

# PDF information and navigation elements with hyperref, pdfTeX, and thumbpdf

Heiko Oberdiek

The PDF format offers additional possibilities for information and navigation through paper-less on-line documents. This paper shows, how the navigation features bookmarks and thumbnails can be created automatically or manually by powerful packages like `hyperref` and `thumbpdf`. The problems and solutions are described that arise from converting TeX strings to the PDF ones used in the general document information or in the outlines.

## Use of colors

- Red is used for internal links.
- Magenta denotes web links and email addresses.
- Often blue, sometimes green, and seldom red are used to put emphasis on words or to illustrative the text.

The table of contents is on page 21

## 1 Introduction

The PDF format provides elements for navigation and information that do not go to the page content and are not printed. But they are nevertheless useful, because they provide the user additional informations about the document and enlighten the navigation while viewing the document on-line. The elements described in this article:

*General document information:* With following programs the entries in the document information can be made visible:

*AcrobatReader* (Figure 1 right bottom):

- Hotkey: <CTRL>—D
- Menu: File→Document Info→General

*pdfinfo:* The program `pdfinfo` produces text output and can easily piped by other programs. So documents with special keywords can be looked for.

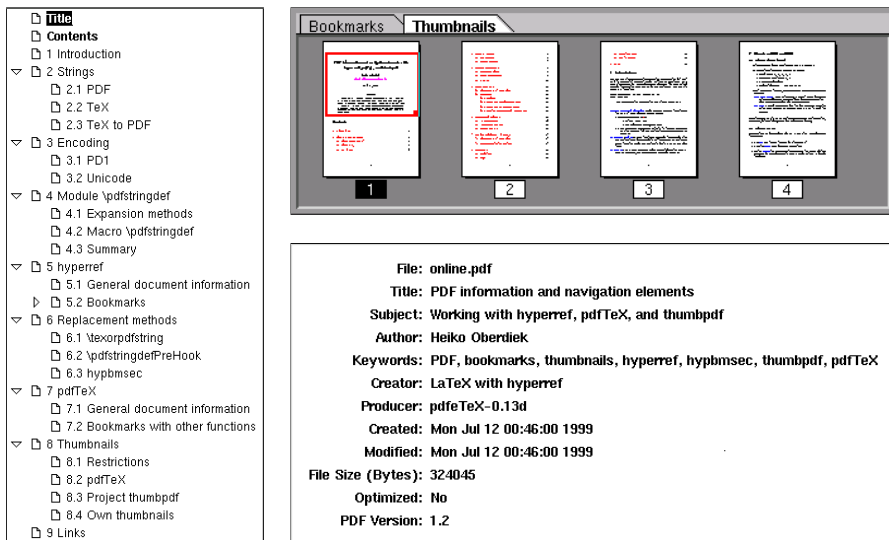


Figure 1: Bookmarks, thumbnails and general information of AcrobatReader

**Bookmarks (outlines):** The outline tree (Figure 1 left) mirrors often the logical structure of a document like a table of contents. But they can also be used to offer access to another related documents or to start a sound file and other things.

**Thumbnails:** Thumbnails (Figure 1 right top) provide an optical representation of the pages. A page can be chosen, and a page area can be selected for viewing.

## 2 Strings in T<sub>E</sub>X and PDF

### 2.1 Strings of PDF format

- Strings are delimited by parentheses: (This is a string.)
- As in PostScript escape sequences can be used for:
  - ▷ White space: \b, \f, \n, \r, \t
  - ▷ Unbalanced parentheses: \(, \)
  - ▷ Escape character itself: \\
  - ▷ Octal notation: äöüß = \344\366\374\337
- Font changes are not allowed, possible encodings are PDFDocEncoding or Unicode:

*PDFDocEncoding* is an 8bit encoding, a superset of ISOLatin1 (characters 160–255 added), and compatible with Unicode (characters < 256). There are differences with WinAnsiEncoding.

*Unicode* is a 16bit encoding and tries to match all possible characters in the world. AcrobatReader versions below four do not support this encoding.

PDF strings are used at many different places: general document information, bookmarks, text annotations, information dictionary of threads, ...

