

# **Настройка FO-процессоров для вывода символов кириллицы в среде Microsoft Windows**

## **Рекомендации по освоению**

*Михаил Острогорский*

Введение.....	1
Начальные сведения о цифровых шрифтах.....	2
Шрифт как явление культуры.....	2
Символы и глифы.....	3
Кодировки символов.....	4
Гарнитур и начертания.....	6
Семейства шрифтов.....	6
Форматы и файлы шрифтов в среде Microsoft Windows.....	7
Шрифт как цифровой продукт.....	8
Просмотр шрифта.....	9
Настройка FO-процессоров.....	14
RenderX XEP.....	14
Apache FOP.....	18
Литература.....	19

## **Введение**

Первое затруднение, которое обычно испытывает отечественный пользователь FO-процессора, — в сформированном документе не видно русских букв. Их место занято пустотой, квадратиками или символами других алфавитов.

Как правило, проблемы с кириллицей вызваны не принципиальным отсутствием ее поддержки в FO-процессоре, а необходимостью предварительной настройки

используемой программы. Мы должны указать ей, где у нас в системе расположены подходящие для вывода кириллицы шрифты. Но прежде мы сами должны в этом как следует разобраться, иначе мы не найдем нужных шрифтов и не поймем, чего от нас хотели разработчики FO-процессора, предлагая указать те или иные параметры его конфигурации.

В предлагаемом обзоре приводятся начальные сведения о цифровых шрифтах и даются рекомендации по настройке кириллицы в популярных FO-процессорах: платном RenderX XEP и бесплатном Apache FOP.

Обзор не претендует на полноту изложения фактов как о шрифтах, так и о FO-процессорах, но дает указания, достаточные для настройки последних в большинстве случаев.

Автор будет чрезвычайно признателен читателям за сообщения о замеченных ими ошибках и неточностях, которые можно прислать по адресу: [misha@philosoft.ru](mailto:misha@philosoft.ru).

## **Начальные сведения о цифровых шрифтах**

### **Шрифт как явление культуры**

Печатный шрифт — сложное культурное явление, история которого восходит ко времени Иоганна Гутенберга<sup>1</sup>. Шрифтам посвящены многочисленные научные и практические труды, существуют объемные каталоги шрифтов различного назначения. Даже общий обзор современного состояния этой области потребовал бы написания обстоятельной искусствоведческой статьи. Вдобавок повсеместное использование цифровых шрифтов приводит к необходимости вникать еще и в технологическую сторону дела.

Поэтому мы не можем здесь говорить о шрифтах подробно. Заинтересованный читатель сам без труда найдет необходимые ему источники<sup>2</sup>.

Нам, простым смертным, для подключения кириллических шрифтов к FO-процессору достаточно знания нескольких важных фактов. Более того, если читатель очень торопится,

---

<sup>1</sup> Промышленная эксплуатация первого книгопечатного станка в Европе началась приблизительно в 1450 г. Хотя идея покрыть образец краской и приложить к нему бумагу приходила людям в голову и прежде, информационную революцию вызвало именно изобретение Гутенберга. Возможно, оно состояло не в том, о чем мы обычно думаем. Он первый понял, что печать — это не способ сделать красиво, а способ сделать много и быстро.

<sup>2</sup> Словарь терминов (или даже небольшая энциклопедия) по шрифтам есть на сайте фирмы «ПараТайп», которая занимается разработкой шрифтов: <http://www.paratype.ru>.

он может пропустить этот раздел и вернуться к нему, только если процедуры, приведенные в следующем разделе, не приведут к желаемому результату.

## **Символы и глифы**

Говоря о шрифтах, необходимо различать *символы* и их *глифы*.

Под *символом* обычно понимают минимальную осмысленную часть сообщения. Символ отличается от других символов из того же набора своим значением для читателя и только им. Если мы заменим в каком-нибудь тексте каждую букву пляшущим человечком, его смысл сохранится. Главное, чтобы каждой букве используемого алфавита соответствовала определенная отличная от других фигурка, и читатель был в состоянии их различать<sup>3</sup>.

*Глифом* называют конкретное изображение, рисунок определенной формы, которая позволяет читателю опознать символ.

Однозначного соответствия между глифами и символами в общем случае не наблюдается.

В некоторых языках для одного и того же символа может потребоваться несколько глифов. Так, в арабском языке и в иврите внешний вид буквы может зависеть от ее позиции в слове.

Бывают и обратные ситуации, когда один глиф отражает целую последовательность символов. Обычно потребность в таких глифах, называемых *лигатурами*, продиктована типографской традицией, присущей культуре определенной страны. Лигатура позволяет изобразить некоторую последовательность символов наиболее эстетичным, компактным, привычным для образованного читателя способом. Например, лигатура *ij* делает возможным размещение подстрочного элемента буквы *j* под буквой *i*.

Под *шрифтом* обычно понимается набор глифов, каждый из которых сопоставлен определенному символу (а не просто «альбом с картинками»).

Типичный *алфавитный шрифт* содержит глифы для цифр, знаков препинания, букв того или иного алфавита и ряда специальных символов: «решетки», «собаки», «звездочки»,

---

<sup>3</sup> С китайскими иероглифами дело обстоит не совсем так. Во-первых, китайский иероглиф может состоять из нескольких значимых частей, во-вторых, форма многих китайских иероглифов прямо или косвенно отражает их смысл. Таким образом, китайские иероглифы — не символы в строгом значении этого термина. Тем не менее, в большинстве современных операционных систем с ними обращаются как с символами. Есть примеры и поближе: шрифт *Webdings*, поставляемый в составе операционной системы *Microsoft Windows*, содержит не символы, а пиктограммы, технически реализованные как символы. Короче говоря, для программиста символ — не совсем то, что для филолога или культуролога.

