Русские Блоги



Введение в инструмент wkhtmltopdf из HTML в PDF

Статьи по Теме

Введение в инструмент wkhtmltopdf из HTML в PDF

Значение терминов

Объект документа

«Объект документа» относится к объекту документа в документе PDF. Существует три типа «объекта документа»: «объект страницы», «объект

Объект страницы

«Объект страницы» относится к объекту, представленному в PDF-документе в форме страницы, который относится к «объекту обложки» и «об

Обложка объекта

«Объект обложки» относится к объекту, представленному в PDF-документе в виде обложки. Такие объекты станут лицевой обложкой PDF-доку

Объект каталога

«Объект оглавления» - это объект, представленный в документе PDF в форме оглавления, также называемый «объектом оглавления». Такие с

Контур

«Схема» относится к навигации по закладкам при чтении документов PDF.

Внешняя ссылка

«Внешние ссылки» относятся ко всем гиперссылкам на этой странице, которые не указывают на якоря на его собственной странице.

Внутренняя цепь

«Внутренняя ссылка» относится к гиперссылке на этой странице, которая указывает на точку привязки на самой странице.

Формат команды

wkhtmltopdf [GLOBAL OPTION]... [OBJECT]... <output file>

Приведенный выше код представляет собой формат командной строки для wkhtmltopdf. Это кажется простым, но на самом деле [GLOBAI медленно слушайте меня.

Введение в объекты документа

wkhtmltopdf может объединять несколько «объектов» для создания PDF-документа. Эти «объекты» могут быть «объектами страницы», «объекументе можно указать с помощью параметров командной строки. Параметры командной строки состоят из двух частей: одна - это парам параметры командной строки. И глобальные параметры командной строки могут быть указаны только в области глобальных параметров ([GLC

Введение в объекты страницы

«Объект страницы» используется для вывода веб-страницы в качестве содержимого документа PDF.

```
(page)? <input url/file name> [PAGE OPTION]...
```

Параметры «объекта страницы» могут быть помещены в «поле глобального параметра ([GLOBAL OPTIONS])» и «поле параметра страниць параметров в соответствии с реальной ситуацией и применит их к странице, верхнему и нижнему колонтитулам.

Введение в объекты обложки

«Объект обложки» используется для вывода веб-страницы в качестве титульной страницы документа PDF. Страница вывода не будет отображ

```
cover <input url/file name> [PAGE OPTION]...
```

Все параметры, которые могут использоваться в «объекте страницы», могут использоваться в «объекте обложки».

Введение в объекты каталога

Функция «объект каталога» - вывод каталога в файл PDF.

```
toc [TOC OPTION]...
```

Все параметры, которые можно использовать в «объекте страницы», можно использовать в «объекте оглавления», и есть много параметров д Каталог создается с помощью XSLT, что означает, что его можно определить как все, что вы хотите видеть. Вы можете передать парамет умолчанию, пройти--dump-outline Параметры командной строки могут указывать выходной каталог обрабатываемого в данный мс дополнительных сведений см. Следующее введение Параметры объекта каталога

Параметры команды

Параметры команды состоят из пяти частей, а именно «глобальные параметры», «параметры параметры структуры», «параметры объек «параметры объекта каталога».

