SCS, AFP или LINE, вы не можете индексировать их другими индексаторами, которые предусмотрены в программе OnDemand. Вы можете задать информацию об индексировании этих файлов в формате, применяемом стандартным индексатором, и загрузить данные об индексе и файлы в систему. После этого пользователи смогут осуществлять поиск и получать файлы с помощью OnDemand клиентской программы.

Для использования Стандартный индексатор необходимо задать все данные индекса для каждого входного файла или документа, который вы хотите хранить и получать из системы. Данные индекса определяются в файле параметров. Этот файл параметров содержит поля индекса, значения индекса и сведения о входных файлах или документах, которые нужно обрабатывать. Стандартный индексатор берет данные индексирования из этого файла параметров и создает информацию об индексе, которая загружается в базу данных. OnDemand создает одну запись индекса для каждого входного файла (или документа), указанного в файле параметров. Запись индекса содержит значения индекса, которые однозначным образом идентифицируют файл или документ в OnDemand.

Стандартный индексатор поддерживает индексы для групп. Индексы групп хранятся в базе данных и служат для поиска документов. Для каждого входного файла или документа, который нужно обрабатывать с помощью стандартного индекса, необходимо задать один набор индексов группы.

Загрузка данных

Большинство пользователей загружают данные в систему с помощью программы ARSLOAD. При необходимости индексирования входных данных программа ARSLOAD вызывает программу индексации (основанную на типе входных данных или для Стандартный индексатор на достоверном файле параметров). Например, программа ARSLOAD может вызывать Стандартный индексатор для обработки файла параметров и создания данных индексирования. Затем программа ARSLOAD может добавить информацию об индексе в базу данных и загрузить входные файлы или документы, указанные в файле параметров, в тома хранения.

Программа ARSLOAD выполняется двумя способами:

- Режим демона. Программа ARSLOAD выполняется как демон (сервера UNIX) или как сервис (сервера Windows) для периодической проверки указанного каталога на наличие входных файлов для обработки. При выполнении программы ARSLOAD в режиме демона для запуска процедуры загрузки требуется фиктивный файл с расширением типа .ARD. Кроме того, файл параметров стандартного индексатора (.IND) должен находиться в

указанном каталоге. Параметр GROUP_FILENAME: в файле .IND задает полное имя реального входного файла для обработки.

- Ручной режим. Программа ARSLOAD для обработки конкретного файла запускается из командной строки. При запуске программы ARSLOAD в ручном режиме указывается только *имя* файла для обработки. Программа ARSLOAD добавляет расширение имени файла .IND к заданному вами имени файла. Например, если вы задали arslod ... ро3510, где ро3510 - имя входного файла, программа ARSLOAD будет обрабатывать файл параметров стандартного индекса ро3510.ind. Параметр GROUP_FILENAME: в файле параметров стандартного индекса задает полное имя фактически обрабатываемого входного файла.

После успешной загрузки данных система удаляет входной файл, заданный параметром GROUP_FILENAME:, если расширением файла является .OUT, и для обработки в режиме демона остальная часть имени входного файла совпадает с именем файла .ARD. Система также удаляет файл .IND (файл параметров стандартного индекса) и файл .ARD (фиктивный файл, использовавшийся для запуска процедуры загрузки при выполнении программы ARSLOAD в режиме демона).

Далее приводится пример имен файлов в режиме демона:

```
MVS.JOBNAME.DATASET.FORM.YYYYDDDD.HHMMSS.T.ARD
MVS.JOBNAME.DATASET.FORM.YYYYDDDD.HHMMSS.T.ARD.IND
MVS.JOBNAME.DATASET.FORM.YYYYDDDD.HHMMSS.T.ARD.OUT
```

Файл MVS.JOBNAME.DATASET.FORM.YYYYDDDD.HHMMSS.T.ARD является фиктивным файлом, который запускает процедуру загрузки в режиме демона. Файл MVS.JOBNAME.DATASET.FORM.YYYYDDDD.HHMMSS.T.ARD.IND является файлом параметров стандартного индекса и содержит параметр GROUP_FILENAME:, который задает входной файл для обработки: MVS.JOBNAME.DATASET.FORM.YYYYDDDD.HHMMSS.T.ARD.OUT. После успешной загрузки данных система удаляет все три файла.

