Comparator

О программе	2
Порядок работы	
настройка источников данных	
База данных	
Excel	
XML	
Текст с разделителями	
Настройка полей	10
Параметры обработки	11
Просмотр результатов сверки	12
Режимы запуска	
Настройки приложения	15
Настройки консольного режима	15
Общие настройки	16

О программе

Эта программа может пригодиться, если нужно сравнить табличные данные из двух источников, а именно:

- определить соответствие строк из заданных источников по совпадению ключевых полей (поиск пары по ключу);
- сравнить содержимое определенных полей в найденных парах строк;
- выявить строки, для которых не было найдено соответствие в другом источнике по ключу;
- выявить вхождение строки более чем в одну пару (неуникальность ключа);
- проанализировать полученный результат, сохранить его в общедоступных форматах;
- проводить регулярные сверки на основе предварительно настроенных параметров (профилей) и делать рассылку результатов.

Системные требования

Windows XP/7/8/10/Server .NET 4.5

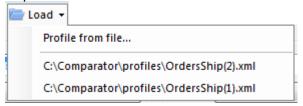
Порядок работы

Сначала...

Делаем параметры сверки «с нуля» или на основе уже загруженного профиля настроек:

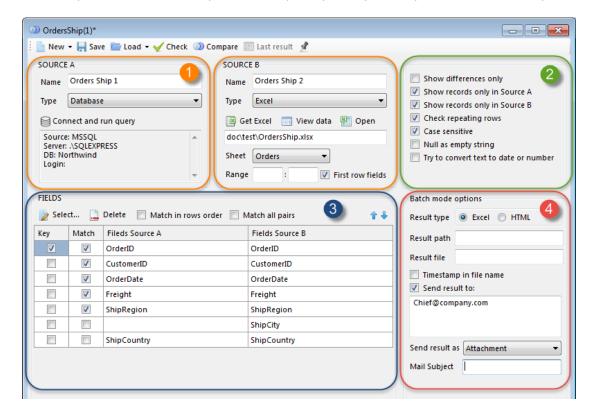


Или открываем существующий профиль, выбирая файл с диска или из списка недавно открытых:



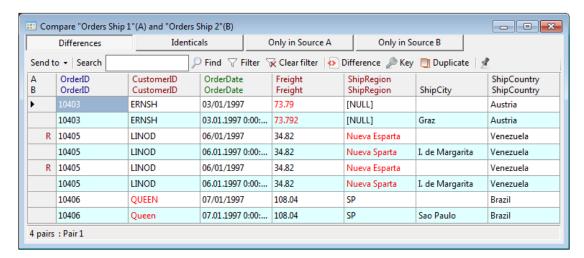
Затем...

- 1. Определяем два источника данных;
- 2. Устанавливаем параметры обработки данных и формирования результата;
- 3. Задаем поля источников для сверки и отображения, определяем условия отбора пар:
- 4. При необходимости определяем параметры запуска сверки из командной строки.

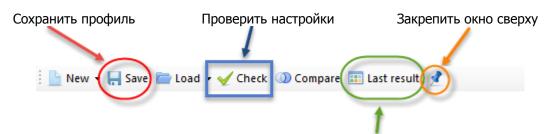


И наконец...

Запускаем сверку по кнопке ОС Сомраге и получаем результаты в отдельном окне:



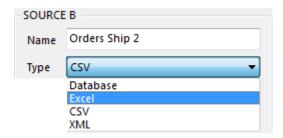
Кроме того можно:



Показать результаты последней сверки

Настройка источников данных

Имя источника будет использовано в отображении результатов сверки, наименовании профиля и теме рассылки, предлагаемых по умолчанию.



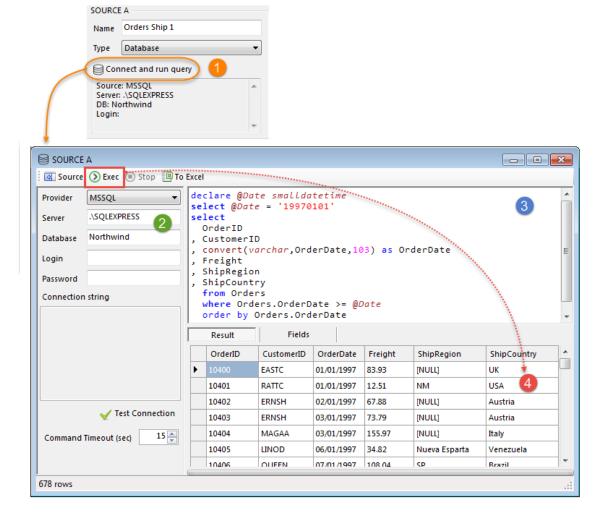
В зависимости от выбранного типа источника - база данных, файл Excel, текст с разделителями (CSV) или файл XML - необходимо выполнить определенные настройки.

