

HTML to ODT to XML to PDF to ...

FrOSCon 2010

Tobias Schlitt <toby@qafoo.com>

August 23, 2010



License

Copyright by Tobias Schlitt, Qafoo GmbH

Licensed under Creative Commons
Attribution-NonCommercial-ShareAlike 2.0 Generic



About me

- ▶ Tobias Schlitt
 - ▶ Apprenticed IT specialist
 - ▶ Waiting for diploma to be approved (TU Dortmund)
 - ▶ Open source enthusiast
 - ▶ Co-founder of Qafoo - passion for software quality
 - ▶ PMC member of Apache Zeta Components
 - ▶ Contributor to various other OSS projects

Outline

Introduction

The Document component

Getting into the code

End



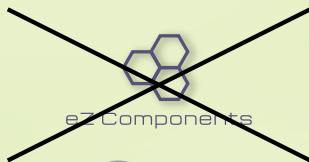
eZ Components



eZ Components



Apache Zeta Components



Apache Zeta Components

- ▶ Previously developed by eZ Systems
 - ▶ <http://ez.no>
- ▶ Originally named eZ Components
- ▶ Code donated to the Apache Software Foundation
 - ▶ <http://apache.org>
- ▶ Currently incubating
 - ▶ Re-organizing / re-gathering community
 - ▶ Join us!

Apache Zeta Components

- ▶ General purpose component library for PHP 5.1+
- ▶ Open source (Apache 2.0 license)
- ▶ Focus
 - ▶ High code quality
 - ▶ Excellent docs
 - ▶ Backwards compatibility
- ▶ Professional support available
- ▶ <http://zetacomponents.org>

The components

49 components, an extract ...

- ▶ Archive
- ▶ ConsoleTools
- ▶ Graph
- ▶ Mail
- ▶ MvcTools
- ▶ Webdav
- ▶ Workflow



Outline

Introduction

The Document component

Getting into the code

End



Goal

► Applications need to deal with

► Different input mechanisms

- WYSIWYG editor (HTML)
- Simple text editor (wiki markup)
- Emails (ReST)

► Different output formats

- Web front end (HTML)
- Technical documentation management (Docbook)
- Print (PDF)

► Document component converts markup formats

Goal

- ▶ Applications need to deal with
 - ▶ Different input mechanisms
 - ▶ WYSIWYG editor (HTML)
 - ▶ Simple text editor (wiki markup)
 - ▶ Emails (ReST)

Input formats

Web front end (HTML)

Technical documentation management (Docbook)

Print (PDF)

Document component converts markup formats

Goal

- ▶ Applications need to deal with
 - ▶ Different input mechanisms
 - ▶ WYSIWYG editor (HTML)
 - ▶ Simple text editor (wiki markup)
 - ▶ Emails (ReST)
 - ▶ Different output formats
 - ▶ Web front end (HTML)
 - ▶ Technical documentation management (Docbook)
 - ▶ Print (PDF)

document component converts markup formats

Goal

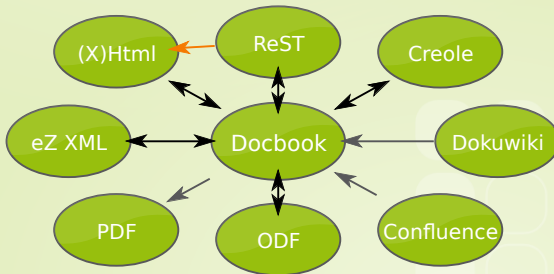
- ▶ Applications need to deal with
 - ▶ Different input mechanisms
 - ▶ WYSIWYG editor (HTML)
 - ▶ Simple text editor (wiki markup)
 - ▶ Emails (ReST)
 - ▶ Different output formats
 - ▶ Web front end (HTML)
 - ▶ Technical documentation management (Docbook)
 - ▶ Print (PDF)
- ▶ The Document component converts markup formats

Supported formats

- ▶ Currently supported formats
 - ▶ Docbook
 - ▶ (X)Html
 - ▶ eZ XML
 - ▶ ReST
 - ▶ Wiki
 - ▶ Dokuwiki, popular PHP based wiki (wiki.php.net) (read-only)
 - ▶ Creole, wiki markup standardization initiative
 - ▶ Confluence, Apache Atlassian wiki dialect (read-only)
 - ▶ PDF (write only)
 - ▶ ODF (only flat)