Обработка данных AFP

Для входных файлов, которые содержат ресурсы и документы AFP, можно задать файл параметров и обработать с помощью стандартного индекса. Однако при задании файла параметров необходимо соблюдать следующее:

- Когда входные файлы содержат ресурсы AFP, начальное расположение (сдвиг байтов) первого документа AFP во входном файле всегда должно быть равно 0 (нулю), даже если фактическое начальное расположение ненулевое. Ресурсы AFP всегда располагаются в начале входного файла. Фактическое начальное положение первого документа во входном файле нулевое плюс число байтов, занятых ресурсами. Однако для обработки документов AFP с помощью стандартного индекса вам не нужно вычислять число байтов, занятых ресурсами.
- Начальные положения других документов во входном файле должны вычисляться на основании длины сдвига от предыдущего документа во входном файле.

Стандартный индекс определяет, где в файле заканчиваются ресурсы AFP, и обрабатывает документы, применяя заданные вами сдвиги и длины, относительно окончания ресурсов.

Глава 10. Задание файла параметров

Для Стандартный индексатор требуются один или несколько входных файлов, которые нужно загрузить в систему, и файл параметров с информацией об индексировании этих файлов. Для применения Стандартный индексатор необходимо создать файл параметров, содержащий информацию об индексировании входных файлов. В этом разделе описывается файл параметров для Стандартный индексатор.

В файле параметров можно определить три типа операторов:

- **Комментарии.** В любом месте файла параметров можно разместить строку комментариев.
- **Кодовая страница.** В самом начале файла параметров необходимо задать строку кодовой страницы.
- **Группы.** Каждая группа представляет один документ для индексации. Группа содержит имена полей группы приложений и значения индексов для них, расположение документа во входном файле, число байтов (символов) в документе и имя входного файла, содержащего документ.

Важная информация!:

1. Имена файлов в файле параметров должны представляться символами верхнего регистра. Например, правильным будет имя `GROUP_FIELD_NAME:account`, а `group_field_name:account` - неправильным.
2. При загрузке данных с помощью Стандартный индексатор локаль должна быть задана соответственно параметру `CODEPAGE:`. Например, если задается `CODEPAGE:954`, установите переменную среды локали на `ja_JP`, или определите другую локаль, которая правильно идентифицирует прописные и строчные символы на кодовой странице 954.

CODEPAGE:

Задаёт кодовую страницу входных данных. Должна быть задана только одна кодовая страница. Строка **CODEPAGE:** должна находиться в самом начале. Строка **CODEPAGE:** необходима.

Важная информация: При загрузке данных с помощью Стандартный индексатор локаль должна быть задана соответственно параметру `CODEPAGE:`. Например, если задается `CODEPAGE:954`, установите переменную среды локали на `ja_JP`, или определите другую локаль, которая правильно идентифицирует прописные и строчные символы на кодовой странице 954.

Синтаксис

CODEPAGE:`cpgid`

Опции и значения

Строка символов **CODEPAGE:** представляет строку, определяющую кодовую страницу входных данных. Строка `cpgid` может быть любой доступной кодовой страницей, зарегистрированным IBM трех-пятибайтовым идентификатором или пользовательской кодовой страницей.

Необходим параметр **CODEPAGE:**.

Пример

Далее описывается, как задать кодовую страницу 37 для входных данных:

```
CODEPAGE:37
```

Параметр COMMENT:

Задаёт строку комментариев. Строки комментариев можно размещать в любом месте файла параметров.

Синтаксис

COMMENT: - текст в отдельной строке

Опции и значения

Строка символов **COMMENT:** обозначает строку комментариев. Все содержимое после двоеточия до конца строки игнорируется.

