

Complex Spatial Digitization Tasks for the SAT Project

Shigeki Moro

The Association for Computerization of Buddhist Texts(ACBUT), Japan

0. Abstract

The SAT project, which is digitizing the *Taishō Shinshū Daizōkyō* 大正新脩大藏經 in Japan, aims, in cooperation with the Chinese digital Buddhist canon society CBETA (located in Taiwan), at the construction of a new highly accurate electronic Buddhist canon. In the work of digitizing the *Taishō*, we have already dealt with many of the basic problems, such as encoding and missing characters. But there are also very often more complex issues involved, such as in the case where the actual printed source has a wide range of spatial, graphically-oriented styles, which play a seminal role in expressing the author's theme. In this paper, I would like to offer a means for resolving the complexities that arise in the digitization of a text such as Euisang's *Chart of the Dharma-realm of the Single Vehicle of the Huayan* (화엄일승법계도, 華嚴一乘法界圖), and documents that include scores and other complex shapes found in the *Taishō*. This will be done by using the markup method named SVG.

1. Introduction: Limitation of Plain Text

The SAT project, which is digitizing the *Taishō Shinshū Daizōkyō* 大正新脩大藏經 in Japan, aims at the construction of a highly accurate new electronic Buddhist canon. In the work of digitizing the *Taishō*, we have already dealt with many of the basic problems, such as that of encoding¹ and missing characters.² But there are also very often more complex issues involved, such as in the case where the actual printed source has a wide range of spatial, graphically-oriented styles, which often play a seminal role in expressing the author's thoughts.

¹ Shigeki Moro. “Tag-tsuki gengo to moji-code タグ付き言語と文字コード”. *Internet-jidai no moji-code* インターネット時代の文字コード. Tokyo: Kyoritsu-shuppan, 2001.

² Shigeki Moro. “On the Missing-Characters (GAIJI) of the Taisho Tripitaka Text Database Published by SAT”. *Proceedings of 1999 EBTI, ECAI, SEER & PNC Joint Meeting* 太平洋鄰里協會一九九九年會論文集. Taiwan: Computer Center of the Academia Sinica, 1999.

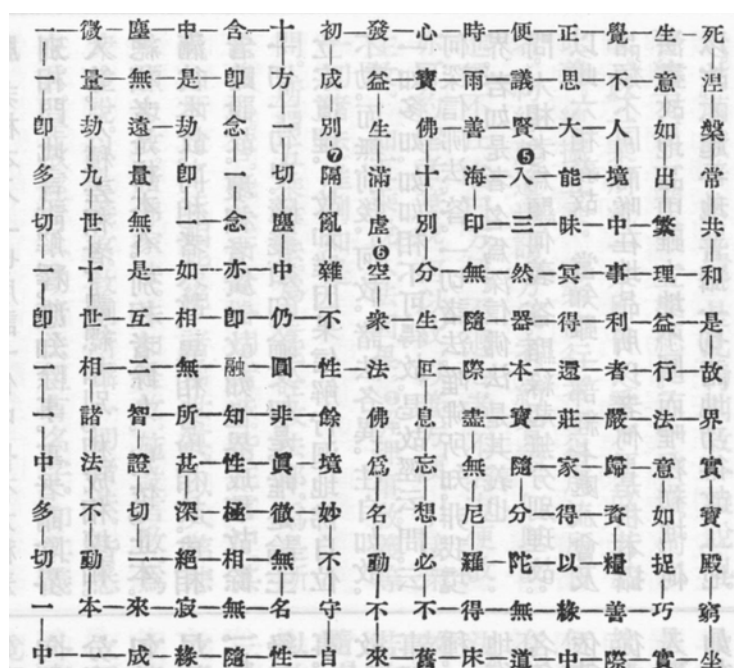


fig. 1: the Chart of the Dharma-realm of the Single Vehicle of the Huayan [T45: 711a]

For instance, the *Chart of the Dharma-realm of the Single Vehicle of the Huayan* (화엄일승 법계도, 華嚴一乘法界圖) written by Euisang 義湘 (625-702), one of the most eminent early Silla新羅 scholar-monks, is a graphical figure as well as a text (fig. 1). For this reason, it is very difficult to digitize it in plain text format. The following is a thinkable solution to arrange the *Chart* in plain text format.