Approach

- ▶ Docbook as central conversion format
 - ▶ Possible conversion shortcuts
 - ▶ Conversions always configurable and extensible



Outline

Introduction

The Document component

Getting into the code

End



Reading ReStructured Text

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26

PHP @ FrOSCon 2010

For the fifth time we will be at the 'Free and Open Source Conference' (FrOSCon) in St. Augustin, near Bonn, organizing a track full of PHP related talks. We also offer space to discuss PHP related topics, or just hack with other open minded people around you. We would love to welcome you in the PHP room.

We are currently looking for talks for the PHP room, and the 'Call For Papers' will last until the 23.05.2010, just like the 'Call For Papers' of the FrOSCon. Please submit a talk, if you got something interesting to talk about.

You might also want to submit talks to the 'main schedule' of the conference, which also accepts PHP related talks. For talks in the main conference your costs will be covered as usual, details are on the 'dedicated website'.

```
-- http://phpugdo.de/  
-- http://froscon.de/  
-- /call\_for\_papers.html  
-- http://www.froscon.de/index.php?id=15&mid=119&ret=15&L=0&L=0  
-- http://www.froscon.de/index.php?id=15&mid=119&ret=15&L=0&L=0  
-- http://www.froscon.de/index.php?id=15&mid=119&ret=15&L=0&L=0
```

Reading ReStructured Text

```
1 <?php
2
3 require './autoload.php';
4
5 $document = new ezcDocumentRst();
6 $document->loadFile( 'froscon.txt' );
7
8 echo $document->getAsDocbook();
```

Reading ReStructured Text

```
1 <?php
2
3 require './autoload.php';
4
5 $document = new ezcDocumentRst();
6 $document->loadFile( 'froscon.txt' );
7
8 echo $document->getAsDocbook();
```

Reading ReStructured Text

```
1 <?php
2
3 require './autoload.php';
4
5 $document = new ezcDocumentRst();
6 $document->loadFile( 'froscon.txt' );
7
8 echo $document->getAsDocbook();
```

Reading ReStructured Text

```
1 <?php
2
3 require './autoload.php';
4
5 $document = new ezcDocumentRst();
6 $document->loadFile( 'froscon.txt' );
7
8 echo $document->getAsDocbook();
```

Reading ReStructured Text

```
1 <?xml version="1.0" ?>
2 <!DOCTYPE article PUBLIC "-//OASIS//DTD_DocBook_XML_V4.5//EN" "http://www.oasis-
  open.org/docbook/xml/4.5/docbookx.dtd">
3 <article xmlns="http://docbook.org/ns/docbook">
4   <section ID="php-froscon-2010">
5     <title>PHP @ FrOSCon 2010</title>
6     <para>For the fifth time <ulink url="http://phpugdo.de/">we</ulink> will be
      at the <ulink url="http://froscon.de/">Free and Open Source Conference<
      /ulink> (FrOSCon) in St. Augustin, near Bonn, organizing a track full
      of PHP related talks. We also offer space to discuss PHP related topics
      , or just hack with other open minded people around you. We would love
      to welcome you in the PHP room.</para>
7     <para>We are currently looking for talks for the PHP room, and the <ulink
      url="/call_for_papers.html">Call For Papers</ulink> will last until the
      23.05.2010, just like the <ulink url="http://www.froscon.de/index.php?
      id=15&amp;amp;mid=119&amp;amp;ret=15&amp;amp;L=0&amp;amp;L=0">Call For
      Papers</ulink> of the FrOSCon. Please submit a talk, if you got
      something interesting to talk about.</para>
8     <para>You might also want to submit talks to the <ulink url="http://www.
      froscon.de/index.php?id=15&amp;amp;mid=119&amp;amp;ret=15&amp;amp;L=0&
      amp;amp;L=0">main schedule</ulink> of the conference, which also
      accepts PHP related talks. For talks in the main conference your costs
      will be covered as usual, details are on the <ulink url="http://www.
      froscon.de/index.php?id=15&amp;amp;mid=119&amp;amp;ret=15&amp;amp;L=0&
      amp;amp;L=0">dedicated website</ulink>.</para>
9   </section>
10 </article>
```

