

Annotate

Dominik Brettnacher

Annotate

Dominik Brettnacher

Published September 5, 2005

Table of Contents

What is Annotate?	viii
Features	viii
1. How to install Annotate	1
Prerequisites	1
Requirements on the authoring station	1
Requirements on the web server	1
Getting the software	2
Configuring Annotate	2
Converting your document	3
Configuring the CGI part	3
Check the result	4
2. How Annotate works	5
Introduction	5
3. The annotate.conf configuration file	6
Introduction	6
The right place for the configuration file	6
The XSL part	7
The CGI part	7
A. Example configuration file	10
B. Further reading	11
DocBook	11
XML, XSL and XSLT	11
Perl	11
Related projects	11
Other	11

List of Figures

1. How Annotate Works	viii
-----------------------------	------

List of Tables

3.1. CGI part configuration parameters	8
--	---

List of Examples

1.1. Installing the DBI module with CPAN	2
1.2. Converting a document with xsltproc	3
1.3. Converting a document with Saxon	3
3.1. Stylesheet configuration	7
3.2. Stylesheet configuration for chunked output	7
3.3. Configuration of the annotatable elements	7
3.4. Making only chapters annotatable	7
3.5. DBI configuration using MySQL	8
3.6. DBI configuration using CSV files	8

What is Annotate?

Annotate is a system that provides an annotation facility for DocBook [<http://www.docbook.org/>] documents. Annotate enables visitors of an online version of a DocBook document to add comments to any paragraph or chapter of the document.

Annotate extends the DocBook XSL stylesheets [<http://docbook.sourceforge.net/projects/xsl/>], leading to a modified HTML output which contains anchors at those places where comments can be made. Comments and notes are stored in a DBMS. A CGI program then merges the DocBook document and the comments to produce the output for the visitor.

Figure 1. How Annotate Works

Comments can be made similar to, for example, the online version of the PHP manual [<http://www.php.net/manual/en>]. The Annotate interface can be customized: a user can decide which parts of the DocBook document (e.g. figures, chapters, sections or paragraphs) can be annotated.

You can find an online demonstration of Annotate at the online version of this document at <http://www.brettnacher.org/users/dominik/annotate/>. [<http://www.brettnacher.org/users/dominik/annotate/>]

Features

Annotate provides the following:

- A modified XSL stylesheet, based on the DocBook XSL stylesheets by Norman Walsh.
- The possibility to decide which DocBook parts may be annotated. Each paragraph may be annotated, or only each chapter as a whole.
- Different possibilities to show the visitors' comments. Comments can be displayed in-line with the original DocBook document, they can be shown on a separate page or they can be left out.
- E-Mail notification for new comments.
- An interface for the Administrator to manage existing comments.
- An easily modifiable look and feel through the use of templates.
- Support for a variety of database backends (such as MySQL or SQLite) using DBI [<http://search.cpan.org/~timb/DBI/DBI.pm>].
- An interactive perl script for easy installation.

Chapter 1. How to install Annotate

Prerequisites

Annotate does two things:

1. It converts your DocBook document into modified HTML. This can be done on any machine, for example on a workstation where the document itself is edited. In the following, this machine is called the *authoring station* (see also the upper part of Figure 1, “How Annotate Works”).
2. It merges the HTML documents and the comments into a new HTML document. This is done on the *web server* (see also the lower part of Figure 1, “How Annotate Works”).

The authoring station and the webserver can be the same machine, but they can also be separate machines. Their requirements of course differ.

Requirements on the authoring station

- You need an XSLT processor (such as Saxon [<http://saxon.sourceforge.net/>] or xsltproc [<http://xmlsoft.org/XSLT/>]).
- You need the DocBook XSL style sheets [<http://docbook.sourceforge.net/projects/xsl/>].
- Of course you need a valid DocBook XML document to do the actual annotation.

Requirements on the web server

- The web server must support the CGI interface and the `PATH_INFO` variable. Apache [??] will do just fine, for example. Microsoft IIS however does not seem to support `PATH_INFO`, so Annotate will not run with this web server.
- Perl [<http://www.perl.com>] 5 or higher
- A few CPAN [<http://www.cpan.org>] modules. CPAN, the Comprehensive Perl Archive Network is an archive of software modules written in Perl. Annotate's installation script will check the availability of these modules. The modules needed are the following:
 - CGI
 - Text::Wrap
 - HTML::Template
 - HTML::Parser
 - MIME::Lite
 - XML::Simple
 - DBI

- a DBD driver for the database to be used. You can use either of these
 - DBD::Mysql [<http://search.cpan.org/dist/DBD-mysql/lib/DBD/mysql.pm>] for MySQL [<http://www.mysql.com>]
 - DBD::Pg [<http://search.cpan.org/dist/DBD-Pg/Pg.pm>] for PostgreSQL [<http://www.postgres.org>]
 - DBD::SQLite [<http://search.cpan.org/dist/DBD-SQLite/lib/DBD/SQLite.pm>] for SQLite [<http://www.sqlite.org>]
 - or any other Perl DBD module [<http://www.cpan.org/modules/by-module/DBD/>]

Note

Some of these modules depend on additional modules. If you use the **install** facility of the CPAN module (or the packaging system of your operating system), the dependencies will be handled automatically:

Example 1.1. Installing the DBI module with CPAN

```
perl -MCPAN -e 'install DBI'
```

- Any database system which is supported by the Perl DBI [<http://dbi.perl.org/>], for example MySQL or Postgres.