Глобальные параметры

```
--collate check при выводе нескольких копий (это настройка по умолчанию)
       --no-collate не проверять при выводе нескольких копий
       --cookie-jar <путь> Чтение и запись данных cookie из предоставленного файла JAR
       --copies <number> Установить количество выходных копий (по умолчанию основная 1), фактически достаточно 1
-d, -dрi < dрi > Укажите желаемое разрешение (это не полезно в системах X11)
-H, --extended-help Относительно параметра -h, отобразить более подробную документацию
-g, --grayscale Задает создание PDF-документов в оттенках серого. Занимает меньше места
-h, --help показать справочную информацию
       --htmldoc вывести html-файл справки программы
       --image-dpi <integer> Когда на странице есть встроенное изображение,
                                              Будет загружать изображения размера, указанного в этом параметре ком
       --image-quality <interger> Использовать качество, указанное в этом параметре, при использовании алгоритма jpeg
       --license вывести информацию об авторизации и выйти
-1, --lowquality Создать PDF / PS низкого качества, что может сэкономить место для хранения окончательно сгенерирова
       --manpage вывести справочную страницу программы
-B, --margin-bottom <unitreal> Установить нижнее поле страницы
-L, --margin-left <unitreal> Установить левое поле страницы (по умолчанию 10 мм)
-R, --margin-right <unitreal> Установить правое поле страницы (по умолчанию 10 мм)
-T, --margin-top <unitreal> Установить верхнее поле страницы
-0, --orientation <orientation> установить в режим «Пейзаж» или «Портрет»,
                                                         По умолчанию используется портретный модуль (Portrait).
       --page-height <unitreal> высота страницы
-s, --page-size <Size> Установить размер страницы, например: А4, Letter и т. д., по умолчанию: А4
       --page-width <unitreal> ширина страницы
       --no-pdf-compress He использовать алгоритмы сжатия, которые теряют небольшой объем информации об объектах PDF
                                                       Потому что сгенерированный PDF-файл будет очень большим.
-q, --quiet static mode, не выводить информацию на стандартный вывод
       --read-args-from-stdin читает параметры командной строки со стандартного ввода, подробное введение в эту кома
                                                       См. ** Получение параметров из стандартного ввода **
       --Readme output program's readme-документ
```

```
31 --title <text> Заголовок сгенерированного PDF-документа, если не указан, используется заголовок первого документа.
```

Вышеупомянутая область кода - это все глобальные параметры и комментарии. Давайте кратко поговорим о значении и использовании отдель

-copies N

N - целое положительное число.

Вам не стоит переживать по поводу этого варианта, потому что он может вам не понадобиться в вашей жизни. Его роль - многократно выводи получите документ PDF, размер и содержание документа будут в N раз больше без этого параметра. Однако повторяющийся контент вам ни к

Если не использовать--copies Параметры, тогда--collate с участием --no-collate Параметры не нужно понимать, потому что они нахо больше 1.

-g, -grayscale

Этот параметр очень полезен. Использование этого параметра может эффективно сжимать пространство хранения, занимаемое созданным в цену, то есть окончательно сгенерированный PDF-документ будет в оттенках серого без какого-либо цвета. Если вы можете принимать д используйте этот параметр. Чем больше размер сгенерированного PDF-документа, тем больший сюрприз вы получите при использовании этог

-I. -lowquality

Этот параметр такой же, как-**g** Параметры разные одинаково,-**1** Параметры также значительно сжимают пространство для хранения, заним качества PDF-документа. Этот параметр также стоит порекомендовать. Вам лучше сначала попробовать и принять решение, увидев разницу тем, что он не используется. Что я могу вам сказать, так это то, что в случае простого текста между ними нет большой разницы. Этот парами влияет на чтение. Если вы ищете чувственное наслаждение или в создаваемом вами PDF-документе много изображений, не используйте этот

-no-pdf-compression

Этот параметр настоятельно не рекомендуется использовать, лучше не понимать его пользы в этой жизни, ведь он точно не используется для приведет к тому, что выходной PDF-документ будет чрезвычайно большим, а качество будет без потерь, но для людей изменение качества обычных людей, и сжатый опыт для вас неприемлем, то я верну предыдущие слова, и вы можете использовать этот параметр сколько угодно.

-q, -quiet

После использования этого параметра вы получите чистый вывод командной строки, даже не ход и статус обработки программы. Этот парамет время рабочего процесса программы. Рекомендуется не использовать этот параметр, потому что программа выводит некоторый прогресс, а и где-то умерла (хе-хе), в-**q** Вы не можете сказать, умерла ли программа в режиме.

Варианты параметров схемы

```
--dump-default-toc-xsl вывести таблицу стилей TOC xsl по умолчанию на стандартный вывод
--dump-outline <файл> Вывести "схему" в указанный файл (содержимое файла - xml)
--outline выводит "контур" в созданный PDF-документ (это настройка по умолчанию)
--no-outline не выводить схему в PDF-документ
--outline-depth <уровень> Установить глубину генерируемого контура (по умолчанию 4)
```

Единственное, что требует особого упоминания в параметрах схемы, это--outline-depth, Остальные параметры по умолчанию в порядке.

