

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ



Тема **2.** Технологии
и средства обработки текста

Текстовый редактор...

... самостоятельная компьютерная программа или компонент интегрированной среды разработки, предназначенная для создания и изменения текстовых данных в общем и текстовых файлов в частности.

Текстовые редакторы предназначены для работы с текстовыми файлами в интерактивном режиме. Они позволяют просматривать содержимое текстовых файлов и производить различные действия — вставку, удаление и копирование текста, контекстный поиск и замену, сортировку строк, просмотр кодов символов и конвертацию кодировок, печать и т. п.

Блокнот

Простейший текстовый редактор в Windows. Используется для заметок, небольших фраз, пометок. Удобно копировать различные коды, которые в нем не изменяются, более продвинутые редакторы могут их распознавать и преобразовывать визуально, в итоге теряя часть кода.

Используется для копирования паролей, ссылок, консольных команд. Поставляется со стандартным пакетом программ ОС.

Для ОС Linux есть аналог блокнота – gedit. Преимущество: просты и компактны, идеальны для пометок и хранения элементов кода. Недостаток: излишняя простота, которая не даёт производить оформление текста.

WordPad

Стандартный текстовый редактор, поставляется с Windows, что-то среднее между блокнотом и Microsoft Word. Прост как Блокнот, но содержит некоторые элементы оформления текста как в Word. Простота + минимальный набор функций для оформления, делают его привлекательным для набора простого текста, не требующего особого оформления.

Учитывать, что правописание программа проверяет плохо.

Notepad++

Notepad++ - бесплатный редактор текстовых файлов с поддержкой синтаксиса большого количества языков программирования. Имеет широкий набор опций и отличается минимальным потреблением ресурсов процессора.

Удобна опция подсветки текста и возможность сворачивания блоков, согласно синтаксису языка программирования. Можно самостоятельно определять синтаксис языка программирования. Доступно выделение цветом директив и операторов языка программирования.

Позволяет одновременно просматривать и редактировать нескольких документов. Также можно просматривать и редактировать в двух окнах отображения один и тот же документ в разных местах. Изменение документа в одном окне просмотра будет автоматически перемещено во второе окно просмотра (т. е. редактируется один документ, который имеет клона во втором окне просмотра).

Текстовые процессоры (ТП)

Используются для написания и модификации документов, компоновки макета текста и предварительного просмотра документов в том виде, в котором они будут напечатаны (свойство [WYSIWYG](#)).

Современные ТП, кроме форматирования шрифтов, абзацев, проверки орфографии, включают возможности, ранее присущие лишь настольным издательским системам (создание таблиц, вставка графических изображений).

Примеры ТП: [Microsoft Word](#), [OpenOffice.org Writer](#), [LibreOffice Writer](#).



AbiWord

автор [Эрик Синк](#) Версия 3.0.2 от 20 .10 2016 Написан на [C++](#)
[Кроссплатформенное ПО. Сайт \[abisource.com\]\(http://abisource.com\)](#)

Текстовый редактор. Базируется на библиотеке [GTK+](#). Имеется для Windows, Linux и Mac OS X, а также исходный код.

Работает по принципу [WYSIWYG](#) и поддерживает основные функции работы с текстом:

- форматирование и стили текста;
- создание таблиц и списков;
- вставка колонтитулов и сносок;
- вставка и масштабирование рисунков;
- проверка орфографии (поддерживается русский язык);
- составление оглавлений;
- печать.

Редактор не имеет: а) встроенных инструментов для создания графики; б) средств проверки грамматики и расстановки переносов.



AbiWord

автор [Эрик Синк](#) Версия 3.0.2 от 20 .10 [2016](#) Написан на [C++](#)

Кроссплатформенное ПО. Сайт abisource.com

Поддержка форматов

Родной *.abw, а также форматы *.doc ([Word](#)), *.rtf, *.html.

Возможен импорт форматов, как *.odt ([OpenDocument](#)), *.wpd ([WordPerfect](#)), *.sdw ([StarOffice](#)) и др.

