

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Локтионова Оксана Геннадьевна

Должность: проректор по учебной работе

Дата подписания: 03.02.2021 18:07:00

Уникальный программный ключ:

0b817ca911e6668abb13a5d426d39e5f1c11eabbf73e943df4a4851fda56d089

## МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования  
«Юго-Западный государственный университет»  
(ЮЗГУ)

Кафедра информационных систем и технологий



## ТЕХНОЛОГИЯ СОРТИРОВКИ, ЗАМЕНЫ И ФИЛЬТРАЦИИ ИНФОРМАЦИИ В DEDUCTOR STUDIO

Методические указания к лабораторной работе № 5  
для студентов направления 09.03.02 и 09.03.03

Курск 2016

УДК 004  
Составитель А.В. Ткаченко

Рецензент  
Кандидат технических наук, доцент С.Ю. Сазонов

**Технология сортировки, замены и фильтрации информации в Deductor Studio:** методические указания к лабораторной работе № 5 по дисциплине «Технологии обработки информации» / Юго-Зап. гос. ун-т; сост. А.В. Ткаченко. Курск, 2016. 12 с. Библиогр.: с. 11.

Приводится последовательность операций при подготовке информации для ее визуализации в Deductor Studio.

Методические указания соответствуют требованиям утвержденной рабочей программы дисциплины.

Предназначены для студентов, обучающихся по направлениям: 09.03.02 «Информационные системы и технологии» и 09.03.03 «Прикладная информатика».

Текст печатается в авторской редакции.

Подписано в печать . Формат 60x84 1/16.  
Усл.печ. л. . Уч.-изд. л. . Тираж 100 экз. Заказ.  
Бесплатно.

Юго-Западный государственный университет.  
305040, г. Курск, ул. 50 лет Октября, 94.

**Цель:** Освоить технологию сортировки, замены и фильтрации информации

## Сортировка

Обработчик **Сортировка** предназначен для изменения порядка следования записей в наборе данных в соответствии с выбранным типом сортировки.

Результатом выполнения сортировки является новый набор данных, записи в котором следуют в соответствии с заданными параметрами сортировки.

Если сортировка производится по одному полю, то все записи исходного набора данных располагаются в порядке возрастания или убывания его значений. Если сортировка производится по двум или более полям, то действуют следующие правила:

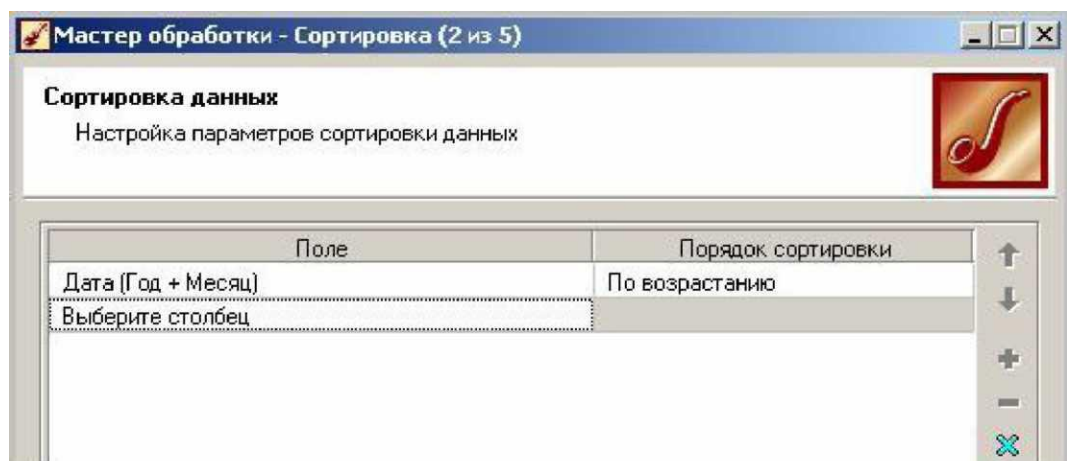
1. Сначала записи сортируются в заданном порядке для первого поля.
2. В каждом наборе одинаковых значений первого поля записи располагаются в заданном порядке для второго поля. И так далее для всех полей, подлежащих сортировке.

