Data Specification for Patent Abstracts of Japan

February, 2012

1. Media and the file recording method

(1) The type of media is CD-ROM.
(2) Data comprise of text data including bibliographic data and the text body and image data including drawings.
(3) “UTF-82 is used for the character code of the text data.
(4) The format of image data is “TIFF”.

2. Each file’s element

2.1 Text file

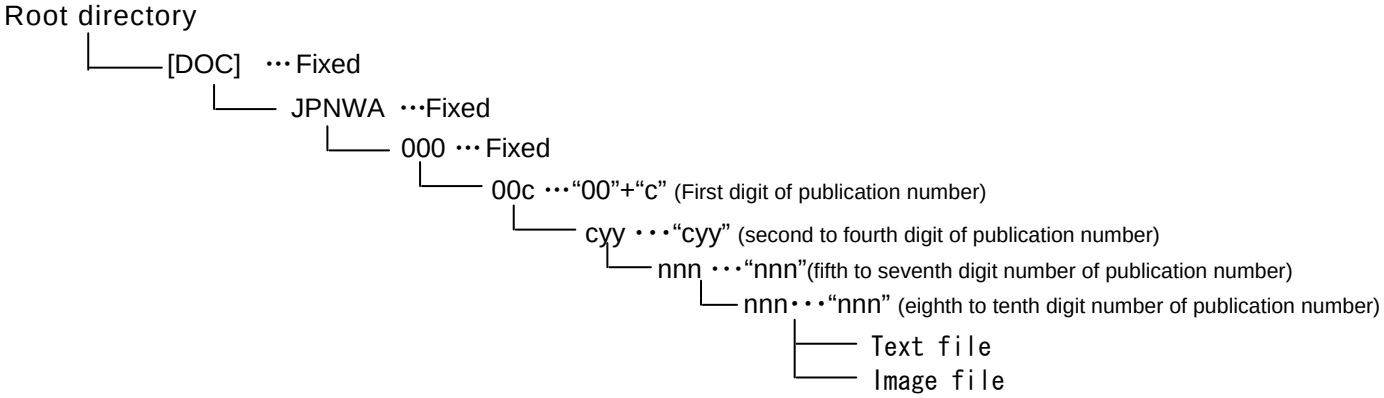
(1) It is described in XML format.
(2) The recorded subdocument comprises of three parts, bibliographic data, an abstract and drawings. However, if the drawing does not exist, the subdocument of drawing is not recorded.
(3) Refer to editing standards of PAJ text data for more details of the text file.

2.2 Image file

(1) Refer to editing standards of PAJ image data for more details of the image file.

3. Directory structure

The directory structure is as follows.



* “ccyynnnnnn” is corresponding to “Publication Number ccyy-nnnnnn”.

4. Title of each file

The title of each file is as follows.

・Text file: JP00000ccyynnnnnnA00000000
"ccyynnnnnn" is a 10 digit publication number.
Other parts are fixed.

・Image file: 00000001
Fixed

5. Volume identifier of CD-ROM

Basic volume descriptor of CD-ROM is as follows.

(a)Definition of Disk Contents
・Volume Identifier is defined for an individual title of CD-ROMs as follows.

P	A	J	2	0	1	1	0	1	1
(1)			(2)				(3)	(4)	

(1)Fixed as "PAJ"
(2)Issuance year of PAJ/CD-ROM
Example: Issuance of February, 2012 is described as 2012.
(3) "0(zero)"
(4) Serial number starting from 1(one) for every issuance year (fiscal year) of PAJ/CD-ROM
Example: Issuance of February, 2012 (11th CD-ROM of 2011FY)
is described as 11.

(b) Volume Set Identifier is fixed as "PAJ_CD_ROM".

(c) Publisher Identifier is fixed as "JPO & INPIT".

(d) Application Identifier is fixed as "MIMOSA V6.0".

