

NAME

`bib2xhtml` – BibTeX bibliography to XHTML converter

SYNOPSIS

bib2xhtml [**-a**] [**-b** *bibtex-options*] [**-B** *bibtex-executable*] [**-c**] [**-D** *mappings*] [**-d** *delimiter*] [**-e** *extended-information*] [**-h** *heading*] [**-i**] [**-k**] [**-m** *macro-file*] [**-n** *author-name*] [**-r**] [**-R**] [**-s** *style*] [**-t**] [**-u**] [**-U**] [**-v**] *sourcefile* [*htmlfile*]

DESCRIPTION

bib2xhtml converts a bibtex-format bibliography file to XHTML. *sourcefile* may be either a BibTeX-format bibliography (**.bib**) file or a LaTeX-formatted bibliography auxillary file (**.aux**). *htmlfile* will often be just part of a larger XHTML file; **bib2xhtml** attempts to update the XHTML file in place. It does this by searching *htmlfile* for two special strings, which delimit the bibliography part of the file. These strings are the XHTML comments

```
<!-- BEGIN BIBLIOGRAPHY delimiter -->
```

```
<!-- END BIBLIOGRAPHY delimiter -->
```

If these delimiters are found, the old bibliography between them is replaced with the new bibliography. Otherwise the new bibliography, with the appropriate delimiters, is appended to the end of *htmlfile*, and it will be necessary to move it to the proper location by hand. Note that it is possible for *htmlfile* to contain any number of bibliographies, and they may be updated independently. If no *htmlfile* is specified, the bibliography is written to standard output.

The tag of each BibTeX bibliography entry is converted into a name anchor (``) in *htmlfile*, and may be used to construct a link to this particular entry.

By default the contents of the bibliography are determined by the contents of the *sourcefile*. It is also possible to specify the contents of the bibliography in the *htmlfile* itself by setting the **-i** option and using two special strings to delimit the citations to be used for generating the bibliography. These strings are the XHTML comments

```
<!-- BEGIN CITATIONS delimiter -->
```

```
<!-- END CITATIONS delimiter -->
```

If these delimiters are found, the contents between them are searched for XHTML comments containing the citations (one citation per line), for example

```
<!-- BEGIN CITATIONS delimiter -->
```

```
<!-- \citation{knuth:1997}
```

```
\citation{knuth:1984} -->
```

```
<!-- END CITATIONS delimiter -->
```

These citations are then used for choosing the bibliography entries from the *sourcefile*.

Special Field Processing

Several BibTeX fields are treated specially. The filenames specified in the **postscript**, **pdf**, and **dvi** fields are relative to the *htmlfile* directory. In addition, **bib2xhtml** recognizes compressed or gzipped files (those with filenames ending in **.Z**, **.gz**, or **.z**).

url If the **url** field exists, a link (“A” tag) is created. The destination of the link is the **url** tag’s value. The bibliography entry’s title is the source of the link. This can be used to create a link to a XHTML version of the paper, or to a version that is available by ftp.

mailto If the **mailto** field exists, a link (“A” tag) is created. The destination of the link is the **mailto** tag’s value. The bibliography entry’s author name is the source of the link.

postscript

If the **postscript** field exists, its value is used as the name of a PostScript file. **bib2xhtml** creates a link to the file, and updates *htmlfile* with the number of pages and size of the PostScript file.

- doi** If the **doi** field exists, its value is used as the designation of a digital object identifier. **bib2xhtml** creates a link to the object through <http://dx.doi.org> and updates *htmlfile* with the **doi** as a hyperlink. The **doi** should normally be in a canonical form like the following: *doi:10.1145/363347.363387*. However, if the **doi** designator is formed as an HTTP hyperlink, it will be converted to the canonical form by replacing the HTTP protocol designation and the host name of the hyperlink with the *doi:* literal string.
- pdf** If the **pdf** field exists, its value is used as the name of a PDF (portable document format) file. **bib2xhtml** creates a link to the file and updates *htmlfile* with the size of the PDF file.
- dvi** If the **dvi** field exists, its value is used as the name of a TeX DVI file. **bib2xhtml** creates a link to the file and updates *htmlfile* with the number of pages and size of the DVI file.
- author, keywords**
The full **author** and **keywords** fields of each BibTeX entry are included in *htmlfile* as XHTML comments, and may be used by CGI scripts (such as **bibsearch**, which is supplied) to select only certain entries.

Hints

bib2xhtml's LaTeX to XHTML conversion is not very smart. If something isn't being converted correctly, try doing it in a different way.

bib2xhtml does not currently understand `\verb`. To get a tilde into a filename or an embedded URL, use `\{ }`. On the other hand, entries containing *URL* fields or `\url{ }` commands, are correctly converted into hyperlinks. Leave such hyperlinks as they are; do not embed any LaTeX commands in them.

The lists that **bib2xhtml** generates contain CSS hooks of the class *bib2xhtml*. You can therefore modify the appearance of the lists by adding style entries like the following in the web page's header.

```
<style type="text/css">
    dl.bib2xhtml dd { padding-bottom: 1em }
</style>

<style type="text/css">
    ul.bib2xhtml li { padding-bottom: 1em }
</style>
```

The value of the named attributes can contain arbitrary HTML markup commands, and can thus be used for substituting the text with icons. For example, the following Unix command can be used for creating stand-alone PDF icons.

```
perl bib2xhtml -e 'nosize,nopages,PDF:' example.bib example.html
```

Under the Windows shell the corresponding command would be.

```
perl bib2xhtml -e "nosize,nopages,PDF:<img src='pdficon_small.gif'
alt='PDF' border='0' />" example.bib example.html
```

A couple of PDF icons and the conditions for their use are available at <http://www.adobe.com/misc/linking.html#pdficon>.

