

Романизация имён файлов

Ключевые слова: **UTF-8,romanize, cyrillic, latin, convert, filename**

При обновлении с версий, которые ещё не имели функции «romanize», вы столкнетесь с полностью нечитаемой структурой директорий.

Например:

%D0%BA%D1%8B%D1%80%D0%B3%D1%8B%D0%B7%D1%81%D1%82%D0%B0%D0%BD.txt — тоже самое, что и «кыргызстан.txt».

Это произошло из-за того, что имена файлов в кодировке UTF-8 подвергаются преобразованию urlencode.

В последних версиях была добавлена опция «романизации» для того, чтобы обойти эту проблему.¹⁾

Скрипт, приведённый ниже, конвертирует нечитаемое содержание директории в «романизированные» имена файлов.

Вам необходимо включить файл [UTF8.php](#), который входит в инсталляцию «ДокуВики».

Замечание: этот скрипт не безупречен: например, есть определенные кириллические буквы, которые приведут к вставке символа «'» в имя файла. Все из-за того, что в UTF-8.php транслитерацией буквы 'ъ' является символ «'».

Пожалуйста, проверьте свою структуру статей после преобразования некорректных имен файлов.

Я надеюсь, что это будет полезно кому-то. Любые улучшения приветствуются.

Update: UTF8.php был переписан, код, приводимый ниже, был протестирован с этой²⁾ версией [UTF8.php](#)

```
<?php

include("utf8.php"); //to be found in the \inc directory of the default
dokuwiki install

/**
 * Copy a file, or recursively copy a folder and its contents, and clean up
the filenames according to the dokuwiki UTF-8
 *
 * @original_author      Aidan Lister <aidan@php.net>
 * @link                  http://aidanlister.com/repos/v/function.copyr.php
 * @param                string   $source    Source path
 * @param                string   $dest      Destination path
 * @return               bool     Returns TRUE on success, FALSE on failure
 */
function copyr($source, $dest)
{
```

```

$dest2=cleanID($dest);
echo $source."->".$dest." ->$dest2<br/>\n";
// Simple copy for a file
if (is_file($source)) {
    return copy($source, $dest2);
}

// Make destination directory
if (!is_dir($dest)) {
    mkdir($dest2);

}

// Loop through the folder
$dir = dir($source);
while (false !== $entry = $dir->read()) {
    // Skip pointers
    if ($entry == '.' || $entry == '..') {
        continue;
    }

    // Deep copy directories
    if ($dest !== "$source/$entry") {
        copyr("$source/$entry", "$dest/$entry");
    }
}

// Clean up
$dir->close();
return true;
}

copyr("/dokuwiki/data/pages/", "/dokuwiki/data/pagesnew/");

function cleanID($id,$ascii=false){
    $id = trim(urldecode($id));
    $id = utf8_strtolower($id);
    $id = utf8_romanize($id);
    utf8_deaccent($id,-1);
    $id = preg_replace('#\'+#', '_',$id);
    return($id);
}

?>

```

¹⁾

См. подробнее в «deaccent» и romanization

²⁾

переводчик: непонятно, какой

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



Permanent link:
<https://wwoss.ru/doku.php?id=wiki:tips:romanize>

Last update: **2025/04/13 11:53**