База данных

Это может быть любой источник, доступный через OleDB или ODBC для получения данных с помощью SQL.

Для настройки необходимо:

1. Открыть форму настройки параметров подключения и запроса:



- 2. Настроить параметры подключения. Для удобства основные параметры подключения через OleDB для MSSQL, Sybase и Oracle вынесены в отдельные поля. В общем случае (провайдеры OleDB или ODBC) необходимо указать строку подключения. В файле настроек профиля пароль и строка подключения
 - хранятся в зашифрованном виде.
- 3. Ввести скрипт SQL.
- Выполнить скрипт. Это необходимо сделать до настройки полей для сверки, чтобы иметь актуальный перечень полей этого источника.

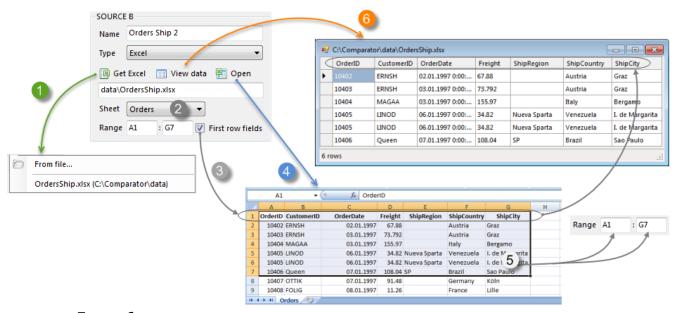
Примечание:

- если скрипт возвращает несколько наборов данных, рассматривается только первый из них.
- для отладки возможно выполнение только выделенной части скрипта.



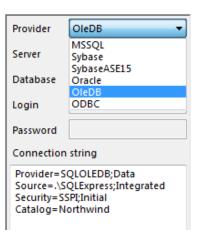
Для настройки получения данных из файла Excel необходимо:

- Выбрать файл Excel или одну из книг, открытых в данный момент. Можно указывать путь к файлу относительно папки приложения.
- Выбрать лист книги.



При необходимости можно:

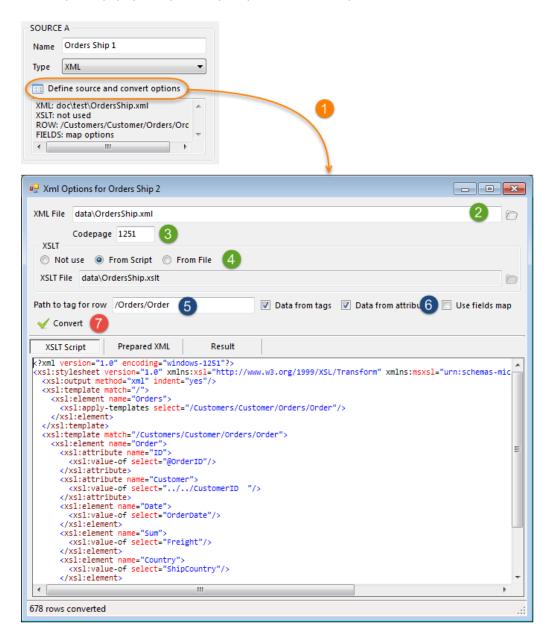
- 3. Отметить, что имена полей должны определяться из заполненных ячеек первой строки установленного диапазона (или всего листа, если диапазон не установлен). В противном случае имя поля соответствует букве столбца.
- 4. Открыть/сделать активной выбранную книгу.
- 5. Установить границы диапазона или вручную, или выделив его на листе. Если какие-то из границ диапазона не заданы, рассматривается используемый диапазон
- 6. Просмотреть данные, которые будут получены в соответствии с выполненными настройками.