Editing Standard of PAJ Image Data

Item Number	Level	Item Name	Attribute	Number of Bytes	Item Value	Hexadecimal Form	Content
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)

(1)	Item Number	Serial number provided to each standard item
(2)	Level	Level representation to each standard item
(3)	Item Name	Name of each standard item
(4)	Attribute	Attribute of each standard item
		C: 1 byte code (alphanumeric characters or katakana characters)
		N: right-align numeral without leading 0 (zero)
		K: 2 byte code (Kanji character)
		X: depending on the value content (no change in attribute)
		B: binary
(5)	Number of Bytes	Number of byte for each standard item
(6)	Item Value	Value of each item
(7)	Hexadecilam Form	Value of each item
(8)	Content	Representing the meaning of each item and value

Editing Standard of PAJ Image Data

Item Number	Level	Item Name	Attribute	Number of Bytes	Item Value	Hexadecimal Form	Content
1	05	Byte order	C	2	'II'		Intel Byte order where the least significant bit comes first
2	05	TIFF version number	B	2	'42'	X '2A00'	TIFF version number
3	05	Pointer to 1st directory	B	4	'8'	X '08000000'	Pointer to 1st directory
4	05	Number of TAGS	B	2	'17'	X '1100'	Number of tags
5	05	TAG1 (New Subfile Type)					Type of image data
	10	TAG ID	B	2	'254'	X 'FE00'	New subfile type
	10	Data Type	B	2	'4'	X '0400'	Data type (1=byte, 2=ASCII, 3=short, 4=long, 5=rational number)
	10	Number of Units	B	4	'1'	X '01000000'	Number of units (showing the amount of data)
	10	Tag Data or Pointer to Tag Data	B	4	'0'	X '00000000'	Tag data or pointer to tag data
6	05	TAG3 (Image Width)					Number of horizontal dots
	10	TAG ID	B	2	'256'	X '0001'	Width of Image
	10	Data Type	B	2	'4'	X '0400'	Data type (1=byte, 2=ASCII, 3=short, 4=long, 5=rational number)
	10	Number of Units	B	4	'1'	X '01000000'	Number of units
	10	Tag Data or Pointer to Tag Data	B	4	'XX'		Tag data or pointer to tag data (Number of horizontal dots in selected drawing of electric publication)
7	05	TAG4 (Image Length)					Number of vertical lines

Editing Standard of PAJ Image Data

Item Number	Level	Item Name	Attribute	Number of Bytes	Item Value	Hexadecimal Form	Content
	10	TAG ID	B	2	'257'	X '0101'	Length of Image
	10	Data Type	B	2	'4'	X '0400'	Data type (1=byte, 2=ASCII, 3=short, 4=long, 5=rational number)
	10	Number of Units	B	4	'1'	X '01000000'	Number of units
	10	Tag Data or Pointer to Tag Data	B	4	'XX'		Tag data or pointer to tag data (Number of vertical lines in selected drawing of electric publication)
8	05	TAG5 (Bits Per Sample)					Black and white identification
	10	TAG ID	B	2	'258'	X '0201'	Bits per sample
	10	Data Type	B	2	'3'	X '0300'	Data type (1=byte, 2=ASCII, 3=short, 4=long, 5=rational number)
	10	Number of Units	B	4	'1'	X '01000000'	Number of units
	10	Tag Data or Pointer to Tag Data	B	2	'1'	X '0100'	Tag data or pointer to tag data (1 bit/1 pixel)
	10	Others	B	2	'0'	X '0000'	Not used
9	05	TAG6 (Compression)					Compression format
	10	TAG ID	B	2	'259'	X '0301'	Compression method
	10	Data Type	B	2	'3'	X '0300'	Data type (1=byte, 2=ASCII, 3=short, 4=long, 5=rational number)
	10	Number of Units	B	4	'1'	X '01000000'	Number of units
	10	Tag Data or Pointer to Tag Data	B	2	'4'	X '0400'	Tag data or pointer to tag data (MR2)

Editing Standard of PAJ Image Data

Item Number	Level	Item Name	Attribute	Number of Bytes	Item Value	Hexadecimal Form	Content
	10	Others	B	2	'0'	X '0000'	Not used
10	05	TAG7 (Photometric Interpretation)					Identification of white (0) and black (1)
	10	TAG ID	B	2	'262'	X '0601'	Photometric interpretation
	10	Data Type	B	2	'3'	X '0300'	Data type (1=byte, 2=ASCII, 3=short, 4=long, 5=rational number)
	10	Number of Units	B	4	'1'	X '01000000'	Number of units
	10	Tag Data or Pointer to Tag Data	B	2	'0'	X '0000'	Tag data or pointer to tag data
	10	Others	B	2	'0'	X '0000'	Not used
11	05	TAG8 (Document Name)					Pointer to documentation number of Item number 25
	10	TAG ID	B	2	'269'	X '0D01'	Document name page number Frame number
	10	Data Type	B	2	'2'	X '0200'	Data type (1=byte, 2=ASCII, 3=short, 4=long, 5=rational number)
	10	Number of Units	B	4	'22'	X '16000000'	Number of units
	10	Tag Data or Pointer to Tag Data	B	4	'218'	X 'DA000000'	Tag data or pointer to tag data (Relative position from TIFF header (No.22 pointer))
12	05	TAG9 (Image Description)					Pointer to Imagedata of Item number 27
	10	TAG ID	B	2	'270'	X '0E01'	Image description
	10	Data Type	B	2	'2'	X '0200'	Data type (1=byte, 2=ASCII, 3=short, 4=long, 5=rational number)