If you do not have a RDBMS system at hand (or do not want to use one), you should use the DBD::SQLite [<http://search.cpan.org/dist/DBD-SQLite/lib/DBD/SQLite.pm>] module which provides a self-contained embedded SQL database.

Getting the software

If you have not already done so, you will have to download Annotate. You can find the latest version at <http://www.brettnacher.org/users/dominik/annotate/>. Unpack the downloaded archive. A directory named `annotate-0.1` will be created which contains the Annotate distribution.

Configuring Annotate

Annotate relies on a configuration file, `annotate.conf`, for everything that is supposed to be customized by the user. To generate an initial configuration, Annotate is shipped with an interactive Perl script. It will ask you a few questions about the most important things to configure and will create a working configuration. You can decide whether you want to configure Annotate for the authoring station, for the web server or for both.

So the first thing to do is to run the installation script:

```
perl install.pl
```

Put the resulting file (`annotate.conf`) into the `/usr/local/etc` directory. If you cannot put it there, please refer to the section called “The right place for the configuration file” for details.

For a complete list of configuration options, please have a look at Chapter 3, *The annotate.conf configuration file*.

Converting your document

The configuration file that you have just created will serve as the new stylesheet for your DocBook document. So basically everything you have to do is to instruct your XSLT processor to convert your document using this new stylesheet. Here are some examples how to do this:

Example 1.2. Converting a document with xsltproc

```
xsltproc /usr/local/etc/annotate.conf example.xml
```

Example 1.3. Converting a document with Saxon

```
java -jar saxon.jar /usr/local/etc/annotate.conf example.xml
```

Note

The file `saxon.jar` must be in your CLASSPATH.

The output of the XSLT processor is one or several HTML files, depending on whether you chose to produce chunked output or not. These HTML files now have to be moved to the web server. The CGI part expects them in the directory you configured as the document root.

Note

Although the concept is the same, the `DocumentRoot` of Annotate is probably a different directory than the `DocumentRoot` of your web server.

Configuring the CGI part

The Annotate distribution contains a directory named `cgi-bin`. The contents of this directory should be moved to a location where the web server is going to treat them as CGI scripts, i.e. it should execute them. Make sure that the `annotate` file is executable.

Note

Maybe you have to adjust the first line of `annotate`

```
#!/usr/bin/perl
```

to point to your Perl binary.

I assume that you already have a working RDBMS system at hand. You have to create a table with the columns described below.

If possible, the installation script will ask you whether it should create the table automatically.

Check the result

FIXME

In the new directory, you will find a subdirectory named `cgi-bin`. Move the contents of this directory to a location where the contents can be executed as CGI by your web server. Make sure that every file in there is readable and that the `annotate` file is executable.

Chapter 2. How Annotate works

Introduction

Annotate gives your readers the possibility to add online comments (*annotations*) to paragraphs, chapters, and other parts of your documents published on the web. In order to achieve this, Annotate's work consists of two steps:

1. Annotate extends the DocBook XSL style sheets so that the exported HTML document contains anchors at every point where an annotation should be possible.
2. The extended HTML file is then processed by a CGI script when a document is requested by a visitor. The anchors are replaced with the respective annotations already made and with links that allow the visitor to add further annotations.

Chapter 3. The `annotate.conf` configuration file

Introduction

The `annotate.conf` configuration file controls nearly every aspect of Annotate. Features you can control from the configuration file include:

- which elements of your DocBook document are annotatable.
- where the annotations shall be displayed.
- where the annotations shall be stored.

The configuration file itself is a valid XSL stylesheet. It imports the original DocBook XSL stylesheet, making a few changes. For convenience, the run-time configuration for the Annotate CGI scripts is also made in this file. Because it serves two purposes, the configuration file consists of two parts:

1. The first part (The XSL part) deals with the XSL processing, defines templates for annotation and says which parts of your document may be annotated.
2. The second part (The CGI part) controls how the Annotate CGI scripts work.

You can find an example configuration file in ???.

The right place for the configuration file

When it comes to the right location for the configuration file, you are basically free to choose one. You have three possibilities:

- Per default, Annotate will look at `/usr/local/etc/annotate.conf`.
- If you have access to the environment of your CGI scripts, you can set the environment variable `ANNOTATE_CONF` to point to the configuration file. This can be done with Apache's `SetEnv` [http://httpd.apache.org/docs/1.3/mod/mod_env.html#setenv] directive in `httpd.conf` (or similar configuration files):

`SetEnv ANNOTATE_CONF "/path/to/annotate.conf"`

- The third possibility is to set the path directly in `ConfigFile.pm`. If neither of the above methods work for you, this is the way to go.