доллара и т. п. Многие современные шрифты включают в себя глифы для нескольких алфавитов и наборов специальных символов.

Существуют полностью *неалфавитные шрифты*, предназначенные для набора пиктограмм или, например, элементов штрих-кода. В поставку операционной системы Microsoft Windows входит несколько таких шрифтов.

### **Кодировки символов**

Современные компьютеры, по крайней мере, если говорить о более-менее распространенных архитектурах, не способны хранить никаких данных кроме чисел. Поэтому при обработке текстов и строк приходится каждый символ обозначать (иными словами, *кодировать*) определенным числом. Например, символ пробела в большинстве операционных систем имеет код 32.

Соответствие между символами из определенного набора и обозначающими их числами называется *кодировкой*. Кодировки прошли извилистый и долгий путь развития. Сегодня они выглядят весьма причудливо, особенно в том, что касается кириллицы.

Пока запоминающие устройства стоили сравнительно дорого, а пропускная способность каналов связи была дефицитным ресурсом, на каждый символ отводили семь бит, что позволяло поместить в один набор не более 128 символов с кодами от 0 до 127.

Следовательно, на экран в текстовом режиме, который использовался в те времена практически повсеместно, можно было вывести не более 127 разных символов одновременно. Этого было мало. Например, напечатать таблицу с рамками, в которой левый столбец занимают английские слова, а правый — их русские переводы, не получалось.

От семибитных кодировок вскоре перешли к восьмибитным. Они дали возможность включать в один набор 256 символов, в том числе, латинские буквы, псевдографику и еще какой-нибудь национальный алфавит, например, греческий или кириллицу. Это сильно облегчило жизнь пользователям «нелатинского» мира, хотя работа, предположим, с греческо-русским текстом требовала применения специальных средств наподобие редактора ChiWriter (мир праху его).

Поскольку в мировой культуре намного больше разных символов, для каждого национального алфавита приходилось создавать изолированную кодировку. Выходило, что в разных кодировках символам разных национальных алфавитов были присвоены

одни и те же коды. И программистам, и пользователям работать в таких условиях было довольно неудобно. Прежде, чем вывести на экран или на принтер какой-нибудь текст, приходилось, во-первых, тем или иным способом сообщить системе его кодировку и, во-вторых, загрузить в систему шрифт, в котором имеются нужные глифы.

Существование множества кодировок, в которых разные символы были обозначены одними и теми же кодами, сильно затрудняло работу с многоязычными текстами и локализацию приложений для разных языков и стран. Путаница в кодировках и шрифтах приводила к обескураживающим результатам. Например, если русский текст по ошибке печатали шрифтом с глифами от другого алфавита, пользователь получал «кракозябры»: чепуху, набранную греческими или диакритическими символами. Написание экранных, клавиатурных и принтерных драйверов, изготовление «на коленке» срочно необходимых шрифтов и борьба с «кракозябрами» на местах были постоянным занятием программистов и системных администраторов. Советским гражданам (оказавшимся в дальнейшем россиянами) хлопот добавляло и то, что разнобой в действиях вендоров привел к появлению, по крайней мере, четырех разных кириллических кодировок.

Как только с компьютерной памятью и со связью стало лучше, в ход пошла так называемая система Юникод. В ней каждый символ, изобретенный когда-либо человеком и занявший свое место в культуре, получил уникальный числовой код. Например, в восьмибитных кодировках Microsoft Windows русская буква У и греческая Σ имели совпадающий код 211. В Юникоде же букве У присвоен код 1059, а букве Σ код 931. Есть разница? Юникод позволяет без лишних ухищрений сочетать в одном тексте символы разных алфавитов, например, набрать русско-армяно-японский словарь.

Существует несколько кодировок, основанных на Юникоде. При общем наборе кодов они подразумевают разные по ресурсоемкости и по совместимости со старыми кодировками способы хранения текста в памяти. Для технического писателя, работающего с XML, важнее всего кодировка UTF-8.

Ну, хорошо. Что же случится, если мы попробуем напечатать русский текст шрифтом, в котором нет глифов для кириллицы? «Кракозябров» мы не увидим, а увидим строки белых квадратов, обведенных черными рамками. Поэтому проблема выбора шрифта, с нужными глифами, вообще говоря, остается. Другое дело, что шрифты, поставляемые вместе с Microsoft Windows, содержат глифы для большинства языков: глобализация.

## Гарнитуры и начертания

У каждого шрифта есть две основных характеристики: гарнитура и начертание.

*Гарнитурой* называют характерный рисунок шрифта, заложенную в него графическую идею, которая делает его узнаваемым и отличным от всех остальных шрифтов. Например, Times и Arial — это две разные гарнитуры. Человек, который хотя бы пару раз видел набранный ими текст, никогда их не перепутает.

*Начертаниями* называются различные способы видоизменения глифов при сохранении их гарнитуры. В быту и на производстве наиболее распространены следующие начертания: нормальное начертание, курсив, жирное или полужирное начертание и жирный или полужирный курсив. Кроме того, существуют наклонное (отличное от курсива!), сверхсветлое, сверхжирное и ряд других начертаний, но рассмотреть их в небольшой технической статье у нас, к сожалению, не получится.

