1 HTML to LaTeX (version 2.7)

This article describes version 2.7 of `html2tex`, a program which can be used to converts a single HTML file or a collection of related HTML files into a single LaTeX file. Such a LaTeX file can be processed into a PostScript file. To generate a single LaTeX file from a collection of HTML files, the user needs to give a skeleton LaTeX file, and indicate where translated versions of the HTML files should be included. The user also has to specify for each HTML file at which level (chapter, section, subsection, ..) it should be included. Links between the different HTML files are mapped to references in the LaTeX file. External links can be included as footnotes or as a bibliography.

The generation of LaTeX is configurable. The mapping of each HTML tag to LaTeX commands can be specified. (This mapping can even be changed dynamically during the processing of the HTML file.) It is also possible to exclude certain parts from the HTML files from the generated LaTeX file, or to include LaTeX parts in HTML comment lines, which are ignored by HTML viewers. This makes it possible to maintain sources for both HTML and LaTeX in the same HTML files.

The program performs certain checking of the HTML files, in order to be able to generate correct LaTeX output, but this checking is not guaranteed to conform any HTML standard. At some places the checking might be more relaxed, while at other places more restrictive then HTML 2.0. So far, there is not much support for extensions beyond HTML 2.0.

The program does extensive checking of links between the different files. Because of this reason it can also be used as a link checking program, by giving it a single HTML file, and the option `-c`. In order to also check all referenced pages in the local directory (and its sub-directories), the option `-s` should be used as well.

Links to excluded HTML files (and other URL's) can either be reported as footnotes, or as a sorted bibliography in the LaTeX file.

Error messages are reported on the standard output file. The program can also generate an extensive cross-references file mentioning all the anchor tags.

1.1 Functionality

The HTML to LaTeX conversion program is implemented by the C program `html2tex.c`, which needs to be compiled first. The program is developed with the popular `gcc` compiler, which is freely available under the GNU public license. Under UNIX the program can be compiled with the command: '`make html2tex`'.

The program can be either used to convert a single HTML file into a LaTeX file, or a collection of related HTML files into a single LaTeX file. These two modes of operation will be described below.

1.1.1 Converting a single HTML file

If the program is executed with a single HTML file, a LaTeX file will be generated. For example, the command '`html2tex test.html` will generate the file `test.tex` However, files generated in this manner, are not a complete LaTeX files. To make them complete some LaTeX commands have to prefixed and appended to the file. A LaTex file starts with commands to specify the document style, the title page, and such.

Instead of adding the required LaTeX commands manually, it is also possible to place them inside comments in the HTML file. See below (cf. Section 1.1.3) for a description of the commands which are recognized by `html2tex` inside HTML files.

1.1.2 Converting a collection of HTML files

To produce a single LaTeX file from a collection of linked HTML files, a skeleton LaTeX file has to be provided. In this skeleton there are commands embedded in comments which specify which HTML files should be included at which place.

When `html2tex` is executed with a skeleton file on the command line, a LaTeX file with the same name as the skeleton file, but with the extension `.tex` added to it, will be created.

A real life example of a skeleton file is
`transcoop`, which includes pages from the
original TransCoop pages[1] (which are gone now). The LaTeX file
`transcoop.tex` was generated when the following command was executed in the TransCoop home directory: '`html2tex transcoop`'. The file `transcoop.ps` contains the PostScript file which was produced after executing latex and dvips.

1.1.3 The skeleton file

The skeleton input file should contain valid LaTeX commands. In the file all lines starting with `%html` will be interpreted as special lines by the conversion program. These are used to indicate which HTML files should be included, and to set the various options. The following special commands are recognized by the `html2tex`:

- `%html` *fn*`.html` *level*

  Causes the the file *fn*`.html` to be included as LaTeX at the given input line. The *level* should be an integer to specify the indentation depth of the headers. A value of 1 indicates that the file should be included on the level using `\section` (or to `\chapter` for the book document style).

[1]See URL http://www.cs.utwente.nl/.transcoo/transcoop.html

- `%html -r` *URL*

  Specifies the URL of the directory of the input file. This is needed to detect if any given URL's in the HTML files map to local HTML files. This command should be given before any HTML file is included as LaTeX.