1887A, 45, 0711a11: 死—涅—槃—常—共—和 是—故 界—實—實—殿—窮—坐
 1887A, 45, 0711a12: | | | | | | | | | | | |
 1887A, 45, 0711a13: 生 意—如—出—繁 理 益 行 法 意—如—捉—巧 實
 1887A, 45, 0711a14: | | | | | | | | | | | |
 1887A, 45, 0711a15: 覺 不 人—境 中 事 利 者 嚴 歸 資—檀 善 際
 1887A, 45, 0711a16: | | | | | | | | | | | |
 1887A, 45, 0711a17: 正 思 大 能 昧 冥 得 還 莊 家 得 以 錄 中
 1887A, 45, 0711a18: | | | | | | | | | | | |
 1887A, 45, 0711a19: 便 議 賢 入 三 然 器 本 實 隨—分 陀 無 道
 1887A, 45, 0711a20: | | | | | | | | | | | |
 1887A, 45, 0711a21: 時 雨 善 海—印 無 隨 際 盡—無—尼—羅 得 床
 1887A, 45, 0711a22: | | | | | | | | | | | |
 1887A, 45, 0711a23: 心 寶 佛—十—別—分 生 匠—息—忘—想—必—不 舊
 1887A, 45, 0711a24: | | | | | | | | | | | |
 1887A, 45, 0711a25: 發 益—生—滿—虛—空—衆 法 佛—爲—名—勤—不—來
 1887A, 45, 0711a26: | | | | | | | | | | | |
 1887A, 45, 0711a27: 初—成—一—別—隔—亂—雜—不 性 餘—境 妙—不—守—自
 1887A, 45, 0711a28: | | | | | | | | | | | |
 1887A, 45, 0711a29: 十—方—一—切—塵—中 仍 圓 非 眞 徹 無—名 性
 1887A, 45, 0711a30: | | | | | | | | | | | |
 1887A, 45, 0711a31: 含 即—念—一—念 亦 即 融 知 性 極 相 無 隨
 1887A, 45, 0711a32: | | | | | | | | | | | |
 1887A, 45, 0711a33: 中 是 劫—即—一 如 相 無 所 甚—深 絕 寂 錄
 1887A, 45, 0711a34: | | | | | | | | | | | |
 1887A, 45, 0711a35: 塵 無 遠—量—無—是 互 二 智—證—切—一 來 成

1887A,45,0711a36: | | | | | |
 1887A,45,0711a37: 微 量—劫—九—世—十一—世 相—諸—法—不—動—本 —
 1887A,45,0711a38: | | | | | |
 1887A,45,0711a39: ————即—多—切———即—————中—多—切———中

The above example is similar to that of CBETA. However, using this format, we cannot retrieve terms such as “法性” or “一即一切” which are very important for the Huayan/Hwaeom school, because the beginning character of the *Chart* “法” is not next to the second character “性”, and the box-drawing symbols also separate the characters from each other. As I have already noted, it is unnatural to represent two-dimensional layout through a one-dimensional format like plain text.³

In this paper, I would like to offer a means for resolving the complexities that arise in the digitization of a text such as the *Chart*, the musical scores for Shōmyō 聲明, and other complex shapes as found in the *Taishō*. This will be done by using the markup method named *Scalable Vector Graphics* (SVG), which is an open technology for describing two-dimensional vectors and mixed vector/raster graphics in XML.

2. Complex text encoding with SVG

2-1 SVG

SVG is an open-standard vector graphics language created as the XML vector-graphics format for the next-generation Web under the World Wide Web Consortium (W3C) and its members include Adobe, Apple, Corel, HP, IBM, Macromedia, Microsoft, Netscape, OASIS, Open Text, Quark, Sun, Xerox, etc. along with staff from the W3C. SVG is currently at the stage of Candidate Recommendation.⁴

A number of SVG viewers and editors exist.⁵ Two major browsers, Netscape 4.x/6 and Microsoft Internet Explorer 5.x, can display SVG vector-graphics with the Adobe SVG Viewer plug-in. The Mozilla project also supports SVG. Adobe Illustrator 9 would be the easiest tool to encode SVG.