XML

Для сверки данные XML необходимо преобразовать в табличный вид. Для настройки процесса преобразования необходимо:

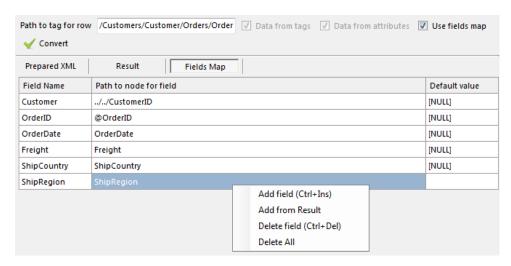
1. Открыть форму настройки преобразования ХМL-файла:



- 2. Выбрать файл ХМL (можно указывать путь относительно каталога приложения).
- 3. При необходимости указать номер кодовой страницы (1251, 866 и т.д.) или принятое обозначение (UTF8, ср866 и т.п.). По умолчанию кодировка определяется по атрибуту encoding в заголовке xml.
- 4. Выбрать вариант предварительного использования XSLT перед преобразованием:
 - Not use не используется;
 - From Script по кнопке XSLT Script становится доступным редактор, в котором необходимо ввести скрипт XSL для предварительного преобразования исходного XML.
 - From File необходимо указать файл XSL для тех же целей (можно указывать путь относительно каталога приложения);

- 5. Ввести путь (в формате XPath) к тэгу, относительно которого будет определяться строка данных таблицы, формируемой в результате преобразования;
- 6. Выбрать способ определения полей таблицы, формируемой в результате преобразования:
 - Data from tags поля определяются по тэгам, дочерним по отношению к тэгам, выбранным в соответствии с Xpath, указанным в п.5;
 - Data from attributes поля определяются по атрибутам тэгов, выбранных в соответствии с Храth, указанным в п.5;
 - Use fields map если выбрать этот вариант, предыдущие варианты определения полей будут не доступны и появится возможность настроить пути (в формате XPath) к тэгам или атрибутам, определяющим поля. Пути должны быть заданы относительно тэга, определенного в п.5. В этом случае, по кнопке Fields Map, становится доступным список, в котором для каждого поля нужно обязательно указать имя и путь, а также, при необходимости, значение по умолчанию для случая, если по указанному пути не будут найдены данные.

Если результат преобразования уже был получен ранее, то поля этого результата можно добавить в список, выбрав соответствующий пункт меню.



- 7. Для проверки процесса преобразования XML нужно нажать кнопку «Convert». В результате:
 - указанный ХМL-файл будет прочитан и, при необходимости, предварительно преобразован через ХSLТ. Подготовленный ХМL доступен для просмотра по кнопке «Prepared XML»;
 - на основании настроек п.5 и п.6 из данных подготовленного XML будет сформирована таблица с данными, доступная для просмотра по кнопке «Result»;

```
Prepared XML
                       Result
                                        Fields Map
k?xml version="1.0" encoding="utf-8" standalone="yes"?>
Customers>
  <Customer:
    <CustomerID>ALFKI</CustomerID>
    <Orders>
      <Order OrderID="10643":
        <OrderDate>25/08/1997</OrderDate>
        <Freight>29.4600</preight>
        <ShipCountry>Germany</ShipCountry>
      </order>
<order OrderID="10692";
        <OrderDate>03/10/1997</OrderDate>
        <Freight>61.0200</preight;</pre>
         <ShipCountry>Germany</ShipCountry>
      </order>
```

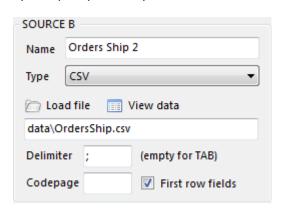
Pro	epared XML	Result	Fields Map				
	Customer	OrderID	OrderDate	Freight	ShipCountry	ShipRegion	
>	ALFKI	10643	25/08/1997	29.4600	Germany		
	ALFKI	10692	03/10/1997	61.0200	Germany		
	ALFKI	10702	13/10/1997	23.9400	Germany		

Текст с разделителями

Для выбранного файла необходимо указать разделитель полей (по умолчанию - символ табуляции) и кодировку - номер кодовой страницы (1251, 866 и т.д.) или принятое обозначение (UTF8, ср866 и т.п.). Если кодировка не указана - используется текущая кодовая страница ANSI операционной системы.