To display the number of pages in local PDF files, the Perl module *PDF::API2* must be installed on the system where **bib2xhtml** is running. A command like the following will accomplish this.

```
perl -MCPAN -e "install PDF::API2"
```

OPTIONS

- a** Include the BibTeX file's abstract field in the bibliography.
- b *bibtex-options***
Provide the specified options to the BibTeX program.
- B *bibtex-executable***
Specify the name or path of the bibtex executable program. By default this is *bibtex*. In some installations specifying *bibtex8* can prevent out of memory conditions.

- c** Sort chronologically. Normally, the bibliography is ordered first by author and then by date. With this option, the bibliography is ordered by year, month, day, and then by author.
 - d delimiter**
Specify the delimiter that sets the bibliography off from the rest of the XHTML file. The default delimiter is the name of the bibliography file with the **.bib** extension stripped.
 - D mappings**
Define URL to directory mappings. The *mappings* argument contains a sequence of comma-separated values. Each value contains a local file path and the corresponding URL, separated by an @ sign. The resulting URL is used for linking the various paper file format types.
 - e extended-information**
Specify how an entry's extended information will be displayed. The extended information is specified as a series of comma-separated attributes. The following plain attributes are supported.
 - notype** Do not display the file's type (PDF, Postscript, DVI).
 - nosize** Do not display the file's size in bytes.
 - nopages**
Do not display the file's number of pages.
 - nocompression**
Do not display the file's compression type.
 - nodoi** Do not display the file's DOI.
 - nobrackets**
Do not enclose the DOI link or the file's extended information in brackets.
- A number of valued attributes allow the specification of the text to display for various file types. These are written as a *name:value* pair. The following names are supported for valued attributes.
- PostScript
 - PDF
 - DVI
 - DOI
- h heading** Use the string *heading* instead of the default title when creating a new *htmlfile*. If updating an existing *htmlfile*, this option is ignored.
 - i** The citations are included in the *htmlfile*. When this option is set, the *sourcefile* must be a .bib file.
 - k** In labeled styles (**alpha**, **named**, **plain**, and **unsort**) append to the label of each entry its BibTeX key. Thus, the label of each entry will consist of the original label, followed by an em-dash, followed by the BibTeX key.
 - m macro file**
Specify a BibTeX macro file to supply to BibTeX when processing the bibliography file. Macro files typically contain local abbreviations and other macro definitions.
 - n author name**
Specify an author name to highlight. Any bibliography entry author names that match the name specified as a regular expression will be set in HTML *strong* tags and will therefore be displayed in a bold (or similar) font.
 - r** Reverse the chronological sorting order. Normally, the bibliography is sorted (by year; BibTeX ignores the month) from the earliest entry to the latest entry. With this option, the order is reversed. Note that to obtain a bibliography ordered by chronological order you must also specify the **-c** option.
 - R** Reference numbers increase from bottom to top, not from top to bottom. This option is useful, for example, in the following situation. When one keeps one's own publication list over many years

or decades, one often prefers that (i) the new ones appear on top of the list, to give them more immediate attention, and (ii) reference numbers, once assigned, do not change, so as not to mess up old documents that use reference numbers. Using options **-c -r** and **-R** together satisfies this need.

- s style** Create a *style*-style bibliography. Supported styles are **empty** (the default), **plain**, **alpha**, **named**, **paragraph**, **unsort**, and **unsortlist**. An **empty** bibliography is a bulleted list. A **plain** bibliography is a numbered list. **paragraph** produces just paragraphs (no bullets). **unsort** is like **plain** except that the entries in the XHTML file are in the same order as as they are in the source file. **unsortlist** is like **unsort** but entry labels are bullets instead of numbers. An **alpha** bibliography has labels that are in the BibTeX alpha style. A **named** bibliography has labels of the form [name, year].
- t** Write a timestamp with the date at which the bibliography was updated to *htmlfile*.
- u** Convert LaTeX special characters into the corresponding Unicode characters, and output XML coded in UTF-8. The output produced by this option may be easier to parse and validate with some XML parsers. By default LaTeX special characters are converted into HTML character entities.
- U** Treat input file as Unicode-coded document.
- v** Report version number on stderr.

FILES

html-a.bst	alpha XHTML BibTeX style file
html-n.bst	named XHTML BibTeX style file
html-u.bst	unsort XHTML BibTeX style file
html-aa.bst, html-na.bst, html-ua.bst	versions of the above style files with abstracts
bibsearch	a CGI script for performing bibliography searches

SEE ALSO

perl(L), **bibtex(L)**.

BUGS

The LaTeX to XHTML translation, while decent, is not perfect.

Requires bibtex and perl. However, I think that most sites that would want to convert BibTeX to XHTML will already have both programs installed.

Relies on dviselect to count the number of pages in a DVI file.

AUTHORS

David Hull while at the University of Illinois at Urbana-Champaign.

Diomidis Spinellis (*dds@aub.gr*), Athens University of Economics and Business.

HISTORY

The program was originally written as **bib2html** by David Hull in 1996, who maintained it until 1998 (version 1.33). In 2002, due to the lack of visible updates on the web, the program was adopted for maintenance, distribution, and further evolution by Diomidis Spinellis. Changes made by him include support for XHTML 1.0 and documentation bug fixes. The first public release of the maintenance effort was in 2004 (version 2.1). On March 2004 the program was renamed into **bib2xhtml** to avoid confusion with projects using the name **bib2html**. Panos Louridas (*louridas@aub.gr*) added the functionality for including the citations in the *htmlfile*.

See <http://www.spinellis.gr/sw/textproc/bib2xhtml> for the latest version. This is free software, and may be modified or redistributed under the terms of the GNU Public License.