### 2.2 Strings in T<sub>E</sub>X

There are two main types of strings with T<sub>E</sub>X. A look at T<sub>E</sub>X's digestion shows the difference:

1. The *eyes* read the input line and the catcodes are set.
2. Then the *mouth* forms tokens and expands them.
3. By the *stomach* the unexpandable tokens are executed, assignments take place, and the characters are set in boxes. The stomach is T<sub>E</sub>X's typesetting engine.

Let us now consider an argument of \section that is typeset in the text, appears in the table of contents and in the bookmarks: In the text the string is fully digested. For the table of contents the argument is written to the .aux file and later to the .toc file. At this stage the argument is only expanded (mouth), unexpandable tokens are written verbatim. But in the last step the text in the .toc file is read and digested

by T<sub>E</sub>X's stomach. For the bookmarks, however, the argument is only written to the output file directly, only the expansion done by the mouth. These tokens do not reach T<sub>E</sub>X's typesetting stomach. Therefore the string must be in a format that fits the PDF specification and is accepted by the PDF viewer.

The following table illustrates the important differences between oral and stomach strings:

	<b>mouth</b>	<b>stomach</b>
“zero”:	<code>\empty</code>	<code>\relax</code>
<code>{...}</code> :	parameter	group
variable:	reading	assignment
character:	command names	boxes
	<code>\csname</code>	ligatures
commands:	<code>\string</code> , <code>\number</code>	<code>\def</code> , <code>\let</code> , <code>\long</code>
	<code>\expandafter</code>	<code>\uppercase</code> , <code>\accent</code> , <code>\par</code>
	<code>\meaning</code> , <code>\the</code>	<code>\bgroup</code> , <code>\begingroup</code>
	<code>\if</code> , <code>\ifx</code> , <code>\ifcase</code>	<code>\hbox</code> , <code>\vtop</code> , <code>\kern</code>

### 2.3 T<sub>E</sub>X to PDF

Arguments or strings, written for T<sub>E</sub>X's **stomach**, will not see the **stomach**, if they are used as PDF strings. This leads to many problems:

- Stomach commands are not executed, they are written verbatim (`\hbox`, `\kern`, ...). But a PDF viewer does not know something about T<sub>E</sub>X commands.
- No font features such as changing fonts or font attributes, ligatures, or hyphenation.
- No manipulation of boxes, no mathematics, no colors, no graphics, ...
- No assignments and definitions.
- Only PDFDocEncoding or Unicode can be used, the encoding cannot be changed within strings.

### 3 Encoding handling

For the conversion from the current  $\text{\TeX}$  font encoding to one of the two PDF encodings `\pdfstringdef` (see [section 4](#)) uses the benefits of  $\text{\LaTeX}$ 's font encoding mechanism. The data, which position has a glyph in an encoding, are stored in macros:

`\OT1\ss`:  $\ss$  in OT1 encoding (`\char"19`  $\rightarrow$   $\beta$ ).

`\T1\ss`:  $\ss$  in T1 encoding (`\char"FF`  $\rightarrow$   $\beta$ ).

The macro name `\T1\ss` is obtained by `\csname T1\string\ss\endcsname`.

In plain- $\text{\TeX}$  (OT1) `\ss` is directly defined as `\char"19`, in  $\text{\LaTeX}$  the definition consists of three parts: `\OT1-cmd \ss \OT1\ss`. The first command `\OT1-cmd` performs the encoding stuff:

- In a protected environment it expands to `\noexpand \ss`.
- **Warning** if the text command is used in [math mode](#).
- Use of `\OT1\ss` ( $= \text{\char"19}$ ) if the currently active encoding is the same as in the name.
- With a different current encoding the data of the default encoding are used: `\?\ss` expands to `\UseTextSymbol{OT1}\ss`.

$\text{\LaTeX}$ 's font encoding mechanism works with appropriately defined [macros](#). So characters with catcodes 11 (letter) and 12 (other) selects the font slot directly, circumventing the font encoding mechanism.

#### 3.1 PD1 encoding

The [PD1 encoding](#) is defined in the file `pd1enc.def` and defines the ordering of the glyphs of the PDFDocEncoding. Most  $\text{\TeX}$  names obey the following rules:

- `\text<(glyph name)>`:  $\text{\textregistered}$   $= \text{\textregistered}$ ,  $\text{\textexclamdown}$   $= \text{\textexclamdown}$
- **Traditional names**:  $\text{\text{Æ}}$   $= \text{\text{Æ}}$ ,  $\text{\ss}$   $= \text{\ss}$
- **Accents**:  $\text{\text{Ü}}$   $= \text{\text{U}}$ ,  $\text{\text{ç}}$   $= \text{\text{c}}$ ,  $\text{\text{Å}}$   $= \text{\text{A}}$ ,  $\text{\text{r A}}$   $= \text{\text{r A}}$

Glyph codes in PD1 encoding are mostly octal sequences (`\textmu`  $\Rightarrow$  `\265`) to allow safe output.