Пример

Следующая строка является примером строки комментариев:

```
COMMENT:  
COMMENT: это комментарий
```

Параметр GROUP_FIELD_NAME:

Задаёт имя поля группы приложений. Каждая группа, задаваемая в файле параметров, должна содержать одну строку **GROUP_FIELD_NAME:** для каждого поля группы приложений. (В группе приложений сохраняется ваш файл или документ функции OnDemand. Имя группы приложений задаётся для программы ARSLOAD.) Для одной группы приложений функция OnDemand поддерживает до 32 полей. Если задаваемые имена полей отличаются от имен полей группы приложений, то на странице Информация о загрузке заданные имена полей необходимо сопоставить именам полей группы приложений.

Для каждого поля группы приложений определите пару строк **GROUP_FIELD_NAME:** и **GROUP_FIELD_VALUE:**. Например, если группа приложений содержит два поля, то каждая группа, задаваемая в файле параметров, должна содержать две пары строк **GROUP_FIELD_NAME:** и **GROUP_FIELD_VALUE:**. Ниже приведен пример группы с двумя полями группы приложений:

```
GROUP_FIELD_NAME:rdate  
GROUP_FIELD_VALUE:05/31/00  
GROUP_FIELD_NAME:ИД учащегося  
GROUP_FIELD_VALUE:0012345678
```

Строки группы должны находиться после строки **CODEPAGE:**.

Синтаксис

GROUP_FIELD_NAME:имя_поля_группы_приложений

Опции и значения

Строка символов **GROUP_FIELD_NAME:** обозначает строку, содержащую имя поля группы приложений. Строка `имя_поля_группы_приложений` задает имя поля группы приложений. Функция OnDemand не учитывает регистр имен полей группы приложений.

Пример

Ниже приведены примеры имен полей группы приложений:

```
GROUP_FIELD_NAME:rdate
GROUP_FIELD_NAME:ИД учащегося
GROUP_FIELD_NAME:номер_счета
```

Параметр **GROUP_FIELD_VALUE:**

Задаёт значение индекса для поля группы приложений. Каждая группа, задаваемая в файле параметров, должна содержать одну строку **GROUP_FIELD_VALUE:** для каждого поля группы приложений. (В группе приложений сохраняется ваш файл или документ функции OnDemand. Имя группы приложений задается для программы ARSLOAD.) Для одной группы приложений функция OnDemand поддерживает до 32 полей. Строка **GROUP_FIELD_VALUE:** должна идти после строки **GROUP_FIELD_NAME:**, для которой задается значение индекса.

Для каждого поля группы приложений определите пару строк **GROUP_FIELD_NAME:** и **GROUP_FIELD_VALUE:**. Например, если группа приложений содержит два поля, то каждая группа, задаваемая в файле параметров, должна содержать две пары строк **GROUP_FIELD_NAME:** и **GROUP_FIELD_VALUE:**. Ниже приведен пример группы с двумя полями группы приложений:

```
GROUP_FIELD_NAME:rdate
GROUP_FIELD_VALUE:05/31/00
GROUP_FIELD_NAME:ИД учащегося
GROUP_FIELD_VALUE:0012345678
```

Строки группы должны находиться после строки **CODEPAGE:**.

Синтаксис

GROUP_FIELD_VALUE:значение

Опции и значения

Строка символов **GROUP_FIELD_VALUE:** обозначает строку, содержащую значение индекса для поля группы приложений. Строка `значение` определяет фактическое значение индекса для поля.

Пример

Ниже приведены примеры значений индекса:

```
GROUP_FIELD_VALUE:05/31/00
GROUP_FIELD_VALUE:0012345678
GROUP_FIELD_VALUE:0000-1111-2222-3333
```

Параметр **GROUP_FILENAME:**

Имя файла или полное имя входного файла. Если путь для файла не указывается, стандартный индекатор найдет текущий каталог для данного файла; однако всегда следует задавать полное имя входного файла.