2-2 The *Chart* in SVG

According to Ishii Kosei, the *Chart* was written under the influence of a popular style of

³ Japan Association for East Asian Text Processing (JAET) ed. *Dennō Chūgoku-gaku* 電腦中國學. Tokyo: Kōbun-shuppan, 1998. p. 197.

⁴ <http://www.w3.org/TR/2000/CR-SVG-20001102/>

⁵ For more information on the implementations of SVG, see the W3C official list on web: <http://www.w3.org/Graphics/SVG/SVG-Implementations>

Chinese poems in the Tang period.⁶ The following is an example of encoding the *Chart* in SVG.

```
<?xml version="1.0" encoding="utf-8"?>
<!DOCTYPE svg PUBLIC "-//W3C//DTD SVG 20000303 Stylable//EN"
"http://www.w3.org/TR/2000/03/WD-SVG-20000303/DTD/svg-20000303-
stylable.dtd">
<svg xml:space="preserve" width="4.8in" height="4.2in"
viewBox="0 0 480 420">

<!-- Text Area -->
<g>
  <text x="230" y="230">法</text>
  <text x="200" y="230">性</text>
  <text x="170" y="230">圓</text>
  <text x="140" y="230">融</text>
  <text x="110" y="230">無</text>
  <text x="80" y="230">二</text>
  <text x="50" y="230">相</text>
  (...)
  <text x="260" y="410">舊</text>
  <text x="230" y="410">來</text>
  <text x="230" y="380">不</text>
  <text x="230" y="350">動</text>
  <text x="230" y="320">名</text>
  <text x="230" y="290">爲</text>
  <text x="230" y="260">佛</text>
</g>

<!-- Graphics Area -->
<g style="stroke:#990000;stroke-width:1.5;">
<path d="M 228 226 h -14 m -16 0 h -14 m -16 0 h -14 m -16 0
h -14 m -16 0 h -14 m -16 0 h -14 z"/>
<path d="M 56 234 v 14 z"/>
<path d="M 56 264 v 14 m 0 16 v 14 m 0 16 v 14 m 0 16 v 14 m 8
8 h 14 m 16 0 h 14 m 16 0 z"/>
<path d="M 124 376 h14 z"/>
  (...)
</g>
</svg>
```

Each character of the *Chart* is tagged by a <text> tag with attributes representing its location. It is especially important that the characters in the example above were arranged in the order in which we read it in the *Taishō*. Consequently we can retrieve the terms or phrases of the *Chart*. The bars connecting the characters are no longer the barriers, because they are described by the empty element tags named <path>.

An SVG file can be embedded not only in a XML file but also in a HTML file using an

⁶ Ishii Kosei. *Kegon shisō no kenkyū* 華嚴思想の研究. Tokyo: Shunjū-sha, 1996. pp. 217-222.

<OBJECT> tag or <EMBED> tag. The following shows the HTML example of the *Chart* in which <OBJECT> includes the text of the *Chart* as its element for the user agents that cannot handle the <OBJECT> tag.

```
<OBJECT data="hokkaizu.svg" type="image/svg+xml" height=480
width=420>
法性圓融無二相
諸法不動本來寂
(...)
窮坐實際中道床
舊來不動名為佛
</OBJECT>
```

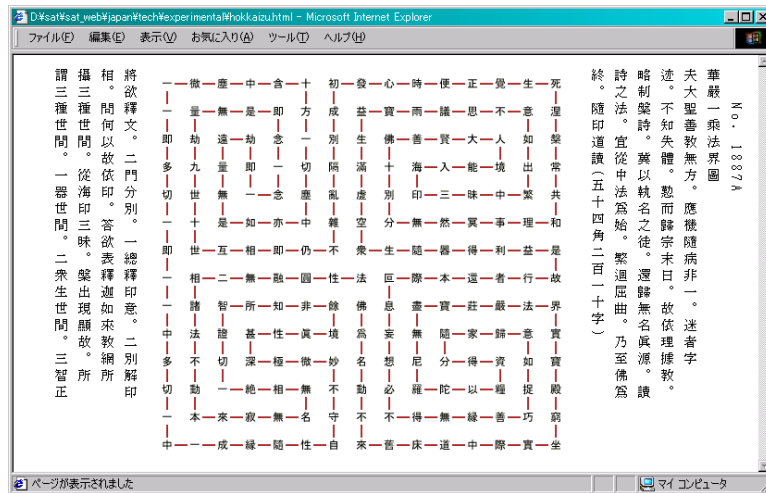


fig. 2: The *Chart* on Internet Explorer 5.5 with the Adobe SVG Viewer⁷

2-3 Is this a character?

