

IdeoType User Guide

2008 年 2 月 13 日

目次

Introduction	iii
Summary	iii
Objective	iii
Philosophy	iv
License	v
Acknowledgments	v
第 1 章 Setting Up	1
1.1 Hardware	1
1.2 Software	1
1.3 IdeoType Distribution	3
1.4 Installation	3
第 2 章 Usage	5
2.1 Manuscript Preparation	5
2.2 Generating PDF from Manuscript	6
2.3 Cleaning Up Files	6
2.4 Advanced Usage	7
2.5 Troubleshooting	7
第 3 章 Customization	9
3.1 Output Style Customization	9
付録 A Manuscript Notation	11
A.1 Title and Headings	11
A.2 Block and Inline Text	12
A.3 Lists	15

ii 目次

A.4	Unsupported	16
A.5	Graphics	16
A.6	Tables	18
A.7	Mathematical Expressions	19
A.8	Extensions	21

Introduction

Summary

Welcome. IdeoType is a software that helps you write, edit, and publish quality books easy.

Technically, it is a converter that converts XHTML-based manuscript to PDF. You can get real-time preview of your manuscript using web browser, and complete preview of print output on-the-fly using PDF viewer.

Fast feedback leads to better quality of the content. With IdeoType, you can iterate write-preview-edit cycle as much as, as fast as you like, unlike traditional book authoring process. IdeoType thus helps authors develop quality books.

In this document, we will describe how to use the software.

Objective

IdeoType's objective is simple: Help people make good books easy.

We love books. We want to make quality books easy.

Good books are:

- Easy to read
- Easy to understand

So we need:

- High quality typesetting
- Aid for extensive iterative elaboration (rewriting)

But existing tools do not meet our needs, though they work great for their purposes.

- Interactive DTP software are easy to use, but require human operators, thus consume time and money for each change.
- LaTeX or XML are batch-processable, but difficult to learn.

IdeoType tries to solve this by:

- Easy to use, open interface (XHTML).
- Automatic on-the-fly typesetting with batch-processing backend (LaTeX, XML, etc.)

IdeoType provides quality typesetting and ease of iterative elaboration by cherry-picking features from existing tools.

(Note that collaboration and other important things in creative process are not in its scope. They are achieved by other tools such as concurrent version control systems.)

Philosophy

Our principles are:

- Help humans concentrate on creative things. Let machines do routine things.
- More freedom and less cost.

Hence we consider the following as important.

- Open data, open code.
 - We can do whatever we want. Easy to use with other tools.
- Follow conventions. Use common interfaces.
 - Reduce learning/development/maintenance cost.
- Do not write code or invent rules unless really necessary.
 - Use existing tools. (= less work, less bugs)
 - Follow standards. (= better compatibility and portability)
 - Fight features. (= less burden and more manageability)
- Avoid dependency. Keep modularity.
 - Adaptable to future changes.

License

Copyright (c) 2005-2008, Project IdeoType. All rights reserved.
IdeoType is free software. See LICENSE for detail.

Acknowledgments

This software was being developed with grant from IPA ESP program, under the supervision by Professor Yoshiaki Mima (2006-Q3..2007-Q1, 2007-Q3..2008-Q1, see http://www.ipa.go.jp/jinzai/esp/*1). We appreciate those who helped us develop this software.

*1 <http://www.ipa.go.jp/jinzai/esp/>

第 1 章

Setting Up

Install required software first, then install IdeoType.

1.1 Hardware

1.1.1 Requirement

No special hardware needed. Any ordinary PC is fine.

