

The Mandoku text format

Christian Wittern

March 12, 2014

Conventions for the text format used by Mandoku are made up of three parts:

- Emacs org-mode format
- convention about how texts are organized into files
- Additional syntactical conventions singular to Mandoku within the files

1 The text format is based on Emacs org-mode

The text editor Emacs has the concept of "modes", which allows for specialized modes of editing depending on the type of text to be edited. This works for various programming languages and helps the programmer to avoid mistakes by giving graphical hints through formatting and colors and in general support the task of writing in this language.

Emacs org-mode (see Org-Mode: Your Life in Plain Text[orgmode.org]) builds on this concept, but the target is not a specific programming language but plain text with some outlining, formatting and hyperlinking conventions. This makes it a lightweight text format that covers some middle ground between absolutely complete plain text (if such a thing exists) and complex markup languages like XML. Since it even supports footnotes and is easily converted to HTML, Latex and OpenOffice format, it is useable even for research articles.

To make this possible, org-mode introduces a few special markup constructs. Some of the assumptions of this plain text according to org-mode are as follows:

- Headings of all kinds are expressed by placing stars "*" in the leftmost column of the text. The number of stars indicates the level of the

heading in the overall outline of the document. The stars have to be followed by a space in order to be recognized as a part of the outline. Structurally, a heading defines a subsection.

* Outline level one レベル 1 の見出し
** Outline level two レベル 2 の見出し
*** Outline level three レベル 3 の見出し

- following a heading a so-called "property drawer" may appear; this is a container for additional information associated with the subsection.

* Introduction
:PROPERTIES:
:CUSTOM_ID: intro
:END:

- Headings can also have "tags" attached to them, which marks them for example for various analytical purposes, they appear on the headline towards the right margin.

* Outline level one レベル 1 の見出し :important:

- Paragraphs are indicated by at least one empty line.
- Distinctive parts of the text, for example quotations or parts in verse can be set apart using a specific section marker:

即於佛前以偈頌曰。

#+BEGIN_VERSE
目淨脩廣如青蓮
心淨已度諸禪定
久積淨業稱無量
導眾以寂故稽首
#+END_VERSE

- Lines beginning with a pound sign '#' are treated specially, not as part of the text:
 - if followed by a "+" they introduce special keywords or syntax used for org-mode's own purposes, for example #+TITLE, #+AUTHOR or #+DATE.

- the pound sign followed by a space character "# " introduces a comment which is not considered part of the document.
- Links to other files target.txt or places within the current file are constructed like this: `[[target.txt][link to target.txt]]`

2 Additional conventions for Mandoku

2.1 Text identifiers

As of September 2013, Mandoku switched to a new type of text identifiers. This identifier, such as ZB6i0076 is composed of three parts:

- "ZB", which is the repository identifier – this indicates the origin of the text
- "6i", which indicates in which subcollection the text is found, in this case "經集部類", which is part of the top-level grouping "6 佛部".
- "0076" which is the serial number of the text in the collection.

Most texts of the part "6 佛部" of the ZB repository do derive from the CBETA collection of Buddhist texts, which has a different set of text identifiers, in this case, the text in question has the title "維摩詰所說經" and the CBETA identifier T14n0475, this was also used in Mandoku in previous versions.

2.2 About texts, versions and editions

In Mandoku every text has a text identifier, for example ZB6i0076=, which uniquely identifies the text within a larger collection and makes it possible to refer to it unambiguously. This identifier is usually also the name of the folder, where all files that make up a text are stored. The text is usually split up so that one file has the content of one scroll, and they are numbered in sequence. An additional file may give the table of contents with links to the appropriate location in the files.

```
ZB6i0076_000.txt
ZB6i0076_001.txt
ZB6i0076_002.txt
ZB6i0076_003.txt
ZB6i0076.org
```

The source of the text has to be given (at the beginning of the file, this is recorded with the property keyword "WITNESS") and there might also be a base edition (keyword "BASEEDITION"). The base edition provides the navigation grid for all editions of a text.

Several editions of a text might be recorded; this allows the documentation of textual witnesses for a text and provides the ground for critical editions of the text. Within **Mandoku**, different editions (witnesses) of a text are stored as "branches" in a version control system.

In order to provide a common reference system common to all editions of a text, page numbers and locations of the line break of the base edition are recorded in all files. Editions that have a different layout may additionally record the page numbers according to the source edition.

2.3 Syntactic conventions

2.3.1 Base edition and text witness

In **Mandoku** some import information about the source and edition of a text can be given at the beginning of a file in machine readable form, for example:

```
#+PROPERTY: BASEEDITION T
#+PROPERTY: ID ZB6i0076
#+PROPERTY: CBETA_ID T14n0475
#+PROPERTY: WITNESS 【CBETA】
#+PROPERTY: JUAN 1
```

Here the base edition is identified as "T" and the witness (the edition in this file) is given as "【CBETA】". In addition, the identification number of the text is recorded and the number of the juan. As mentioned above, the text identifiers have been changed recently, therefore the previously used CBETA_ID is also given for reference.

2.3.2 Page numbers

The page number is given in the following form: There are three parts to this number, separated by the "_" character, surrounded by the page break indicator "":

- "ZB6i0076" is the identification number of the text
- "T" is the identifier of the edition

- "0537a" is the page number ("a" stands for the upper part of the text or the first half of a page in woodblock prints that are separated in two halves) At the moment **Mandoku** requires the page number to end in a letter.

The page break indicator ": on a line by themselves and are ended with :END: again on a line by themselves. In interactive use, the content of drawers can be hidden and made to appear only when needed.