Обработчик **Сортировка** находится в группе узлов **Трансформация данных** мастера обработки.

В единственном окне настройки параметров сортировки мастера обработки представлен список условий сортировки, в котором содержатся две графы:

- Имя поля - содержит имена полей, по которым следует выполнить сортировку.

Порядок сортировки - содержит порядок сортировки данных в соответствующем поле - по возрастанию или по убыванию.



## Замена данных

Обработчик **Замена данных** предназначен для замены значений набора данных по таблице подстановок, которая содержит пары, состоящие из исходного значения и результирующего значения.

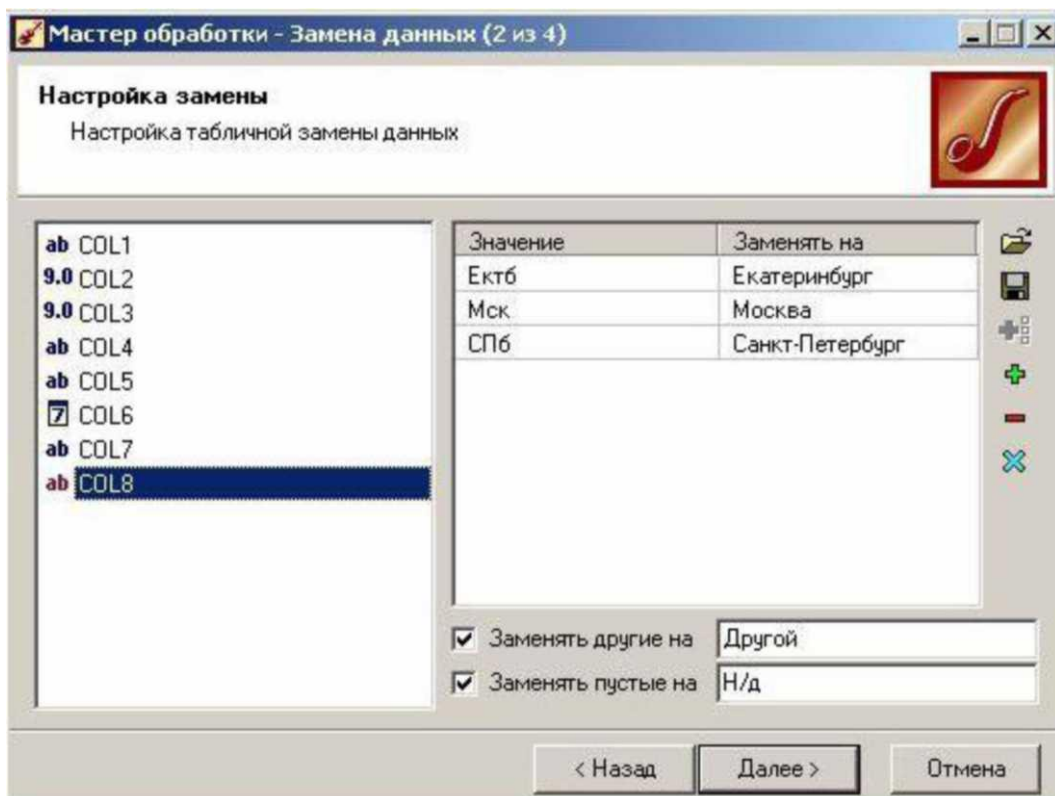
Пример таблицы подстановок.