The file `testbmg1.tex` of package `hyperref` lists available glyphs and the  $\text{\TeX}$  macro names to produce them.

### 3.2 Unicode encoding

General problems:

- **Characters** having catcode 11 (**letter**) and 12 (**other**) are problematic, if the glyph position is not the same as in PDFDocEncoding. It is easy to write code, that detects this characters and replaces them with the Unicode slots.
  - ▷ But T<sub>E</sub>X does not know the glyph name of the current font slot addressed by the character code. Therefore great arrays with font names and encoding tables of that fonts would be required.
  - ▷ Checking 256 characters for both catcodes takes a lot of time.

Memory consumption and computing time would be very high.

- Systems that use **ligatures** to select the glyphs do not work, because this stomach feature are simulated by expanding commands. This would require huge tables and lots of time. It can only done for exceptions like the spanisch ligatures !' and ?'. `\pdfstringdef` detects and converts them to `\exclamdown (¡)` and `\questiondown (¿)`. With an empty group `{}` or the italic correction `\/` ligatures can be prevented.
- If the glyphs are produced by **direct macros**, so `\pdfstringdef` has to redefine them to produce the two bytes for Unicode output. Disadvantage: large redefinitions tables cost time.
- In the easiest case the glyph macros obey **L<sup>A</sup>T<sub>E</sub>X's font encoding mechanism**. Then only an encoding definition file for Unicode has to be provided.

So the **Unicode** support of `hyperref` is implemented using the later method and relies on an input encoding that uses active characters.

Another possibility is reported for AcrobatReader on Russian windows: Because a Cyrillic font is used for the bookmarks, the 8bit positions of the PDFDocEncoding show Cyrillic glyphs now. But this solution is not portable and violates the PDF specification.

`\pdfstringdef` supports the **Unicode route**:

- Because the Unicode macros use up lots of memory, Unicode support is only loaded, if the option `unicode` is selected.
- `\hypersetup{unicode=false}` and `\hypersetup{unicode}` switches between PDFDocEncoding and Unicode.

- In the encoding definition files the two byte sequences are masked in order to be able to distinguish between letters, one and two byte octal sequences.
- Converting of letters and one byte octal sequences into the two byte ones; it is assumed that the one byte glyphs obey the PDFDocEncoding and the higher byte is set to zero.
- Removing the mask.
- Adding the Unicode marker in front of the string (\376\377)
- Currently (early September 1999) only Cyrillic, Greek and some other glyphs of the range 0x0000–0x04FF are supported.

The file `testbmu.tex` of package `hyperref` shows the available glyphs and supported  $\TeX$  macro names to produce them.

## 4 Module `\pdfstringdef`

Package `hyperref` contains a module with the command `\pdfstringdef` that tries the conversion of  $\TeX$  to PDF strings:

- $\TeX$  should not stop with an error.
- Detecting of parts that cannot be converted correctly; so forbidden commands should give a comprehensive warning (`\kern`, `\hbox`, ...).
- Redefining macros that produces text or characters without forbidden commands (`\TeX`, `\P`).
- Simulating forbidden but useful commands (`\xspace`).
- Conversion between  $\LaTeX$  and PDF encoding (PD1).
- Disabling some forbidden commands (`\def`, `\discretionary`) with and without warning.
- `\pdfstringdef` can do local redefinitions, but catcodes are untouched, because catcode changes do not work, if the argument has already been seen by  $\TeX$ 's eyes.
- Avoiding a processing step by an external program.

## 4.1 Expansion methods

T<sub>E</sub>X provides two main methods to expand strings:

- `\edef` defines a macro with the expansion of the argument. Tokens that cannot be expanded further are included unchanged.
- The `\csname` method converts the text to a macro name. Using `\string`, any macro name can be converted back into a sequence of characters, and with `\@gobble` the first backslash can be removed.