Что такое набросок







Контур

Как показано на картинке выше, я предпочитаю называть это каталогом или навигацией. Схема создается автоматически на основе заголовка

-outline-depth

--outline-depth Используется для указания глубины создаваемого контура. Значение по умолчанию - 4. Вы можете указать большее чис соответствующие элементы в структуре, что удобно для быстрого перехода при чтении.

Когда указано--no-outline Параметр, контур не будет выводиться в документ PDF, поэтому укажите--outline-depth Het смысла.

Параметры объекта страницы

```
--allow «путь» Укажите каталог для загрузки файла относительного пути в HTM. (этот параметр можно использовае Directory), этот параметр будет объяснен более подробно в 3 --background вывести фон страницы в PDF-документ (это настройка по умолчанию)

4 --no-background Не выводить фон страницы в PDF-документ (это настройка по умолчанию)

5 --checkbox-top «путь» кеш-каталог веб-страницы в PDF-документ (--checkbox-top «путь» отноразить выбранный флажок с использованием указанного файла SVG --checkbox-top «путь» отноразить выбранный флажок с использованием указанного блока фильтра --cookie «пить» «значение» Установить заголовок НТР при доступе к веб-страницы (этот параметр можно использовать повторно, чтобы указа --cookie-header «шин» «значение» Установить заголовок НТР при доступе к веб-странице (этот параметр можно использовать повторно, чтобы указа --custom-header-aggation фодавляет заголовок НТР при доступе к веб-странице (этот параметр можно использовать повторно, чтобы указа --custom-header-aggation фодавляет заголовок НТР при доступе к веб-странице (этот параметр можно использовать повторно, чтобы указаний параметром --custom-header, для каждого загружаемот 12 --no-custom-header-aggation фодавляет вижорпие данише отнорите и при нестранице (этот параметр можно использовать повторно, чтобы указателя (этот параметром --custom-header, для каждого загружаемот --default-header добавить заголовок по умолчанию, отображать ини страницы и певой части страницы, что--deday-javascript не отображать выжорпие рашене отпарать участи (это заголовок по умолчанию, отображать ини отображать в правой части страницы, что--header-left="(вераде1" --header-left="(вераде1" --header-left="(вераде1" --header-left="(вераде1" --header-line --enedding <-enedding <-enedd
```

Приведенный выше фрагмент кода предназначен для всехПараметры объекта страницы Дается общее объяснение, а ниже приводится боле

-allow

Этот параметр действителен, только когда «объект страницы» является файлом, и недействителен, когда «объект страницы» является URL-аг

Функция этого параметра - указать базовый каталог для загрузки файлов для файлов, на которые ссылаются относительные пути на стран HTML, будут из--allow Каталог, указанный параметром, загружен. Собственно указать в HTMLbase Этикетки могут служить той же цели. находится обрабатываемый в данный момент файл HTML, используется в качестве базового каталога для загрузки файла, указанного относит

-background AND -no-background

Эти два параметра представляют собой пару и используются для указания, применять ли фон веб-страницы в сгенерированном PDF-файле, созданный по умолчанию документ PDF имеет фоновое изображение или цвет фона HTML-страницы. Если открыто--no-backgroupd Парам цвет фона на странице HTML.