Обратите внимание, что в языках XSL-FO и CSS реализован немного другой подход к видоизменению глифов одной гарнитуры. Вариант формы глифа и степень преобладания черного цвета над белым на отведенной ему площади предлагается задавать с помощью двух независимых атрибутов: стиля и веса. Иногда это различие приводит к обескураживающим результатам. Предположим, пользователь (автор, верстальщик, программист) указывает в разметке стиль курсив и вес жирный, но реально в системе для выбранной им гарнитуры жирного курсива нет. Тогда при выводе документа отсутствующий шрифт будет заменен шрифтом по умолчанию, который может выглядеть совсем иначе. Хуже того, в некоторых системах оформленный отсутствующим шрифтом текст может вовсе не отображаться.

## Семейства шрифтов

Гарнитуры, предназначенные для набора более-менее объемных текстов, делятся на два крупных *семейства*: шрифты с засечками и шрифты без засечек. Шрифты с засечками также называются серифными шрифтами или *антиквами*, а шрифты без засечек — сансерифными шрифтами или *гротесками*. Считается, что первыми следует набирать основной текст, а вторыми заголовки, однако, в любом книжном шкафу легко найти образцы, опровергающие это правило.

Из наиболее популярных среди пользователей офисных программ к шрифтам с засечками относятся Times и Courier, а к шрифтам без засечек Arial, Tahoma, Verdana.

Помимо антиквы и гротесков существуют другие семейства шрифтов, в частности, *акцидентные* шрифты, которые больше подходят для плакатов, рекламы, заставок и других случаев, когда необходимо «зацепить» читателя, привлечь его внимание.

Одновременно с этим шрифты подразделяют на пропорциональные и моноширинные. В пропорциональном шрифте ширина символа зависит от его формы, например, буква Ш занимает больше места в строке, чем цифра 1. В моноширинном шрифте, как недвусмысленно следует из этого названия, все символы<sup>4</sup> одной ширины.

Из наиболее распространенных к моноширинным шрифтам относятся вездесущий Courier и хорошо знакомый пользователям макинтошей Fixedsys.

Моноширинные шрифты полезны при наборе таблиц, потому что тогда у чисел в столбцах единицы оказываются под единицами, десятки под десятками, сотни под сотнями и т. д.: читателю проще их сравнивать. Кроме того, моноширинными шрифтами по традиции набирают листинги программ, листинги входных и выходных данных и тому подобный «компьютерный» текст.

Языки XSL-FO и CSS позволяют вместо конкретной гарнитуры указывать семейство шрифтов. При формировании разметки не всегда заранее ясно, какие шрифты окажутся в той системе, где документ будут просматривать или печатать, поэтому выбор наиболее подходящего шрифта из имеющихся осуществляется системой автоматически при выводе документа. Вопреки типографике в языках XSL-FO и CSS значения *serif*, *sans-serif*, *cursive*, *fantasy* и *monospace*, используемые для обозначения шрифтовых семейств, образуют единый набор взаимоисключающих вариантов.

Остается добавить, что все сказанное относится к современным латинским и кириллическим шрифтам. Шрифты, которыми пользовались в XIX веке и ранее<sup>5</sup>, а также алфавиты восточных языков мы здесь не обсуждаем.

## **Форматы и файлы шрифтов в среде Microsoft Windows**

В среде Microsoft Windows шрифт, вообще говоря, является принадлежностью системы. Иными словами, все приложения, функционирующие внутри одной системы, пользуются

---

<sup>4</sup> Именно символы, а не глифы. В моноширинном шрифте глиф, обозначающий последовательность из двух символов, должен быть ровно вдвое шире глифа, обозначающего один символ.

<sup>5</sup> Немецкие газеты традиционно набирали *фрактурой*, определенной разновидностью готического шрифта. Фрактурой же часто делали надписи на вывесках и указателях. Как ни странно, во время второй мировой войны гитлеровское правительство запретило фрактуру несмотря на ее отчетливо национальный характер. См. Ян Чихольд, «Новая типографика», издание Студии Артемия Лебедева.

общим набором шрифтов. Конечно, каждый конкретный разработчик располагает возможностью написать приложение, которое будет пользоваться встроенными в него шрифтами, недоступными другим приложениям, но обычно так не делают.

Существуют разные форматы представления шрифтов.

Прежде всего, шрифты подразделяются на растровые и векторные. В растровом шрифте глиф представляет собой прямоугольную матрицу пикселей, каждый из которых может быть черным или белым. В векторном шрифте глиф — это набор команд и данных, «объясняющих» системе, как его начертить.

Существуют разные форматы представления шрифтов. В современных версиях операционной системы Microsoft Windows мы чаще всего имеем дело со шрифтами форматов TrueType и OpenType (последний представляет собой усовершенствованную версию первого). Технически шрифт любого из этих двух типов представляет собой один файл, который содержит глифы определенной гарнитуры в определенном начертании, например Arial жирный курсив.

Шрифты, входящие в поставку Microsoft Windows или установленные на компьютер позже стандартными средствами этой операционной системы, находятся в каталоге **Windows\Fonts**<sup>6</sup>. Файл шрифта в формате TrueType или OpenType стандартно имеет расширение **ttf**. Например, Arial жирный курсив хранится в файле **arialbi.ttf**.

### **Шрифт как цифровой продукт**

Записав себе на компьютер файл цифрового шрифта, вы получаете в распоряжение как саму гарнитуру, так и ее машинное представление в определенном формате. У каждой из этих ипостасей шрифта может быть свой автор и свой правообладатель. Ваши права на использование цифрового шрифта в общем случае зависят от того, в какие отношения с этими субъектами вы вступили.

Даже если гарнитура за давностью лет или по воле художника сделалась общественным достоянием, фирма или персона, которая реализовала эту гарнитуру в конкретном продукте, файле формата TrueType (допустим), имеет право требовать плату за его использование. То есть, разные разработчики могут выпускать свои шрифтовые продукты на основе одной и той же гарнитуры, примерно так же, как разные издательства могут

---

<sup>6</sup> Обычно этот каталог находится на диске C:. В общем случае он находится на том диске, куда установлена операционная система. В операционных системах линейки NT шрифты располагались в каталоге WinNT\Fonts.