HTML to RST conversion

```
1 <?php
2
3 require 'autoload.php';
4
5 $xhtml = new ezcDocumentXhtml();
6 $xhtml->setFilters( array(
7     new ezcDocumentXhtmlElementFilter(),
8     new ezcDocumentXhtmlContentLocatorFilter(),
9 ) );
10
11 $xhtml->loadFile( 'whitepaper.html' );
12 // $xhtml->loadFile( 'http://qafoo.com/
13     whitepaper.html' );
14
15 $rst = new ezcDocumentRst();
16 $rst->createFromDocbook( $xhtml->getAsDocbook()
17 );
18
19 echo $rst;
```



HTML to RST conversion

```
1 <?php
2
3 require 'autoload.php';
4
5 $xhtml = new ezcDocumentXhtml();
6 $xhtml->setFilters( array(
7     new ezcDocumentXhtmlElementFilter(),
8     new ezcDocumentXhtmlContentLocatorFilter(),
9 ) );
10
11 $xhtml->loadFile( 'whitepaper.html' );
12 // $xhtml->loadFile( 'http://qafoo.com/
13     whitepaper.html' );
14
15 $rst = new ezcDocumentRst();
16 $rst->createFromDocbook( $xhtml->getAsDocbook()
17 );
18
19 echo $rst;
```



HTML to RST conversion

```
1 <?php
2
3 require 'autoload.php';
4
5 $xhtml = new ezcDocumentXhtml();
6 $xhtml->setFilters( array(
7     new ezcDocumentXhtmlElementFilter(),
8     new ezcDocumentXhtmlContentLocatorFilter(),
9 ) );
10
11 $xhtml->loadFile( 'whitepaper.html' );
12 // $xhtml->loadFile( 'http://qafoo.com/
13     whitepaper.html' );
14
15 $rst = new ezcDocumentRst();
16 $rst->createFromDocbook( $xhtml->getAsDocbook()
17 );
18
19 echo $rst;
```



HTML to RST conversion

```
1 <?php
2
3 require 'autoload.php';
4
5 $xhtml = new ezcDocumentXhtml();
6 $xhtml->setFilters( array(
7     new ezcDocumentXhtmlElementFilter(),
8     new ezcDocumentXhtmlContentLocatorFilter(),
9 ) );
10
11 $xhtml->loadFile( 'whitepaper.html' );
12 // $xhtml->loadFile( 'http://qafoo.com/
13     whitepaper.html' );
14
15 $rst = new ezcDocumentRst();
16 $rst->createFromDocbook( $xhtml->getAsDocbook() );
17
18 echo $rst;
```



HTML to RST conversion

```
1 <?php
2
3 require 'autoload.php';
4
5 $xhtml = new ezcDocumentXhtml();
6 $xhtml->setFilters( array(
7     new ezcDocumentXhtmlElementFilter(),
8     new ezcDocumentXhtmlContentLocatorFilter(),
9 ) );
10
11 $xhtml->loadFile( 'whitepaper.html' );
12 // $xhtml->loadFile( 'http://qafoo.com/
13     whitepaper.html' );
14
15 $rst = new ezcDocumentRst();
16 $rst->createFromDocbook( $xhtml->getAsDocbook()
17     );
18
19 echo $rst;
```



HTML to RST conversion

```
1
2 Your softwares quality matters
3
4
5 Contents
6
7
8 - 'Introduction'__
9   - 'Business goals'__
10  [...]
11
12 Raising and continuously monitoring the quality of software allows you to
13 improve the Return On Investment, to reduce the time to market and to increase
14 customer satisfaction for your software product. No matter if you develop for
15 company internal use only, on a customer basis or standard software.
16
17 .. introduction_
18 .. business-goals_
19 [...]
20
21 .. figure:: /images/whitepaper/stairways_to_quality.png
22    :width: 500
23    :height: 225
24    :alt: Stairways to quality
25
26 Introduction
27
28
29 Quality impacts various areas of your company\'s software development [...]
```