-debug-javascript ADN -no-debug-javascript

Эти два параметра используются для указания, выводить ли отладочную информацию JavaScript в стандартный вывод, по умолчанию информация JavaScript не выводится по умолчанию. Картинка ниже открыта -- debug-javascript Демонстрация параметров.

```
D:\\wkhtmltopdf —debug-javascript http://www.baidu.com 1.pdf
Loading pages (1/6)
Warning: https://ssl.bdstatic.com/beNlbjg8AAUYm2zgoY3K/r/www/cache/static/protocol/https/global/js/ail_async_search_dd9e
5445.js:1 一张网页,要经历怎样的过程,才能抵达用户面前?
一位新人,要经历怎样的成长,才能站在技术之巅?
探寻这里的秘密;
体验这里的挑战;
成为这里的主人;
加入白度,加入网页搜索,你,可以影响世界。

Warning: https://ssl.bdstatic.com/5eNlbjg8AAUYm2zgoY3K/r/www/cache/static/protocol/https/global/js/all_async_search_dd9e
5445.js:1 请将简历发送至 %c ps_recruiter@baidu.com ( 邮件标题请以 "姓名一文聘双职位一来自console" 命名》
Warning: https://ssl.bdstatic.com/5eNlbjg8AAUYm2zgoY3K/r/www/cache/static/protocol/https/global/js/all_async_search_dd9e
5445.js:1 职位介绍:http://dwz.com/sen/bjg8AAUYm2zgoY3K/r/www/cache/static/protocol/https/global/js/all_async_search_dd9e
5445.js:1 职位介绍:http://dwz.com/hr2013
Counting pages (2/6)
Resolving links (4/6)
Loading headers and footers (5/6)
Printing pages (6/6)
Done
```

-disable-external-links AND -enable-external-links

Эти два параметра используются, чтобы установить, будут ли внешние ссылки на странице отображаться в документе PDF в виде гипє **терминов**. По умолчанию **--enable-external-links** Параметр включен, поэтому по умолчанию внешняя ссылка на странице находится в страницу.

-exclude-from-outline AND -include-in-outline

Эти два параметра используются, чтобы установить, включен ли объект текущей страницы в оглавление и структуру.

По умолчанию--include-in-outline Параметр открыт. Другими словами, сгенерированный каталог и структура документа PDF включают забавлялась в каталог и структуру, вы можете открыть ee.--exclude-from-outline параметр.

-post AND -post-file

Эти два параметра можно использовать, когда целевая страница должна принять форму POST для правильного получения ответа. Оба эти па

Другой сценарий приложения используется в автоматизированном тестировании веб-приложений. Вы можете получить PDF-файл в качестве с

--post-file Его также можно использовать для автоматической загрузки файлов партиями.

-run-sript

Когда страницу необходимо предварительно обработать для создания сцены документа PDF, использование этого параметра не может быть б нескольких кодов JS, которые необходимо выполнить после загрузки страницы. Вы можете обработать структуру и содержимое страницы в после выполнения JS

-disable-internal-links AND -enable-internal-links

Эти два параметра используются для определения, будет ли внутренняя ссылка на странице отображаться в документе PDF в виде гиперссытерминов. По умолчанию --enable-internal-links Параметр включен, поэтому по умолчанию внутренняя ссылка на странице находится точке привязки в текущем PDF-документе.

-enable-toc-back-links AND -disable-toc-back-links

Этот набор параметров используется для определения необходимости создания гиперссылки в теге H в содержимом PDF. После нажат соответствующей тегу H в таблице содержания и структуре. по умолчанию --disable-toc-back-links Когда параметр включен, гиперссыли

Если вам нужно быстро вернуться к оглавлению при чтении содержимого PDF-документа, вы можете открыть -- enable-toc-back-links пар

-user-style-sheet

Этот параметр используется для загрузки пользовательской таблицы стилей для изменения исходного стиля HTML-страницы. Студенть использовать этот параметр для достижения своих целей.

Параметры параметров верхнего и нижнего колонтитула

```
--footer-center \langletext\rangle Показать текст нижнего колонтитула \langletext\rangle в центральной части нижнего колонтитула
--footer-font-name <имя> Установить шрифт нижнего колонтитула (по умолчанию Arial)
 --footer-font-size <<mark>размер> Установить размер шрифта нижнего колонтитула (по умолчанию</mark> 12)
 --footer-html <url> добавить html как нижний колонтитул
 --footer-left <text> Отображать текст нижнего колонтитула в левой части нижнего колонтитула <text>
 --footer-line отображает прямую линию над нижним колонтитулом для разделения текста
 -no-footer-line Не использовать прямую линию для разделения нижнего колонтитула и тела (это настройка по ум
  -footer-right <text> Показать текст нижнего колонтитула <text> в правой части нижнего колонтитула
--footer-spacing <real> Расстояние между нижним колонтитулом и телом (по умолчанию ноль)
--header-center <text> Показать текст заголовка <text> в центральной части заголовка
--header-font-name <umm> Установить шрифт заголовка (по умолчанию Arial)
--header-font-size <pasмep> Установить размер шрифта заголовка (по умолчанию 12)
--header-html <url> добавить html в качестве заголовка
--header-left <text> Отображать текст заголовка <text> в левой части заголовка
--header-line отображает прямую линию под заголовком для разделения текста
--no-header-line Не использовать прямую линию для разделения заголовка и тела (это настройка по умолчанию)
--header-right <text> Отображать текст заголовка <text> в правой части заголовка
--header-spacing <real> Расстояние между заголовком и телом (по умолчанию ноль)
```