печатать Шекспира или Джоан Роулинг и продавать книжки. Могут. А имеют ли право, другой вопрос.

Художник, который придумал и нарисовал гарнитуру, получает авторские права на нее. Имущественные права на гарнитуру он может оставить за собой, передать своему заказчику или работодателю по договору или отрегулировать какой-нибудь лицензией вроде GPL. Можно представить себе много вариантов, какой из них выберет автор, зависит от того, в какой юрисдикции он находится и насколько искушен в юридических вопросах. Важно, что факт существования гарнитуры не дает нам с вами автоматически прав на ее использование. Например, если какой-нибудь умелец изготовит цифровой шрифт для гарнитуры, которой написано слово «Volvo» на корпусах автомобилей соответствующей марки, а вы наберете этим шрифтом свой рекламный буклет, компания Volvo может предъявить вам обоим претензии, если, конечно, заметит.

С другой стороны, цифровой шрифт, даже если он основан на общедоступной гарнитуре, очевидно, возник не из воздуха. У него есть изготовитель или другой правообладатель, который задает права на его использование примерно так же, как это происходит с программными продуктами.

Конкретные условия использования гарнитуры и ее цифровой реализации задает правообладатель<sup>7</sup>. Лицензионные соглашения, предлагаемые разными правообладателями, могут различаться. Таким образом, перед тем, как использовать цифровой шрифт, следует изучить лицензионное соглашение и обеспечить выполнение его условий (либо убедиться в том, что шрифт бесплатный). В противном случае вы действуете на свой страх и риск.

## **Просмотр шрифта**

В операционных системах семейства Microsoft Windows есть, по крайней мере, две встроенные программы, которые позволяют просматривать шрифты. Возможно, они не удовлетворят профессиональных дизайнеров, но пользователю, выбирающему шрифт для оформления документа, их возможностей будет вполне достаточно.

Программа просмотра шрифтов, как правило, ассоциирована с расширением ttf. Для того чтобы посмотреть шрифт:

1. В проводнике Windows перейдите в каталог Windows\Fonts (или в другой каталог, содержащий файлы шрифтов).

---

<sup>7</sup> Изложение подхода к лицензированию шрифтов компании «ПараТайп»: <http://www.paratype.ru/help/license/>

2. Дважды щелкните мышью по значку файла того шрифта, который вы намерены посмотреть. После этого на экран будет выведено окно с примером (рис. 1).

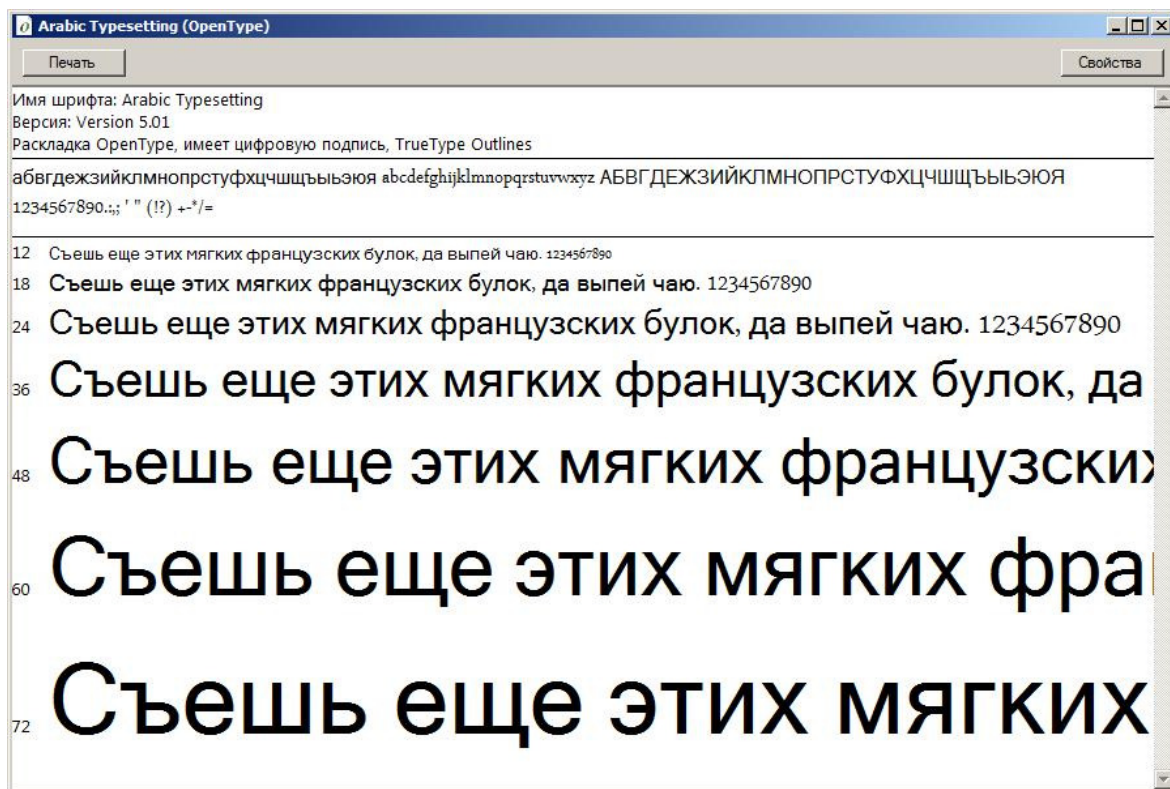


Рисунок 1. Программа просмотра шрифтов

Фраза-пример встроена в программу просмотра. Обычно она подобрана таким образом, чтобы в нее хотя бы один раз входила каждая буква алфавита.