Значение	Заменять на
Мск	Москва
Спб	Санкт-Петербург
Ектб	Екатеринбург

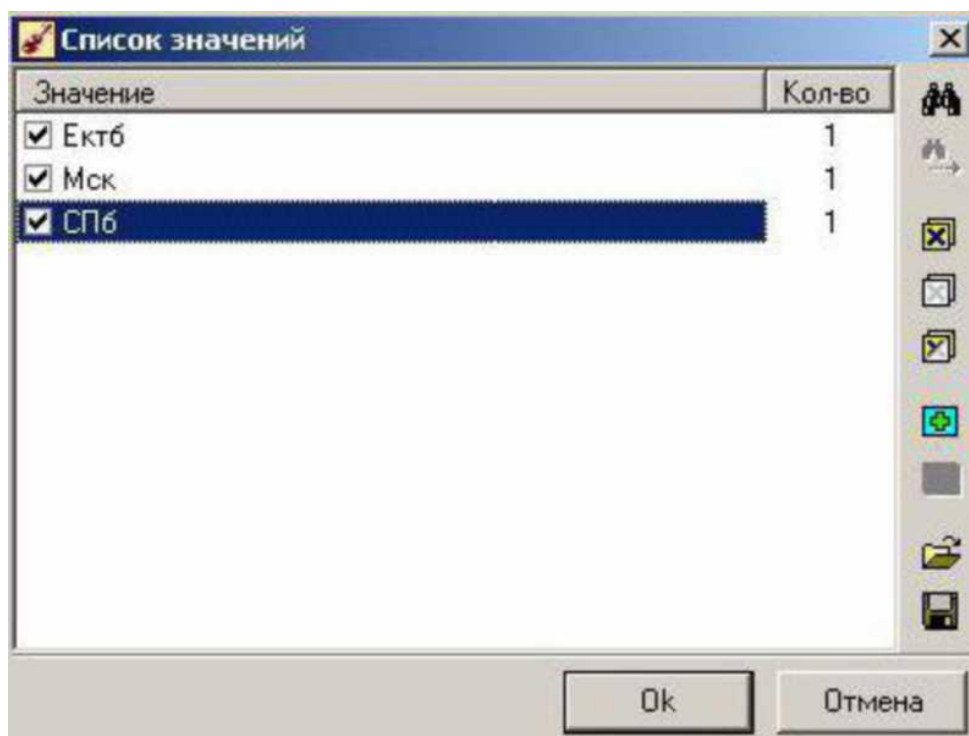
Для каждого значения исходного набора данных ищется соответствие среди исходных значений таблицы подстановки. Если соответствие найдено, то значение меняется на соответствующее выходное значение из таблицы подстановки. Если значение не найдено в таблице, оно может быть либо заменено значением, указанным для замены «по умолчанию», либо оставлено без изменений (если такое значение не указано).

В результате замены для каждого поля, которое в нем участвует, создается новое поле с префиксом REPLACE как к имени, так и к метке поля. Например, для поля Город после узла **Замена данных** появится новое поле Город\_REPLACE.

Обработчик **Замена данных** находится в группе узлов **Трансформация данных** мастера обработки. В окне настройки параметров замены для каждого поля можно ввести таблицу подстановок. Добавление новой строки в таблицу подстановок производится нажатием кнопки удаление существующей - \* .



В таблице подстановок должны быть заполнены два поля: **Значение** - заменяемое значение поля исходной таблицы. Если поле дискретное, то для ввода значения можно воспользоваться кнопкой выбора, где флажками отметить нужные значения. При этом откроется диалоговое окно:



**Заменять на** - значение для замены того, что указано в поле Значение. Внизу таблицы подстановок расположены еще два флага, которые при необходимости можно включить:

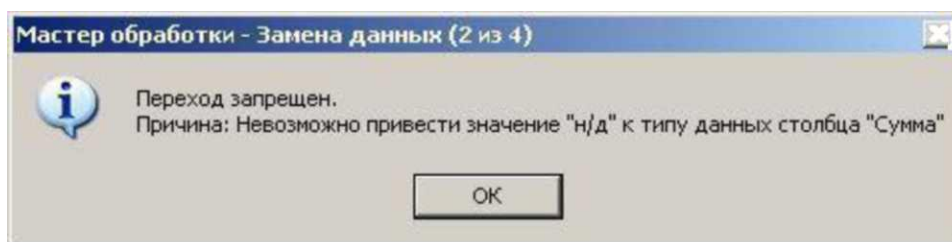
**Заменять другие на** - на какое значение следует заменить значения, не указанные в таблице замены. Для этого установите флажок и в поле напротив введите значение для замены.