The result of the `\csname` method is a safe string without T<sub>E</sub>X commands. But stomach commands are not allowed and will result in an error message (“`\endcsname` expected”), that confuses the user. Because the error handling of T<sub>E</sub>X is very old, modern methods like catch-try-blocks are missing. Therefore `hyperref` uses the `\edef` method and makes great efforts to detect forbidden commands.

## 4.2 Macro `\pdfstringdef`

`\pdfstringdef` takes two parameters: a macro name and a T<sub>E</sub>X string. As result the macro is defined containing the PDF string. (Currently this is done globally, but do not rely on it.)

All the following tasks, definitions and redefinitions are made in a group to keep them local:

- Switching to `PD1` or `PU` encoding
- definition of the “octal sequence commands” (`\345`):  
`\edef\3{\string\3}`
- Special glyphs of T<sub>E</sub>X: `\{`, `\%`, `\&`, `\space`, `\dots`, ...
- National glyphs (`german.sty`, `french.sty`)
- Logos: `\TeX`, `\eTeX`, `\MF`, ...
- Disabling commands that do not have a useful function in bookmarks:  
`\label`, `\index`, `\glossary`,  
`\discretionary`, `\def`, `\let`, ...  
L<sup>A</sup>T<sub>E</sub>X’s font commands like `\textbf`, ...
- Support for `\xspace` of package `xspace`



- **Parentheses** are protected to avoid the danger by unsafe unbalanced parentheses in the PDF string. (see **subsection 2.1**).

A very simple example is the redefinition of `\discretionary` to get the non-break text:

```
\let\discretionary\@gobbletwo
```

More difficult are the `\def` commands. They do not work in T<sub>E</sub>X's mouth and undefined command names will cause errors. A tricky redefinition of the `\def` commands avoids them by *defining* the undefined macros with `\csname` to be `\relax` (This is the only possible assignment in an expanding context).

After the redefinitions the string is expanded by `\edef`. Because the result can contain command tokens and other unwanted things, each token has now to be checked by `\pdfstringdef`:

- **Grouping characters** are discarded silently.
- Also unexpandable commands and active characters with the meaning of `\relax` are removed silently (`\relax` itself, `\protect`).
- All the other **unexpandable** ones and **special characters** (`$`, `&`, ...) are cancelled with a comprehensive warning.

### 4.3 Summary

Allowed are:

- **Expandable** macros,
- `\relax`, `\protect`, grouping characters `{` and `}`,
- `\xspace`, `\discretionary`, `\,`, `\penalty`.

The most other unexpandable commands give **warnings**. Cryptic error messages should be seldom, but they *can* occur, especially if the code of the T<sub>E</sub>X string depends on assignments.

## 5 Package hyperref

### 5.1 General document information

Package hyperref knows the following options to fill the elements of the general document information:

```
pdftitle, pdfsubject, pdfauthor,
pdfkeywords, pdfcreator, pdfproducer
```

The normal place for options is the optional argument of the `\usepackage` command. However  $\LaTeX$  expands the options before passing them to the package. So the switching of the encoding and all the redefinitions of `\pdfstringdef` are not active while expanding. Therefore it is better to set the information options, after package hyperref is loaded. For this purpose the package provides `\hypersetup`:

```
\usepackage{hyperref}
\hypersetup{
  pdftitle={About umlauts: \~a \~o \~u \ss},
  pdfauthor={\textcopyright\ Heiko Oberdiek},
  ...
}
```

Another possibility is to reuse the data of `\maketitle`:

```
\newcommand\org@maketitle{}
\let\org@maketitle\maketitle
\def\maketitle{%
  \hypersetup{
    pdftitle={\@title},
    pdfauthor={\@author}
  }%
  \org@maketitle
}
```

### 5.2 Bookmarks

#### 5.2.1 Options

There are four places for hyperref options:

1. Global: `\documentclass[...]` (e.g. driver)

2. Package: `\usepackage[...]`
3. Configuration file: `hyperref.cfg` with `\hypersetup`
4. After package: `\hypersetup{...}`

The following options affect bookmarks:

**bookmarks**: Make bookmarks (Default: `true`). This option cannot be used after package has been loaded.

**bookmarksnumbered**: Put section numbers in bookmarks (Default: `false`).

**bookmarksopen**: Open up the bookmark tree (Default: `false`).

**bookmarksopenlevel**: Level (see [subsection 5.2.4](#)) to which bookmarks are open (Default: `\maxdimen`).

**bookmarkstype**: To specify which ‘toc’ file to mimic (Default: `toc`).

**pdfpagemode**: How document starts when opened (Default: `None`):

**None**: Neither outlines nor thumbnails are visible.

**UseOutlines**: Outlines are visible.

**UseThumbs**: Thumbnails are visible.

**FullScreen**: Full-screen mode without bars, outlines, and thumbnails.

**unicode**: Creating Unicode bookmarks (Default: `false`). After package has been loaded, it switches between Unicode and PDFDocEncoding.