В программе предусмотрена возможность просмотра основных свойств шрифта. Для того чтобы ознакомиться со свойствами шрифта, нажмите кнопку **Свойства** в правом верхнем углу окна программ. После этого будет открыто диалоговое окно просмотра свойств шрифта (рис. 2).

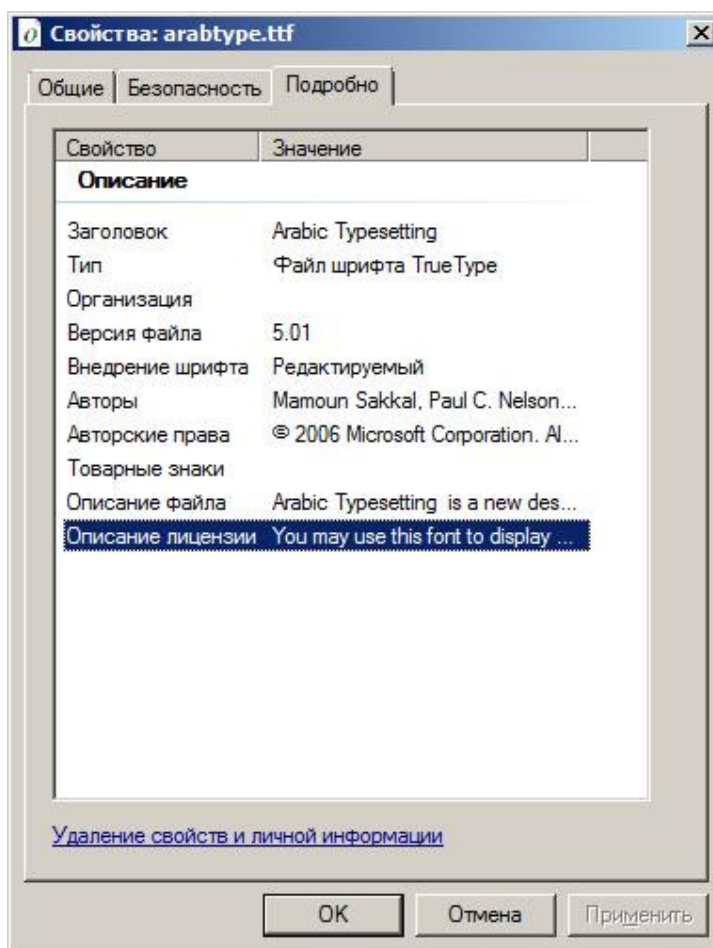


Рисунок 2. Программа просмотра шрифтов

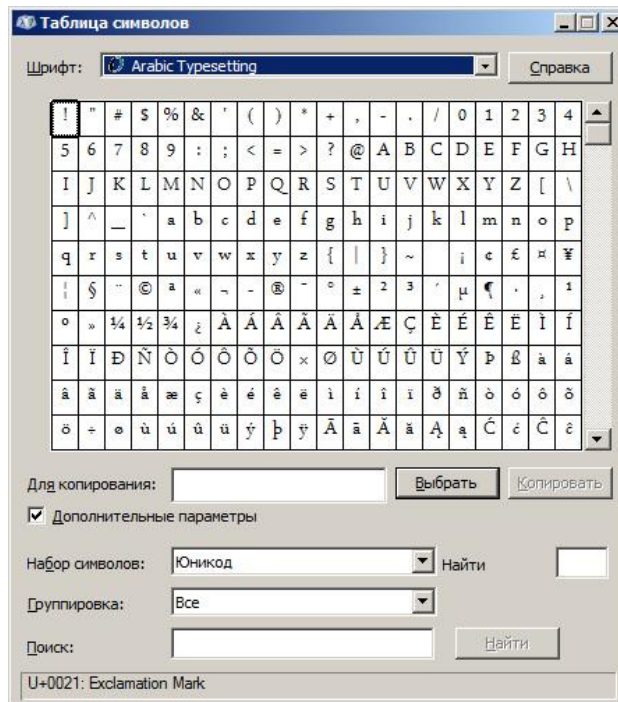
Особое внимание рекомендуем обратить на текст в поле **Описание лицензии**. Там должно быть написано, какие права и на каких основаниях правообладатель предоставляет пользователю.

Программа просмотра шрифтов не позволяет выяснить, для каких национальных алфавитов предусмотрены глифы в том или ином шрифте. Если в ее окне отображается фраза на русском языке, это не значит, что в шрифте нет, например, латинских букв с диакритическими знаками или греческих букв. По крайней мере, в доступных автору версиях операционных систем Microsoft это так, хотя не исключено, что в дальнейшем ее как-нибудь усовершенствуют.

Другая программа для просмотра шрифтов — «Таблица символов». Обычно она входит в набор так называемых стандартных программ.

Для того чтобы запустить программу «Таблица символов»:

1. Нажмите кнопку **Пуск** на панели задач Windows. После этого будет открыто главное меню системы.
2. В главном меню системы выберите пункт **Все программы**. Будет открыт список установленных в системе программ.
3. В списке программ найдите и откройте группу **Стандартные**. В этой группе найдите и откройте группу **Служебные**.
4. В группе **Служебные** выберите пункт **Таблица символов**. После этого на экране будет отображено главное окно этой программы (рис. 3).