- `%html -s` *style*

  Indicate the style that should be used. By default the book document style is used. Currently, the following values for `style` are supported:

  - `book`: for the book document style.
  - `report`: for the report document style.
  - `article`: for the article document style.
  - `plain`: for an article style without section numbering like (most) HTML browsers do.

  The command causes the mapping of the `H1` to `H7` tags to be set correctly for the given document style. This command should be given before all commands to include HTML files as LaTeX.

- `%html -b`

  Causes LaTeX bibitems to be generated file for all excluded HTML files (and other URL's), at the current location of the skeleton file. If this command is not given anywhere in the input file (and also not the `-b` command line option), all external URL's are given as footnotes.

- `%html -d` *tag-name options* `"`*LaTeX-open*`"` `"`*LaTeX-close*`"`

  Changes the mapping of the *tag-name* HTML tag to the given LaTeX formating commands. See below (cf. Section 1.1.4) for a complete description.

- `%html -l` *from-URL to-URL*

  To indicate that the *from-URL* is a (symbolic) link to *to-URL*. To be used when there are two (or more) URL's for the same physical file. The given URL's should be relative to the root-URL.

- `%html -m` *rel-URL comp-URL*

To display a different URL then the one found in the HTML files, if for example, one wants an `ftp` URL instead of an `http` URL, or if one wants to reference the original source, in case one has a local mirror of certain files found at external URL's.

- `%html -i` *URL*

  To indicate that the *URL* should be ignored. To be used when there are additional HTML pages (for navigation purposes) that you do not want to be referenced in the document. The given URL should be an relative URL to the root-URL.

- `%html -o` *option-name option-value*

  Setting various LaTeX generation options. The various options are explained below (cf. Section 1.1.5).

1.1.4 Special command in the HTML files

The following special commands (inside HTML comments) are recognized in the HTML files:

- `latex` *latex-commands*

  Causes the latex-commands to be copied to the LaTeX output file. Use '`&`', '`<`', '`>`', and '`‐`' for the characters '`&`', '$<$', '$>$', and '`-`'respectively.

- `latex-off`

- `latex-on` *latex-commands*

  Causes the HTML text and tags to be omitted from the generated LaTeX files. These special commands are recognized as tags and should be placed at the proper places with respect to the other tags. They can be nested.

  `latex-on` may be followed by additional commands which are copied into the LaTeX file just like `latex` command described above.

- `latex-def` *tag-name options* `"`*LaTeX-open*`"` `"`*LaTeX-close*`"`

Changes the mapping of the *tag-name* HTML tag to the given LaTeX formating commands. Follows the same rules as the special command '`%html -d`' in the input file, except that '`&`', '`<`', '`>`', and '`‐`' should be used for the characters '`&`', '<', '>', and '`-`' respectively. See below (cf. Section 1.1.4) for a detailed description.

- `latex-rep` *latex-commands*

  Causes the latex-commands to be copied to the LaTeX output file, just like '`latex` *latex-commands*', but if it occurs inside a normal HTML tag, it replaces the LaTeX output that would otherwise have been generated.

- `latex-opt` *option-name option-value*

  Causes the LaTeX generation option *option-name* to be set to the value of *option-value*. The various options are explained below (cf. Section 1.1.5).

- `latex-fmt` *format-string*

  The given format string is used for the generation of the next reference. *(This is an experimental feature which has not been fully tested.)*

- `latex-style` *document-style*

  With this command the document style to be used is specified. By default the book document style is used. Currently, the following values for `document-style` are supported:

  - `book`: for the book document style.
  - `report`: for the report document style.
  - `article`: for the article document style.
  - `plain`: for an article style without section numbering like (most) HTML browsers do.

  The command causes the mapping of the `H1` to `H7` tags to be set correctly for the given document style. This command should appear at the start of the HTML file, and should appear at most once. It is only usefull to use this when generating a LaTeX file from a single HTML file (cf. Section 1.1).