When creating bookmarks `hyperref` writes them into the file `\jobname.out`. At the second run the bookmarks are set. A “rerun” warning is not implemented.

## 5.2.2 Bookmarks by section commands

The section commands (`\part`, `\chapter` down to `\subparagraph`) are automatically used for the outline tree, unless option `bookmarks` is disabled. The behaviour can be modified with two [L<sup>A</sup>T<sub>E</sub>X](#) counters:

**tocdepth**: This counter sets the level to which the section commands appear in the table of contents. It is interpreted for the outline tree, too.

**secnumdepth:** This counter specifies the level, to which the section commands are numbered. If option `bookmarksnumbered` is in force, this counter is also used for the bookmarks.

The bookmarks build a *rigide tree structure*. Intermediate levels cannot be omitted, because the next subentry after an omitted entry will start leftmost at the highest level in the bookmark tree.

```
\chapter{five}\subsection{one}
```

In the table of contents the `\subsection` is formatted as `\section` with number 5.0.1. But in the outline tree there is no `\section` entry which the `\subsection` can be attached to. So the `\subsection` entry hangs directly on the steam.

### 5.2.3 Bookmarks by `\addcontentsline`

For starred section commands `\addcontentsline` adds an entry for the table of contents. Also a bookmark is added.

There are problems, if the link does not point to the page in general, but to the coordinates on the page (`/XYZ` view):

- Package `hyperref` does not want to redefine the section commands too deeply in order to remain compatible with most of the packages. The disadvantage is, that the destination is set by the overloaded section command *below* the title.

```
\section*{Starred section}
\addcontentsline{toc}{section}{Starred section}
```

The macro `\addcontentsline` cannot be used before, because the anchor name for the destination is only known *after* the section command.

- Package `hypbmsec` supports an automatic inclusion of starred section commands into the table of contents and the outline tree with option `startotoc`. With option `prefix` it tries to solve the problem by putting `\addcontentsline` before the section command. Therefore it defines its own destination. The great disadvantage is a possible page break between the destination and the section command.
- With option `infix` package `hypbmsec` uses another method: It puts the destination definition and `\addcontentsline` into the title itself to avoid page breaking problems. This works for left aligned titles, with right aligned or centered titles the horizontal offset goes wrong.

### 5.2.4 Creating bookmarks with \pdfbookmark

As explained above the direct children can be attached to a parent, but no grandchildren. Each section command has an **internal level number** to check the relationships:

Documentclass	\part	\chapter	\section	\subsection	...
book/report:	-1	0	1	2	...
article:	0		1	2	...

This level is also needed in own bookmarks. These can be created by the macro `\pdfbookmark`:

```
\pdfbookmark[level]{bookmark text}{anchor name}
```

This command creates an anchor and adds this entry to the bookmarks. The anchor name should be unique. For the internal name the argument with the anchor name is appended by a dot and the level. Examples:

```
\pdfbookmark[0]{Beginning of Document}{beg}% beg.0
\pdfbookmark[1]{Titlepage}{tit}% tit.1
```

It is possible to add an outline entry that points to another target, e. g.: Somewhere in the text the **target (anchor)** is defined:

```
\hypertarget{place.1}{}%
```

Then the bookmark can be added at another location as follows:

```
\begingroup
\makeatletter
\def\hyper@anchorstart #1\hyper@anchorend{}%
\pdfbookmark[1]{Go to the place}{place}%
\endgroup
```

The current bookmark level of the last command is stored in the internal macro `\Hy@currentbookmarklevel`. The both commands `\currentpdfbookmark` and `\subpdfbookmark` omits the level argument of `\pdfbookmark`, because they use the internal stored value of the current bookmark level. With this two commands an user can create bookmarks at the same level or a level below without looking after a level number.