1.2 Software

1.2.1 Requirement

IdeoType requires the following software to run (Debian package names we used for testing are in parentheses):

- Unix-like operating environment (Debian GNU/Linux 4.0 (etch))
- xsltproc (xsltproc 1.1.19)
- XHTML plus Math 1.1 DTD (w3c-dtd-xhtml 1.1, w3-dtd-mathml 2.0.0.0)
- Ruby (ruby1.8 1.8.5)
- Rake (rake 0.7.1)
- pLaTeX (ptex-bin 3.1.10^{beta3}+0.04b, ptex-jisfonts 2-19)
- Adobe CMaps (cmap-adobe-japan1 0+20040605, cmap-adobe-japan2 0+20020208)
- mendex (mendexk 2.6d)

2 第1章 Setting Up

- dvi_{pdf}mx (dvi_{pdf}mx 20050831)
- jsbook/jsclasses LaTeX class (okumura-clsfiles 2006.11.06)
- Computer Modern font (cm-super 0.3.3)
- Ghostscript (gs-esp 8.15.3, gs-cjk-resource 1.20021122)
- ImageMagick (imagemagick 6.2.4.5)
- RMagick (librmagick-ruby 1.15.8)
- pdffinfo (xpdf-utils 3.01)
- epstool (epstool 3.08)
- nkf (nkf 2.07)
- GNU wget (wget 1.10.2)
- HTML Tidy (tidy 20051018)
- unzip (unzip 5.52)
- time (time 1.7)
- Apache FOP (fop 1:0.94)

See `Depends:` field in `debian/control` in the distribution tarball for detail.

We use the following for development:

- Debian GNU/Linux unstable (sid)
- GNU make (make 3.81)
- Subversion (subversion 1.4)

We recommend you use OS's with packaging system to avoid dependency problem. Excuse us for inconvenience.

Note for Windows users: While IdeoType does not directly support Windows yet, it runs on Unix-like OS in VMware, coLinux or similar virtualization software.

Note for Mac OS X users: While IdeoType does not packaged for Mac OS X yet, it is runnable if required software are properly set up using MacPorts, Fink, etc. It is not practically usable yet, though. Help for packaging are welcome.

1.2.2 Recommended

These are not requirement, but recommended over other options.

- Debian GNU/Linux^{*1}

^{*1} <http://www.debian.org>

- Mozilla Firefox (1.5 and above, which supports MathML by default)
- MIT MathML fonts

1.2.3 Suggested

These are not required to run the software, but may be useful when used in conjunction with.

- nxml-mode, psgml-mode / Emacs (for editing XHTML/XML)
- Web Developer extension for Firefox (for validating XHTML with w3 validator service)
- OpenOffice.org Draw (for drawing diagrams)
- Adobe Acrobat (Professional, for making commercial-print ready PDF/X)
- Decent fonts

1.3 IdeoType Distribution

Everything is available at the following site, including binary packages and source code:

<http://sourceforge.net/projects/ideotype>^{*2}

1.4 Installation

1.4.1 Debian Package

Debian package is easy to install and also easy to uninstall. Download the .deb package and install it using dpkg.

```
$ sudo dpkg --install ideotype_x.y.z-*.deb
$ sudo apt-get install -f
$ sudo dpkg --install ideotype_x.y.z-*.deb
```

Do not forget to point to unstable and include contrib and non-free in your apt lines. If you would not want to use unstable and like to stay on stable, you can configure APT^{*3} to partly use packages from unstable.

^{*2} <http://sourceforge.net/projects/ideotype>

^{*3} <http://wiki.debian.org/AptPinning>

4 第1章 Setting Up

1.4.2 Tarball

We do not recommend installation from tarball. Try to use package management system. If you dare to install using tarball, use `make`. Use `DESTDIR` to change installation directory.

```
$ tar zxvf ideotype-x.y.z.tar.gz
$ cd ideotype-x.y.z
$ make
$ make install DESTDIR=/home/jdoe
```

第 2 章

Usage

The workflow is simple. Just get your manuscript ready and run the software.

1. Prepare manuscript.
2. Run IdeoType to generate PDF.

You can edit the manuscript and re-generate PDF as many times as you want. Clean up files when you are done. See the following sections for instruction.

2.1 Manuscript Preparation

IdeoType accepts XHTML-like manuscript format^{*1}. Technically, it is a subset of XHTML + MathML, expanded using class, id, title and other attributes. (A kind of microformats^{*2}.)