Editing Standard of PAJ Image Data

Item Number	Level	Item Name	Attribute	Number of Bytes	Item Value	Hexadecimal Form	Content
	10	Number of Units	B	4	'8'	X '08000000'	Number of units
	10	Tag Data or Pointer to Tag Data	B	4	'240'	X 'F0000000'	Tag data or pointer to tag data (Relative position from TIFF header (No.28 pointer))
13	05	TAG10 (Strip Offset)					Pointer to image data
	10	TAG ID	B	2	'273'	X '1101'	Strip offset
	10	Data Type	B	2	'4'	X '0400'	Data type (1=byte, 2=ASCII, 3=short, 4=long, 5=rational number)
	10	Number of Units	B	4	'1'	X '01000000'	Number of units
	10	Tag Data or Pointer to Tag Data	B	4	'264'	X '08010000'	Tag data or pointer to tag data (Relative position from TIFF header (No.33 pointer))
14	05	TAG11 (Orientation)					Orientation of image
	10	TAG ID	B	2	'274'	X '1201'	Orientation
	10	Data Type	B	2	'3'	X '0300'	Data type (1=byte, 2=ASCII, 3=short, 4=long, 5=rational number)
	10	Number of Units	B	4	'1'	X '01000000'	Number of units
	10	Tag Data or Pointer to Tag Data	B	2	'1'	X '0100'	Tag Data or Pointer to Tag Data (direction code of image data)
	10	Others	B	2	'0'	X '0000'	Not used
15	05	TAG12 (Samples Per Pixel)					Number of bits per pixel
	10	TAG ID	B	2	'277'	X '1501'	Samples Per Pixel

Editing Standard of PAJ Image Data

Item Number	Level	Item Name	Attribute	Number of Bytes	Item Value	Hexadecimal Form	Content
	10	Data Type	B	2	'3'	X '0300'	Data type (1=byte, 2=ASCII, 3=short, 4=long, 5=rational number)
	10	Number of Units	B	4	'1'	X '01000000'	Number of units
	10	Tag Data or Pointer to Tag Data	B	2	'1'	X '0100'	Tag data or Pointer to tag data (black and white)
	10	Others	B	2	'0'	X '0000'	
16	05	TAG13 (Rows Per Strip)					Setting the same value as that of item number 7
	10	TAG ID	B	2	'278'	X '1601'	Rows per strip
	10	Data Type	B	2	'4'	X '0400'	Data type (1=byte, 2=ASCII, 3=short, 4=long, 5=rational number)
	10	Number of Units	B	4	'1'	X '01000000'	Number of units
	10	Tag Data or Pointer to Tag Data	B	4	'XX'		Tag Data or pointer to tag data (number of vertical lines in selected drawing of electric publication)
17	05	TAG14 (Strip Byte Count)					Number of bytes of image data
	10	TAG ID	B	2	'279'	X '1701'	Strip byte count
	10	Data Type	B	2	'4'	X '0400'	Data type (1=byte, 2=ASCII, 3=short, 4=long, 5=rational number)
	10	Number of Units	B	4	'1'	X '01000000'	Number of Units
	10	Tag Data or Pointer to Tag Data	B	4	'XX'		Data type (1=byte, 2=ASCII, 3=short, 4=long, 5=rational number)
18	05	TAG15 (X Resolution)					Density of horizontal lines