Каждая группа, задаваемая в файле параметров, должна содержать одну строку **GROUP_FILENAME_NAME:**. Строка **GROUP_FILENAME:** должна идти за строками **GROUP_FIELD_NAME:** и **GROUP_FIELD_VALUE:**, которые составляют группу. Ниже приведен пример группы:

```
GROUP_FIELD_NAME:rdate
GROUP_FIELD_VALUE:05/31/00
GROUP_FIELD_NAME:ИД учащегося
GROUP_FIELD_VALUE:0012345678
GROUP_OFFSET:0
GROUP_LENGTH:0
GROUP_FILENAME:/tmp/statements.out
```

Если строка **GROUP_FILENAME** не содержит никакого значения (пусто), то стандартный индекатор для обработки текущей группы берет значение из строки **GROUP_FILENAME** предыдущей группы. В следующем примере входные данные для второй и третьей групп берутся из входного файла, который задан для первой группы.

```
GROUP_FIELD_NAME:rdate
GROUP_FIELD_VALUE:05/31/00
GROUP_FIELD_NAME:ИД учащегося
GROUP_FIELD_VALUE:0012345678
GROUP_OFFSET:0
GROUP_LENGTH:8124
GROUP_FILENAME:/tmp/statements.out
GROUP_FIELD_NAME:rdate
GROUP_FIELD_VALUE:06/30/00
GROUP_FIELD_NAME:ИД учащегося
GROUP_FIELD_VALUE:0012345678
GROUP_OFFSET:8124
GROUP_LENGTH:8124
GROUP_FILENAME:
GROUP_FIELD_NAME:rdate
GROUP_FIELD_VALUE:07/31/00
GROUP_FIELD_NAME:ИД учащегося
GROUP_FIELD_VALUE:0012345678
GROUP_OFFSET:16248
GROUP_LENGTH:8124
GROUP_FILENAME:
```

Если первая строка **GROUP_FILENAME** в файле параметров пустая, то на момент выполнения программы ARSLOAD необходимо задать имя входного файла.

Строки группы должны находиться после строки **CODEPAGE:**.

После успешной загрузки данных система удаляет входной файл, заданный параметром **GROUP_FILENAME:**, если расширением файла является **.OUT**, и для обработки в режиме демона остальная часть имени входного файла совпадает с именем файла **.ARD**. Система также удаляет файл **.IND** (файл параметров стандартного индекатора) и файл **.ARD** (фиктивный файл, использовавшийся для запуска процедуры загрузки при выполнении программы ARSLOAD в режиме демона). За дополнительной информацией обратитесь к разделу “Загрузка данных” на стр. 57.

Синтаксис

GROUP_FILENAME:имя_файла

Опции и значения

Строка символов **GROUP_FILENAME:** обозначает строку, содержащую входной файл для обработки. Строка имя_файла определяет полное имя входного файла. Следует всегда задавать полное имя входного файла для обработки. Например:

```
GROUP_FILENAME:/tmp/ondemand/входные_файлы/f1b0a1600.out
```

Пример

Ниже приведены допустимые строки имен файлов:

```
GROUP_FILENAME:/tmp/операторы
GROUP_FILENAME:D:\ARSTMP\операторы
GROUP_FILENAME:/tmp/ondemand/входные_файлы/f1b0a1600.out
GROUP_FILENAME:
```

Параметр **GROUP_LENGTH:**

Задаёт количество смежных байтов (символов), составляющих документ для индексации. Для задания всего входного файла или остальной части файла укажите 0 (нуль). Каждая группа, задаваемая в файле параметров, должна содержать одну строку **GROUP_LENGTH:**. Строка **GROUP_LENGTH:** должна идти за строками **GROUP_FIELD_NAME:** и **GROUP_FIELD_VALUE:**, которые составляют группу. Например:

```
GROUP_FIELD_NAME:rdate
GROUP_FIELD_VALUE:05/31/00
GROUP_FIELD_NAME:ИД учащегося
GROUP_FIELD_VALUE:0012345678
GROUP_OFFSET:0
GROUP_LENGTH:0
```

Строки группы должны находиться после строки **CODEPAGE:**.