In esoteric Buddhist texts, we often encounter complex figures in which texts and graphics are mixed together. The following is the figure of *The Foshuo Beidoujixing Yanming Jing* 佛說北斗七星延命經 (T. No. 1803) describing The Big Dipper, in which each star has its name, personified figure, and talismanic drawing (*zhoufu* 呪符) which seems to be relevant to Daoism, with the exception of the sixth star who is attended with the smaller star.

⁷ Vertical layout is a property only of Internet Explorer 5.5 for Windows (see <http://msdn.microsoft.com/workshop/author/dhtml/reference/properties/writingMode.asp>). Vertical text is currently on the agenda for the next level of CSS (see <http://www.w3.org/TR/WD-i18n-format/>).

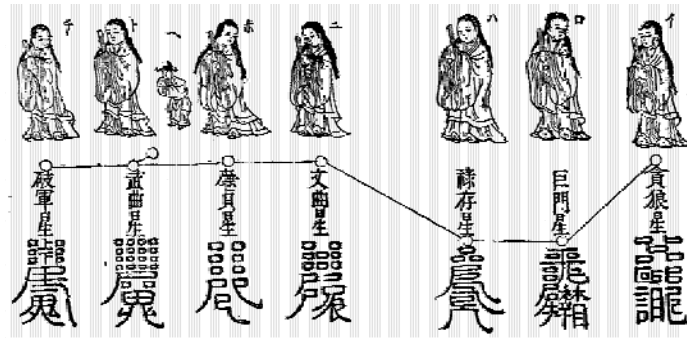


fig. 3 the Esoteric drawing of the Big Dipper (T21: 425b)

The names of the stars (“貪狼星” etc.) and Katakana letters (“イ”, “口”, “ハ”, etc.) for representing the order of the stars should be encoded using <text> tags; on the other hand, the drawing of the constellation, personified figures and Taoist-like *zhoufu* might be encoded as vector or raster graphics.

Looking at *zhoufu* in detail, however, we can find that they are constructed with some Chinese characters. For example, in the seventh *zhoufu*, we can find two Chinese characters, “上” and “下” (fig. 4).



fig. 4 the seventh talismanic drawing of the Big Dipper

It would be difficult to regard these parts as Chinese characters or not. Sakade Yoshinobu interprets the words “不空” in the twelfth *zhoufu* on p. 3874 as representing the author of the *zhoufu*.⁸ The digitization of *zhoufu* awaits further studies.

2-4 Order of Characters

The following quote from *The Jiyao Rangzai Jue* 七曜攘災決 (T. No. 1804) are figures in which Chinese characters are arranged in the shapes of humans. The characters in the right figure are the stars (e.g. *mao* 昴 means the Pleiades). This shows the correspondence between an astronomical body and a human body.

⁸ Sakade Yoshinobu. “Shoki-mikkyō to dōkyō tonō kōshō 初期密教と道教との交渉”. *Series Mikkyō 3: Chūgoku-Mikkyō*. Ed. Tachikawa Musashi and Yoritomi Motohiro. Tokyo: Shunjūsha, 1999. p. 167.



fig. 5 Figure-like Characters [T21: 428b]



fig. 6 Figure-like Characters in SVG

At first sight, the order of the characters in the figures seems very simple: head to foot. The text, however, shows an unexpected order.