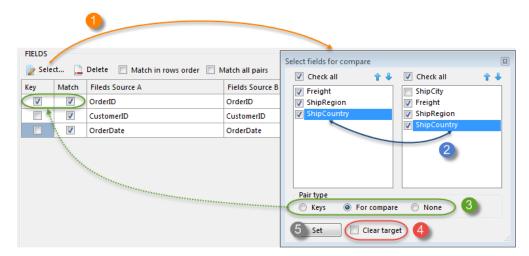
Аналогично, как и для Excel-источника, можно отметить, что имена полей должны определяться по данным первой строки. Также можно просмотреть разобранные (по разделителю) данные файла в отдельном окне, в указанной кодировке, по кнопке "View data".

Путь к файлу можно указать относительно папки приложения.



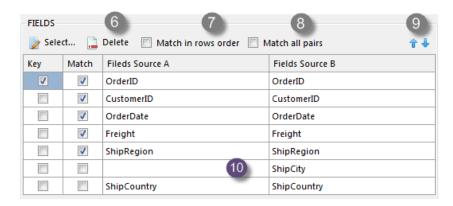
Настройка полей

После настройки источников необходимо определить поля, данные из которых будут участвовать в сверке. Для этого:



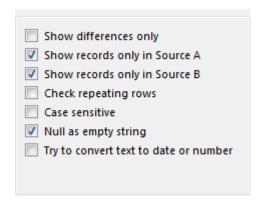
- 1. Открываем списки еще не отобранных полей (слева поля источника A, справа источника B).
- 2. Отмечаем поля, которые должны участвовать в сверке. Соответствие полей (пара) определяется по порядку сверху вниз. Поле из соседнего списка подсвечивается при перемещении. С целью подобрать нужную пару можно изменить порядок следования полей. Полей в одном из источников может быть отмечено больше, чем в другом. Данные из поля, не имеющего пару, не будут участвовать в сверке, но будут отображаться в результатах.
- 3. Определяем тип отмеченных полей ключи, поля для сравнения или только для отображения (None).
- 4. Определяем нужно ли очистить список ранее отобранных полей.
- 5. Добавляем отмеченные поля в список.

Список полей для сверки можно при необходимости отредактировать:



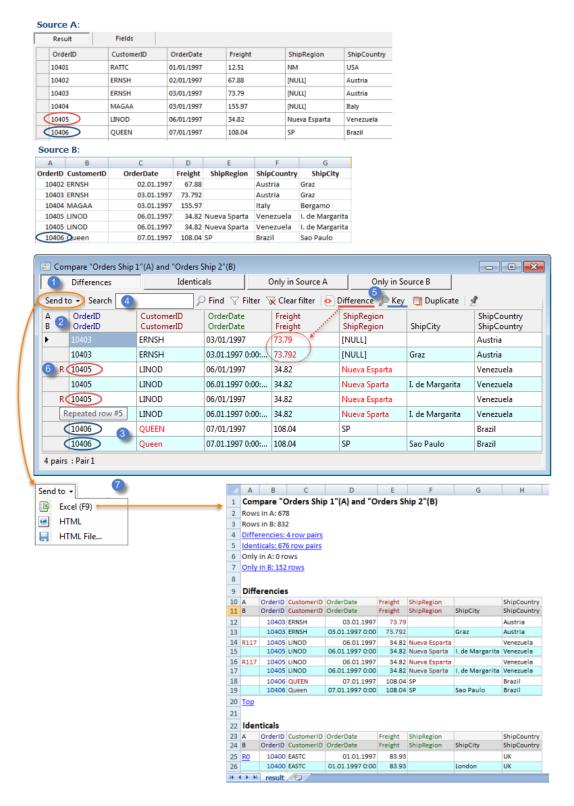
- 6. Удалить строки, ячейки которых выделены в данный момент.
- 7. Определить, что пары строк будут определяться по порядку их следования в источнике, а не по ключам.
- 8. Определить, что сравниваться будут данных всех полей.
- 9. Изменить порядок полей в результатах сверки.
- 10. Указать тип обработки пары полей:
- "Кеу" поля являются ключевыми и по их совпадению будет определяться пара строк из источников;
 - "Match" данные из этих полей будут сравниваться в найденных парах строк.