Синтаксис

GROUP_LENGTH:значение

Опции и значения

Строка символов **GROUP_LENGTH:** обозначает строку, содержащую число байтов индексируемых данных. Строка значение задаёт фактическое число байтов. Значением по умолчанию является 0 (нуль) для всего (или остальной части) файла.

Пример

Ниже приведены примеры задания значений длин:

```
GROUP_LENGTH:0
GROUP_LENGTH:8124
```

Параметр **GROUP_OFFSET:**

Задаёт начальное положение (сдвиг байтов) во входном файле индексируемых данных. Для первого байта (начала) файла задается 0 (нуль). (Обработка документов и ресурсов AFP с помощью стандартного индексируемого описывается в разделе “Обработка данных AFP” на стр. 58.) Каждая группа, задаваемая в

файле параметров, должна содержать одну строку **GROUP_OFFSET_NAME:**. Строка **GROUP_OFFSET:** должна идти за строками **GROUP_FIELD_NAME:** и **GROUP_FIELD_VALUE:**, которые составляют группу. Например:

```
GROUP_FIELD_NAME:rdate
GROUP_FIELD_VALUE:05/31/00
GROUP_FIELD_NAME:ИД учащегося
GROUP_FIELD_VALUE:0012345678
GROUP_OFFSET:0
```

Строки группы должны находиться после строки **CODEPAGE:**.

Синтаксис

GROUP_OFFSET:значение

Опции и значения

Строка символов **GROUP_OFFSET:** обозначает строку, содержащую сдвиг (положение) байтов индексируемых данных. Строка значение задает фактический сдвиг байтов. Для начала файла задается 0 (нуль).

Пример

Ниже приведены примеры значений сдвигов для трех документов из одного входного файла. Документы имеют длину 8 Кб.

```
GROUP_OFFSET:0
GROUP_OFFSET:8124
GROUP_OFFSET:16248
```

Глава 11. Примеры файлов параметров

Ниже приведен пример задания информации об индексировании для трех групп (документов). Для индексирования каждого документа применяются два поля. Входные данные для каждого документа содержатся в разных файлах.

```
COMMENT:
COMMENT: Стандартный индексатор: Пример 1
COMMENT: Документы в разных входных файлах
COMMENT:
COMMENT: Задание кодовой страницы данных индекса
CODEPAGE:37
COMMENT: Документ No.1
COMMENT: Поле индекса No.1
GROUP_FIELD_NAME:rdate
GROUP_FIELD_VALUE:07/13/99
COMMENT: Поле индекса No.2
GROUP_FIELD_NAME:ИД учащегося
GROUP_FIELD_VALUE:0012345678
COMMENT: данные документа начинаются с начала файла
GROUP_OFFSET:0
COMMENT: данные документа занимают весь файл до конца
GROUP_LENGTH:0
GROUP_FILENAME:/arstmp/statement7.out
COMMENT: Документ No.2
COMMENT: Поле индекса No.1
GROUP_FIELD_NAME:rdate
GROUP_FIELD_VALUE:08/13/99
COMMENT: Поле индекса No.2
GROUP_FIELD_NAME:ИД учащегося
GROUP_FIELD_VALUE:0012345678
GROUP_OFFSET:0
GROUP_LENGTH:0
GROUP_FILENAME:/arstmp/statement8.out
COMMENT: Документ No.3
COMMENT: Поле индекса No.1
GROUP_FIELD_NAME:rdate
GROUP_FIELD_VALUE:09/13/99
COMMENT: Поле индекса No.2
GROUP_FIELD_NAME:ИД учащегося
GROUP_FIELD_VALUE:0012345678
GROUP_OFFSET:0
GROUP_LENGTH:0
GROUP_FILENAME:/arstmp/statement9.out
COMMENT:
COMMENT: Конец примера 1 для стандартного индексатора
```