- `latex-biblio`

  With this command the place where the bibliography should be included is specified. It causes LaTeX bibitems to be generated for all excluded

HTML files (and other URL's), at the current location of the HTML file. This should appear at the end of the HTML file, and should appear at most once. It is only usefull to use this when generating a LaTeX file from a single HTML file (cf. Section 1.1).

The program recognizes comments inside a pair of double dashes (`--`), in any of the HTML tags including <`!` >. It also recognizes any text in a <`!` > tag not surrounded by double dashes as comment, but not without generating a warning message for it.

1.1.5 Defining mappings

As we wrote above the various mappings of HTML tags to LaTeX can be changed in both the input file (cf. Section 1.1.2) (as a line of the form **`%html -d`** *tag-name options* `"`*LaTeX-open*`"` `"`*LaTeX-close*`"`), and inside comments (cf. Section 1.1.3) in the HTML files (in the form of **`latex-def`** *tag-name options* `"`*LaTeX-open*`"` `"`*LaTeX-close*`"`).

They change the mapping of the *tag-name* HTML tag to the given LaTeX formating commands. The strings *LaTeX-open* and *LaTeX-close* are put around the text that is marked by the HTML tag. (The string in *LaTeX-close* is generated at the proper place, in case the closing tag is not obligatory in the HTML syntax.) If the LaTeX command has to include a double quote one should use two double quotes in the string. If a real newline (the '\\`n`' character) has to be included, use '\\`nl`' instead. (There is no LaTeX command starting with this sequence, but there are many starting with '\\`n`'.)

The options are used for some special kind of translating. The following options are possible:

- `-math`

  To be used for math mode. This mode assumes that everything that is inside the tags, is correct for the LaTeX math environment. The contents is copied literally, except for `#` and `%` which are quoted.

- `-iim`

  To be used in combination with `-math` to ignore the HTML tags for italics as LaTeX math mode uses italics by default.

- `-off`

  Causes the text inside the HTML tags to be excluded from the generated LaTeX file. The *LaTeX-open* and *LaTeX-close* are both outputted to the LaTeX file (if not inside another tag with `-off`).

- `-on`

  Causes the text inside the HTML tags to be included from the generated LaTeX file. At the start of the file generation is switched off (one-level). In case of nested TAGS with `-off`, the `-on` does only cancel one level. The *LaTeX-open* and *LaTeX-close* are both outputted to the LaTeX file (if not inside another tag with `-off`). If both `-on` and `-off` are used (in an environment with one level off), then only the LaTeX code for the tags is generated.

- `-verb`

  To be used for the verbatim LaTeX environment. Ignores all nested HTML tags that would conflict with the LaTeX verbatim environment.

- `-alltt`

  To be used for the `alltt` LaTeX environment, which is like verbatim, but allows some additional formating.

- `-br`

  To be used for HTML tags that produce an error message when generated on an empty line (like `\newline`).

- `-igh`

  To be used for HTML tags which do not allow section commands inside their generated LaTeX output.

- `-l1` to `-l6`

  To be used to indicate to which section-level a tag should be mapped in LaTeX. The level at which the file is included is added. If this option is used, then *LaTeX-open* and *LaTeX-close* are ignored, except when the tag occurs in an environment where an section heading cannot be generated.

The pseudo HTML tags (which cannot occur in the HTML files) `L1` to `L9` specify what LaTeX commands should be generated for which section level. The definition of these pseudo-tags is changed by the command `%html -s` *style* for setting the document style.

