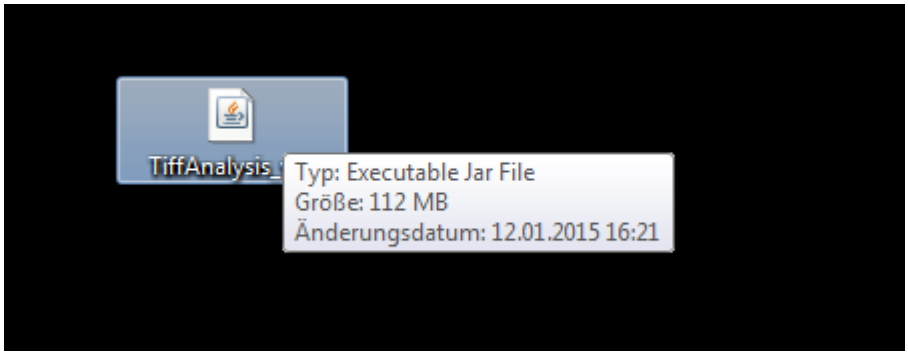
Tiff Analysis

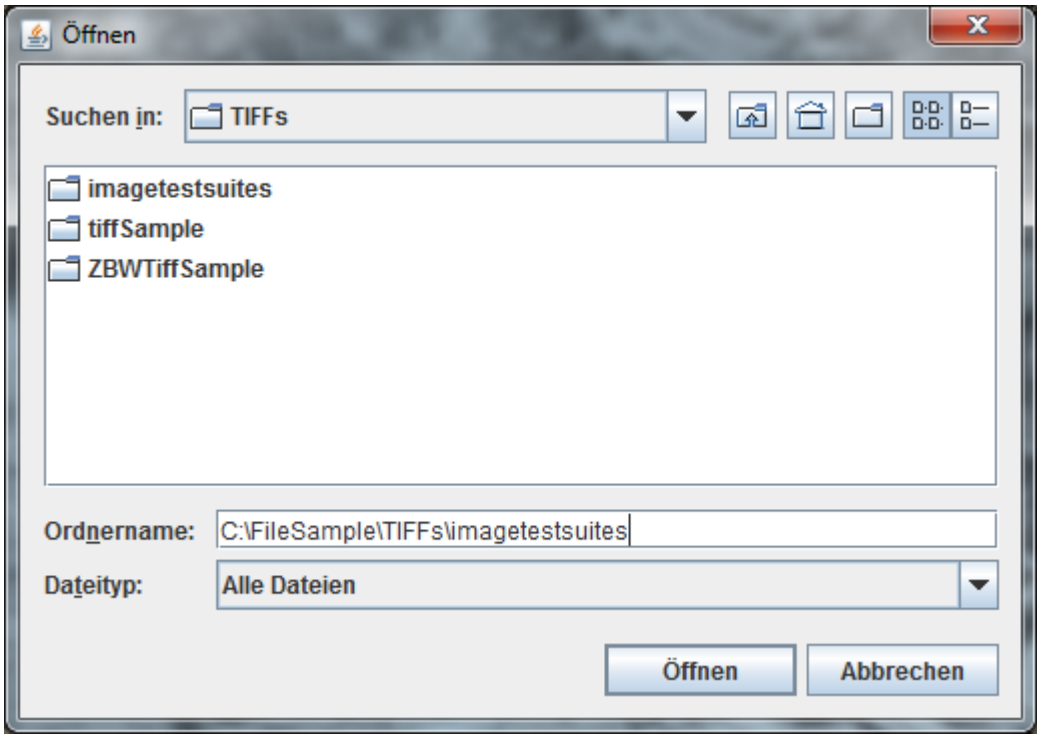
Documentation for the user

Currently, you only need to download (and maybe unzip) the jar-file and then double-click on the jar.



It's java and platform-agnostic.

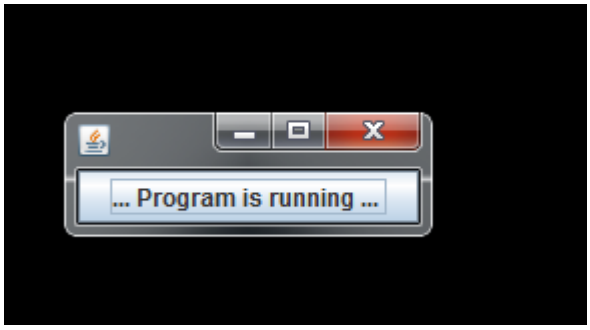
Next, a Folder Browser Dialog opens. Navigate to the folder you want to examine.