Ниже приведен пример задания информации об индексировании для трех групп (документов). Для индексирования каждого документа применяются два поля. Входные данные всех документов содержатся в одном входном файле.

```
COMMENT:
COMMENT: Стандартный индексатор: Пример 2
COMMENT: Все документы в одном входном файле
COMMENT:
COMMENT: Задание кодовой страницы данных индекса
CODEPAGE:37
COMMENT: Документ No.1
GROUP_FIELD_NAME:rdate
GROUP_FIELD_VALUE:07/13/99
GROUP_FIELD_NAME:ИД учащегося
GROUP_FIELD_VALUE:0012345678
COMMENT: первый документ начинаются с начала файла (с байта 0)
GROUP_OFFSET:0
COMMENT: длина документа 8124 байта
GROUP_LENGTH:8124
GROUP_FILENAME:/arstmp/accounting.student information.loan.out
COMMENT: Документ No.2
GROUP_FIELD_NAME:rdate
GROUP_FIELD_VALUE:08/13/99
GROUP_FIELD_NAME:ИД учащегося
GROUP_FIELD_VALUE:0012345678
COMMENT: второй документ начинается с байта 8124
GROUP_OFFSET:8124
COMMENT: длина документа 8124 байта
GROUP_LENGTH:8124
COMMENT: используется до GROUP_FILENAME:
GROUP_FILENAME:
COMMENT: Документ No.3
GROUP_FIELD_NAME:rdate
GROUP_FIELD_VALUE:09/13/99
GROUP_FIELD_NAME:ИД учащегося
GROUP_FIELD_VALUE:0012345678
COMMENT: третий документ начинается с байта 16248
GROUP_OFFSET:16248
COMMENT: длина документа 8124 байта
GROUP_LENGTH:8124
COMMENT: use prior GROUP_FILENAME:
GROUP_FILENAME:
COMMENT:
COMMENT: Конец примера 2 для стандартного индексатора
```

Глава 12. Дополнительные разделы по индексированию

Здесь приводятся те стороны индексирования, которые не описываются больше нигде в этом руководстве и которые относятся ко всем индексаторам (OS/400, PDF и стандартному), если другое не оговаривается специальным образом.

Программа постпроцессора

Если требуется программа постпроцессора для дальнейшей обработки данных индексирования ваших входных данных, то можно создать специальную программу, которую OnDemand будет применять ко всем записям полей индекса непосредственно перед занесением их в базу данных. За новейшими инструкциями и примерами программ обратитесь на Web-сайт сопровождения OnDemand для i5/OS <http://www.ibm.com/software/data/ondemand/400/support.html> и найдите "постпроцессор."

Файлы индексов (.ind) в IFS

Файлы в IFS можно найти на вашем сервере i5/OS, которые могут выглядеть примерно следующим образом:

```
/SP_QPRLR133_QPRTJOB_TKRUPA_067503_000003_MYSYSTEM_1040629_083851.ind
```

Они являются результатом выполнения команды **Добавить отчет** в OnDemand (ADDRPTOND) с параметром **Только индекс (IDXONLY)**, установленным на ***YES** или создания командой ADDRPTOND, заданием монитора OnDemand или одним из ARSxxx API из поврежденного архива. Если для пользовательского профайла, выполняющего задание обработки архива домашний каталог существует, то эти файлы будут вставлены в этот домашний каталог. В противном случае эти файлы будут занесены в корневой каталог и найти их будет немного труднее.

Эти файлы предназначены для того, чтобы помочь определить, почему обработка архива не удалась. Файлы .ind содержат данные индексирования, полученные в процессе обработки файла, и могут упростить выявление причины неполадки. В случае большого количества таких файлов в вашей системе необходимо исследовать причину этого (если тестирование не выполнялось командой IDXONLY(с установкой *YES), упомянутой выше).

Если эти файлы не нужны для определения неполадки или тестирования, то их можно удалить.