Warning

Do not put the configuration file to a location that is readable from the outside. The configuration file will include details about your setup (i.e. a password for the database) which you most probably do not want to disclose. So be careful.

The XSL part

The first two sections of the configuration file tell the XSLT processor where to find the DocBook XSL stylesheets (these are imported) and which parts of your documents should be marked with comment anchors.

The most important element is the `xsl:import`: with its `href` attribute, you tell the XSLT processor the path to your DocBook XSL style sheet. You can choose chunked output here if you want.

Example 3.1. Stylesheet configuration

```
<xsl:import
  href="/usr/local/share/xsl/docbook/html/docbook.xsl"/>
```

Example 3.2. Stylesheet configuration for chunked output

```
<xsl:import
  href="/usr/local/share/xsl/docbook/html/chunk.xsl"/>
```

The other thing to be configured is the list of elements that shall be annotatable. To change them, change the value of the `match` attribute, separating the tag names with a `"|"` character:

Example 3.3. Configuration of the annotatable elements

```
<xsl:template
  match="para|simpara|programlisting|note|important|warning|caution|tip">
  <xsl:apply-imports /> <xsl:call-template
    name="annotate"/> </xsl:template>
```

Example 3.4. Making only chapters annotatable

```
<xsl:template
  match="chapter">
  <xsl:apply-imports /> <xsl:call-template
    name="annotate"/> </xsl:template>
```

The CGI part

The third section of the configuration file configures Annotate's CGI script. This part of the configuraton

decides where your document and its comments are stored and what the comments should look like. The following is a list of all directives:

Table 3.1. CGI part configuration parameters

Parameter	Description
DocumentRoot	Like the Apache directive, this specifies the directory where Annotate can find your DocBook document. This path can be given relative or absolute.
TemplateRoot	This should point to the directory containing the templates.
NavigationTemplate	Points to the template describing the navigation bar which is displayed at bottom and top of each delivered page.
LinkTemplate	Points to the template describing the links which are displayed at the comment anchors. They contain links to the "View comment" and "Write new comment" pages.
AnnotateTemplate	Points to the template containing the entry form for a new comment.
ViewTemplate	Points to the template which displays one or more comments.
MailTemplate	Points to the template which contains the body of a notification about a new comment.
Database	<p>This is the data source string for DBI. The database is used to store the comments of your visitors. Any database which is supported by Perl DBI can be used. Please refer to the Perl documentation [http://search.cpan.org/~timb/DBI/DBI.pm] for more information.</p> <p>Example 3.5. DBI configuration using MySQL</p> <pre>DBI:mysql:databasename:hostname[:port]</pre> <p>Example 3.6. DBI configuration using CSV files</p> <pre>DBI:CSV:f_dir=comments</pre>
DatabaseUsername	Contains the user name (i.e. login) for the database. Can be left empty if not needed.
DatabasePassword	Contains the password for the database. Can be left empty if not needed.

Parameter	Description
DefaultOutputMode	<p>Controls how comments are displayed by default.</p> <p>Valid modes are:</p> <p>Hide Do not show comments.</p> <p>Inline Display comments inline, directly below the element they belong to.</p> <p>Link Create links which display the comments in a new window.</p> <p>Each visitor can choose his favorite mode, which is automatically stored in a cookie by Annotate. The default value is therefore only used if the user visits the document for the first time or there is no cookie.</p>
eMailAddress	<p>This can be set to one (or more, separated by comma) eMail address. A notification will be sent to this address if a new comment is made to any part of the document.</p> <p>If left empty, no mail is sent.</p>
eMailSubject	<p>The subject of the notification eMail.</p>

Appendix A. Example configuration file

FIXME

Appendix B. Further reading

DocBook

- DocBook [<http://www.docbook.org/>]
- DocBook: The Definitive Guide [<http://www.docbook.org/tdg/>]
- DocBook XSL: The Complete Guide [<http://www.sagehill.net/book-description.html>]
- docbook-apps mailinglist [<http://lists.oasis-open.org/archives/docbook-apps/>]

XML, XSL and XSLT

- The XML Bible [<http://www.ibiblio.org/xml/books/bible/>]
- XSL Transformations (XSLT) [<http://www.w3.org/TR/xslt>]
- libxslt / xsltproc [<http://xmlsoft.org/XSLT/>]
- Saxon [<http://saxon.sourceforge.net/>]

Perl

- Perl [<http://www.perl.com>]
- Comprehensive Perl Archive Network [<http://www.cpan.org>]

Related projects

- Annotea Project [<http://www.w3.org/2001/Annotea/>]
- Annotatio [<http://sourceforge.net/projects/annotatio/>]

Other

-