**Рисунок 3. Программа «Таблица символов»**

Выберите тот шрифт, который вы намерены изучить, в раскрывающемся списке **Шрифт**. Установите флажок **Дополнительные параметры**. После этого окно программы расширится вниз, а элементы, расположенные ниже флажка, сделаются видимыми. Теперь вы получили возможность просматривать все глифы, которые имеются в шрифте. В программе предусмотрено несколько способов просмотра, вы можете пользоваться тем, который сочтете наиболее удобным для себя.

Самый простой способ просмотра — прокрутка таблицы символов с помощью полосы прокрутки, однако, если в шрифте много глифов, он может оказаться слишком трудоемким.

Кроме того, вы можете отбирать и просматривать отдельные наборы символов. Для этого служат раскрывающиеся списки **Набор символов** и **Группировка**. Ими можно по-разному фильтровать и сортировать таблицу. Покажем один из возможных способов просмотра шрифта.

Выберите в раскрывающемся списке **Набор символов** пункт **Юникод**. После этого выберите в раскрывающемся списке **Группировка** пункт **Диапазоны Юникода**. После этого будет открыто дополнительное окно **Группировка**. Выберите в нем тот набор символов, который представляет для вас интерес. Если шрифт не содержит подходящих глифов, таблица символов в основном окне будет пустой. Если такие глифы есть, они будут отображены в таблице символов (рис. 4).

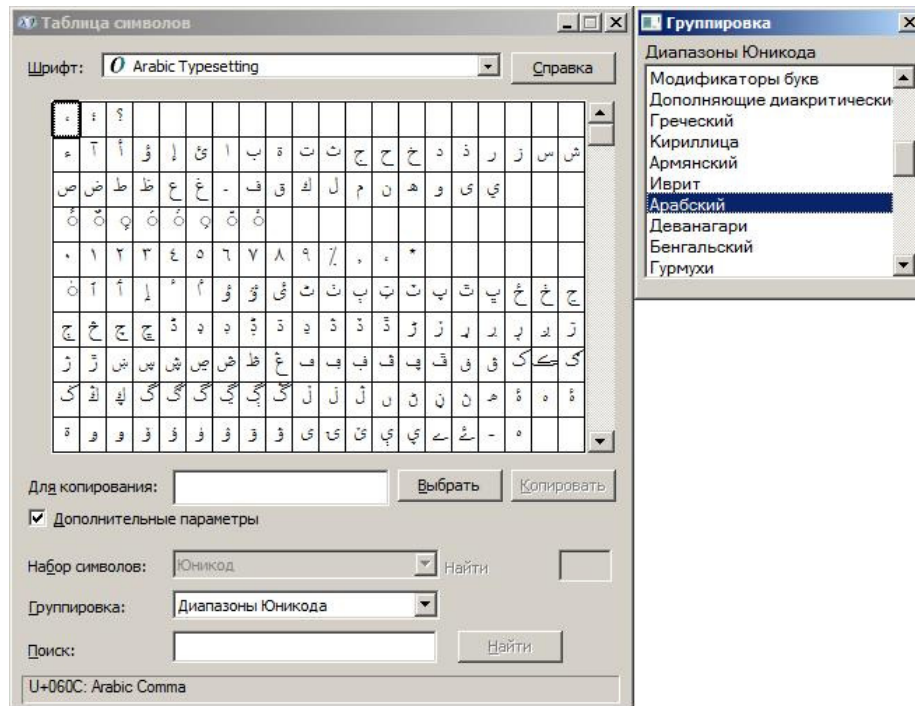


Рисунок 4. Просмотр символов в кодировке Юникод по отдельным наборам

## Настройка FO-процессоров

### RenderX XEP

Здесь рассматривается настройка XEP версии 4.xx., чтобы привести в чувство другие версии этой программы, могут потребоваться другие действия.

Настройка XEP для работы с кириллическими шрифтами в среде Microsoft Windows не представляет никакой сложности. Для формирования русскоязычных PDF-документов, как правило, достаточно тех шрифтов, которые уже имеются в системе. Для того чтобы XEP правильно использовал их, необходимо решить две простые задачи:

- зарегистрировать в конфигурационном файле XEP те шрифты, которыми вы планируете пользоваться в документах;
- указать, какие именно шрифты из числа зарегистрированных следует использовать, если в документе указан не конкретный шрифт, а тип (например, sens-serif).

Обе задачи решаются внесением изменений в конфигурационный файл XEP. Он располагается в том каталоге, куда установлена программа, и называется всегда xer.xml.

Технически xer.xml представляет собой XML-файл, который можно редактировать в XML-редакторе или в текстовом редакторе, лишь бы он правильно обрабатывал переводы строк в стиле Unix, поэтому обычный «Блокнот» будет не самым удачным выбором.

Параметры настройки XEP и предназначенные для их указания теги подробно описаны в справочнике, который поставляется вместе с программой. Он находится в подкаталоге doc и называется reference.html или reference.pdf.

XEP — кроссплатформенное приложение, написанное на языке Java. Поэтому в конфигурационном файле содержатся данные для всех официально поддерживаемых сред. Участок конфигурационного файла, который отвечает за шрифты в среде Windows, при поставке закрыт комментариями. Первое, что необходимо сделать, снять с него комментарии. Найти этот фрагмент можно поиском по комментарию: «Sample configuration for Windows TrueType fonts», который расположен непосредственно перед ним.

Фрагмент начинается элементом font-group, внутри которого находятся регистрационные записи шрифтов. В атрибуте xml:base этого элемента укажите каталог, где хранятся шрифты. Обычно для хранения шрифтов используется каталог c:\Windows\Fonts, но

возможны и другие варианты, если операционная система установлена на другом диске, в каталоге с другим именем, или вы вообще используете какой-то другой набор шрифтов.