The default settings are the ones given below, using the format to be used in the input file:

```
%html -d html    ""  ""
%html -d head    ""  ""
%html -d title   ""  ""
%html -d body    -on ""  ""
%html -d address ""  ""
%html -d h1      -l1 "{\\LARGE \\bf " "}"
%html -d h2      -l2 "{\\Large \\bf " "}"
%html -d h3      -l3 "{\\large \\bf " "}"
%html -d h4      -l4 "{\bf " "}"
%html -d h5      -l5 "{\\small \\bf " "}"
%html -d h6      -l6 "{\\footnotesize \\bf " "}"
%html -d p       "\nl\nl"  ""
%html -d ul      -igh "\nl\begin{itemize}"  "\nl\end{itemize}\nl"
%html -d menu    -igh "\nl\begin{itemize}"  "\nl\end{itemize}\nl"
%html -d dir     -gnh "\nl\begin{itemize}"  "\nl\end{itemize}\nl"
%html -d ol      -igh "\nl\begin{enumerate}"  "\nl\end{enumerate}\nl"
%html -d li      "\nl\item "  ""
%html -d lh      "\nl\item "  ""
%html -d dl      -igh "\nl\begin{description}"  "\nl\end{description}\nl"
%html -d dt      "\nl\item["  "]"
%html -d dd      ""  ""
%html -d a       ""  ""
%html -d q       "``"  "''"
%html -d i       -iim "{\em "  "}"
%html -d em      "{\em "  "}"
%html -d b       "{\bf "  "}"
%html -d strong  "{\bf "  "}"
%html -d tt      "{\tt "  "}"
%html -d samp    "{\tt "  "}"
%html -d kbd     "{\tt "  "}"
%html -d var     "{\sl "  "}"
%html -d dfn     "{\sc "  "}"
%html -d code    "{\tt "  "}"
%html -d blink   ""  ""
%html -d cite    "{\em "  "}"
%html -d blockquote  -igh "\begin{quotation} "  "\end{quotation}\nl"
%html -d bq      -igh "\begin{quotation} "  "\end{quotation}\nl"
%html -d u       "\underbar{"  "}"

%html -d pre     -verb "\begin{verbatim} "  "\end{verbatim}\nl"%html -d xmp     -verb "\begin{verbatim} "  "\end{verbatim}\nl"
%html -d listing -verb "\begin{verbatim} "  "\end{verbatim}\nl"
```

Suggested alternative settings for the various tags are:

```
%html -d title -on "\newpage\thispagestyle{myheadings}\markright{\sc{}" "}\pagenumbering{arabic}\nl\nl"
%html -d h1 -l1 "{\nl\nl\smallskip\LARGE\bf\noindent " "}\nl\nl\noindent{}"
%html -d h2 -l2 "{\nl\nl\smallskip\Large\bf\noindent " "}\nl\nl\noindent{}"
%html -d h3 -l3 "{\nl\nl\smallskip\large\bf\noindent " "}\nl\nl\noindent{}"
%html -d h4 -l4 "{\nl\nl\smallskip\bf\noindent " "}\nl\nl\noindent{}"
%html -d h5 -l5 "{\nl\nl\smallskip\small\bf\noindent " "}\nl\nl\noindent{}"
%html -d h6 -l6 "{\nl\nl\smallskip\footnotesize\bf\noindent " "}\nl\nl\noindent{}"
%html -d code -math
%html -d blockquote "\nl{\parindent=2em\narrower\nl" "\nl}\nl"
```

The default setting for the pseudo tags for the book and report styles are:

```
%html -d l1      "\nl\nl\chapter{"  "}\nl\nl"
%html -d l2      "\nl\nl\section{"  "}\nl\nl"
%html -d l3      "\nl\nl\subsection{"  "}\nl\nl"
%html -d l4      "\nl\nl\subsubsection{"  "}\nl\nl"
%html -d l5      "\nl\nl\paragraph{"  "}\nl"
%html -d l6      "\nl\nl\subparagraph{"  "}\nl"
%html -d l7      ""  ""
%html -d l8      ""  ""
%html -d l9      ""  ""
```

The default setting for the pseudo tags for the article styles is:

```
%html -d l1      "\nl\nl\section{"  "}\nl\nl"
%html -d l2      "\nl\nl\subsection{"  "}\nl\nl"
%html -d l3      "\nl\nl\subsubsection{"  "}\nl\nl"
%html -d l4      "\nl\nl\paragraph{"  "}\nl"
%html -d l5      "\nl\nl\subparagraph{"  "}\nl"
%html -d l6      ""  ""
%html -d l7      ""  ""
%html -d l8      ""  ""
%html -d l9      ""  ""
```