Editing Standard of PAJ Image Data

Item Number	Level	Item Name	Attribute	Number of Bytes	Item Value	Hexadecimal Form	Content
	10	TAG ID	B	2	'282'	X '1A01'	X resolution
	10	Data Type	B	2	'5'	X '0500'	Data type (1=byte, 2=ASCII, 3=short, 4=long, 5=rational number)
	10	Number of Units	B	4	'1'	X '01000000'	Number of units
	10	Tag Data or Pointer to Tag Data	B	4	'248'	X 'F8000000'	Data type (1=byte, 2=ASCII, 3=short, 4=long, 5=rational number)
19	05	TAG16 (Y Resolution)					Density of vertical lines
	10	TAG ID	B	2	'283'	X '1B01'	Y resolution
	10	Data Type	B	2	'5'	X '0500'	Data type (1=byte, 2=ASCII, 3=short, 4=long, 5=rational number)
	10	Number of Units	B	4	'1'	X '01000000'	Number of Units
	10	Tag Data or Pointer to Tag Data	B	4	'256'	X '00010000'	Tag Data or Pointer to Tag Data (Relative position from TIFF header (No.31 pointer))
20	05	TAG17 (Group 4 Options)					
	10	TAG ID	B	2	'293'	X '2501'	Group 4 options
	10	Data Type	B	2	'4'	X '0400'	Data type (1=byte, 2=ASCII, 3=short, 4=long, 5=rational number)
	10	Number of Units	B	4	'1'	X '01000000'	Number of units
	10	Tag Data or Pointer to Tag Data	B	4	'0'	X '00000000'	Tag data or pointer to tag data
21	05	TAG18 (Resolution unit)					Representation unit

Editing Standard of PAJ Image Data

Item Number	Level	Item Name	Attribute	Number of Bytes	Item Value	Hexadecimal Form	Content
	10	TAG ID	B	2	'296'	X '2801'	Resolution unit
	10	Data Type	B	2	'3'	X '0300'	Data type (1=byte, 2=ASCII, 3=short, 4=long, 5=rational number)
	10	Number of Units	B	4	'1'	X '01000000'	Number of units
	10	Tag Data or Pointer to Tag Data	B	2	'2'	X '0200'	Tag Data or Pointer to Tag Data (Inches)
	10	Others	B	2	'0'	X '0000'	Not used
22	05	Pointer to Next Directory	B	4	'0'	X '00000000'	Pointer to next directory or termination '0''
23	05	Publication Office	C	2	'JP'	X '4A50'	Names of states, other entities and intergovernmental organizations (Corresponding to codes of WIPO
24	05	Kind of Document	C	2	'A '		Kind of document (corresponding to codes of WIPO ST.16) In PAJ, "A "(second digit is blank) for publication of unexamined applications and "B1" for granted gazettes
25	05	Document Number	C	10	YYYNNNNNN		Document number (publication of unexamined applications ("YYYYNNNNNN"), granted gazettes(" NNNNNNN") (the first three digits are blank.))
26	05	Subdocument Sequence NO.	C	8	'0000XXXX '		Sequence number of subdocuments (Sequence number of subdocuments for text records)
27	05	Imagedata ID	C	8	0000XXXX '		Imagedata ID (EMI ID in text)
28	05	X-Axis Resolution Numerator	B	4	'200'	X 'C8000000'	X-axis resolution numerator
29	05	X-Axis Resolution Denominator	B	4	'1'	X '01000000'	X-axis resolution denominator

Editing Standard of PAJ Image Data

Item Number	Level	Item Name	Attribute	Number of Bytes	Item Value	Hexadecimal Form	Content
30	05	Y-Axis Resolution Numerator	B	4	'200'	X 'C8000000'	Y-axis resolution numerator
31	05	Y-Axis Resolution Denominator	B	4	'1'	X '01000000'	Y-axis resolution denominator
32	05	Image Data	B	indeterminate			Main body of image data

Editing Standard for PAJ Text Data (Explanatory Note)

Item Number	Repeat	Mandatory Items	Tag Name	Item Content	Attribute	Number of Digits
(1)	(2)	(3)	(4)	(5)	(6)	(7)

(1) Item Number
Serial number provided to each standard item

(2) Repeat
Display of existence or non-existence of repeat item
● :Key item for repeat
○: Repeat item based on the above-mentioned key item

(3) Mandatory Items
○ means mandatory items

(4) Tag Name
Tag name provided to each item
Same names as those in the output data layout

(5) Item Content
Name of each item

(6) Attribute
Attribute of each item
C: 1 byte code(alphanumeric characters and symbols)
N: Right-align numeral without leading 0 (zero)
K: 2 byte code(Kanji character)
* The attribute information is not necessary in the present text file, however, refer to this for your information.