Внимание: путь необходимо указывать в формате URI, т. е. для каталога `c:\Windows\Fonts` значение атрибута `xml:base` должно быть таким: `c:///Windows/Fonts`.

Внутри элемента `font-group` находятся элементы `font-family`. Каждый из них представляет собой регистрационную запись определенной шрифтовой гарнитуры. Вы можете отредактировать уже имеющиеся записи или заменить их другими по своему усмотрению. При модификации или замене регистрационной записи шрифтовой гарнитуры следует уделять особое внимание значению атрибута `name` элемента `font-family`. XEP выбирает шрифт по его значению. Например, если в документе XSL-FO, который вы преобразуете в PDF, указан шрифт Times New Roman, а в конфигурационном файле XEP отсутствует регистрационная запись для такой гарнитуры, зато упомянута гарнитура "Times", программа «не сообразит», что вообще-то имелось в виду то же самое<sup>8</sup>.

Внутри элемента `font-family` расположены ссылки на файлы конкретных шрифтов.

В качестве примера разберем одну такую ссылку.

```
<font-family name="Times New Roman" ligatures="&#x2013; &#x2014;">
  <font style="italic">
    <font-data ttf="timesi.ttf"/>
  </font>
</font-family>
```

Элемент `font` описывает шрифт как таковой. В нем можно указать, должен ли шрифт использоваться в качестве жирного и наклонного начертания гарнитуры, указанной в родительском элементе `font-family`. В приведенном примере шрифт зарегистрирован для наклонного начертания гарнитуры Times New Roman. Дочерний элемент `font-data` содержит ссылку на сам файл шрифта. В приведенном примере он ссылается на файл `timesi.ttf`. Поскольку в атрибуте `xml:base` элемента `font-group` уже указано расположение каталога шрифтов, здесь достаточно указать только имя файла.

Подчеркнем еще раз, присвоение атрибуту `style` элемента `font` значения `italic` не приводит к тому, что прямой шрифт за счет какого-то автоматического преобразования становится наклонным. Когда XEP обнаружит во входном документе XSL-FO предписание

---

<sup>8</sup> При наборе текста в формате DocBook/XML или DITA автор обычно не указывает шрифт напрямую. Имена или типы шрифтов подставляются XSLT-стилями в процессе преобразования входного документа в промежуточный документ формата XSL-FO. Если вы не знаете, какие именно это шрифты, посмотрите значения параметров XSLT-стилей, которые вы используете для преобразования. Другой способ — открыть сформированный этими стилями документ формата XSL-FO и попробовать найти в нем названия шрифтов. О XSLT-стилях и языке XSL-FO см. «Настройка оформления выходных документов в DocBook/XML», <http://www.philosoft.ru/dbcust.zhtml>.

использовать наклонное начертание шрифта Times New Roman, он просто будет брать глифы из файла, зарегистрированного в этом качестве. Короче говоря, вам вменяется в обязанность знать, какие шрифты хранятся в каких файлах, а не FO-процессору автоматически разыскивать их или, тем более, преобразовывать необходимым образом.

Напомним, что кроме конкретных шрифтовых гарнитур существуют еще их типы, объединяющие шрифтовые гарнитуры, имеющих сходные черты. В документах XSL-FO могут быть указаны как названия гарнитур, так и названия типов. Например, для какого-нибудь текстового блока может быть указана гарнитура Tahoma, а может тип sans-serif.

В первом случае XEP найдет (если, конечно, он там есть) в своем конфигурационном файле элемент font-family с атрибутом name="Tahoma" и обратится к соответствующему шрифтовому файлу.

Во втором случае XEP будет нуждаться в указании, какой именно шрифт ему следует использовать в качестве шрифта без засечек (как и любого из остальных четырех типов). Для этого каждому типу шрифтов необходимо сопоставить гарнитуру из числа зарегистрированных в конфигурационном файле XEP. Сопоставление гарнитуры типу осуществляется с помощью элемента font-alias. Рассмотрим следующий пример.

```
<font-alias name="sans-serif" value="Arial"/>
```

Эта строка велит XEP использовать в качестве шрифта без засечек гарнитуру Arial. Таким же образом надо задать гарнитуры для типов serif (с засечками), monospace (моноширинный), cursive (для выделения) и fantasy (для привлечения внимания).

Об отсутствии строгой системы в такой классификации мы уже говорили выше, но современные языки разметки не предлагают нам ничего иного.

Итак, повторим сказанное в двух словах, специально для тех читателей, которые не стремятся разобраться в деталях, а хотят поскорее получить результат.

Предположим, что вы только что установили к себе на компьютер FO-процессор RenderX XEP. Для того чтобы указать ему расположение кириллических шрифтов, выполните следующую процедуру.

1. Откройте каталог, в котором установлен FO-процессор XEP. Обычно по умолчанию это каталог c:\Program Files\RenderX\XEP. Если при установке XEP вы указали другой каталог, например, F:\XEP, откройте его.
2. Найдите в этом каталоге файл xep.xml. Откройте этот файл в любом текстовом

редакторе или XML-редакторе.

3. Найдите в файле хер.xml строку Sample configuration for Windows TrueType fonts. Для этого удобнее всего использовать функцию поиска, которая есть наверно в любом редакторе.
4. Снимите комментарий с элемента font-group, расположенного ниже этой строки. Не забудьте, что необходимо удалить не только открывающий, но и закрывающий тег комментария.
5. В атрибуте xml:base узла font-group укажите URI каталога, в котором находятся шрифты, имеющиеся в системе. Обычно это каталог c:///Windows/Fonts.
6. Ниже, после элементов font-group, укажите, какие именно шрифты из числа фактически имеющихся в системе следует использовать в качестве шрифта с засечками (serif), шрифта без засечек (sans-serif) и моноширинного шрифта (monospace).
7. Сохраните файл хер.xml.