PDF generation

```
1 <?php
2 require 'autoload.php';
3
4 // Convert some web page to PDF
5 $xhtml = new ezcDocumentXhtml();
6 $xhtml->setFilters( array(
7     new ezcDocumentXhtmlElementFilter(),
8     new ezcDocumentXhtmlXPathFilter( '//div[@class="content"]' ),
9 ) );
10 $xhtml->loadFile( 'consulting.html' );
11
12 // Load the docbook document and create a PDF from it
13 $pdf = new ezcDocumentPdf();
14 $pdf->options->errorReporting = E_PARSE | E_ERROR | E_WARNING;
15
16 // Load a custom style sheet
17 $pdf->loadStyles( 'custom.css' );
18
19 // Add a customized header
20 $pdf->registerPdfPart( new ezcDocumentPdfHeaderPdfPart(
21     new ezcDocumentPdfFooterOptions( array(
22         'showPageNumber' => false ,
23         'height'          => '10mm',
24     ) )
25 ) );
26
27 $pdf->createFromDocbook( $xhtml->getAsDocbook() );
28 file_put_contents( __FILE__ . '.pdf', $pdf );
```

PDF generation

```
1 <?php
2 require 'autoload.php';
3
4 // Convert some web page to PDF
5 $xhtml = new ezcDocumentXhtml();
6 $xhtml->setFilters( array(
7     new ezcDocumentXhtmlElementFilter(),
8     new ezcDocumentXhtmlXPathFilter( '//div[@class="content"]' ),
9 ) );
10 $xhtml->loadFile( 'consulting.html' );
11
12 // Load the docbook document and create a PDF from it
13 $pdf = new ezcDocumentPdf();
14 $pdf->options->errorReporting = E_PARSE | E_ERROR | E_WARNING;
15
16 // Load a custom style sheet
17 $pdf->loadStyles( 'custom.css' );
18
19 // Add a customized header
20 $pdf->registerPdfPart( new ezcDocumentPdfHeaderPdfPart(
21     new ezcDocumentPdfFooterOptions( array(
22         'showPageNumber' => false ,
23         'height'          => '10mm',
24     ) )
25 ) );
26
27 $pdf->createFromDocbook( $xhtml->getAsDocbook() );
28 file_put_contents( __FILE__ . '.pdf', $pdf );
```

PDF generation

```
1 <?php
2 require 'autoload.php';
3
4 // Convert some web page to PDF
5 $xhtml = new ezcDocumentXhtml();
6 $xhtml->setFilters( array(
7     new ezcDocumentXhtmlElementFilter(),
8     new ezcDocumentXhtmlXPathFilter( '//div[@class="content"]' ),
9 ) );
10 $xhtml->loadFile( 'consulting.html' );
11
12 // Load the docbook document and create a PDF from it
13 $pdf = new ezcDocumentPdf();
14 $pdf->options->errorReporting = E_PARSE | E_ERROR | E_WARNING;
15
16 // Load a custom style sheet
17 $pdf->loadStyles( 'custom.css' );
18
19 // Add a customized header
20 $pdf->registerPdfPart( new ezcDocumentPdfHeaderPdfPart(
21     new ezcDocumentPdfFooterOptions( array(
22         'showPageNumber' => false ,
23         'height'          => '10mm',
24     ) )
25 ) );
26
27 $pdf->createFromDocbook( $xhtml->getAsDocbook() );
28 file_put_contents( __FILE__ . '.pdf', $pdf );
```


PDF generation

```
1 <?php
2 require 'autoload.php';
3
4 // Convert some web page to PDF
5 $xhtml = new ezcDocumentXhtml();
6 $xhtml->setFilters( array(
7     new ezcDocumentXhtmlElementFilter(),
8     new ezcDocumentXhtmlXPathFilter( '//div[@class="content"]' ),
9 ) );
10 $xhtml->loadFile( 'consulting.html' );
11
12 // Load the docbook document and create a PDF from it
13 $pdf = new ezcDocumentPdf();
14 $pdf->options->errorReporting = E_PARSE | E_ERROR | E_WARNING;
15
16 // Load a custom style sheet
17 $pdf->loadStyles( 'custom.css' );
18
19 // Add a customized header
20 $pdf->registerPdfPart( new ezcDocumentPdfHeaderPdfPart(
21     new ezcDocumentPdfFooterOptions( array(
22         'showPageNumber' => false ,
23         'height'          => '10mm',
24     ) )
25 ) );
26
27 $pdf->createFromDocbook( $xhtml->getAsDocbook() );
28 file_put_contents( __FILE__ . '.pdf', $pdf );
```