**Заменять пустые на** - на какое значение заменять пустые значения поля.

Таблицу подстановки, кроме непосредственного ввода, можно заполнить, загрузив ее из текстового файла (кнопка ^). Формат текстового файла должен быть следующим:

<заменяемое значение> <символ табуляции> <значение для замены> И наоборот, список подстановок можно сохранить в текстовый файл (кнопка H).

Если по полю настроена таблица подстановок, иконка типа данных меняет свой цвет на красный. Попытка сделать замену данных с числового типа на строковый может потерпеть неудачу с выдачей соответствующего сообщения. Например, заменить все пустые значения на «н/д» в вещественном поле Сумма не получится, т.к. поле уже становится не вещественным, а строковым. Поэтому предварительно необходимо преобразовать поле Сумма в строковый тип при помощи обработчика **Настройка набора данных**.



## Фильтрация

Обработчик **Фильтрация** предназначен для исключения из набора данных записей, не удовлетворяющих условиям фильтрации.

Обработчик **Фильтрация** находится в группе узлов **Очистка данных** мастера обработки.

Параметры фильтрации задаются в виде списка условий, который содержит следующие столбцы.

1. Операция - позволяет установить функцию отношения «И» или «ИЛИ» между полями, для каждого из которых выполняется фильтрация. Возможна фильтрация по нескольким условиям для нескольких полей одновременно. В результате фильтрации по каждому из полей или условий будет получено отдельное множество значений. Функция в поле Операция устанавливает отношение между этими множествами. Если используется отношение «И», то в результирующий набор будут включены записи, удовлетворяющие условиям фильтрации по обоим полям. Если используется отношение «ИЛИ», то в выходной набор будут включены данные, удовлетворяющие хотя бы одному из условий. Установка отношений возможна, только если настроены два или более условия фильтрации. Для выбора операции следует дважды щелкнуть левой кнопкой мыши в столбце Операция для соответствующего условия и из списка, открываемого кнопкой, выбрать нужную функцию отношения. По умолчанию устанавливается отношение «И».
2. Имя поля - позволяет выбрать поле, по значениям которого должна быть выполнена фильтрация. Одно и то же поле может быть использовано в нескольких условиях.
3. Условие - указывается условие, по которому нужно выполнить фильтрацию для данного поля. Для выбора условия достаточно дважды щелкнуть мышью в соответствующей ячейке и в списке условий, открываемом кнопкой, выделить нужное условие. Доступны следующие условия фильтрации:
  - (равно), < (меньше), <= (меньше или равно), > (больше), >= (больше или равно), <> (не равно) - отбираются только те записи, значения которых в данном поле удовлетворяют заданному выражению;
  - пустой - отбираются только те записи, для которых в данном поле содержится пустое значение. В этом случае поле Значение не используется;
    - не пустой - отбираются только те записи, для которых в данном поле не содержится пустое значение. В этом случае поле Значение не используется;
  - содержит - отображаются только те записи, которые в данном столбце содержат указанное значение;

