

Any to any: Convert your documents

International PHP Conference 2010

Tobias Schlitt <toby@qafoo.com>
Kore Nordmann <kore@qafoo.com>

October 13, 2010



About us

- ▶ Long time FLOSS developers
- ▶ Open source enthusiast
- ▶ Authors, conference speakers
- ▶ PMC members of Apache Zeta Components
- ▶ Recently founded Qafoo GmbH - passion for software quality
 - ▶ We provide commercial support for Apache Zeta Components

Outline

Introduction

The Document component

Getting into the code

End



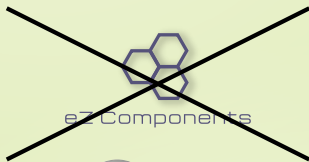
eZ Components



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

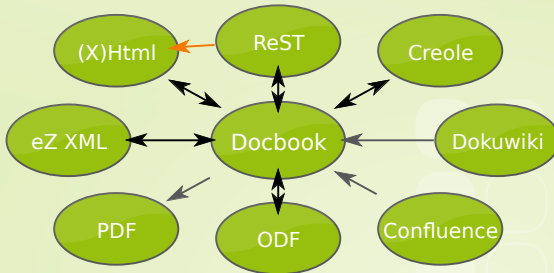
- ▶ Applications need to deal with
 - ▶ Different input mechanisms
 - ▶ WYSIWYG editor (HTML)
 - ▶ Simple text editor (wiki markup)
 - ▶ Emails (ReST)
 - ▶ Different output formats
 - ▶ Web front end (HTML)
 - ▶ Technical documentation management (Docbook)
 - ▶ Print (PDF)
- ▶ The Document component 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 PHP @ FrOSCon 2010  
3 =====  
4
```

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

```
10  
11 We are currently looking for talks for the PHP room, and the 'Call For  
12 Papers'__ will last until the 23.05.2010, just like the 'Call For Papers'__ of  
13 the FrOSCon. Please submit a talk, if you got something interesting to talk  
14 about.
```

```
15  
16 You might also want to submit talks to the 'main schedule'__ of the  
17 conference, which also accepts PHP related talks. For talks in the main  
18 conference your costs will be covered as usual, details are on the 'dedicated  
19 website'__.
```

```
20  
21 -- http://phpugdo.de/  
22 -- http://froscon.de/  
23 -- /call\_for\_papers.html  
24 -- http://www.froscon.de/index.php?id=15&mid=119&ret=15&L=0&L=0  
25 -- http://www.froscon.de/index.php?id=15&mid=119&ret=15&L=0&L=0  
26 -- 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 @ FosCon 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> (FosCon) 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 FosCon. 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 

---


3 Your softwares quality matters
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



The screenshot shows a web browser window with the title "Consulting - Qafoo GmbH - passion for software quality". The browser's address bar shows "http://www.qafoo.com/". The page content is a PDF document with the following sections:

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 IT. 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 comprehensible 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 spend the minimum amount. 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 $sodt = $converter->convert( $xhtml->getAsDocbook() );
17 file_put_contents( __FILE__ . '.fodt', $sodt );
```

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 $sodt = $converter->convert( $xhtml->getAsDocbook() );
17 file_put_contents( __FILE__ . '.fodt', $sodt );
```

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 $sodt = $converter->convert( $xhtml->getAsDocbook() );
17 file_put_contents( __FILE__ . '.fodt', $sodt );
```

ODT generation

05_01_write_foed.php.odt - OpenOffice.org Writer

File Edit View Insert Format Table Tools Window Help

Default Sans-Serif 10

Quality acceptance report

Usage of third party software and software components in a project is meant to speed up development and raise overall reliability. However, in many cases, the opposite is the case: Due to bugs, scalability problems, inflexible APIs and many other issues, an external software component can make your complete project fail.

The same issue holds for commissioned software, especially for customers. If you sell software developed by a third party, it is even more crucial, that the software conforms to certain quality standards. The Qafoo experts can take over the tasks of a quality acceptance test for you. Similar to a quality state report, we analyze the delivered software in respect to its quality, assessing if it is safe for you to accept its state, depending on your individual decision criteria. This can for example be:

- Expected stability
- Possibility for future enhancements
- Future-safety of technology

Quality engineering

Beside the quality process and quality report (systems), which both are arranged around software projects, Qafoo experts can also assist you inside a project, actively.

Concept reviews

It is said that open source works so well, due to the **thousand eyes principle**. The more people review a concept, the more secure and stable it can be considered. You can seize this principle now for your confident projects, too. Let Qafoo experts review your concept drafts, planning and design documents. That way, you can safeguard yourself and even more your project against predictable issues. Let a Qafoo consultant give you an independent view and fresh insights about your planning, in order to run a successful project. Among others, we will review the following document types for you:

- Requirements engineering
- Architecture planning
- Object-oriented design specifications

The most essential factor for software quality is your **crew**. Maybe you want a professional Qafoo training for them?

Engineering support

Page 2 / 3 Default English (USA) NORS S10 von der comnet.net GmbH © 2008

Outline

Introduction

The Document component

Getting into the code

End



Thanks for listening

Are there any questions left?



Thanks for listening

Please rate this talk:
<http://joind.in/2168>



Thanks for listening

Please rate this talk:
<http://joind.in/2168>

Stay in touch

- ▶ Tobias Schlitt
- ▶ toby@qafoo.com
- ▶ @tobySen
- ▶ Kore Nordmann
- ▶ kore@qafoo.com
- ▶ @koredn

Need help with Apache Zeta Components? Rent us!