宿度法

東方七十五度

北方九十八度

西方八十度

南方百一十三度

角 後井六至井十八
前女初至虛初

斗 後翼三至十五
前婁七至胃七

奎 後箕三至斗四前
井三十二至柳七

井 後危十五至室初
前九四至輪七

亢 後井十八至三十
前女十至虛八

牛 後軫十一至角六
前胃三至昴二

婁 後斗八至斗二十
前柳七至星四

鬼 後奎六至婁二前
九八至輪十一

鶉 後井二十七至柳三
前虛十一至危十二

女 後角二至亢二
前昴初至畢二

胃 後斗二十至牛六
前星二至張八

柳 後奎十至婁二前
輪十四至尾初

房 後柳五至星三
前危五至危末

虛 後九二至六輪五
前畢初至十二

昴 後朱末至女末
前張十至翼二

星 後婁九至胃九前
心初至尾十八

心 後柳十至星末
前危十至室五

危 後輪四至房初
前嘴末至井三

畢 後女十一至虛末
前翼七至軫初

張 後胃四至昴二前
尾十五至箕九

尾 後柳末至張五前
室十一至壁七

室 後心初至尾八
前井七至十九

嘴 後危四至十六
前翼九至軫三

翼 後胃九至畢十
前斗四至十六

箕 後張十一至翼四
前壁六至奎九

壁 後尾十二至箕二
前十六至二十八

參 後尾六至室初
前翼末至軫三

軫 後畢末至井初前
斗二十一至牛七

(T21:427b18-427c6)

This shows that the order begins from *jiao* 角 and ends in *zhen* 軫. Thus the characters should be encoded in the following order:

```

<text transform="matrix(1.0002 0 0 1 120.8242 69.4648)"><tspan
x="0" y="0" style="&st29; &st33;">角</tspan></text>
<text transform="matrix(0.7071 -0.7071 0.7071 0.7071 119.6113
94.1313)"><tspan x="0" y="0" style="&st29;">斗</tspan></text>
<text transform="matrix(0.6589 -0.7522 0.7522 0.6589 105.748
148.8755)"><tspan x="0" y="0" style="&st29;">奎</tspan></text>
<text transform="matrix(0.6771 0.7359 -0.7359 0.6771 139.6387
139.6978)"><tspan x="0" y="0" style="&st29;">奎</tspan></text>
(...)
<text transform="matrix(0.552 -0.5535 0.7081 0.7061 121.1201
63.209)"><tspan x="0" y="0" style="&st29; &st35;">翼
</tspan></text>
<text transform="matrix(0.6286 0.6518 -0.7198 0.6942 125.25
87.4233)"><tspan x="0" y="0" style="&st29; &st12;">箕
</tspan></text>
<text transform="matrix(0.6007 -0.6213 0.7189 0.6951 112.4795
139.4077)"><tspan x="0" y="0" style="&st29; &st16;">壁
</tspan></text>
<text transform="matrix(0.9444 0 0 1 127.0391 30.5591)"><tspan
x="0" y="0" style="&st29;">參</tspan></text>
<text transform="matrix(0.6259 -0.7799 0.7799 0.6259 105.2861
77.3389)"><tspan x="0" y="0" style="&st29;">軫</tspan></text>

```

The characters in fig. 6 should be also arranged the same as the stars (頤→脇→膝→賴→胸→脊→脛→脛→鼻→臆→腸→足→齒→臂→臀→頭→頂→臂→膀→膀→額→肩→心→腿→眉→肩→脇→腿→賴→手).

3. Conclusion

To sum up the characteristics of SVG we have seen thus far, we can say that SVG is a powerful solution for digitizing documents including both texts and graphics that are constructed in a two-dimensional layout.

However, attention must be paid to the fact that the relationship between the texts or letters in an example is linear, in other words, one-dimensional. SVG does not suit non-hierarchical figures, such as word puzzles, palindromes, or some kind of map in which place names are written, and overlapping-hierarchical figures, such as some kind of family tree or Dharma-lineage chart, etc. The mark-up scheme for these kinds of documents still remains a matter to be discussed.⁹

⁹ David T. Barnard et al. proposed some SGML-based solutions for complex structures (“Hierarchical

Encoding of Text: Technical Problems and SGML Solutions”. *Computers and the Humanities* 29.3 (1995): 211-231). They include a solution using CONUR, which is not available in XML. C.M. Sperberg-McQueen and Claus Huitfeldt also proposed a solution named GODDAG which is not a SGML-style scheme in order to markup overlapping structures (“GODDAG: A Data Structure for Overlapping Hierarchies”. <http://jefferson.village.virginia.edu/ach-allc.99/proceedings/sperberg-mcqueen.html>).