

SanselanProposal

Proposal Draft

Sanselan, a Pure-Java Image Library

Abstract

Sanselan is a pure-java image library for reading and writing a variety of image formats.

Proposal

The sanselan image library will provide a portable toolkit for reading and writing a variety of image formats. This includes parsing of the image info (size, color space, icc profile etc.), metadata (ie. EXIF) and image data.

Common operations (such as reading an image) should be simple and require little code, but every operation should also allow fine-grained control through optional arguments.

Correctness is preferred over performance. Completeness (ie. support of all features/variants of the image file formats) is preferred. Flexibility (ie. ability to treat files, byte arrays and input streams interchangeably) is preferred.

Though hiding the differences between the file formats in common usage, the library should also provide the means to explore the internals of the file formats (for example, what jfif segments or png chunks are present).

Background

The initial work on Sanselan was begun in 2004 by Charles M. Chen, and was open sourced soon after. Much of the code was finished by the end of 2004, and work since then has been primarily been bug fixes, simplification of the API, and addition of optional parameters to allow finer-grained control.

Since its release, Sanselan has been used by a variety of projects from around the world.

Definitions:

In the context of Sanselan, "Image Info" refers to things like image size, bits per pixel, color space, transparency, etc. "Image Metadata" refers to structured metadata (ie. EXIF) embedded in an image format (ie. JFIF), for example, Geocoding, time taken, encoder info, etc. "Image Data" refers to the raw data that is interpreted to decode pixel info.

Rationale

There are many libraries dealing with image formats in the Java World, but still, each of them has problems when it comes to portability, specification conformance and functionality. Some of the libraries require non-portable native-code, others support reading of specific formats but not writing etc.

Sanselan offers all of the following for its core file formats:

1. file format identification.
2. fast extraction of image info (such as size, color type, etc.) in a format-neutral structure, without reading the image data.
3. extraction of icc profiles without reading image data.
4. extraction of image metadata without reading image data.
5. simple, concise syntax for common usages.
6. optional fine-grained control over reading and writing images.
7. color-correctness by applying icc profile, gamma and color space color metadata.
8. reading and writing images.

For those formats which Sanselan cannot read & write image data (ie. jpeg/jfif, photoshop/psd and windows icon/ico), Sanselan can still read image info and metadata.

Sanselan's code will be available under the flexible Apache license.

The Sanselan project attempts to streamline this duplication of efforts. We believe that by starting the Sanselan project with an existing codebase, this will create a library without the defects mentioned above and might also create enough interest and tension to draw in other image libraries/code to get an even bigger functionality set.

Initial Goals

The initial goals of the proposed project are:

- Viable community around the Sanselan codebase
- Active relationships and possible cooperation with related projects and communities
- Initial generic code base dealing with image formats and metadata

- Implementation of a variety of image formats.

Current Status

The current code base has been developed by Charles M. Chen (<http://www.fightingquaker.com/sanselan/>) in his spare time. It provides a very good basis. The code has to be (and will be) donated to Apache by Charles. It is already licensed under the Apache 2.0 license. The further development will be based on this code base taking it wherever the community wants it to be.

The project has been refactored to remove any external dependencies. It has been loosely tested, and deployed in a variety of production environments.

No patent issues obtain. The file formats in question are well documented and stable.

Meritocracy

All the initial committers are familiar with the meritocracy principles of Apache, and have already worked on the various source codebases. We will follow the normal meritocracy rules also with other potential contributors.

Community

There is not yet a clear Sanselan community. The current code base has a number of interested users. The primary goal of the incubating project is to build a self-sustaining community around this code base.

Core Developers

The initial set of developers comes from various backgrounds, with different but compatible needs for the proposed project.

Charles Chen has written all of the current code in the project, though others have helped point out specific bugs. Charles continues to patch bugs as he becomes aware of them, as well as continuing work on improving the API.

Alignment

Apache contributes a strong development environment together with a solid brand to help make this project a success. There are several existing libraries, each with their own advantages and disadvantages. Bringing the project to Apache will help gather the community around a single project.

There will also be connections to existing Apache projects like the Tika project and perhaps commons.

Known Risks

By adopting this project, the Apache project would place itself in implicit competition with the other available image libraries.

Orphaned products

There is a high need in quality image libraries for Java. Sanselan currently has a strong user base and among this user base is a very strong interest in this project.

Inexperience with Open Source

The project's original developer, Charles Chen, has contributed in small ways to open source projects for years. However, he has never been actively involved in an open source project with a thriving community and doesn't have any experience in fostering or coordinating such a community.

The other developers have a big experience with open source projects, especially with Apache projects and are long time users of Sanselan. However, we look forward to cultivating that community under the guidance of the Apache organization.

Homogenous Developers

We will see... 😊

Reliance on Salaried Developers

Actually, no one is paid to work on this project. Charles Chen has continued to work on this project for 3 years without being paid.

Some of the developers are paid to work on this or related projects, but the proposed project is not the primary task for anyone.

Relationships with Other Apache Products

Sanselan is related to at least the following Apache projects. None of the projects is a direct competitor for Sanselan.

- [Apache Tika](#) - Tika provides a framework to extract metadata out of documents. The plan is to develop Tika parsers using Sanselan.
- [Apache XML Graphics](#) - Batik and FOP both make extensive use of image libraries. The Commons subproject even contains some image codecs.

- [Apache Harmony](#) - The ImageIO API is part of the class library and Harmony has to provide implementations (currently only JPEG?).

A Excessive Fascination with the Apache Brand

All of us are familiar with Apache and we have participated in Apache projects as contributors, committers, and PMC members. We feel that the Apache Software Foundation is a natural home for a project like this.

Documentation

- Sanselan (<http://www.fightingquaker.com/sanselan/>)

Initial Source

Sanselan will start with the contributed code base:

- The [Sanselan](#) project

Source and Intellectual Property Submission Plan

All seed code and other contributions will be handled through the normal Apache contribution process.

We will also contact other related efforts for possible cooperation and contributions.

External Dependencies

None.

Cryptography

Sanselan itself will not use cryptography, but it is possible that at a later time support for image formats is developed that requires cryptography. Currently there is no such support/code.

Required Resources

Mailing lists

- sanselan-dev@incubator.apache.org
- sanselan-commits@incubator.apache.org
- sanselan-private@incubator.apache.org

Subversion Directory

- <https://svn.apache.org/repos/asf/incubator/sanselan>

Issue Tracking

- JIRA Sanselan (SANSELAN)

Other Resources

- none

Initial Committers

Name	Email	CLA
Charles M. Chen	charlesmchen at gmail dot com	no
Carsten Ziegeler	ctiegeler at apache dot org	yes
Philipp Koch	pkoch at day dot com	no

Affiliations

Name	Affiliation
------	-------------

Charles M. Chen	
Carsten Ziegeler	Day Management AG
Philipp Koch	Day Management AG

Sponsors

Champion

- Carsten Ziegeler (cziegeler at apache dot org)

Nominated Mentors

- Craig Russell (clr@apache.org)
- Yoav Shapira (yoavs[at]a.o)
- Jeremias Maerki (jeremias@a.o, not available before Oct 2007)
- others TBD

Sponsoring Entity

- We are asking the Incubator PMC to sponsor this proposal.