Настройка верхнего и нижнего колонтитула относительно проста.Объяснение в приведенном выше фрагменте кода уже очень четкое, поэтог нижних колонтитулов.

Параметры объекта каталога

```
--disable-dotted-lines Не использовать пунктирные линии в каталоге
--toc-header-text <text> Установить текст заголовка таблицы содержания
--toc-level-indentation <ширина> Ширина отступа заголовка уровня в таблице содержания (по умолчанию 1em)
--disable-toc-links не генерируют гиперссылки на якоря контента в каталоге
--toc-text-size-shrink <real> Коэффициент масштабирования каждого уровня заголовка в каталоге (по умолчанию 0,
--xsl-style-sheet <файл> Использовать настраиваемую таблицу стилей XSL для отображения содержимого каталога
```

Обычно мы не используем «объекты каталога», и пояснения в приведенных выше фрагментах кода нетрудно понять, поэтому мы не даем подг

Об инструкциях по размеру страницы

Размер страницы по умолчанию - A4, вы можете использовать --page-size В параметрах указывается нужный размер страницы, напри поддерживаемые этой программой, посетите http://qt-project.org/doc/qt-4.8/qprinter.html#PaperSize-enum

Вы также можете использовать -- page-height с участием -- page-width Для более точного управления размером страницы.

Получить параметры из стандартного ввода

Если вам нужно обработать много страниц в пакетном режиме и вы чувствуете, что wkhtmltopdf открывается медленно, вы можете попробовати

Команда wkhtmltopdf будет--read-args-from-stdinДля каждой строки отправленных параметров выполняется отдельный вызов коман каждый раз при чтении строки. Параметры в последней выполненной команде представляют собой комбинацию параметров в команди параметром.

Следующий фрагмент кода является примером:

```
1  echo "http://qt-project.org/doc/qt-4.8/qapplication.html qapplication.pdf" >> cmds
2  echo "cover google.com http://en.wikipedia.org/wiki/Qt_(software) qt.pdf" >> cmds
3  wkhtmltopdf --read-args-from-stdin --book < cmds</pre>
```

Проинструктировать агента

По умолчанию информация о прокси будет считывать переменные среды: proxy, all_proxy и http_proxy, параметры прокси также можно указать

Определение агента, использующего BNF, следующее:

```
1 | <type> := "http://" | "socks5://" | <serif> := <username> (":" <password>)? "@" | <proxy> := "None" | <type>? <serif>? <host> (":" <port>)?
```

Если вы не знакомы с BNF, следующие фрагменты кода являются тремя примерами:

```
1 http://user:password@myproxyserver:8080
2 socks5://myproxyserver
3 None
```

Верхний и нижний колонтитулы

Верхний и нижний колонтитулы могут использовать параметры--header-* с участием --footer-* Добавить в файл. Некоторые параме значения параметра. Вы можетеtextВставьте следующие переменные, и они будут заменены соответствующими значениями.

```
| [раде] Номер страницы, выводимой в данный момент.
| [frompage] Номер первой страницы документа.
| [topage] Номер последней страницы в документе.
| [веб-страница] URL-адрес выводимой в данный момент страницы
| [section] Имя экспортируемого раздела.
```

```
6 [подраздел] Имя выводимого в данный момент подраздела 7 [date] Текущая дата в формате локальной системы.
8 [isodate] Текущая дата в формате ISO 8601.
9 [время] Текущее время в формате локальной системы
10 [title] Заголовок текущего объекта
11 [doctitle] Заголовок выходного документа.
12 [sitepage] Номер текущей страницы обрабатываемого объекта
13 [sitepages] Общее количество страниц обрабатываемого объекта
```

Приведите пример для иллюстрации, --header-right "Page [page] of [toPage]", Будет генерировать похожий на верхний правый у последней страницы текущего документа.