Note:

- Manuscript should be valid XHTML.
- Encoding must be UTF-8.
- Filename suffix has to be .xhtml.
- Table for layout, complex table, GIF images, ruby, etc. are not supported (yet).

Manuscript should look like this, for example:

^{*1} We use XHTML as manuscript format since (X)HTML is well supported by many people and software, and it can express rich information necessary for books. You can import HTML output from other tools such as word processors, blogs, Wikis, etc. (We will discuss it later (§2.4.1, p. 7).)

^{*2} <http://www.microformats.org/>

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1 plus MathML 2.0//EN"
"http://www.w3.org/TR/MathML2/dtd/xhtml1-math11-f.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
  <head>
    <meta http-equiv="Content-type" content="application/xhtml+xml;" />
    <meta name="author" content="John Doe" />
    <meta name="date" content="2038-02-31" />
    <title>Title</title>
  </head>
  <body>
    ...
  </body>
</html>
```

See the notation section (§A, p. 11) for detail.

2.2 Generating PDF from Manuscript

As of writing this, command-line interface is the default.

When your manuscript is *foo.xhtml*, type **ideotype build *foo.pdf*** to produce PDF.

You can try it out with the test data included in the distribution.

```
% cp -r /usr/share/doc/ideotype ./
% cd ideotype
% ls test/testdata.xhtml
test/testdata.xhtml
% ideotype build test/testdata.pdf
...
% ls test/testdata.pdf
test/testdata.pdf
%
```

2.3 Cleaning Up Files

To clean up intermediate files, type **ideotype clean *bookname***.

```
% ideotype clean test/testdata
```

To clean up all the generated files including PDF, type **ideotype distclean *bookname***.

```
% ideotype distclean test/testdata
```

2.4 Advanced Usage

2.4.1 Using Existing HTML Documents as Manuscript

To download HTML document accompanying images, etc. from the Web, type **ideotype download URL**.

```
% ideotype download http://example.org/foo.html
```

To import HTML document and accompanying images, etc., type **ideotype import filename**.

```
% ideotype import example.org/foo.html
```

Note that only clean HTML can be imported. This feature is still preliminary.

2.4.2 Render Specific Language Only

langs option lets you extract elements with specific `xml:lang` attribute (and language-independent common elements with no `xml:lang` attribute).

```
% ideotype --langs="en" build test/testdata.pdf
% ideotype --langs="all" build test/testdata.pdf
```

Note that invalid `xml:lang` value in the manuscript leads to improper result.

2.4.3 Per-Project Customization

You can configure your project setting via project configuration files (still an experimental feature).

```
% ideotype initproject .
% cp .book/html2latex.xsl.example .book/html2latex.xsl
% $EDITOR .book/html2latex.xsl
```

2.5 Troubleshooting

- Q.It does not convert my manuscript.

8 第2章 Usage

A. Make sure the manuscript is valid XHTML using validator^{*3}.

A. Building intermediate files helps troubleshooting.

% **ideotype build test/testdata.tex**

- Q. Is feature *X* supported?

A. Type **ideotype --help** to see the help message. If not found, sorry, not yet.

Your help would be highly appreciated.

^{*3} <http://validator.w3.org/>

第 3 章

Customization

3.1 Output Style Customization

You can customize the structure and style of your book, using project customization mechanism. If you like endnote better than footnote, overwrite the translation template. If you want fancier headings, add your own definition.

Project customization mechanism is under development. If you are interested, look into `$PROJECT/.book/`, which are generated when you **ideotype initproject**.

付録 A

Manuscript Notation

Manuscript notation for IdeoType is not much different from HTML, other than a few exceptions and extensions. (All these mappings are preliminary and may change in future.)

A.1 Title and Headings

A.1.1 head/title

`head/title` is title (for colophon, no style allowed).