Following a Dialog opens and you are asked to name your xml Outputfile. It can be "outputXML", "Test", "auntDebbie" or whatever. The .xml-Extension will be added automatically and the File will be put in the same folder you want to examine. If the filename already exists in the folder, you are asked if you want

to stop and try again (“Ja”) or overwrite the file and continue (“Nein”). The program is not a 100% English yet.

You see that the program is running:



This window might be at the corner of your monitor up in the left side. It will vanish when the program is done.

You will find two new files in your examined folder now:

Go to your favorite Browser (e. g. firefox, I think there are issues with the internet explorer) and open the xml-file there, in this case the path would be: file:///C:/FileSample/TIFFs/ZBWTiffSample/test.xml

A – hopefully – nicely looking page with many tables will open.

Table 1: Summary of Tiff Tags

Examination of Tiff Tags

Summary of Tiff Tags

TiffTag	Occurance	Description	SourceOfTag
Bits Per Sample	575	Number of bits per component.	Baseline
Compression	575	Compression scheme used on the image data.	Baseline
Fill Order	192	The logical order of bits within a byte.	Baseline
Image Height	575	The number of rows of pixels in the image.	Baseline
Image Width	575	The number of columns in the image, i.e., the number of pixels per row.	Baseline
Orientation	575	The orientation of the image with respect to the rows and columns.	Baseline
Photometric Interpretation	575	The color space of the image data.	Baseline
Planar Configuration	575	How the components of each pixel are stored.	Baseline
Preview Image Length	575	For each strip, the number of bytes in the strip after compression.	Baseline
Preview Image Start	575	For each strip, the byte offset of that strip.	Baseline
Resolution Unit	575	The unit of measurement for XResolution and YResolution.	Baseline
Rows Per Strip	575	The number of rows per strip.	Baseline
Samples Per Pixel	575	The number of components per pixel.	Baseline
XResolution	575	The number of pixels per ResolutionUnit in the ImageWidth direction.	Baseline
YResolution	575	The number of pixels per ResolutionUnit in the ImageLength direction.	Baseline
Page Number	575	The page number of the page from which this image was scanned. Extended	Extended
T4 Options	192	Options for Group 3 Fax compression. Extended	Extended
T6 Options	192	Options for Group 4 Fax compression. Extended	Extended
Subfile Type	575	null	null

You can see the names of the Tiff Tags – which are usually known and can be displayed. The description, though, can only be seen if I personally have put it into the program, which I have not done with nearly all the possible Tiff Tags yet. The same is true for the SourceOfTag, but here, at least, it is known if a tag is private, if the dec value is higher than a certain value (32768), so at least this should always be displayed.

If an information is not there, it's "null". Null means no more than "I do not have any clue".

SubFileType, however, is very baseline and should be known to the program. In the next release, this might show the description: "A general indication of the kind of data contained in this subfile."

Table 2: Tiff Statistic

This table is short and just shows how many Tiff files are in the folder and how many of them could be examined. Some are so difficult, that the program cannot get the metadata about the Tiff Tags out of them.

Other files, however, are ignored. The program deals with the ".tif"-Extension (not context-sensitive) and with the file header, which has to start with "II" or "MM".

Table 3: Examination Failed for following Tiff Files

All Tiff-Files that could not be examined are displayed by file path and the error message is shown.

FilePath	ErrorMessage
C:\FileSample\TIFFs\imagetestsuites\imagetestsuideTiff\problematicTiffs\TiffPreviewNotPossible\part1\sm1-2743db7c68f091ddc61dcd582febff2d.tif	java.lang.NegativeArraySizeException
C:\FileSample\TIFFs\imagetestsuites\imagetestsuideTiff\problematicTiffs\TiffPreviewNotPossible\part1\sm1-f0d7bcb90496c323c880a9773dfe93ff.tif	java.io.IOException: Could not read value from file
C:\FileSample\TIFFs\imagetestsuites\imagetestsuideTiff\problematicTiffs\TiffPreviewNotPossible\part1\sm14-42c19f8e79e582bef107f372f18a074b.tif	java.io.IOException: Could not read value from file
C:\FileSample\TIFFs\imagetestsuites\imagetestsuideTiff\problematicTiffs\TiffPreviewNotPossible\part1\sm15-42c19f8e79e582bef107f372f18a074b.tif	java.io.IOException: Could not read value from file
C:\FileSample\TIFFs\imagetestsuites\imagetestsuideTiff\problematicTiffs\TiffPreviewNotPossible\part1\sm2-76d5d8fd02d58b774f2bae6f7b763e3e.tif	java.io.IOException: Could not read value from file
C:\FileSample\TIFFs\imagetestsuites\imagetestsuideTiff\problematicTiffs\TiffPreviewNotPossible\part1\sm3-76d5d8fd02d58b774f2bae6f7b763e3e.tif	org.apache.sanselan.ImageReadException: Couldn't find image data.
C:\FileSample\TIFFs\imagetestsuites\imagetestsuideTiff\problematicTiffs\TiffPreviewNotPossible\part1\sm6-f0d7bcb90496c323c880a9773dfe93ff.tif	java.io.IOException: Could not read value from file
C:\FileSample\TIFFs\imagetestsuites\imagetestsuideTiff\problematicTiffs\TiffPreviewNotPossible\part1\sm7-f0d7bcb90496c323c880a9773dfe93ff.tif	org.apache.sanselan.ImageReadException: offsets.length(70) != byteCounts.length(24)
C:\FileSample\TIFFs\imagetestsuites\imagetestsuideTiff\problematicTiffs\TiffPreviewNotPossible\part1\sm8-f0d7bcb90496c323c880a9773dfe93ff.tif	java.io.IOException: Could not read value from file

