

R-Dorgen 2.x

R-Dorgen – это абсолютно новый генератор статических дорвеев. Он отличается хорошей скоростью, большим количеством настроек и множеством методов генерации текста.

Возможности:

1. 7 методов обработки текста.
2. 39 макросов для вставки в html шаблоны.
3. Быстрая генерация(1000 страниц ~15-30 секунд, с морфологией – 40-90сек).
4. Возможность заливки по ftp протоколу.
5. Заливка и распаковка по ftp zip архивами
6. Генерация проектов для Filezilla
7. Генерация “белых страниц”.
8. Парсинг и использование картинок с images.yandex.ru.
9. Парсинг видео с Youtube
10. Xml/html карта сайта .
11. Пакетная генерация.
12. Очистка контента от лишних символов.
13. Генерация robots.txt
14. Возможность задать шаблон только для главной страницы.
15. Возможность подключать свои плагины.
16. Пинг google,yandex (плагин).
17. Обработка и морфология как в ru контенте, так и в eng.
18. Возможность использования всех имеющихся макросов в файле пакетки и структуре страниц.
19. Генерация favicon

Методы генерации текста:

1. Цепи Маркова
2. Синонимизация
3. Синонимизация + смена предложений местами
4. Текст без изменений
5. Смена предложений местами в тексте без изменений.
6. Синонимизация + смена предложений местами с удалением использованных предложений.
7. Смена предложений местами в тексте без изменений с удалением использованных предложений.

Методы вставки кейвордов:

1. Обычная вставка
2. Соблюдая морфологию
3. Соблюдая морфологию + разбивка кейворда на слова
4. Приписывая к началу или концу предложения
5. Упрощенная версия 3 способа
6. Упрощенная pWn морфология
7. Сложная pWn морфология

Структура страниц (ЧПУ):

1. Кейворд.html
2. Кейворд-случайное число.html
3. Подпапка/кейворд.
4. 1.html,2.html
5. Часть кейворда (“Скачать музыку” -> “Скачать.html” или “музыку.html”). Стоп слова учитываются.
6. Возможность указать свою, из имеющихся макросов.

Полезные ссылки:

<http://rushter.com/otzyvy-o-r-dorgen/> - Отзывы

<http://dm.rushter.com> – пример дорвея #1

<http://dm2.rushter.com> – пример дорвея #2

Макросы:

Макросы могут быть как статические, так и динамические. Динамические указываются в [], статические в {}. Разница между статическими и динамическими заключается в том, что статические заменяются только 1 раз и всегда имеют 1 значение.

Все макросы так же работают при создании своего шаблона для ссылок, во всех файлах которые лежат в папке с шаблоном и файле пакетки.

Описание:

1. [RANDNAME:file.txt:] – Случайное имя или слово из files/file.txt.
2. [RANDINT-x-y] – Случайное число, от x до y.
3. [RANDKEYWORDURL] – Случайная ссылка на страницу дорвея(без href).
4. [RANDHREF] – Случайная ссылка на страницу дорвея(с href).
5. [RANDKEYWORD] – Случайный кейворд .
6. [RANDOM:x#y#..z:END] – Случайный выбор из списка.
7. [TEXT-x-y] – Текстовый блок случайным размером от x до y.
8. [CTEXT-x-y] – Текстовый блок случайным размером от x до y. Без примеси кейвордов.
9. [MTEXT-(x|y)-(x2|y2)-z] - Генерация текстового блока случайным размером от x до y, со случайным количеством кейвордов от x2 до y2. z=1 если обрамлять кейворды html тегами, иначе 0.
10. [RTEXT-(x|y)-(x2|y2)-z] – Тоже самое что и MTEXT, только каждый раз вставляется случайный кейворд.
11. [KTEXT-(x|y)-(x2|y2)] – Сгенерировать текст размером от x до y, и вставить один кейворд после случайного кол-ва (x2-y2) символов.
12. [KEYWORD] – Текущий кейворд на странице.
13. [UKEYWORD] – Текущий кейворд с большой буквы.
14. [NEXTKEYWORD] – Следующий кейворд
15. [LASTKEYWORD] – Предыдущий кейворд.
16. [NEXTKEYWORDURL] – Ссылка на следующую страницу.
17. [UNEXTKEYWORD] – Ссылка на следующую страницу с кейвордом с большой буквы.
18. [LASTKEYWORDURL] - Ссылка на предыдущую страницу.
19. [ULASTKEYWORD] - Ссылка на предыдущую страницу с кейвордом с большой буквы.
20. [DATE] – Случайная дата. Чем больше вызовов, тем дата старше.
21. [N] – Номер страницы (1, 2, 3).
22. [PAGE] – Имя страницы (keyword.html).
23. [PAGES-10] – Навигация по дору из 10 ссылок.
24. [DOMAIN] – Url сайта.
25. [MDOMAIN] – Url сайта без http
26. [SNAME] – Имя сайта.
27. [JS] – Заменяется на значение из конфига. Подойдет для вставки скриптов, редиректов на платник и т.д
28. [IMG] – Ссылка на случайную картинку.
29. [SMAP] – Вставить все ссылки (только для sitemap.html).
30. [LOOP:N: TEXT] – Вставляет TEXT N-раз. В качестве n можно указывать интервал значений (“1-6” – случайное число от 1 до 6).
31. [YOUTUBE-100-500] – вставить видео с youtube размером 100 на 500.
32. [TAGS-10] – Вставить на страницу облако тэгов состоящее из 10 ссылок.
33. [RMK-1] – вставить случайное слово из текущего кейворда. 1 – фильтровать от стопслов, 0 – не фильтровать.
34. [KEYM-мр,ед,им] – вставить кейворд в указанной словоформе. (мр, жр, ср - род; им,рд,дт,вн,тв,пр – падежи; ед, мн – число)
35. [KEYDM-мр,ед,им] – тоже самое что KEYM, только с перемешиванием слов в кейворде.
36. [ARTICLE-arc.txt-\$\$\$\$] – вставлять поочередно из файла(files/arc.txt) статьи, разделителем для статей является \$\$\$\$, если статьи кончились макрос просто удаляется.
37. [SPLIT] – слова текущего кейворда через запятую.
38. [RDIR-data/lnk] – рандомная строка из рандомного файла, который находится в папке data/lnk.
39. [LDIR-data/lnk] – текст из рандомного файла, который находится в папке data/lnk.

Настройки(config.cfg):

Описание настроек обновляется редко, все самые последние параметры можно посмотреть в config.cfg.

Settings

- **site name** - Имя сайта.
- **template** – имя шаблона. Сам шаблон кидать в папку “files/templ/имя шаблона/html.html”. Все файлы лежащие в папке с шаблоном будут автоматически скопированы в дорвей.
- **domain** – Доменные имена для дорвеев. Если генерация пакетная, то домены указываются через запятую.
- **pages** – количество страниц.
- **struct** – структура ссылок. См структура ссылок.
- **kw in text min** – минимальное количество кейвордов в тексте.
- **kw in text max** – максимальное количество кейвордов в тексте.
- **href in text min** – минимальное количество ссылок в тексте.
- **href in text max** – максимальное количество ссылок в тексте.
- **main href keyword** – делать ссылки с кейвордами в тексте, ссылающимися сами на себя. (1/0)
- **random href keyword** – делать ссылки в тексте ссылающиеся на случайные страницы. (1/0)
- **clear pages** – через сколько “чёрных” делать “белые” страницы . Если указать 10, то в дорвее на 1000 страниц будет 100 белых. -1 отключает белые страницы.
- **out encoding** – кодировка дорвея на выходе.
- **js** – см макрос js.
- **tags** – тэги через запятую, которые будут выделять кейворды в тексте.
- **type** – тип генерации контента, типы описаны выше.(1,2,3,4,5)
- **use xml map** – использовать xml карту сайта. (1/0)
- **savelinks** – сохранять ссылки дорвея в файл. (1/0)
- **zip** – зиповать (1/0)
- **delete** – удалять после генерации(1/0)
- **rss** – использовать rss
- **exitonerr** – выходить при ошибке.
- **custom** – свой шаблон ссылок.
- **Inkreg** – шаблон для сохранения ссылок на дор. %url% - заменяется на ссылку, %key% - заменяется на кей.
- **savelinks** – сохранять ссылки (1/0).
- **useplug** – подключать плагины (1/0).