## 6 Replacement methods

Because of limitations of T<sub>E</sub>X's mouth (see [subsection 2.2](#)), many things like mathematics, colors, or font changes cannot be used in bookmarks. But package `hyperref` uses the argument of section commands for the bookmarks (see [subsection 5.2.2](#)) that can contain these forbidden things. Therefore a way is needed to replace or disable these for the bookmarks.

### 6.1 Providing an alternative with `\texorpdfstring`

Package `hyperref` sets a switch while expanding a string in `\pdfstringdef`. This is used by the full expandible macro `\texorpdfstring`. It expects two arguments, the first contains the string that will be set and processed by T<sub>E</sub>X's stomach, the second contains the replacement for PDF strings, e. g.:

```
\section{Pythagoras:
  \texorpdfstring{$ a^2 + b^2 = c^2 $}{%
    a\texttt{two}superior\ + b\texttt{two}superior\ =
    c\texttt{two}superior}}
\section{\texorpdfstring{\textcolor{red}}{}{Red} Mars}
```

Also `\texorpdfstring` can be used in own definitions perhaps of logos, so that they work in all contexts.

### 6.2 Hook for own macro redefinitions

Macro `\pdfstringdef` executes the hook `\pdfstringdefPreHook` before it expands the string. Here the user can disable additional commands:

```
\expandafter\def\expandafter\pdfstringdefPreHook
\expandafter{%
  \pdfstringdefPreHook
  \renewcommand{\mycommand}[1]{}%
}
```

An easier way is the new command `\pdfstringdefDisableCommands`, that adds its argument to the definition of `\pdfstringdefPreHook` ('@' can here be used as letter in command names):

```
\pdfstringdefDisableCommands{%
  \let~\textasciitilde
  \def\url{\pdfstringdefwarn\url}%
}
```

```

\let\textcolor\@gobble
}

```

### 6.3 Package hypbmsec

The second method offered by package `hypbmsec` extends the syntax of the section commands. A second argument in square brackets or an argument within parentheses is interpreted as string for the bookmarks:

```

\section{toc / head = bookmark = text}
\section[toc / head = bookmark]{text}
\section[toc / head][bookmark]{text}
\section(bookmark){toc / head = text}
\section[toc / head](bookmark){text}
\section(bookmark)[toc / head]{text}

```

The package `hypbmsec` works with packages that do not change the syntax of the section commands, it should only be loaded as last package. Incompatible are packages that also change the syntax of the section commands.

For bookmarks `hypbmsec` uses package `hyperref`, but if this package is not loaded, it ignores the bookmark arguments. So the same text can be used for various output formats.

The use of the optional parameter `delimiters` `inside` the optional parameter should be protected by `braces`:

```

(...(...{...})...) or [{...[...]}]

```

The next version of package `hypbmsec` will know the option `starttotoc`: The starred section commands are included in the table of contents and in the outline tree automatically. For options `infix` and `prefix` see [subsection 5.2.3](#).

## 7 Additional features of PDF format

Package `hyperref` does not support all possibilities of the PDF format. For [special effects](#) the low level commands of the driver must be used. As an example this is shown with the primitive commands of `pdfTeX`.

### 7.1 General document information

There are no `hyperref` options for `/CreationDate` and `/ModDate`. `pdfTeX` sets them to the current date. With `pdfTeX`'s primitive command `\pdfinfo` they can be set explicitly. The special format of the date strings is defined in the PDF specification. The following shows an example that explicitly sets `/CreationDate` and `/ModDate`:

```
\pdfinfo{/CreationDate (D:19990901000000-01'00')}
\begingroup
  \def\twodigits#1{\ifnum#1<10 0\fi\the#1}%
  \count0=\time \divide\count0 by 60
  \edef\x{\twodigits{\count0}}%
  \multiply\count0 by 60
  \count1=\time \advance\count1 by -\count0
  \edef\x{\x\twodigits{\count1}}%
  \edef\x{/ModDate (D:\the\year
    \twodigits\month \twodigits\day \x 00-01'00'))%
\expandafter\endgroup
\expandafter\pdfinfo\expandafter{\x}%
```

### 7.2 Bookmarks with other functions

The generation of outlines with package `hyperref` is limited to ones that point to a destination inside the document. But in general outlines can point to other documents, call functions of AcrobatReader, play sound, video, ...