The default setting for the pseudo tags for the plain style is:

```
%html -d l1      "\nl\nl\section*{"  "}\nl\nl"
%html -d l2      "\nl\nl\subsection*{"  "}\nl\nl"
%html -d l3      "\nl\nl\subsubsection*{"  "}\nl\nl"
%html -d l4      "\nl\nl\paragraph*{"  "}\nl"
%html -d l5      "\nl\nl\subparagraph*{"  "}\nl"
%html -d l6      ""  ""
%html -d l7      ""  ""
%html -d l8      ""  ""
%html -d l9      ""  ""
```

1.1.6 Options

The options can be used to configure the LaTeX fragments which are generated by the program for the various kinds of references. The options can be given in the input file (cf. Section 1.1.2) (as a line of the form `%html -o` *option-name option-value*), and inside comments (cf. Section 1.1.3) in the HTML files (in the form of `latex-opt` *option-name option-value*).

There are options that determine the cases in which references should be generated and when not. For example, it will often be the case that an HTML file contains a HREF tag, whenever an email address is given, which can be used to send an email. As the essential information is already provided it is not necessary to include it in a footnote or a bibliographic entry. The following options can be used for this purpose:

- `dni_email` [`on`|`off`]

  This option determine whether email addresses are included in the references/bibliography, if they appear in the text.

- `dni_news` [`on`|`off`]

  This option determine whether news groups are included in the references/bibliography, if they appear in the text.

- `dni_ftp` [`on`|`off`]

  This option determine whether ftp addresses are included in the references/bibliography, if they appear in the text.

- `dni_other` [`on`|`off`]

  This option determine whether all other kind of URL's are included in the references/bibliography, if they appear in the text.

By default all these options are `on`.

The references can be divided into internal and external. The internal references are HREF tags that point to a file that is included in the LaTeX output, and external are those that are not. Internal references can be mapped to phrases, that state to look at the corresponding section. External references have to be given completely, either as a footnote at the bottom of the page or as a bibliographic entry. They are generated as bibliographic entries if the input file contains a line with '`%html -b`' (or if the program option (cf. Section 1.1.6) `-b` is given), otherwise they are generated as footnotes. There are four generation modes:

- `normal`: this means that the internal references are generated with a '(cf. Section)' text, and external references as either a footnote or a citation.
- `cffn`: this is the same as the above, except that the internal references given as a footnote with the a 'See Section' text.
- `fn`: this is the same as the above, except that citations are also given as a footnote. This option generates a footnote for each kind of reference.
- `none`: This option prevents the generation of any references.

These four modes can be set for three different environments, namely: the headers, LaTeX alltt environments, and all the remaining parts. The options for this are:

- `href_in_header` [`normal`|`cffn`|`fn`|`none`]

  This controls the generation of HREF tags inside headers. The default value is `normal`.

- `href_in_alltt` [`normal`|`cffn`|`fn`|`none`]

  This controls the generation of HREF tags inside LaTeX alltt environments. The default value is `none`.

- `href` [`normal`|`cffn`|`fn`|`none`]

  This controls the generation of HREF tags at all other places. The default value is `normal`.

There are also options that determine the format in which the various kinds of references are to be generated (including the format of the bibliographic entries). All these options make use of format strings (like those used in C), where the percentage symbol followed by letter indicates a place holder for a string or number that has to be outputted. A double percentage symbol causes a single percentage symbol to be printed. All these options should contain LaTeX formating commands. Because references can be generated in fragile environments '`%p`' has to be used at places where a '`\protect`' is required in a fragile environment. Also because a '`\footnote`' is not allowed every where, a '`%F`' has to be used instead.

These are the options for internal references:

- `filenr` *format-string* (Default value: `"f%d"`)

  This option is used to specify the format of the file references.

- `label` *format-string* (Default value: `"%p\label{%f}"`)

  This option is used to specify the format of a label without an additional name part. `"%f"` indicates the place of the file part of the label.

- `label_n` *format-string* (Default value: `"%p\label{%f:%n}"`)

  This option is used to specify the format of a label with an additional name part. `"%n"` indicates the place of the name text.

- `cf` *format-string* (Default value: `" (cf.  Section~%p\ref{%f})"`)

  This option is used to specify the format of an internal reference without an additional name part in the running text.