```
<head>
  <title>Foo Tutorial</title>
</head>
```

A.1.2 head/meta

`head/meta` can set some meta info (e.g. `name="author" content="John Doe", name="date" content="2038-02-31"`)

```
<head>
  <meta name="author" content="John Doe"/>
</head>
```

A.1.3 h1 .. h6

`h1` is title (for front page, style allowed).

```
<h1><em>Foo</em> Tutorial</h1>
```

h2 is chapter, h3 section, h4 subsection, ...

A.2 Block and Inline Text

A.2.1 p

p is paragraph. `class="continued"` suppresses indentation of the first line.

```
<p>First paragraph.</p>
<p>Second paragraph.</p>
<p class="continued">Second paragraph continued.</p>
First paragraph.
Second paragraph.
Second paragraph continued.
```

A.2.2 address, blockquote, div

address, blockquote, div are same as HTML.

```
<address>info@example.org</address>
<blockquote>
  <p>If you do not think about the future, you cannot have one.
    -- John Galsworthy</p>
</blockquote>
<div> does not do much without class specified.</div>
info@example.org
  If you do not think about the future, you cannot have one. -- John Galsworthy
div does not do much without class specified.
```

A.2.3 pre

pre is preformatted code block, usually in typewriter font.

```
<pre>#include <stdio.h>

int main(void) {
    printf("hello, world\n");
    return 0;
}</pre>
```

```
#include <stdio.h>

int main(void) {
    printf("hello, world\n");
    return 0;
}
```

A.2.4 abbr, acronym

abbr and acronym make their title to be inserted as footnote.

```
<abbr title="World Wide Web">WWW</abbr> is an abbreviation while
<acronym title="radio detecting and ranging">rader</acronym>
is an acronym.
```

WWW^{*1} is an abbreviation while rader^{*2} is an acronym.

A.2.5 br

br is linebreak. Do not abuse it, as it causes an error in inappropriate places.

```
First line<br/> and the second line.
```

```
First line
```

```
and the second line.
```

A.2.6 cite

cite is a citation. Its title attribute becomes the key to lookup.

```
<cite title="doe2000">John Doe, "Foo Bar"</cite>
[?, John Doe, "Foo Bar"]
```

A.2.7 code

code is inline program code, usually in typewriter font.

```
function <code>foo()</code>.
```

```
function foo().
```

*1 World Wide Web

*2 radio detecting and ranging

A.2.8 dfn

`dfn` automatically generates index entry. If `class` includes `noindexterm`, indexing will be suppressed.

`<dfn>HTML</dfn>` stands for Hypertext Markup language.
HTML stands for Hypertext Markup language.

A.2.9 em

`em` is to emphasize, `strong` puts more emphasis.

`emphasized` or `strong`.
emphasized or **strong**.

A.2.10 kbd

`kbd` is user input (usually emphasized and in typewriter font). `samp` is output from computer (usually in typewriter font).

Type `<kbd>echo foo</kbd>` at the terminal.
Type **echo foo** at the terminal.

A.2.11 q

`q` is inline quotation.

She say `<q>Good bye</q>`, I say `<q>Hello</q>`.
She say “Good bye”, I say “Hello”.

A.2.12 span

`span` plays various roles depending on what `class` is specified.

`class` is used for expanding features.
`class` is used for expanding features.

A.2.13 a (with scheme)

`a href="http://..."` inserts the URL as footnote.

See `Example.org`.

See Example.org^{*3}.

A.2.14 a (with fragment)

`a href="#..."` inserts cross reference.

We will discuss the meaning of foo `later`.

We will discuss the meaning of foo later (§A.2.15, p. 15).

A.2.15 id

* `id="..."` is an identifier for cross reference and document inclusion.

`<p id="sec-foo">Foo is a sample name of anything.</p>`

Foo is a sample name of anything.

A.2.16 var

`var` is variable usually in *italic*.

Variable `<var>fname</var>` refers to a filename.

Variable *fname* refers to a filename.

A.2.17 del

`del` is deletion (strikethrough) and `ins` is insertion (underline).

`Deleted` and `<ins>inserted</ins>`.

Deleted and inserted.

A.3 Lists

Lists (`dl`, `ul`, `ol`) are the same as ones in HTML. Do not nest too deep though.

A.3.1 dl

`dl` is definition list. Multiple terms for single description is not supported yet.

^{*3} `http://example.org/`