Files

- **keywords file** – файл с кейвордами.
- **content file** – файл с контентом.
- **Imgnames** – файл с названиями картинок.

FTP

- **ftp** – Использовать ftp.(1/0).
- **ftphost** – адрес ftp.
- **ftpusr,ftppass** – логин/пароль от ftp.
- **ftkdir** – папка куда заливать. Например(www/public_html/%s) %s заменяется на доменное имя(www/public_html/domain.ru).
- **filezilla** – генерировать проекты для filezilla(1/0)
- **ftpzip** - зиповать ,заливать и распаковывать дор в виде zip файла.(1/0)

Img

- **useimg** - использовать картинки. (1/0)
- **maximg** – максимально количество копируемых картинок.
- **imgdir** – папка с картинками.
- **dorimgdir** – папка с картинками на самом доре.

Парсинг

- **img dir** = папка куда парсить картинки.

- **pages** = сколько парсить картинок.
- **query** = запрос по которому парсить.

Youtube

- **pages** – сколько страниц
- **file** – файл для сохранения
- **query** - запрос

Парсинг

- **pages** – сколько страниц
- **file** – файл для сохранения
- **query** - запрос
- **feedtype** – источник.

Чтобы быстрее разобраться в макросах и настройках советую просто сгенерировать тестовый дорвей со стандартным шаблоном. Т.к в дорах используется много страниц, копировать столько же картинок не очень удобно из за веса. Для это сделан параметр `maximg`, он регулирует количество копируемых картинок. На страницах дора картинки будут просто повторяться. Т.е если `maximg=10`, то на страницах дора будет выводиться только 10 уникальных картинок, остальные будут просто повторяться.

Генерация белых страниц

Подразумевает под собой генерацию некого количества страниц в дорвее, в тексте которых не будет кейвордов. Таким образом дорвей выглядит более “белым” для Яндекса и увеличивается время его жизни .

Парсинг картинок

тут всё просто, указывается ключ, по которому будет производиться парсинг и количество страниц, на выходе получаем картинки, которые потом можно вставить в дорвей.

Основные ошибки при генерации дорвея:

1. Неправильный синтаксис и значения макросов
2. Некорректная кодировка файлов
3. Отсутствие значений в конфиге

При генерации текста с использованием разбивки слов + морфологии в текстовых блоках требуется много контента, от 5 предложений минимум.

Если в шаблоне пропадают какие-то символы, то нужно поменять кодировку файла с шаблоном. Если используется utf-8, то обязательно без BOM.

Вставка кейвордов

Кейворды могут рассчитываться как количеством, так и процентами.

Формула расчета кейвордов процентами:

Кол-во кейвордов = (Кол-во символов в тексте*(процент кеев*0.01))/Кол-во кеев в тесте.

Методы вставки кейвордов:

1. Обычная вставка – случайное слово в тексте заменяется на кейворд
2. Соблюдая морфологию – случайное слово в тексте заменяется на кейворд, кейворд пытается принимать словоформу заменяемого слова.
3. Соблюдая морфологию + разбивка кейворда на слова - тоже самое что во втором случае, только кейворд разбивает на части. Данный способ имеет легкий и сложный режимы.
4. `rWp` - в обоих случаях кейворд разбивается на слова и из него удаляются стоп слова. первом случае в исходном тексте находится слово полностью соответствующее словоформе слова из кейворда (часть речи, род, число падеж и т.д) и слово в текстовом блоке заменяется на слово из кейворда. Во втором случае тоже самое, но только исключая некоторые морфологические признаки, ввиду того, что слова полностью совпадающие по словоформе встречаются не так часто.