Верхний и нижний колонтитулы также могут быть предоставлены через HTML-документы. Также приведите пример с использованием пзаголовок, и содержимое header.html выглядит следующим образом:

Outline (контуры)

wkhtmltopdf можно использовать--outline Параметры командной строки используются для указания, что «схема», такая как оглавление основном HTML-документе. Перейдите на конкретный уровень структуры и попробуйте.оглавление

Если в HTML-документе больше H-тегов, вы можете создать «контур» с глубокой древовидной структурой, и будет сгенерирована истинная глу

оглавление

Вы можете добавить каталог в созданный PDF-документ, добавив в командную строку объект TOC, например следующий фрагмент кода:

```
wkhtmltopdf toc http://qt-project.org/doc/qt-4.8/qstring.html qstring.pdf
```

Созданный каталог также основан на теге H HTML-документа. Процесс состоит в том, чтобы сначала сгенерировать XML-документ, а затем исг Созданный XML-документ может передавать - - dump - out line Просмотр параметров.

```
wkhtmltopdf --dump-outline toc.xml http://qt-project.org/doc/qt-4.8/qstring.html qstring.pdf
```

Если вы хотите использовать собственный документ XSLT, вы можете передать--xsl-style-sheet Спецификация параметра

```
wkhtmltopdf toc --xsl-style-sheet my.xsl http://qt-project.org/doc/qt-4.8/qstring.html qstring.pdf
```

ты можешь использовать --dump-default-toc-xsl Параметры распечатывают документ XSLT по умолчанию на стандартный вывод, а затег

подводить итоги

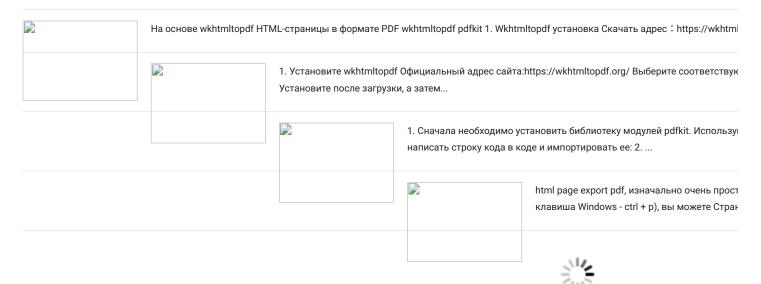
Выше представлен весь контент об инструменте wkhtmltopdf, большая часть которого основана на чтении wkhtmltopdf. - Н Получен английский какие-то недостатки.

Статьи по Теме

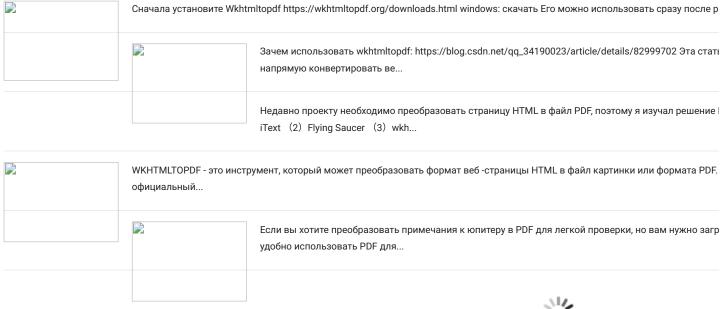
Введение в инструмент wkhtmltopdf из HTML в PDF

Интеллектуальная рекомендация

Go проект с GoLand IDE Название проекта toPDF Добавить файл main.go, добавить основную функцию, Пример взят из:https://github.com/adrg/go-wkhtmltopd



Вам также может понравиться





© 2020-2024 All rights reserved by russianblogs.com