

CommonsCsv

Proposal

Proposal to submit the Neteccera CSV library to the ASF as Jakarta Commons CSV, or potentially as a sub-package of another Commons component (io/codecc/lang).

Rationale

There have been requests for an ASL compatible licensed Java library for parsing Comma Separated Values (CSV) files. The latest user threads (<http://tinyurl.com/7utvn>) lead to <http://csv4j.sourceforge.net/>, which as of this document has not yet seen any activity.

A lot of interest was exhibited in the idea of a Commons CSV, and in discussing the various codebases that could be used initially, a code donation from Neteccera was deemed the best option to start with. Thus the need to go via the Incubator.

CSV parsing is not as basic as it might first seem due to differences in platforms, and wide differences in interpretations of the de-facto specification. Commons CSV would have to publish their own interpretation of said de-facto specification.

Scope

- Java components to read and write CSV files.
- An interpretation specification.
- Variants on the formatting of the CSV files.
- Various parsing APIs (for example SAX style, DOM style)

Initial source

Two basic compatible libraries exist (<http://kasparov.skife.org/csv/> by Brian McCallister and <http://www.osjava.org/genjava/multiproject/gj-csv/> by Henri Yandell), which could be considered (authors willing) as the basis for a CSV library, and Neteccera have offered to donate their internal CSV component, which on paper looks to be far superior.

Current plan is to go with the Neteccera codebase.

Known risks

As a mature component, the Neteccera library may after careful thought be considered to not need extensive development. It would thus become a mature Apache codebase (not in itself a bad thing) and could lack maintenance due to a short development period at the ASF.

The Neteccera developers may not end up getting very involved beyond their initial (very much appreciated) code donation. Folding the skife and gj-csv feature-sets in, and ending those projects should build up as much of a community as a commons component requires.

Other open-source examples

Besides the afore-mentioned libraries (skife + gj-csv), Stephen Ostermiller has long had a GPL'd CSV library (<http://ostermiller.org/utlis/CSV.html>). Going by his website he has no desire to re-licence:

"Could it be licensed under something less restrictive?

Many people request they receive a copy of the utilities licensed under the Library General Public License (LGPL) or a BSD style license. The answer is firmly "no".

Those Licenses do not ensure that derivative works remain liberated. If such a license were given, the libraries could then be used in closed source applications."

There are also libraries out there that approach the problem from a different angle.

- CSV via JDBC API (<http://csvjdbc.sourceforge.net/>)
- CSV via XML API (<http://www.dpawson.co.uk/java/csv2xml.html>)

but these are definite niche concepts and not the general desired use case.

Source submission plan

The following Java packages will be submitted:


Package name	Purpose
org.apache.commons.csv. *	CSV

Resources



svn: /incubator/jakarta/commons/csv

The existing Jakarta Commons mailing lists and bugzilla will be used.

Legal

- Software Grant would be needed from Netcetera .
- Netcetera Corporate CLA already exists, need to find out if a new one would be needed.
- Skife (brianm@apache) and gj-csv (bayard@apache) both have CLAs.

Initial committers

- Steven Caswell (stevencaswell@apache.org)
- Brian [McCallister](#) (brianm@apache.org)
- Henri Yandell (bayard@apache.org)
- Urs Hardegger (original coder) 
- Stefan Rufer (submitter) 

Apache sponsor/champion

Sponsor: Jakarta (specifically Jakarta Commons) Champion: Henri Yandell (bayard@apache.org)