Достоинства и недостатки

Уступает [Microsoft Word](#) или [LibreOffice Writer](#), но превосходит простые редакторы ([WordPad](#)). Функциональность можно расширить подключением плагинов (см. сайт разработчиков).

Достоинства: а) небольшой дистрибутив; б) открытое распространение; в) высокое быстродействие ; г) кроссплатформенность.

Недостатки — ограниченная функциональность. Совместимость с Word ограничена: простые документы импортируются и экспортируются без проблем, в сложных возможно существенное нарушение форматирования.

LibreOffice

Создан для Linux, на смену OpenOfficeOrg. Затем появилась версия под Windows. Бесплатный редактор. Похож на Word 2003, но хуже проверяет правописание

В Windows и Linux, чтобы сохранить и распечатать красиво обрамлённый текст – оптимальный вариант, так как он бесплатный, можно не тратиться на довольно дорогой офисный пакет от Microsoft.

Загрузить офисный пакет для Windows можно со странички [LibereOffice](#), в Linux установлен по умолчанию.

LibreOffice Writer



Разработчик The Document Foundation . Версия 5.3.0 ([1.02.17](#)). Написан на [C++](#), [Python](#), [Java](#),
Для ОС [Linux](#), [Mac OS X](#), [FreeBSD](#) и др. [UNIX-подобные](#), [Microsoft Windows](#).
Сайт ru.libreoffice.org/featu...

ТП и визуальный редактор HTML, входит в офисный пакет [LibreOffice](#). Ответвление ТП [OpenOffice.org Writer](#).

Сообщество OpenOffice.org объявило 28.09.2010 о создании независимого от [Oracle](#) ответвления офисного пакета. Появилась «[The Document Foundation](#)», основная цель которой - разработка независимого офисного проекта с поддержкой открытого формата **ODF** без передачи авторских прав.

В OpenOffice.org от разработчиков требовалась передача авторских прав компании Oracle.

Полнофункциональный ТП и инструмент публикаций, выполняет расстановку переносов, автозамену, поиск и замену, автоматическое составление оглавлений, проверку правописания, имеет тезаурус и пр.



LibreOffice Writer

Разработчик The Document Foundation . Версия 5.3.0 ([1.02.17](#)). Написан на [C++](#), [Python](#), [Java](#),
Для ОС [Linux](#), [Mac OS X](#), [FreeBSD](#) и др. [UNIX-подобные](#), [Microsoft Windows](#).

С первых версий осуществляется:

- методы макетирования страниц (рамки, столбцы и таблицы);
- внедренная или связанная графика, ЭТ и другие объекты;
- работа со стилями и шаблонами;
- встроенные средства рисования;
- работа с составным документом;
- отслеживание изменений в версиях документов;
- интеграция с БД, включая базу данных библиографии;
- экспорт в формат PDF, включая закладки;



LibreOffice Writer

Разработчик The Document Foundation . Версия 5.3.0 ([1.02.17](#)). Написан на [C++](#), [Python](#), [Java](#),
для ОС [Linux](#), [Mac OS X](#), [FreeBSD](#) и др. [UNIX-подобные](#), [Microsoft Windows](#).

Новые возможности с версии 4.0:

- улучшения импорта документов DOCX: плавающие таблицы, объекты OLE в обрамлениях, отступы изображений в тексте;
- уменьшено время загрузки файлов RTF;
- поддержка импорта и экспорта математических формул RTF;
- Импорт рукописных примечаний из документов DOCX и RTF (используется в Word на планшетных ПК);
- добавление комментариев к текстовым областям документа;
- сокращение Java-кода: мастера факсов и писем переписаны на Python.
- возможность выделять и копировать текст в окнах сообщений.
- новое окно управления шаблонами пользователя.



LibreOffice Writer

Разработчик The Document Foundation . Версия 5.3.0 ([1.02.17](#)). Написан на [C++](#), [Python](#), [Java](#),
Для ОС [Linux](#), [Mac OS X](#), [FreeBSD](#) и др. [UNIX-подобные](#), [Microsoft Windows](#).

