Кодировка UTF-8

Все данные «Докувики» сейчас хранит в кодировке UTF-8. Во избежание проблем с файловой системой, имена файлов кодируются посредством URL-encoded. Версии «Докувики», более старые чем 2005-20-06 использовали другие кодировки, так что при обновлении данные необходимо перекодировать (англ.). Переход на другие кодировки не поддерживается.

Настройки браузера для UTF-8

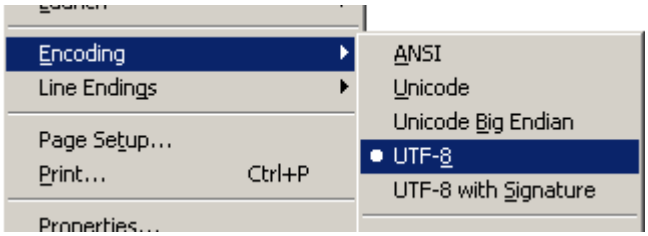
Все современные браузеры поддерживают кодировку UTF-8 — это одна из немногих вещей, которые работают в большинстве браузеров как надо. Если ваш браузер отображает некоторые символы неверно, возможно у вас отсутствуют необходимые шрифты.

Пользователям Windows нужно установить шрифт `Arialuni.TTF`, предлагаемый Микрософтом. Он включён в пакет Microsoft Office.

Пользователи Debian могут узнать на странице о шрифтах (англ.) о том, как правильно ставить юникодовые шрифты.

- Unicode and HTML
- Configuring browsers for Unicode

Редактирование файлов



Если вы собираетесь править данные непосредственно или сделать перевод, вам нужен текстовый редактор, поддерживающий UTF-8. Таких достаточно в мире, однако стоит посоветовать пару маленьких, простых и свободных на случай необходимости[1]:

- TEA — редактор под GNU/Linux на основе GTK2.
- Notepad2 — удачная замена «блокнота» в Windows.

Обратите внимание, «Докувики» *не* использует маркер порядка байтов и вам необходимо убедиться, что применяемое ПО не использует его тоже (особенно, если речь идёт о правке файлов PHP или конфигурации).

Массовая перекодировка файлов

- Под Unix используйте www.gnu.org/software/libiconv/

- Под Windiws используйте recode, порт iconv: http://recode.progiciels-bpi.ca/archives
  - Пример простого преобразования файла:

    ```
    recode lat1..u8 test.txt
    ```

    , где `lat1` — кодировкая источника (замените её на свою, например koi8-r) и u8 — желаемая кодировка UTF-8.

  - Чтобы массово перекодировать файлы под Windows, используйте эту команду (преобразующую фсе файлы в каталоге):

    ```
    FOR /F "tokens=*" %%G IN ('dir/b/S/X ^"C:\yourpath\*.txt^"') DO
    recode -v lat1..u8 %%~sG
    ```

- Подробнее узнать можно по английски там.

Примеры

Далее вы увидите примеры применения UTF-9 и убедиться в его правильной поддержке браузером[2)].

Знаки зодиака: ♈ ♉ ♊ ♋ ♌ ♍ ♎ ♏ ♐ ♑ ♒ ♓

Шахматная доска:

	A	B	C	D	E	F	G	H
8	♜	♞	♝	♛	♚	♝	♞	♜
7	♟	♟	♟	♟	♟	♟	♟	♟
6								
5								
4								
3								
2	♙	♙	♙	♙	♙	♙	♙	♙
1	♖	♘	♗	♕	♔	♗	♘	♖

Текст на русском:

```
По оживлённым берегам
Громады стройные теснятся
Дворцов и башен; корабли
Толпой со всех концов земли
К богатым пристаням стремятся;
```

На древнегреческом:

```
Αρχαίο Πνεύμα Αθάνατον!
Ἰοὺ ἰού· τὰ πάντ’ ἂν ἐξήκοι σαφῆ.
Ὦ φῶς, τελευταῖόν σε προσβλέψαιμι νῦν,
ὅστις πέφασμαι φύς τ’ ἀφ’ ὧν οὐ χρῆν, ξὺν οἷς τ’
```

οὐ χρῆν ὁμιλῶν, οὕς τέ μ’ οὐκ ἔδει κτανών.

На современном греческом:

Η σύγχρονη Ελλάδα, έχει να παρουσιάσει δυναμικό
έργο στον τομέα του πολιτισμού, των τεχνών και
των γραμμάτων. Αντίστοιχα δυναμική είναι η παρουσία
των Ελλήνων επιχειρηματιών στην διεθνή οικονομική
και βιομηχανική σκηνή.

На санскрите:

पशुपतिरपि तान्यहानि कृच्छ्राद्
अगमयदद्रिसुतासमागमोत्कः ।
कमपरमवशं न विप्रकुर्युर्
विभुमपि तं यदमी स्पृशन्ति भावाः ॥

На хинди:

गूगल समाचार हिन्दी में

На корейском:

한글은 아름다운 우리글입니다.
곱고 아름답게 사용하는 것이 우리의 의무입니다.

На китайском:

子曰：「學而時習之，不亦說乎？有朋自遠方來，不亦樂乎？
人不知而不慍，不亦君子乎？」

有子曰：「其為人也孝弟，而好犯上者，鮮矣；
不好犯上，而好作亂者，未之有也。君子務本，本立而道生。
孝弟也者，其為仁之本與！」

На японском:

「秋の田の かりほの庵の 苫をあらみ わが衣手は 露にぬれつつ」　天智天皇
「春すぎて 夏来にけらし 白妙の 衣ほすてふ 天の香具山」　持統天皇
「あしびきの 山鳥の尾の しだり尾の ながながし夜を ひとりかも寝む」　柿本人麻呂

На латвийском:

Iedomu jaukie ideāli,
Vecākie principi, tikla, mīla -
Dienas allažības priekšā
Šķīst kā graudi akmeņstarpā.

Glāžšķūņa rūķīši jautri dziedādami čiepj koncertflīģeļa vāku.

Упрощённым китайским письмом:

```
这是简体字汉语 zhè shì jiǎn t zì hàn yǔ
```

На армянском:

```
Հարգանքներիս հավաստիքը Հայ ժողովրդին:
Ամենալավ օրենքները չեն օգնի, եթե մարդիկ բանի պետք չեն:
```

На иврите:

```
המשפט עם הזכוכית שאפשר לאכול בלי שזה מפריע, לא זוכר איך הוא הולך
```

1)
Это никак не претендует ни на всеобъемлющий список редакторов с поддержкой юникода, ни на идеальный выбор. Просто **не** добавляйте сюда другие.
2)
взято с www.eleves.ens.fr:8080/home/madore/misc/unitest/

From:
https://wwoss.ru/ - **worldwide open-source software**

Permanent link:
https://wwoss.ru/doku.php?id=wiki:utf-8

Last update: **2025/02/16 22:05**