```

<dl>
  <dt>running head</dt>
  <dt>running title</dt>
  <dd>A heading printed at the top of every (other) page of a book.</dd>
  <dt>date</dt>
  <dd>Specific time that can be named.</dd>
  <dd>A person with whom you are dating.</dd>
  <dd>Fruit of the date palm.</dd>
</dl>

```

running head

running title A heading printed at the top of every (other) page of a book.

date Specific time that can be named.

A person with whom you are dating.

Fruit of the date palm.

A.4 Unsupported

- `applet`, `object`, `map` are ignored.
- `form` is ignored.
- Some legacy styles (`b`, `big`, `i`, `small`, `sub`, `sup`, `tt`) are obsolete, though not ignored for now. Use alternatives instead, such as `em`.
- a `name="..."` is obsolete, though not ignored for now. Use `id` attribute instead.

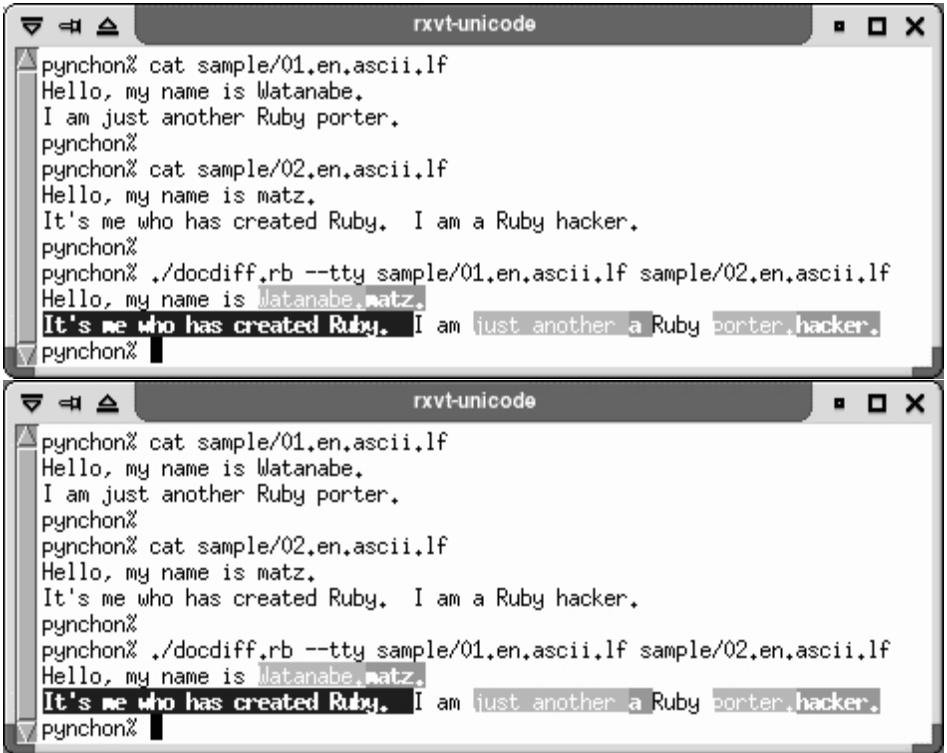
A.5 Graphics

A.5.1 Supported formats

- PNG (`.png`)
- JPEG (`.jpg`)
- EPS (`.eps`): for diagrams. Encoding of the embedded text should be in EUC-JP.
- PDF (`.pdf`): for diagrams. Encoding of the embedded text should be in Unicode (UTF-16?).

Note:

- No color support yet. Only grayscale images are supported for now.
- Set proper resolution to PNG and JPEG images. Or use `--force-conventional-resolution` option in `ideotype.rb`.



☒ A.1 Rxvt screenshots in PNG (former) and JPEG (latter).

A.5.2 img

`img src="..."` is inline graphics.