(7) Number of Digits
Number of digits for each item
Numerical value : Number of digits
n: 1byte code undefined length
m: 2 byte code undefined length

Editing Standard of PAJ Text Data

Item Number	Repeat	Mandatory	Tag Name	Item Content	Attribute	Number of Digits	Note
1		○	<SDOBI lang="en">	Start of bibliographic data			
2		○	<B100>	Start of items related to publication number and document			
3		○	<B110>	Publication number or granted number	C	8 OR 10	Publication number 1. Two years for Emperor's reign year+ up to 6 digits for the number in and before 1999 2. From 2000 onwards, four-digit Western year + up to 6 digits for the number Granted number 3.000 (three zeros) + seven digit number
4		○	<B120> <B121>	Simple indication of a language used in document	C	25	"PATENT ABSTRACTS OF JAPAN"
5		○	<B130>	Kind of document	C	1	"A"
6		○	<B140> <date>	Publication date or granted date	C	8	Set in Western calendar
7		○	<B190>	Publication country	C	2	"JP"
8		○	<B200>	Start of application items			
9		○	<B210>	Application number	C	8 OR 10	1. Two years for Emperor's reign year+ up to 6 digits for the number in and after 1999 2. From 2000 onwards, four-digit Western year + up to 6 digits for the number
10		○	<B220> <date>	Application date	C	8	Set in Western calendar
11	○		<B300>	Start of priority claims			
12	○		<B310>	Priority number	C	8 OR 10 OR 16	1. Where JP is the country claiming priority in and before 1999 Two years for Emperor's reign year+ up to 6 digits for the number 2. Where JP is the country claiming priority in and after 2000 Four-digit Western year + up to 6 digits for the number 3. Where Japan is not the country claiming priority The number is set as it is
13	○		<B320> <date>	Priority date	C	8	Set in Western calendar

Item Number	Repeat	Mandatory	Tag Name	Item Content	Attribute	Number of Digits	Note
14	○		<B330> <ctry>	Country claiming priority	C	2	Two digit country code
15		○	<B500>	Start of International Patent Classification (IPC) and invention items			
16	○	○	<B510JP>	International Patent Classification (IPC)		n	"JP" in <B510JP>: Publishing country ID="n": "n" is a serial number beginning from 1 28 digit alpha-numeral,"A" for advanced level. 29 digit alpha-numeral, "F" for major classification and "L" for subclassification onwards 30 digit alpha-numeral, "I" for invention information and "N" for additional information <TEXT> Each of invention and additional information has 50 digits. * Usually, no end tag is set in bibliographic data. The present tag and its content have an end tag. <B510JP> <classification-ipcr id="1"> <text>invention information</text> </classification-ipcr> <classification-ipcr id="2"> <text>additional information</text> </classification-ipcr> </B510JP>
17		○	<B541>	Language used in title of invention	C	2	"en"
18		○	<B542>	Title of the invention	C	n	

Item Number	Repeat	Mandatory	Tag Name	Item Content	Attribute	Number of Digits	Note
19			<B600>	Start from division representation			
20	○		<B620> <parent> <dnum> <anum>	Division representation	C	8 OR 10	1. Two years for Emperor's reign year+ up to 6 digits for the number in and before 1999 2. From 2000 onwards, four-digit Western year + up to 6 digits for the number
21			<B700>	Start of items related to applicant and inventor			
22	○	○	<B711> <snm>	Applicant	C	n	
23	○	○	<B721> <snm>	Inventor	C	n	
24		○	<abstract lang="en">	Abstract sentence			
25		○	<P>PROBLEM TO BE SOLVED:	Purpose sentence	K	m	・It is described as purpose sentence right after direction word "PROBLEM TO BE SOLVED.
26		○	<P>SOLUTION:	Configuration sentence	K	m	・It is described as configuration sentence right after direction word "SOLUTION".
27		○	<P>COPYRIGHT:	Publisher	C	n	・Publication year and publisher are set right after direction word "COPYRIGHT: ." "COPYRIGHT: (C)2***,JPO&INPIT"
28		○	<drawings lang="en"> <figure>	Drawing			
29		○	<img XX…X>	Attribute of image data	C	n	・Fixex character string: ""