- не содержит - отображаются только те записи, которые в данном столбце не содержат указанное значение;
  - в интервале, вне интервала - для числовых полей и полей типа Дата/время отбираются только те записи, значения которых в данном столбце лежат в выбранном диапазоне (вне выбранного диапазона);
  - в списке, вне списка - отбираются только те записи, которые содержатся в выбранном списке (вне выбранного списка);
  - начинается на, не начинается на - для строковых полей отбираются записи, значения которых в данном столбце начинаются (не начинаются) на введенную последовательность символов.
  - заканчивается на, не заканчивается на - для строковых полей отбираются записи, значения которых в данном столбце заканчиваются (не заканчиваются) на введенную последовательность символов.
  - первый, не первый - для полей типа Дата/время - по данному полю отбираются первые (не первые) N периодов от выбранной даты. Периодом может быть день, неделя, месяц, квартал, год. Например, если выбрать условие первые 3 дня от 29.11.2004, то будут отображены записи, в которых значение данного поля равно 29.11.2004, 30.11.2004, 01.12.2004 - 3 последующих дня.
  - последний, не последний - для полей типа Дата/время отбираются последние (не последние) N периодов от выбранной даты. Периодом может быть день, неделя, месяц, квартал, год. Например, если выбрать условие последние 3 дня от 29.11.2004, то будут отображены записи, в которых значение данного поля равно 29.11.2004, 28.11.2004, 27.11.2004 - 3 предыдущих дня.
4. Значение - указывается значение(я), по которому будет производиться фильтрация записей в соответствии с заданным условием. Способ ввода значения будет различным в зависимости от типа данных и выбранного условия. Допустим, в качестве условия выбрана операция отношения «=», «<>», «>» и т.д. Если данные в поле являются непрерывными (т.е. числовыми), то достаточно дважды щелкнуть мышью в соответствующей ячейке, чтобы появился курсор, затем ввести значение (число). Если поле, по которому выполняется фильтрация, имеет тип «строка» (т.е. является дискретным), то в результате двойного щелчка в столбце Значение появится кнопка выбора, которая откроет окно «Список



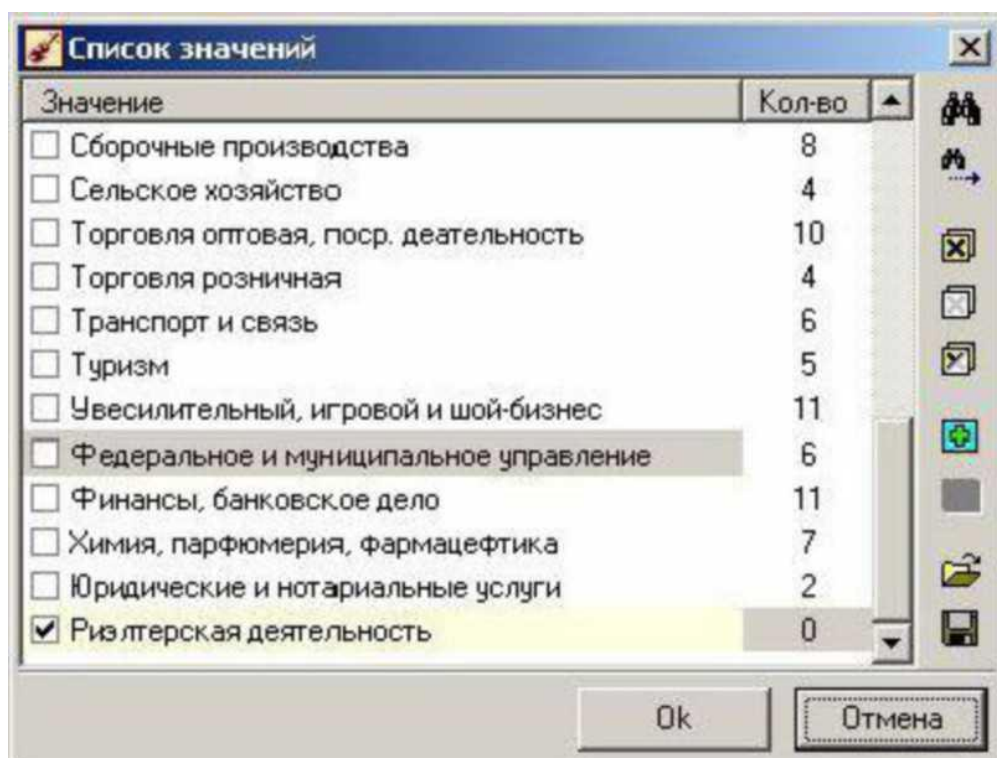
уникальных значений», где будут отображены все уникальные значения поля и их количество. Чтобы выбрать значение для условия отбора, достаточно выделить его и щелкнуть **Ok**, либо просто дважды щелкнуть мышкой на нужном значении. Если выбрано условие между или не между, тогда после щелчка мышки откроется окно, в котором необходимо указать верхнюю и нижнюю границы интервала, и так далее.