- `cf_n` *format-string* (Default value: `" (cf.  Section~%p\ref{%f:%n})"`)

  This option is used to specify the format of an internal reference with an additional name part in the running text.

- `f_cf` *format-string* (Default value: `"%p%F{See also Section~\ref{%f}.}"`)

  This option is used to specify the format of an internal reference without an additional name part inside a footnote.

- `f_cf_n` *format-string* (Default value: `"%p%F{See also Section~\ref{%f:%n}.}"`)

  This option is used to specify the format of an internal reference with an additional name part inside a footnote.

The options for external references as footnotes are:

- `f_news` *format-string* (Default value: `"%p%F{See URL news:%n}"`)

  This option is used to specify the format for a newsgroup. `"%n"` indicates the place of the newsgroup name.

- `f_mailto` *format-string* (Default value: `"%p%F{See URL mailto:%m}"`)

  This option is used to specify the format for an email address. `"%m"` indicates the place of the email address.

- `f_ftp` *format-string* (Default value: `"%p%F{See URL ftp://%s}"`)

  This option is used to specify the format for an ftp site. `"%s"` indicates the place of the site.

- `f_ftp_d` *format-string* (Default value: `"%p%F{See URL ftp://%s/%d}"`)

  This option is used to specify the format for a directory on an ftp site. `"%d"` indicates the place of the directory path.

- `f_ftp_f` *format-string* (Default value: `"%p%F{See URL ftp://%s/%f}"`)

  This option is used to specify the format for a file on an ftp site. `"%f"` indicates the place of the file name.

- `f_ftp_df` *format-string* (Default value: `"%p%F{See URL ftp://%s/%d/%f}"`)

  This option is used to specify the format for a file, in a directory on an ftp site.

- `f_URL` *format-string* (Default value: `"%p%F{See URL %U}"`)

  This option is used to specify the format an URL without an additional name part. `"%U"` indicates the place of the URL.

- `f_URL_n` *format-string* (Default value: `"%p%F{See URL %U\#%n}"`)

  This option is used to specify the format an URL with an additional name part. `"%U"` indicates the place of the URL.

The options for citations are:

- `citenr` *format-string* (Default value: `"b%d"`)

  This option is used to specify the format of the citation labels.

- `cite` *format-string* (Default value: `"%p\cite{%c}"`)

  This option is used to specify the format of a normal citation without an additional name part. `"%c"` indicates the place of the citation label.

- `cite_n` *format-string* (Default value: `"%p\cite[%n]{%c}"`)

  This option is used to specify the format of a normal citation with an additional name part. `"%n"` indicates the place of the name text.

- `f_cite` *format-string* (Default value: `"%p%F{See \cite{%c}}"`)

  This option is used to specify the format of a citation as a footnote without an additional name part.

- `f_cite_n` *format-string* (Default value: `"%p%F{See \cite[%n]{%c}}"`)

  This option is used to specify the format of a citation as a footnote with an additional name part.

The options for the bibliographic entries are:

- `b_news` *format-string* (Default value: `"news:%n"`)

  This option is used to specify the format for a newsgroup. `"%n"` indicates the place of the newsgroup name.

- `b_mailto` *format-string* (Default value: `"mailto:%m"`)

  This option is used to specify the format for an email address. `"%m"` indicates the place of the email address.

- `b_ftp` *format-string* (Default value: `"ftp://%s"`)

  This option is used to specify the format for an ftp site. `"%s"` indicates the place of the site.

- `b_ftp_d` *format-string* (Default value: `"ftp://%s/%d"`)

  This option is used to specify the format for a directory on an ftp site. `"%d"` indicates the place of the directory path.

- `b_ftp_f` *format-string* (Default value: `"ftp://%s/%f"`)

  This option is used to specify the format for a file on an ftp site. `"%f"` indicates the place of the file name.

- `b_ftp_df` *format-string* (Default value: `"ftp://%s/%d/%f"`)

  This option is used to specify the format for a file, in a directory on an ftp site.