Параметры обработки



- **Show differences only** если отмечен, пары строк без расхождений не включаются в результат сверки.
- **Show records only in Source A** если отмечен, в результат сверки включаются строки источника A, для которых не нашлось пар в источнике B.
- **Show records only in Source B** если отмечен, в результат сверки включаются строки источника B, для которых не нашлось пар в источнике A.
- Check repeating rows если отмечен, происходит поиск всех возможных пар для каждого набора ключевых полей и, если одна и та же строка входит более чем в одну пару, она помечается в результатах сверки как повторяющаяся. Эта настройка замедляет процесс сверки. Если не отмечен поиск для каждого набора ключевых полей прекращается после нахождения первой пары.
- Case sensitive определяет чувствительность сверки к регистру символов.
- **Null as empty string** если отмечен, поле со значением NULL при сверке считается равным пустой строке.
- Try to convert text to date or number если отмечен, в процессе сверки будет попытка привести значения текстовых полей к типу дата/время или к числовому типу. Например значения "20151231" и "31.12.2015" в этом случае будут считаться одинаковыми.

Просмотр результатов сверки



- (1) Результаты сверки отображаются в отдельном окне, где в зависимости от параметров сверки доступны несколько наборов результатов, между которыми можно переключаться:
- Differences пары строк, в которых найдены отличия;
- Identicals пары строк, в которых не найдено отличий;
- Only in Source A строки источника A, для которых не нашлось пар в источнике B;

- Only in Source B строки источника B, для которых не нашлось пар в источнике A; (2) Наименования полей результата выделяются красным цветом в случае расхождений; синим цветом выделяются ключевые поля, зеленым поля, в которых расхождений не обнаружено.
- (3) Заголовки и строки в наборах результатов "Differences" и "Identicals" выводятся попарно первая строка из источника А, вторая из источника В (выделяется цветом). Если одна и та же строка входит более чем в одну пару (6), она помечается как повторяющаяся в заголовке строки проставляется «R» с указанием во всплывающей подсказке номера этой строки в источнике (начиная с 0).
- (4) По значению, указанному в поле Search, можно проводить поиск по всей таблице или в пределах текущего столбца, а также фильтровать данные в текущем столбце.
- (5) Кнопка "Difference" позволяет перемещаться по ячейкам с расхождениями, кнопка "Key" по ключевым полям, кнопка "Duplicate" по повторяющимся строкам. Описание комбинаций клавиш можно увидеть во всплывающих подсказках к соответствующим кнопкам.
- (7) Все результаты можно открыть в Excel или HTML. В этом случае в папке результатов формируется файл result.xls или result.htm соответственно (путь к папке результатов определяется в настройках приложения). Также можно сохранить результаты как HTML-файл.

Режимы запуска

1. Запуск в оконном режиме (используются обычные настройки приложения)

```
comparator [-profile:"[путь]имя"]
```

profile: "[путь]имя" - файл профиля, который загрузится сразу. Если не указывать - приложение запустится с новым профилем - пустым или заполненным данными из профиля-шаблона (профиль-шаблон указывается в настройках приложения).

2. Запуск в консольном режиме (используются настройки приложения для консольного режима)

```
comparator -batch -profile:"[путь]имя" [-path:"путь"] [-log:"[путь]имя"] [-type:excel/html] [-sendto:"адрес[;адрес...]"] [-file:"имя"] [-open]
```

обязательные параметры:

-batch - признак консольного режима - будут использованы настройки консольного режима указанного профиля (см.ниже);

```
-profile:"[путь]имя" - файл профиля;
```

необязательные параметры - игнорируются если параметр -batch не указан, перекрывают аналогичные настройки из профиля и настроек приложения:

-path:"путь" - путь для файла результата (в конце не должно быть слэша):

-log:"[путь]имя" - лог-файл;

-type:excel/html - тип результата (excel или html);

-sendto: "appec[;appec...]" - список адресатов для рассылки результата - если задан, но список пустой, результат не рассылается;

-file:"имя" - имя файла результата;

-open - если задан, файл результата будет открыт.

После запуска в консольном режиме происходит проверка данных профиля, поиск/подключение к источникам, сверка и формирование файла результата, его рассылка по почте и/или открытие.