Флажок **Учитывать регистр** учитывает регистр символов при отборе.

Внизу окна настроек в автоматическом режиме формируется выражение для фильтрации, полученное объединением всех условий, например:

([Размер ссуды, руб] в интервале [2000..5000]) И ([Цель ссуды] = 'Покупка товара').

Иногда возникает необходимость построить фильтр для дискретного поля с условием *в списке, вне списка* для значений, которые не существуют в наборе данных (но предполагается, что они могут появиться в будущем). Выходом служит кнопка **®** **Добавить значение** в окне выбора списка значений. Количество записей такого «несуществующего» списочного значения всегда будет равно нулю, а строка подкрашена светло-желтым цветом.



### **Практическая работа:**

1. Создайте новый проект. Импортируйте в него текстовый файл **CreditSample.txt**, идущий в поставке Deductor (по умолчанию расположен в каталоге /Samples директории установки Deductor).
  2. Отсортируйте этот набор данных по следующим полям в порядке возрастания: Срок ссуды, Размер ссуды, Количество иждивенцев.
  3. Сделайте следующую замену (после **Сортировки**) в поле Семейное положение: значение Да измените на Женат/замужем, Нет - на Холост/Не замужем.
  4. Сделайте следующую замену (после предыдущего узла **Замена данных**) в поле Количество иждивенцев: значение 0 - на Нет, 1 - без изменений, 2 и 3 - 2 и более. Используйте два способа - непосредственным вводом в мастере обработки и через файл таблицы соответствий. Файл подстановок предварительно создайте в любом текстовом редакторе, например, в Блокноте.
  5. Старое поле Количество иждивенцев удалите из набора данных, а новое поле Количество иждивенцев\_1 переименуйте в Иждивенцы.
  6. Отфильтруйте набор данных, полученный в п. 5 по полю Иждивенцы так, чтобы в выходной набор попали только строки, у которых значение в поле Иждивенцы не равно Нет. Сколько записей прошло через фильтр?
  7. Отфильтруйте набор данных, полученный в п. 5 по полю Иждивенцы так, чтобы в выходной набор попали только строки, у которых значение в поле Иждивенцы не равно Н/д. Сколько записей прошло через фильтр?
  8. Продолжите фильтровать набор данных, полученный в п. 6. Наложите следующий фильтр, в который попадают все записи, удовлетворяющие условиям а либо условиям б:
    - а. Размер ссуды - от 2000 до 5000, Цель ссуды - Покупка товара.
    - б. Цель ссуды - Иное.Сколько записей прошло через фильтр?
- Отсортируйте последний набор данных по полю Код.

### **Вопросы для самоконтроля:**

1. Как работает обработчик **Сортировка**?

2. Можно ли отсортировать набор данных по нескольким полям?
3. Для чего предназначен узел **Замена данных**?
4. Как определить в мастере обработки, что для поля настроена замена?
5. Как работает **Замена данных**?
6. Какие существуют способы заполнить таблицу подстановок?
7. Для чего предназначен узел **Фильтр**?
8. Какие условия фильтрации существуют?
9. Сколько записей будет отфильтровано в результате фильтра «([Размер ссуды, руб] в интервале [2000..5000]) И ([Цель ссуды] = 'Покупка товара') И ([Цель ссуды] = 'Иное')» ?
10. Что делать, если нужно поставить фильтр по значению, которого в данный момент нет в рассматриваемом наборе данных?

### **Библиографический список**

1. Deductor Studio [Электронный ресурс]:  
[www.basegroup.ru/download/deductor/](http://www.basegroup.ru/download/deductor/).
2. Решения по построению хранилищ данных [Электронный ресурс]: <http://ibarus.ru/solutions/dwh/>?
3. Основные обработчики в Deductor Studio [Электронный ресурс]:  
[http://deductor.org/Deductor\\_help\\_manual/deductor-help.html](http://deductor.org/Deductor_help_manual/deductor-help.html).