For this a driver primitive command is necessary, e. g. `pdfTeX`:

```
\pdfoutline action count n {text}
```

The absolute value of  $n$  is the count of the direct subentries of this outline. If the value is negative, the subentries are closed.

Example for “[named actions](#)”:



```

\newcommand*{\baction}[3][0]{%
  \begingroup
    \pdfstringdef\x{#3}%
    \pdfoutline
      user {<< /S /Named /N /#2 >>} count #1 {\x}%
  \endgroup
}
\baction[3]{NOP}{Navigation}
  \baction[-2]{FullScreen}{Full-screen mode}
  \baction{PageOnly}{Page only}
  \baction{ShowThumbs}{Show thumbnails}
\baction[-6]{NOP}{Page selecting}
  \baction{PrevPage}{Go to previous page}
  \baction{GoBack}{\textless\textless}
...

```

The possible actions and the syntax can be read in the PDF specification. A bookmark that points to the action section of that file on page 96:

```

\pdfoutline
  user {<< /S /GoToR /F (pdfspec.pdf) /D [95 /Fit] >>}
  count 0 {Actions}

```

## 8 Thumbnails

Thumbnails are small images, pictograms or reduced images from the pages of a document. They support the navigation through the document, if the viewer supports it (e. g. AcrobatReader).

### 8.1 Restrictions

Some [restrictions](#) are mentioned in the PDF specification (1.2):

- The maximal supported size of a thumbnail is  $106 \times 106$  samples.
- The size reduction should be done [proportionally](#) in order not to confuse the user at navigating. Recommendation for pages with letter or A4 size is an one-eighth scale.

- The following [color spaces](#) can be used: mono (black and white), gray, direct RGB or indexed RGB (with full palette). AcrobatReader has problems to view some optimized graphics with a reduced palette.

## 8.2 pdfT<sub>E</sub>X

There is a powerful program to generate PDF files: [pdfT<sub>E</sub>X](#) written by Hàn Thế Thành. But there is no direct support for thumbnails in versions up to 0.13d. The only way I found is to include the thumbnails as images with [\pdfimage](#). The guessed object number of this image can be used to specify the thumbnail and it works.

The problem to set the thumbnail on the page can be solved by moving it outside the /MediaBox, e. g. `\maxdimen` apart from the margin). However I do not want to set the thumbnails in the document visible or invisible and the problem of guessing the object numbers remains.

## 8.3 Project thumbpdf

So I developed the project [thumbpdf](#):

1. The thumbnails are created by a program like [Ghostscript](#).
2. These thumbnails are included with [\pdfimage](#) in a T<sub>E</sub>X file that will be processed by pdfT<sub>E</sub>X to get the object representation of the thumbnails.
3. A [Perl script](#) extracts the objects and writes them in a T<sub>E</sub>X readable file (and optimizes the object structure).
4. A T<sub>E</sub>X package reads this file. It gets the object number by [\pdflastobj](#) and adds the thumbnail reference with this number to the page attributes ([\pdf-pageattr](#)).

The main work is done by the Perl script [thumbpdf.pl](#):

1. It calls Ghostscript to [generate the thumbnails](#) ([thumb???.png](#)). This step can be omitted, if the user wants to use his own thumbnail generating program.
2. It calls pdfT<sub>E</sub>X with [thumbpdf.tex](#) to generate the PDF file [thumbpdf.pdf](#). Each page contains a thumbnail included with [\pdfimage](#).
3. It scans the objects to [get the object data](#), optimizes them and writes them to the file [thumbdta.tex](#).

The package `thumbpdf.sty` reads the data in `thumbdta.tex` and inserts the thumbnails in the main document. The package works with plain or  $\text{\LaTeX}$  formats. If the data file `thumbdta.tex` is not created yet, if the number of thumbnails in the file does not suffice, or if the  $\text{\TeX}$  compiler is not  $\text{pdf\TeX}$ , then the package writes warnings only. So the user does not need to comment and uncomment the lines that includes the package to avoid error messages.

By the time of this writing  $\text{pdf\TeX}$  version 0.14a is available. Here it is possible to include an image and get its object number without setting it on the page. But I have not changed the method in `thumbpdf` in order not to get rid of the optimizations.