One day, I might deal with them and look for ways to fix them, as there are some tools out there and Tiff should (in some cases) not be too hard to fix.

Table 4: Mandatory Tiff Tags in Baseline Tiff

These Tags should be there. Sometimes they should even have a certain value (not all of the compression types are allowed or good for longterm archiving), but I have not dealt with that yet.

YResolution and *ResolutionUnit* are not seen on this screenshot, because the table is too wide, but they are there.

12 Mandatory Tiff Tags in Baseline Tiff

FileName	ImageWidth	ImageLength	BitsPerSample	Compression	Photometric	StripOffSets	SamplesPerPixel	RowsPerStrip	StripByteCounts	XResolution
a1008.TIF	7681 Short	5761 Short	8 8	11 Short	21 Short	6261 Long	31 Short	5761 Short	13271041 Long	721 Rational
BBC3.TIF	1431 Short	531 Short	8 8	51 Short	21 Short	86 424	31 Short	81 Long	338 713	3001 Rational
Buch.tif	24791 Long	35081 Long	11 Short	41 Short	01 Short	3421 Long	11 Short	35081 Long	833301 Long	3001 Rational
cramps-tile.tif	8001 Short	6071 Short	81 Short	11 Short	01 Short	8 65544	11 Short	2561 Long	65536 65536	
cramps.tif	8001 Short	6071 Short	81 Short	327731 Short	01 Short	8 176	11 Short	121 Long	168 168	721 Rational
fax2d.tif	17281 Short	10821 Short	11 Short	31 Short	01 Short	81 Long	11 Short	11 Long	325251 Long	2041 Rational
g3test.tif	17281 Short	11031 Short	11 Short	31 Short	01 Short	81 Long	11 Short	11 Long	501101 Long	2041 Rational

Table 5: Nice to have Tiff Tags optional in Baseline Tiff

These are (for now)

- MinSampleValued
- MaxSampleValue
- GrayResponseUnit
- GrayResponseCurve
- Software
- DateTime
- Artist
- ColorMap

- Copyright

Table 6: No Tags for Baseline Tiff but nice to have for Digital Preservation

ExifIdf & ExtraSamples

Table 7: Other known Tiff Tags

Only a few I have come across already and the number might grow a bit in the near future.

The others are only seen in the Table 1 and of course, if you examine the XML-Output in an editor, you will have all the talkative information.

```
<DateTime>'2005:01:24 11:05:43'20 ASCII</DateTime>

<Predictor>21 Short</Predictor>

<ExtraSamples>01 Short</ExtraSamples>

<UnknownTiffTag>Application Notes</UnknownTiffTag>

<IPTC_NAA>469893120 33554944</IPTC_NAA>

<UnknownTiffTag>IPTC- NAA</UnknownTiffTag>

<UnknownTiffTag>Photoshop Settings</UnknownTiffTag>

<UnknownTiffTag>Photoshop Settings</UnknownTiffTag>

<ExifIfd>24540921 Long</ExifIfd>

<UnknownTiffTag>Exif Offset</UnknownTiffTag>

<InterColorProfile>0 0</InterColorProfile>

<UnknownTiffTag>ICC_ Profile</UnknownTiffTag>
```

I will enhance this soon that it outputs the contents of the unknown tags as well. There might be another table as well.

Documentation for the developer

To be really short for now, I have used the `org.apache.sanselan.formats.tiff.*` and `org.apache.sanselan.common.IImageMetadata;` libraries and the program has more or less only one heart:

```
IImageMetadata metadata = Sanselan.getMetadata(file);
TiffDirectory tiffDirectory = ((TiffImageMetadata)
metadata).findDirectory(TiffDirectoryConstants.DIRECTORY_TYPE_ROOT);

ArrayList<TiffField> allEntries = tiffDirectory.getDirectoryEntrys();
tiffDirectory.dump();
```

The rest is just a lot of XML, XSL and running through folders to look for Tiff-Files.