Поддержка форматов

Родной формат **ODF** - стандартный формат файлов офисных документов ([OpenDocument Format](#) - ГОСТ Р ИСО/МЭК 26300-2010, введён 1.06.2011).

Формат текстового документа ODF имеет расширение **odt**, а **шаблон** текстового документа — **ott**.

Поддерживает форматы Microsoft Word различных версий (Microsoft Word 97/2000/XP/2003 (*.doc), а также OOXML *.docx).

Реализована функция экспорта документа в формате переносимого документа PDF.



OpenOffice.org Writer

Разработчик [Apache Software Foundation](#), ранее [Oracle Corporation](#)

Версия 4.1.3 ([12 окт 2016](#)) Написан на [C++](#), [Java](#), Кроссплатформенное ПО.

Сайт openoffice.org/product/writer.html

ТП и визуальный редактор HTML, входит в [OpenOffice.org](#) и является свободным ПО. Аналог [Microsoft Word](#). Имеет возможности, отсутствующие в Word (поддержку стилей страниц).



OpenOffice.org Writer

Разработчик [Apache Software Foundation](#), ранее [Oracle Corporation](#)

Версия 4.1.3 ([12 окт 2016](#)) Написан на [C++](#), [Java](#), Кроссплатформенное ПО.

Сайт openoffice.org/product/writer.html

Поддержка форматов

Позволяет сохранять документы в форматах, включая [Microsoft Word](#), [*.rtf](#), [*.xhtml](#), [*.pdf](#) и [*.odf](#), который является форматом по умолчанию, начиная с OpenOffice.org 2.0, а также в формате [*.odt](#) предыдущих версий (включая и версию SO Writer 5.2).

Можно экспортировать в формат MediaWiki (версии 3.0 и выше — с помощью доп. расширения). Импортируются Corel WordPerfect ([*.wdp](#)), 602 Document ([*.602](#)), WPS Word ([*.wps](#)), файлы некоторых форматов мобильных ТП на платформах PocketPC и Palm и др.

Существует специальная версия OpenOffice.org Writer Portable (часть [OpenOffice.org Portable](#)), которую можно использовать без установки, что позволяет запускать её, например, с [флэш-диска](#).



OpenOffice.org Writer

Разработчик [Apache Software Foundation](#), ранее [Oracle Corporation](#)

Версия 4.1.3 ([12 окт 2016](#)) Написан на [C++](#), [Java](#), Кроссплатформенное ПО.

Сайт openoffice.org/product/writer.html

Наиболее популярные расширения:

CompPad — позволяет делать математические и инженерные расчеты в форме ввода математических выражений с помощью редактора уравнения (см. также [Math](#))^[10].

Language Tool — расширение для проверки грамматики. Реализована возможность проверки для русского, английского, немецкого, польского и других языков^[11].

Типографика для ООО — расширение позволяет привести текст в соответствие с типографскими стандартами (кавычки, тире, лишние пробелы)^[12].

AltSearch — расширение для улучшения параметров поиска и замены текста, поддерживает регулярные выражения.

Текстовый редактор онлайн

Современный вид редакторов, позволяет набирать тексты непосредственно на удалённом сервере в Интернете – онлайн. Преимущества: 1) текст сразу сохраняется в облачном хранилище; 2) текст доступен сразу нескольким пользователям (возможно совместное редактирование); 3) имеется доступ к документам со всех устройств, которые подключены к Интернету. Не страшна потеря данных или отключение света – документ всегда доступен в Интернете.

Наилучший онлайн редактор - сервис GoogleДиск. Зарегистрировав облачный диск, можно создавать на нём текстовые документы. Дизайн документов ограничен, так как это лишняя нагрузка на Интернет-канал, да и главное это набор текста.

Проверка орфографии производится браузером, вариант не идеальный, но и не самый плохой. Преимущество - бесплатность сервиса. Документ можно сразу же сохранить на компьютере, если в этом будет необходимость.