Фрагмент исправленного файла хер.xml показан на листинге ниже.

```
<!-- Sample configuration for Windows TrueType fonts. -->
<font-group xml:base="file:///C:/Windows/Fonts/" label="Windows TrueType"
  embed="true" subset="true">
  <font-family name="Arial">
    <font><font-data ttf="arial.ttf"/></font>
    <font style="oblique"><font-data ttf="ariali.ttf"/></font>
    <font weight="bold"><font-data ttf="arialbd.ttf"/></font>
    <font weight="bold" style="oblique"><font-data ttf="arialbi.ttf"/></font>
  </font-family>

  <font-family name="Times New Roman" ligatures="&#x2013; &#x2014;">
    <font><font-data ttf="times.ttf"/></font>
    <font style="italic"><font-data ttf="timesi.ttf"/></font>
    <font weight="bold"><font-data ttf="timesbd.ttf"/></font>
    <font weight="bold" style="italic"><font-data ttf="timesbi.ttf"/></font>
  </font-family>

  <font-family name="Courier New">
    <font><font-data ttf="cour.ttf"/></font>
    <font style="oblique"><font-data ttf="couri.ttf"/></font>
    <font weight="bold"><font-data ttf="courbd.ttf"/></font>
    <font weight="bold" style="oblique"><font-data ttf="courbi.ttf"/></font>
  </font-family>

  <font-family name="Tahoma" embed="true">
    <font><font-data ttf="tahoma.ttf"/></font>
    <font weight="bold"><font-data ttf="tahomabd.ttf"/></font>
  </font-family>

  <font-family name="Verdana" embed="true">
    <font><font-data ttf="verdana.ttf"/></font>
    <font style="oblique"><font-data ttf="verdanai.ttf"/></font>
    <font weight="bold"><font-data ttf="verdanab.ttf"/></font>
    <font weight="bold" style="oblique"><font-data ttf="verdanaz.ttf"/></font>
  </font-family>

  <font-family name="Palatino" embed="true">
```

```

    ligatures="&#x2010; &#x2013; &#x2014; &#x2018; &#x2019;">
    <font><font-data ttf="pala.ttf"/></font>
    <font style="italic"><font-data ttf="palai.ttf"/></font>
    <font weight="bold"><font-data ttf="palab.ttf"/></font>
    <font weight="bold" style="italic"><font-data ttf="palabi.ttf"/></font>
  </font-family>
</font-group>

<!-- Required aliases -->

<font-alias name="monospace" value="Courier New"/>
<font-alias name="sans-serif" value="Arial"/>
<font-alias name="serif" value="Times New Roman"/>
<font-alias name="cursive" value="Arial"/>
<font-alias name="fantasy" value="Arial"/>

```

## Apache FOP

Здесь рассматривается настройка Apache FOP версии 0.20.5. Настройка других версий этой программы может потребовать других действий.

Бесплатный FO-процессор Apache FOP очень популярен среди разработчиков, поэтому найти в Интернете статьи о его настройке не составляет труда. Тем не менее, мы приводим здесь основные сведения на этот счет, чтобы облегчить читателю дальнейшее освоение разрозненного и не всегда четко изложенного материала.

У FOP имеется конфигурационный файл для пользовательской настройки, который находится в подкаталоге conf и называется userconfig.xml. Как у XEP он представляет собой XML-файл. Как и у XEP, в этом конфигурационном файле необходимо зарегистрировать шрифты, которыми вы собираетесь пользоваться. Элементы, которые используются для регистрации шрифтов, в деталях отличаются от аналогичных в конфигурационном файле XEP, но принципиально устроены примерно таким же образом.

Рассмотрим пример описания шрифта в конфигурационном файле FOP.

```

<font embed-file="c:/windows/fonts/timesi.ttf"
      kerning="yes"
      metrics-file="file:///e:/privat/misha/fop/fonts/timesi.xml">
  <font-triplet name="Times New Roman" style="italic" weight="normal"/>
  <font-triplet name="serif" style="italic" weight="normal"/>
</font>

```

Элемент font служит для регистрации шрифта (именно шрифта, а не гарнитуры). Атрибут embed-file указывает расположение шрифтового файла.

Внутри элемента font находятся так называемые триплеты (да, разрешается, чтобы их было несколько), каждый из которых описан одним элементом font-triplet. Триплет задает соответствие между шрифтом, указанным во входном документе XSL-FO, и шрифтовым файлом, на который ссылается родительский элемент font. Триплет представляет собой комбинацию следующих свойств шрифта:

- гарнитуры либо типа гарнитуры;
- признака жирного начертания;
- признака курсивного начертания.

Теперь вернемся к элементу `embed-file` и обратим внимание на его атрибут `metrics-file`.

В нем указывают расположение так называемого файла *метрики шрифта*. Не вдаваясь в подробности, скажем, что это вспомогательный файл, который требуется FOP для правильного масштабирования глифов.

Вы можете найти готовые файлы метрик поиском в Интернете или изготовить их самостоятельно с помощью утилиты TTFReader, которая входит в поставку FOP. Более подробно эта операция описана в файлах `build\site\fonts.pdf` и `build\site\fonts.html`, которые также входят в поставку FOP.

## Литература

[1] Юрий Гордон. Книга про буквы от Аа до Яя. — М: Издательство Студии Артемия Лебедева, 2006. — 382 с.: ил.

[2] Юлия Сырых. Компьютерные шрифты на 100%. — СПб.: Питер, 2007. — 272 с.: ил.