```
Book logo.
```



Book logo.

A.5.3 `div class="figure"`

`div class="figure"` makes floating figure. `div class="figure" / p class="caption"` is caption.

```
<div class="figure" id="fig-photo">  
    
  <p class="caption">  
    J. Peress' 1-atm dive suit, Tritonia, explored the Lusitania wreck i  
  </p>  
</div>
```

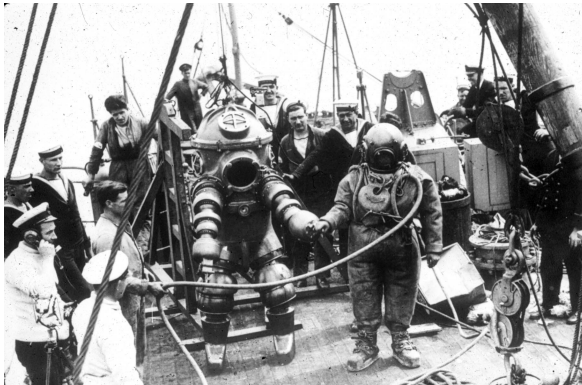


図 A.2 J. Peress' 1-atm dive suit, Tritonia, explored the Lusitania wreck in 1935.

A.6 Tables

Table support is preliminary. Avoid using tables if you can. You can always draw a table in drawing tool and embed it as figure.

- `table` is table.
- `table / caption` is table caption.

```

<table summary="table comparison">
  <caption>Comparison of tables.</caption>
  <tr><th></th><th>float</th><th>caption</th><th>nestable</th></tr>
  <tr><th>HTML table</th><td>no</td><td>yes</td><td>yes</td></tr>
  <tr><th>LaTeX tabular</th><td>no</td><td>no</td><td>yes</td></tr>
  <tr><th>LaTeX table</th><td>yes</td><td>yes</td><td>no</td></tr>
</table>

```

表 A.1 Comparison of tables.

	float	caption	nestable
HTML table	no	yes	yes
LaTeX tabular	no	no	yes
LaTeX table	yes	yes	no

A.7 Mathematical Expressions

Mathematical expressions are supported using MathML. Do not forget to put namespace identifier.

- `math xmlns="http://www.w3.org/1998/Math/MathML" display="block"` is display (block) math.
- `math xmlns="http://www.w3.org/1998/Math/MathML" display="block" id="foo"` is display math with number.
- `math xmlns="http://www.w3.org/1998/Math/MathML"` is inline math.

A.7.1 Math (Block)

```

<math xmlns="http://www.w3.org/1998/Math/MathML"
  display="block" id="eqn-block">
  <mfraction>
    <mrow><mi>d</mi></mrow>
    <mrow><mi>d</mi><mi>x</mi></mrow>
  </mfraction>
  <msubsup><mo>&int;</mo><mi>a</mi><mi>x</mi></msubsup>
  <mi>f</mi>
  <mfenced>
    <msup><mi>x</mi><mrow><mo>&prime;</mo></mrow></msup>
  </mfenced>
  <mi>d</mi><msup><mi>x</mi><mo>&prime;</mo></msup>
  <mo>=</mo>
  <mi>f</mi><mfenced><mi>x</mi></mfenced>
</math>

```

$$\frac{d}{dx} \int_a^x f(x') dx' = f(x) \quad (\text{A.1})$$

A.7.2 Math (Inline)

<p>The same equation can be placed inline:

```

<math xmlns="http://www.w3.org/1998/Math/MathML"
  id="eqn-inline">
  <mfraction>
    <mrow><mi>d</mi></mrow>
    <mrow><mi>d</mi><mi>x</mi></mrow>
  </mfraction>
  <msubsup><mo>&int;</mo><mi>a</mi><mi>x</mi></msubsup>
  <mi>f</mi>
  <mfenced>
    <msup><mi>x</mi><mrow><mo>&prime;</mo></mrow></msup>
  </mfenced>
  <mi>d</mi><msup><mi>x</mi><mo>&prime;</mo></msup>
  <mo>=</mo>
  <mi>f</mi><mfenced><mi>x</mi></mfenced>
</math> as well.</p>

```

The same equation (A.1, p. 20) can be placed inline: $\frac{d}{dx} \int_a^x f(x') dx' = f(x)$ as well.

A.8 Extensions

A.8.1 Frontmatter, Mainmatter, Appendix, Backmatter

`div class="frontmatter"` indicates the beginning of frontmatter. In frontmatter in general, section headings has no numbers and Roman numerals are used for page number.

`div class="mainmatter"` indicates the end of frontmatter and the beginning of frontmatter.

`div class="appendix"` indicates the end of normal chapters and the beginning of appendices. Appendices have section numbering different from normal chapters (e.g. Chapter 1 and Appendix A).

`div class="backmatter"` indicates the end of mainmatter and the beginning of backmatter.

```
<div class="frontmatter"/>
```

Frontmatter (title, colophon, dedication, foreword, preface, acknowledgments, table of contents, etc)

```
<div class="mainmatter"/>
```

Chapters

```
<div class="appendix"/>
```

Appendices

```
<div class="backmatter"/>
```

Backmatter (index, etc.)

A.8.2 Table of Contents

`div class="toc"` inserts table of contents (§, p. i).

```
<div class="toc"/>
```

A.8.3 Index

`div class="index"` inserts index.

```
<div class="index"/>
```

a `class="indexterm"` makes index entry.

```
<a class="indexterm">some term</a>
some term
```

You can explicitly specify sort key by `title` attribute. If you need a hierarchical index, specify parent entry by `href` and `rel` attribute.

```
Sing and play are both <a class="indexterm" title="verb" id="idx-verb">v
<a class="indexterm" title="sing" href="#idx-verb" rel="parent">Singing<
Sing and play are both verbs. Singing is fun.
```

A.8.4 File Inclusion

a with `include` class causes file inclusion.

- a `href="foo" class="include"` inserts the file as plain text.
- a `href="foo#..." class="include"` inserts the file as partial manuscript (inclusion starts at the element with that id).

A.8.5 Exclusion

Elements with `exclude` class causes the element to be excluded from the output. Labels (anchors) for cross reference are omitted regardless of `id` attribute. When you need labels, specify `label` class explicitly.

```
<p>We have apples<span class="exclude"> and bananas</span>.</p>
<p class="label exclude" id="orange">They have oranges.</p>
We have apples.
```

A.8.6 Non-escaped Text

Elements with `noescape` class will be passed on to the backend without escaping. Use with caution.

```
<p>Escaped (default): \fbox{\LaTeX?}.
Not escaped: <span class="noescape">\fbox{\LaTeX?}</span>.</p>
Escaped (default): \fbox{\LaTeX?}. Not escaped: LATEX?.
```

A.8.7 Footnotes

a is also used for footnotes.

a href="#foo" class="footnotemark" inserts a footnote mark. ul class="footnotes" / li class="footnotetext" is footnote text.

```
<p>
  Footnote <a href="#fn-fnmark" class="footnotemark">marks</a> and
  footnote <a href="#fn-fntext" class="footnotemark">texts</a> must be i
</p>
<ul class="footnotes">
  <li class="footnotetext" id="fn-fnmark">Footnote mark.</li>
  <li class="footnotetext" id="fn-fntext">Footnote text.</li>
</ul>
```

Footnote marks^{*4} and footnote texts^{*5} must be in pairs.

A.8.8 Language Translation Support

Manuscript elements can be filtered by `xml:lang` attribute. This is useful for translation project.

* `xml:lang="..."` sets language.

```
<p>
  <span xml:lang="en">English text.</span>
  <span xml:lang="ja">日本語のテキスト.</span>
</p>
```

English text.

^{*4} Footnote mark.

^{*5} Footnote text.

索引

HTML, 14

some term, 22

verbs, 22

 Singing, 22