Работа с FTP

R-dorger поддерживает 3 вида работы с ftp:

1. Обычная заливка. Тут думаю всё итак понятно, все файлы поочередно заливаются на ftp сервер.
2. Заливка zip архива и распаковка его на сервере. Все файлы сжимаются в zip архив, далее архив заливается на сервер и распаковывается там php скриптом. Требования для такого режима: php, возможность указать права файлам.
3. Генерация проекта для filezilla Генерируется проект для fz. Проекты сохраняются в папке `data/lnk/имя домена.xml`. `Packet.xml` – проект сразу со всеми дорами, если используется пакетка.

Свой шаблон для главной страницы

В R-doggen так же имеется возможность указывать свой шаблон для главной страницы, для этого в папке с шаблоном достаточно создать файл index.html и поместить туда шаблон для главной. Для всех остальных страниц будет использоваться шаблон из html.html

Своя структура ссылок:

Для этого в структуре ссылок нужно указать “Свой шаблон” а в поле “Свой шаблон url” из имеющихся макросов указать структуру. Пример: [KEYWORD]-[RANDINT-10-999].php

Все страницы дора в одном файле:

Для этого нужно выбрать свой шаблон для url и указать в нём значение ‘append’ (без кавычек).

Robots.txt

Чтобы сгенерировать свой robots.txt достаточно положить его в папку шаблоном, для подстановки домена используется %s.

Например: если указать http://%s/sitemap.xml, то на выходе в robots.txt окажется http://domain.ru/sitemap.xml.

Использование своего шаблона для файла с ссылками

Для этого требуется поправить параметр lnkreg в config.cfg.

Если требуется несколько файлов, то в шаблоне используем разделитель “|||” (без кавычек).

Пример: %key%||| [URL=%lnk%]%key%[/URL]

Создастся два файла: В одном из них будет Тайтл страницы, а в другом [URL='http://site.com']Тайтл страницы[/URL].

Все файлы сохраняются в data/lnk,

Пакетная генерация:

В новой версии пакетки есть возможность переопределить любой параметр из конфига, за исключением некоторых критических, которые не могут быть изменены по ходу работы.

Все изменяемые параметры так же указываются через |. Чтобы изменить какое либо значение достаточно написать так:

```
settings>domain=site.ru|ftp>ftphost=ftp.ya.ru|settings>clearpages=10
```

Все пробелы в параметрах конфига нужно убирать clear pages = clearpages и т.д.

Не указанные параметры в пакетке автоматически берутся из основного файла конфига. Все задания добавляются в packet.txt, одно задание – одна строка. settings>domain=site.ru|ftp>ftphost=ftp.ya.ru|settings>clearpages=10 settings>domain=site2.ru|ftp>ftphost=ftp.ya.ru|settings>clearpages=10

Разбивка на <p></p>

Если добавить к текстовым макросам параметр “p” ([ТЕХТ-10-15p], [СТЕХТ-10-15p] и т.д), то текстовый блок автоматический разобьётся на параграфы(<p></p>).

Проверка ftp аккаунтов перед пакетной генерацией

В R-doggen так же присутствует чек ftp аккаунтов, которые находятся в файле packet.txt. Перед проверкой все аккаунты сохраняются в файле packet_old.txt, валидные аккаунты остаются в packet.txt

Генерация Rss

Rss генерируется только тогда, когда количество страниц больше 10.

Использование своих плагинов:

Для того чтобы использовать свои плагины достаточно положить их в папку plugins.

Примером плагина служит файл ping.ru, он выполняет пинг яндекса и гугла. Принцип подключения плагинов:

1. В папке plugins производится поиск всех .ru файлов.
2. Во всех найденных файлах производится поиск классов.
3. В каждом классе вызывается метод run(domain,keywords,keyurl,pages)

Domain – домен

Keywords – список кейвордов

Keyurl – список ссылок на страницы(без domain), чтобы получить полную ссылку можно воспользоваться – “http://%s/%s”%(domain,keyurl[0]).

Pages – количество страниц.

Кейворд на странице keyurl[5] равен keywords[5].

Количество страниц равно количеству кеев

Для того чтобы количество страниц было равно количеству кеев в настройках достаточно указать кол-во страниц = 0 .

История изменений программы:
<http://r.rushter.com/changelog.txt>