Adding thumbnails requires three steps, `jobname.tex` includes `thumbpdf.sty` (`\usepackage` or `\input`):

1. `pdf(la)tex jobname % Format required by jobname.tex.`
2. `thumbpdf jobname %`
3. `pdf(la)tex jobname % Here thumbpdf.sty is required.`

Other programs besides Ghostscript can be used:

- 2a. `OtherProgram jobname.pdf`
- 2b. `thumbpdf % without filename`

## 8.4 Other images as thumbnails

Other images, symbols, or pictograms in the formats PNG, TIFF, or JPEG can be used as thumbnails instead of the reduced image of the page. If some symbols or pictograms should be included instead of reduced images, they can be specified with `\thisthumb` in the document  $\text{\TeX}$  file. In the project `thumbpdf` the `thumbpdf.tex` picks up the thumbnail images. The extra symbols are announced to `thumbpdf.tex` by declaring them in the file `thumbopt.tex` that will be scanned, if it exists. This optional file consists of `\thumb` macros with an optional label and the filename as arguments. Although PNG, TIFF, and JPEG files can be used, only the file extension `.png` can be omitted.

The `names`, used in references later, can be the filenames themselves or labels, declared in the optional argument of `\thumb`:

<code>thumbopt.tex</code>		<code>jobname.tex</code>
<code>\thumb{one}</code>	→	<code>\thisthumb{one}</code>
<code>\thumb{one.png}</code>	→	<code>\thisthumb{one.png}</code>
<code>\thumb[two]{one}</code>	→	<code>\thisthumb{two}</code>
<code>\thumb[one]{./extras/one.png}</code>	→	<code>\thisthumb{one}</code>

## 9 Links

- PDF specification (version 1.3):  
<http://partners.adobe.com/asn/developer/PDFS/TN/PDFSPEC.PDF>
- pdf<sub>T</sub>E<sub>X</sub> (Hàn Thế Thành):  
<http://www.tug.org/applications/pdftex/>  
Win32: <ftp://ftp.eset-metz.fr/pub/TeX/win32-beta/>
- hyperref (Sebastian Rahtz):  
<ftp://ftp.ctan.org/tex-archive/macros/latex/contrib/hyperref/>  
If test versions are available, then they can be get from  
<http://www.tug.org/applications/hyperref/hyperref.zip>
- hy<sub>p</sub>bmsec (Heiko Oberdiek):  
<ftp://ftp.ctan.org/tex-archive/macros/latex/contrib/oberdiek/>
- thumbpdf (Heiko Oberdiek):  
<ftp://ftp.ctan.org/tex-archive/support/thumbpdf/>

## Address

Heiko Oberdiek

E-Mail: [oberdiek@uni-freiburg.de](mailto:oberdiek@uni-freiburg.de)

## Contents

1	Introduction . . . . .	1
2	Strings in T <sub>E</sub> X and PDF . . . . .	3
2.1	Strings of PDF format . . . . .	3
2.2	Strings in T <sub>E</sub> X . . . . .	3
2.3	T <sub>E</sub> X to PDF . . . . .	4
3	Encoding handling . . . . .	5
3.1	PD1 encoding . . . . .	5
3.2	Unicode encoding . . . . .	6
4	Module \pdfstringdef . . . . .	7
4.1	Expansion methods . . . . .	8
4.2	Macro \pdfstringdef . . . . .	8
4.3	Summary . . . . .	9
5	Package hyperref . . . . .	10
5.1	General document information . . . . .	10
5.2	Bookmarks . . . . .	10
5.2.1	Options . . . . .	10
5.2.2	Bookmarks by section commands . . . . .	11
5.2.3	Bookmarks by \addcontentsline . . . . .	12
5.2.4	Creating bookmarks with \pdfbookmark . . . . .	13
6	Replacement methods . . . . .	14
6.1	Providing an alternative with \texorpdfstring . . . . .	14
6.2	Hook for own macro redefinitions . . . . .	14
6.3	Package hypbmsec . . . . .	15
7	Additional features of PDF format . . . . .	16
7.1	General document information . . . . .	16
7.2	Bookmarks with other functions . . . . .	16
8	Thumbnails . . . . .	17
8.1	Restrictions . . . . .	17
8.2	pdfT <sub>E</sub> X . . . . .	18
8.3	Project thumbpdf . . . . .	18
8.4	Other images as thumbnails . . . . .	19
9	Links . . . . .	20