Процесс отражается в лог-файле:

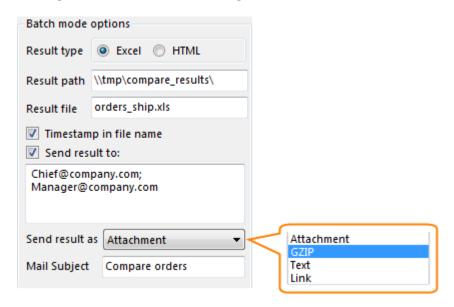
[08.02.2016 20:37:07.776] >>> Start process for "testprofile.xml" [08.02.2016 20:37:08.192] Profile loaded

[08.02.2016 20:37:09.977] Compared

[08.02.2016 20:37:10.002] Save result in "C:\Comparator\results\testprofile__20160208_203709.xls"

Настройки приложения

Настройки консольного режима



- Result type тип результата (excel или html).
- **Result path** путь для файла результата (по умолчанию из настроек приложения).
- **Result file** имя файла результата (по умолчанию имя профиля с расширением в соответствии с типом результата).
- **Timestamp in file name** если отмечено к имени файла результата будут добавлены дата и время запуска сверки.
- Send result to если отмечено, необходимо:
 - о указать адреса рассылки (через точку-с-запятой)
 - о выбрать один из вариантов отправки результата в поле **Send result as**
 - Attachment файл результата во вложении;
 - ZIP файл результата во вложении в архиве в формате Gzip (имя файла.**gz**) этот вариант предпочтителен для больших объемов;
 - Text отправляется содержимое файла результата как HTML в тексте письма;
 - Link отправляется ссылка на файл результата, сам файл не отправляется;
 - по желанию указать тему письма в поле Mail Subject (по умолчанию «Сотраге "имя источника A" and "имя источника B"»)

Если не отмечено или адресаты не указаны - рассылки не будет.

Другие настройки рассылки - подключение к SMTP серверу и т.п. - см. в настройках приложения.

Общие настройки

В качестве глобальных настроек приложения используем файл **Comparator.xml** который должен находится в каталоге приложения.

Если он не существует, то будет создан при первом запуске. Пример:

```
<?xml version="1.0" encoding="windows-1251"?>
<Config xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"</pre>
xmlns:xsd="http://www.w3.org/2001/XMLSchema">
 <optionsList>
    <Options Host="BATCH">
      <ProfileFolder>C:\Comparator\profiles
     <ResultFolder>C:\Comparator\results
     <LogFile>C:\Comparator\Comparator.log</LogFile>
     <HtmlStylesFile>C:\Comparator\styles.css/HtmlStylesFile>
     <SendOptions>
       <Server>smtp.server.ru
       <Port />
       <Ssl>false</Ssl>
       <User />
       <Pwd />
       <PwdEncr />
       <From>Comparator@server.ru</from>
        <FromAlias>Comparator/FromAlias>
     </SendOptions>
    </Options>
   <Options Host="HOST1">
      <RecentFiles>
       <file>C:\Comparator\profiles\Источник1 Источник2.xml</file>
        <file>C:\Comparator\profiles\test.xml</file>
     </RecentFiles>
     <ProfileFolder>C:\Comparator\profiles</ProfileFolder>
     <ResultFolder>C:\Comparator\results
     <HtmlStylesFile>C:\Comparator\styles.css/HtmlStylesFile>
     <PatternFile>C:\Comparator\pattern.xml</PatternFile>
   </Options>
   <Options Host="HOST2">
      . . . . . . . . . . . . . .
   </Options>
 </optionsList>
</Config>
```

Настройки для консольного режима - внутри тэга <Options> с атрибутом + Host="BATCH".

Настройки для обычного (оконного) режима пишутся для каждого хоста внутри тэга - <Options> c атрибутом Host="имя хоста".

При первом запуске с хоста прописываются дефолтные настройки, потом их можно изменить, отредактировав вручную файл.

Предусмотрены следующие тэги для хранения настроек:

<ProfileFolder> - путь к профилям (по умолчанию папка profile в каталоге приложения);

<ResultFolder> - путь для сохранения результата (по умолчанию папка result в каталоге приложения);

<LogFile> - лог-файл (по умолчанию comparator.log в каталоге приложения); <HtmlStylesFile> - файл стилей (CSS) для оформления результата (по умолчанию styles.css в каталоге приложения);

<PatternFile> - файл профиля-шаблона, по данным которого будут по умолчанию

заполнены поля нового профиля;