PDF generation

```
1 <?php
2 require 'autoload.php';
3
4 // Convert some web page to PDF
5 $xhtml = new ezcDocumentXhtml();
6 $xhtml->setFilters( array(
7     new ezcDocumentXhtmlElementFilter(),
8     new ezcDocumentXhtmlXPathFilter( '//div[@class="content"]' ),
9 ) );
10 $xhtml->loadFile( 'consulting.html' );
11
12 // Load the docbook document and create a PDF from it
13 $pdf = new ezcDocumentPdf();
14 $pdf->options->errorReporting = E_PARSE | E_ERROR | E_WARNING;
15
16 // Load a custom style sheet
17 $pdf->loadStyles( 'custom.css' );
18
19 // Add a customized header
20 $pdf->registerPdfPart( new ezcDocumentPdfHeaderPdfPart(
21     new ezcDocumentPdfFooterOptions( array(
22         'showPageNumber' => false ,
23         'height'          => '10mm' ,
24     ) )
25 ) );
26
27 $pdf->createFromDocbook( $xhtml->getAsDocbook() );
28 file_put_contents( __FILE__ . '.pdf', $pdf );
```

PDF generation

```
1 <?php
2 require 'autoload.php';
3
4 // Convert some web page to PDF
5 $xhtml = new ezcDocumentXhtml();
6 $xhtml->setFilters( array(
7     new ezcDocumentXhtmlElementFilter(),
8     new ezcDocumentXhtmlXPathFilter( '//div[@class="content"]' ),
9 ) );
10 $xhtml->loadFile( 'consulting.html' );
11
12 // Load the docbook document and create a PDF from it
13 $pdf = new ezcDocumentPdf();
14 $pdf->options->errorReporting = E_PARSE | E_ERROR | E_WARNING;
15
16 // Load a custom style sheet
17 $pdf->loadStyles( 'custom.css' );
18
19 // Add a customized header
20 $pdf->registerPdfPart( new ezcDocumentPdfHeaderPdfPart(
21     new ezcDocumentPdfFooterOptions( array(
22         'showPageNumber' => false ,
23         'height'         => '10mm' ,
24     ) )
25 ) );
26
27 $pdf->createFromDocbook( $xhtml->getAsDocbook() );
28 file_put_contents( __FILE__ . '.pdf', $pdf );
```

PDF generation

```
1 article {
2     font-family: "sans-serif";
3     font-size: "10pt";
4 }
5
6 title {
7     font-family: "sans-serif";
8     color: #97BF0D;
9
10    border-bottom: 1px solid #555753;
11    border-left: 4px solid #555753;
12    padding-left: 4px;
13 }
14
15 section > section > section > section > title {
16     border-color: #babdb6;
17 }
18
19 page {
20     padding: "15mm 30mm";
21 }
22
23 ulink {
24     color: #97BF0D;
25 }
26
27 link {
28     color: #97BF0D;
29 }
```



PDF generation



Consulting - Qafoo GmbH - passion for software quality

File Edit View Go Help

Previous Next 1 of 31 Read All

Consulting - Qafoo GmbH - passion for software quality

Consulting

The Qafoo consulting team of highly experienced developers, software architects and designers is constantly available to assist you with a broad field of tasks around high-quality IT development. We join your team on-site in order to discuss critical decisions, analyze your environment and help you to implement changes in direction of high-quality PHP. Our consulting services are partitioned in three essential areas. However, if you have individual requirements, do not hesitate to contact Qafoo for an individual solution. Our goal is your software quality.

Quality processes

Establishing efficient, smoothly running and comprehensive processes in order to raise software quality is a crucial task development projects. From the analysis through out designing and initial process implementation, Qafoo consultants are available to your assistance in this area.

Process analysis

The initial step for designing new quality process is an in-depth analysis of the current environment. Our consulting team is eager to visit you on-site in order to analyze existing processes in your project. We will help you to determine a clear goal to be reached and to identify the bottlenecks in your current project environment, which are to be tackled, in order to achieve a maximum quality increase with minimum effort.