Рекомендуемый порядок определения триггеров и полей

Обычно следует определять триггеры и поля от левого верхнего до правого нижнего угла отчета. Это упрощает процедуру создания параметров индексатора.

Индекс

A

AFP

- индексирование с помощью стандартного индеклятора 58
- обработка с помощью стандартного индеклятора 58
- стандартный индеклятор, обработка с помощью 58

ARSPDOC1 19

- параметр COORDINATES 35
- параметр FIELD 35
- параметр FONTLIB 39
- параметр INDEX 39
- параметр INDEXDD 40
- параметр INDEXSTARTBY 41
- параметр INPUTDD 42
- параметр MSGDD 42
- параметр OUTPUTDD 43
- параметр PARMDD 43
- параметр TEMPDIR 43
- параметр TRACEDD 44
- параметр TRIGGER 44
- сообщения 47
- сообщения об ошибках 47
- справочник 35, 49

D

DBCS

- индеклятор PDF 34

N

NLS 59

- индеклятор PDF 34

B

входной файл

- параметр стандартного индеклятора 62, 63

Г

- графический индеклятор 3

Д

документ

- параметр стандартного индеклятора 62, 63
- документы Adobe PDF
- См. индеклятор PDF

З

закладки

- индеклятор PDF 33

И

имена входных файлов

- индеклятор PDF 33

Индеклятор OS/400

- введение 1
- о программе 1
- обзор 1
- применение 1

индеклятор PDF

- Adobe PDF 49
- DBCS 34
- NLS 34
- введение 19
- закладки 33
- имена входных файлов 33
- кодировка страницы 34
- координаты x, y 28
- о программе 19
- обзор 19
- ограничения 33
- передача входных файлов в 33
- печать 33
- поддержка национальных языков (NLS) 34
- применение 19
- принципы 27
- принципы индексирования 27
- система координат 28
- соглашения об именах файлов 33
- сообщения 47
- сообщения об ошибках 47
- справочник параметров 35
- справочник по ARSPDOC1 49
- справочник по ARSPDUMP 51
- ссылки 33
- файл параметров 28
- шрифты 33, 39

Индеклятор PDF

- индексы 39
- индексы групп 39
- маска поля 37
- опция маски 37
- поле константы 38
- поле триггера 35
- поля 35
- стандартное значение индекса 37
- триггеры 44

индексирование

- документы Adobe PDF 19
- индеклятор OS/400 1
- индеклятор PDF 19
- индексы 39
- индексы групп 39
- маска поля 37
- начальные страницы 41
- опция маски 37
- параметры 28
- поле константы 38

индексирование *(продолжение)*
поле триггера 35
поля для Индексатор PDF 35
пропуск начальных страниц 41
стандартное значение индекса 37
стандартный индексатор 55
триггеры 44
индексы
параметр Индексатор PDF 39
параметр стандартного индексатора 61
индексы групп
определение 39, 60
определение для стандартного индексатора 61

К

кодированная страница
DBCS 34
индексатор PDF 34
стандартный индексатор 59
команды
ARSPDOCI 49
ARSPDUMP 51
координаты
параметра FIELD для Индексатор PDF 36
параметра TRIGGER для Индексатор PDF 44
координаты x, y 28

М

маска
опция параметра FIELD 37
мастер отчетов 3

Н

начальные страницы
пропуск 41

О

ограничения
индексатор PDF 33
отладка 53

П

параметр CODEPAGE: 59
параметр COMMENT: 60
параметр COORDINATES 35
флаги и значения 35
параметр FIELD 35
опция маски 37
поле константы 38
поле триггера 35
стандартное значение индекса 37
флаги и значения 35
параметр FONTLIB 39
флаги и значения 39
параметр GROUP_FIELD_NAME: 60