- `b_URL` *format-string* (Default value: `"%U"`)

  This option is used to specify the format an URL without an additional name part. `"%U"` indicates the place of the URL.

- `b_URL_n` *format-string* (Default value: `"%U\#%n"`)

  This option is used to specify the format an URL with an additional name part. `"%U"` indicates the place of the URL.

The following options deal with the formating of all kinds of references. They make it possible to add additional formating around the anchor text or the image tag. The `"%R"` indicates the place where the reference should be placed. This can either be an internal or an external reference, in the running text or as a footnote. In case the `"%R"` appears in an fragile environment, it should be changed into `"%fR"`. In case it appears in a place where a `\footnote` would not be proper, a combination of an `"%mR"` and an `"%tR"` can be used to indicate the place of the footnote marker and the footnote text, respectively. (An `"f"` can be added if they occur in a fragile environment.)

- `t_href` *format-string* (Default value: `"%R"`)

  This option is used to specify the format to be used with a reference in an HREF tag.

- `t_img` *format-string* (Default value: `"\fbox{\tt %n %mR}%tR"`)

  This option is used to specify the format to be used with a reference in an IMG tag, when there is not alternative text specified. `"%n"` indicates the place of the file name (without the path) of the imagine.

- `t_img_r` *format-string* (Default value: `"%r"`)

  This option is used to specify the format to be used with a reference in an IMG tag, when there is an alternative text specified. `"%r"` indicates the place of the alternative text.

1.1.7 Program options

If the program is given an input file with the extension `.html`, it does not generate a LaTeX output file, but only analyse the file, and the files it references (if the `-s` option is given).

The program recognizes the following command line options:

- `-i` : print info.
- `-w` : print warning (and info).
- `-p` : pedantic: does not report omissions of HTML open and close tags.
- `-s` : scan not include HTML files. The program will scan all HTML files that can be reached from the included files, and that are found in the directory (and its sub-directories) of the input file.
- `-r` *URL* : the URL of the directory in which the program is ran. This is needed to find out if any full URL points to a local HTML file.
- `-b` : make bibliography. If this option is not given, references to external URL will appear in footnotes. The input file should contain a line with `%html -b`.
- `-cr` : make a cross-reference file with the extension `.ref`
- `-d`: generate lots and lots of debug statements :-). Only to be used if you want to know what it all does.

1.2 Bugs

There is still a long road to go with respect to bugs. I still cannot process the web testing[2] pages correctly.

Known bugs are:

[2] See URL http://www-dsed.llnl.gov/documents/tests/html.html

- The usage of <`SUP`> and <`SUB`> can cause incorrect LaTeX output to be generated when used within a math-environment, as these open a math-environment.
- The `font` is not supported.
- Any form and table related tags are not supported.

1.3 The source

The source of `html2tex` falls under the GNU General Public License, and thus **no warrants what so ever are implied!** Earlier versions are available on request. I cannot give much support, because I a busy with my kids Annabel and Andy.

1.4 Version history

For all versions: **No warrants what so ever are implied!.** Each version has a version number and a date at the top of the source file. Please use these for bug reports.

- Version 1.0, July, 1995: This is the oldest version, which worked okay for my applications.
- Version 2.0, November 11, 1995: This version is the first step towards a more complete implementation. It included support for more tags, and has improved checking of the HTML input. Warwick Allison[3] made some bug-fixes to this version which can be found in his version[4].
- Version 2.1, March 5, 1996: Contains many improvements, including the corrections by Warwick Allison[5]. This version includes additional HTML checking, and configuration of LaTeX output generation was added.
- Version 2.2, May 8, 1996: This was a *stable* version for bug fixing.
- Version 2.4, May 8, 1996: This version includes many more customization options.
- Version 2.5, May 17, 1996: This version was created to improve link checking.
- Version 2.6, August 28, 1996: This version allows the user to determine which heading tags should be mapped to which level.

[3] See URL http://www.cs.uq.oz.au/personal/warwick/
[4] See URL http://www.cs.uq.oz.au/personal/warwick/product/html2tex.wwa1.c
[5] See URL http://www.cs.uq.oz.au/personal/warwick/