Стандарт предложен в 1991 г. некоммерческой организацией «Консорциум Юникода» (Unicode Consortium, Unicode Inc.). Он позволяет закодировать большое число символов из разных письменностей (китайские иероглифы, математические символы, буквы греческого алфавита, латиницы и кириллицы), при этом становится ненужным переключение кодовых страниц.

Стандарт состоит из двух основных разделов:

- универсальный набор символов (UCS, universal character set);
- семейство кодировок (UTF, Unicode transformation format).

Универсальный набор символов задаёт однозначное соответствие символов кодам — элементам кодового пространства, представляющим неотрицательные целые числа. Семейство кодировок определяет машинное представление последовательности кодов UCS.



ЮНИКОД

Коды в стандарте Юникод разделены на несколько областей. Область с кодами от U+0000 до U+007F содержит символы набора [ASCII](#) с соответствующими кодами.

Далее расположены области знаков различных письменностей, знаки пунктуации и технические символы. Часть кодов зарезервирована для использования в будущем.

Под символы кириллицы выделены области знаков с кодами от U+0400 до U+052F, от U+2DE0 до U+2DFF, от U+A640 до U+A69F.



Проблемы кодировок

Предпосылки создания Юникода

- **«Кракозябры»** — отображение документов в неправильной кодировке. Как решить: а) последовательно внедрять методы указания используемой кодировки, б) внедрить единую (общую) для всех кодировку.
- **Ограниченность набора символов**: Можно либо переключать шрифты внутри документа, либо внедрить «широкую» кодировку. Переключение шрифтов практиковалось в текстовых процессорах, причём часто использовались шрифты с нестандартной кодировкой, т. н. «dingbat fonts». В итоге при попытке переноса документа в другую систему все нестандартные символы превращались в «кракозябры».

Проблемы Юникода

- тексты на китайском, корейском и японском записываются сверху вниз, начиная с правого верхнего угла. Юникод вертикальное написание не предусматривает, осуществляться за счет языков разметки или внутренними механизмами текстовых процессоров.
- в Юникоде возможно разное начертание одного и того же символа в зависимости от языка. Следить, чтобы текст был правильно помечен как относящийся к тому или другому языку.
- перевод из строчных букв в заглавные зависит от языка.
- арабские цифры бывают прописные и строчные, пропорциональные и моноширинные — Юникод это не различает.

Проблемы Юникода

- **Файлы нелатинского текста в Юникоде всегда занимают больше места, так как один символ кодируется не одним байтом, как в различных национальных кодировках, а последовательностью байтов (исключение составляет UTF-8 для языков, алфавит которых укладывается в ASCII, а также наличие в тексте символов двух и более языков, алфавит которых не укладывается в ASCII.**
- **Файл шрифта, для отображения всех символов Юникод, занимает довольно много места в памяти и требует больших ресурсов, чем шрифт только одного национального языка пользователя. С увеличением мощности КС и удешевлением памяти и дискового пространства эта проблема менее существенна; но актуальна, например, для мобильных телефонов.**

UTF-8

Представление, обеспечивающее наилучшую совместимость со старыми системами, использовавшими 8-битные символы. Текст, состоящий только из символов с номером меньше 128, при записи в UTF-8 превращается в обычный текст [ASCII](#). И наоборот, в тексте UTF-8 любой [байт](#) со значением меньше 128 изображает символ ASCII с тем же кодом.

Остальные символы Юникода изображаются последовательностями длиной от 2 до 6 байт (на деле, только до 4 байт, в Юникоде нет символов с кодом больше 10FFFF), в которых первый байт всегда имеет вид 11xxxxxx, а остальные — 10xxxxxx.

Формат UTF-8 был изобретён [2 сентября 1992 года Кеном Томпсоном](#) и [Робом Пайком](#) и реализован в [Plan 9^{\[36\]}](#). Сейчас стандарт UTF-8 официально закреплён в документах [RFC 3629](#) и ISO/IEC 10646 Annex D.