Qafoo experts can provide you consulting in fluent English and German.

A Qafoo process analysis does not only yield to the current state of quality processes in your project. We also identify spots in your processes which can be changed in order to enhance existing processes and recommend changes to you, which will yield the maximum benefit. The Qafoo consultants will not only observe your project, but get into active discussion with your team, because your co-workers are the important parts in any process.

Process design

Once your environment has been analyzed by a Qafoo expert, we can recommend you an individual design for your future quality process. This does not only include the spots in your current processes, which are recommended to be changed but also detailed information on the way through these changes can be achieved and how you see room needed to tackle a change.

Our consultants will also be available for discussion with you and your team in order to have

ODT generation

```
1 <?php
2
3 require 'autoload.php';
4
5 // Convert some input RSTfile to docbook
6 $xhtml = new ezcDocumentXhtml();
7 $xhtml->setFilters( array(
8     new ezcDocumentXhtmlElementFilter(),
9     new ezcDocumentXhtmlXPathFilter( '//div[@class="content"]' ),
10 ) );
11 $xhtml->loadFile( 'consulting.html' );
12
13 $converter = new ezcDocumentDocbookToOdtConverter();
14 $converter->options->styler->addStylesheetFile( 'custom.css' );
15
16 $odt = $converter->convert( $xhtml->getAsDocbook() );
17 file_put_contents( __FILE__ . '.fodt', $odt );
```

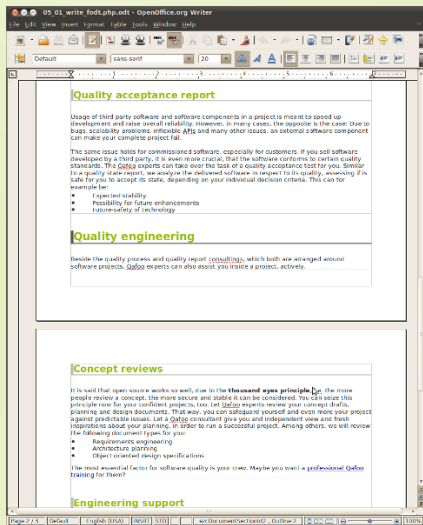
ODT generation

```
1 <?php
2
3 require 'autoload.php';
4
5 // Convert some input RSTfile to docbook
6 $xhtml = new ezcDocumentXhtml();
7 $xhtml->setFilters( array(
8     new ezcDocumentXhtmlElementFilter(),
9     new ezcDocumentXhtmlXPathFilter( '//div[@class="content"]' ),
10 ) );
11 $xhtml->loadFile( 'consulting.html' );
12
13 $converter = new ezcDocumentDocbookToOdtConverter();
14 $converter->options->styler->addStylesheetFile( 'custom.css' );
15
16 $odt = $converter->convert( $xhtml->getAsDocbook() );
17 file_put_contents( __FILE__ . '.fodt', $odt );
```

ODT generation

```
1 <?php
2
3 require 'autoload.php';
4
5 // Convert some input RSTfile to docbook
6 $xhtml = new ezcDocumentXhtml();
7 $xhtml->setFilters( array(
8     new ezcDocumentXhtmlElementFilter(),
9     new ezcDocumentXhtmlXPathFilter( '//div[@class="content"]' ),
10 ) );
11 $xhtml->loadFile( 'consulting.html' );
12
13 $converter = new ezcDocumentDocbookToOdtConverter();
14 $converter->options->styler->addStylesheetFile( 'custom.css' );
15
16 $sdt = $converter->convert( $xhtml->getAsDocbook() );
17 file_put_contents( __FILE__ . '.fodt', $sdt );
```


ODT generation



Outline

Introduction

The Document component

Getting into the code

End



Thanks for listening

- ▶ Apache Zeta Components
 - ▶ <http://zetacomponents.org>
 - ▶ #zetacomponents @ Freenode
- ▶ Stay in touch
 - ▶ toby@qafoo.com
 - ▶ @tobySen
- ▶ Take care for your software quality
 - ▶ <http://qafoo.com>
 - ▶ Also consulting / training / support for Zeta