параметр GROUP_FIELD_VALUE: 61
параметр GROUP_FILENAME: 62
параметр GROUP_LENGTH: 63
параметр GROUP_OFFSET: 63
параметр INDEX 39
флаги и значения 39
параметр INDEXDD 40
флаги и значения 40
параметр INDEXSTARTBY 41
флаги и значения 41
параметр INPUTDD 42
флаги и значения 42
параметр MSGDD 42
флаги и значения 42
параметр OUTPUTDD 43
флаги и значения 43
параметр PARMDD 43
флаги и значения 43
параметр TEMPDIR 43
флаги и значения 43
параметр TRACEDD 44
трассировщик 53
флаги и значения 44
параметр TRIGGER 44
опции и значения 44
параметры 19
CODEPAGE: 59
COMMENT: 60
COORDINATES 35
FIELD 35
FONTLIB 39
GROUP_FIELD_NAME: 60
GROUP_FIELD_VALUE: 61
GROUP_FILENAME: 62
GROUP_LENGTH: 63
GROUP_OFFSET: 63
INDEX 39
INDEXDD 40
INDEXSTARTBY 41
INPUTDD 42
MSGDD 42
OUTPUTDD 43
PARMDD 43
TEMPDIR 43
TRACEDD 44
TRIGGER 44
индексатор PDF 28, 35
индексирование 51
программа ARSPDOCI 35, 49
программа ARSPDUMP 51
стандартный индексатор 59
параметры индексатора
применение break=yes в зависимости от
break=no 6
Переносимый формат документов (PDF)
См. индексатор PDF
печать
индексатор PDF 33
поддержка национальных языков (NLS) 59
индексатор PDF 34
поле константы 38

- поле триггера 35
- поля
 - опция маски 37
 - параметр Индексатор PDF 35
 - параметр стандартного индексатора 60, 61
 - поле константы 38
 - поле триггера 35
 - стандартное значение индекса 37
- примеры
 - стандартный индексатор 65
- программа ARSPDUMP
 - справочник 51
- пропуск начальных страниц 41

Р

- расположение IFS 49

С

- Синтаксис
 - COORDINATES 35
 - FONTLIB 35
 - INDEXDD 35
 - INDEXn 35
 - INDEXSTARTBY 35
 - INPUTDD 35
 - MSGDD 35
 - OUTPUTDD 35
 - PARMDD 35
 - TEMPDIR 35
 - TRIGGER 35
 - Поле 35
 - Поле константы 35
- система координат 28
- сообщения
 - индексатор PDF 47
 - программа ARSPDOC1 47
- сообщения об ошибках
 - индексатор PDF 47
 - программа ARSPDOC1 47
- ссылки
 - индексатор PDF 33
- стандартное значение индекса
 - опция параметра FIELD 37
- стандартный индексатор
 - NLS 59
 - введение 55
 - входной файл 62, 63
 - данные AFP, обработка 58
 - документ 62, 63
 - значения полей 61
 - имена полей 60
 - имена полей группы приложений 60
 - индексы групп, определение 60, 61
 - кодированная страница 59
 - о программе 55, 58
 - обзор 55
 - Параметр CODEPAGE: 59
 - параметр COMMENT: 60
 - параметр GROUP_FIELD_NAME: 60

- стандартный индексатор *(продолжение)*
 - параметр GROUP_FIELD_VALUE: 61
 - параметр GROUP_FILENAME: 62
 - параметр GROUP_LENGTH: 63
 - параметр GROUP_OFFSET: 63
 - поддержка национальных языков (NLS) 59
 - применение 55
 - примеры 65
 - файл параметров 59, 65

Т

- трассировщик 53
- триггеры
 - параметр Индексатор PDF 44
- Триггеры
 - TRIGGER1 35
 - синтаксис полей 35
 - Триггеры группы 35

Ф

- файл параметров
 - индексатор PDF 28, 35
 - программа ARSPDOC1 35
 - стандартный индексатор 65
- файлы
 - индексатор PDF 33

Ш

- шрифты
 - индексатор PDF 33, 39



Номер программы: 5761-RD1

SC89-1059-04

