JenaProposal

Jena, a Semantic Web Framework

Nov 2010: This proposal has been accepted: Results email.

Abstract

Jena is a semantic web framework for Java, based on W3C standards.

Proposal

Jena provides a semantic web framework in Java that implements the key W3C recommendations for the core semantic web technologies of RDF and SPARQL. Jena is a number of components and modules built on this core system. It currently includes:

- an API for working with RDF
- Parsers and writers for the RDF formats (RDF/XML, Turtle, N-triples, NQuads, TriG)
- an implementation of SPARQL, the W3C standard RDF query language
- multiple storage systems for RDF data including in-memory, file-backed, in SQL databases and in custom scalable storage systems
- an API for manipulation of OWL
- a rule-based inference engine
- an implementation of GRDDL for extraction of RDF from XML formats
- a standards compliant IRI library.

The project includes facilities based around this core to encourage the creation of components and contributions both as part of Jena and also as companion open source activities.

This proposal includes the main components of Jena: the main Jena download, ARQ, GRDDL, SDB, TDB, the IRI library and Joseki. Other components may be contributed later - we're just starting with the main part of Jena for now.

Background

The W3C recommendations provide detailed specifications and it is important to follow these standards so that independently built applications can exchange data over the web. Jena provides high quality Java implementations of RDF input/output and storage so that application writers can concentrate on the application, not the low-level details.

W3C Semantic Web: http://www.w3.org/standards/semanticweb/

Jena has been on SourceForge since 2001. http://sourceforge.net/projects/jena/

Rationale

The open source project was originally created as part of a research activity in HPLabs. In building new systems, the researchers identified the value of a common platform that dealt with the low level details of the standards. This lead to engagement with the standards process and the creation of a framework that provided a library to deal with the details of semantic web standards. This work was released as Jena. The developers have contributed implementation experience back to the working groups.

None of the contributors now work for HP. Providing a uniform contributor and licensing framework assists commercial use of Jena.

Current Status

Jena is already an established project with a large user base in industry and academia. It currently uses a BSD-style three-clause license with a number of contributing copyright holders. Support is primarily provided via the jena-dev@groups.yahoo.com mailing list. The majority of the team was employed in HPLabs, and HP holds the majority of the copyright over the code - there are contributions from non-HP companies. HP decided to close the research group as of October 2009 and the people from HPLabs connected with the project have moved on to several different semantic web companies.

This change does not immediately affect Jena because the people who were in HP still remain active contributors to Jena. The project continues to be supported and actively enhanced. There is now the opportunity to become an open source project without a single large organisation involved.

Meritocracy

The Jena team has always been self-determining; there has not been a project manager in charge of the effort. Instead, it has grown through individuals contributing to the codebase as part of their research activities. The team has organised itself to create the framework for builds, releases and public support, and people who had worked on Jena in HP, and moved to other companies and institutions, have continued to contribute.

Core developers

Jena originated within a research activity in HPLabs, starting around 2000. Contributors to jena have been active in W3C working groups including chairing the "RDF Core" working group and acting as document editors on several other working groups. W3C processes are public; jena contributors have been involved in public debate and decision making. People have since moved on from HP to several semantic web forced companies and to university positions.

Alignment

Jena is already in use in many commercial systems as well as widely used in academic research and teaching. We want to continue making this easy and at the same time encourage contribution in a well-known environment.

Jena is already pretty much run in a collaborative open development style with communication on mailing lists.

Known Risks

Orphaned products & Reliance on Salaried Developers

Jena is in use by companies we work for so the companies have an interest in its continued vitality.

The Jena team members are not employed to work on Jena specifically; while there is some development as part of their day-jobs, the team members do also contribute personal time as well.

Inexperience with Open Source

While Jena has been open-source since 2001, the majority of individuals involved do not have wide experience of contributing to other open source projects, so the team members need to develop more skills in participating in open-source communities.

Relationships with Other Apache Products

Jena uses Xerces, Lucene, Apache Commons HttpClient and Apache Commons FileUpload.

Jena is used by Clerezza (in incubation).

A Excessive Fascination with the Apache Brand

Jena has an established community of users and is used in both academic and commercial settings. The Apache environment offers Jena the opportunity to expand the ways that more people can be involved and contribute, and hence to ensure the project is not dependent on the current members. We hope that association with Apache will also encourage other open source projects that use Jena to help develop a healthy and vibrant semantic web open source ecosystem.

Apache offers us a clear licensing framework and support infrastructure which would reassure the many users of Jena who exploit it in commercial environments as well as those in other open source projects.

Documentation

Overview documentation, tutorials, topic-based how-tos and detailed JavaDoc can be found at http://openjena.org/

Initial Source

The majority of the current codebase resides in the Jena project CVS/SVN on SourceForge. Joseki is also on SourceForge; we later decided to put all projects under one SF project so this is a historical anomaly. The modules in the initial source are:

- Jena CVS area on SourceForge
    - jena2 (the core system, include RDF, rules and OWL subsystems)
    - iri (the IRI library)
    - Eyeball and EyeballAcceptance (a checker for RDF)
- Jena SVN area on SourceForge
    - ARQ (SPARQL query and update engine)
    - Fuseki (SPARQL server)
    - grddl (GRDDL implementation for Jena)
    - SDB (SQL database layer for Jena)
    - TDB (customer storage layer for Jena)
    - Ymris (experimental rules engine)
    - Experimental/Jena3 (experiment reorganisation of jena)
- Joseki CVS area on SourceForge
    - Joseki3 module.

Source and Intellectual Property Submission Plan

We are in discussions with HP, the largest copyright holder, about licensing to Apache and currently HP has indicated that it is willing to do so in principle.

The initial committers overtake to resolve all IP and copyright issues that concern the dependencies of the initial source and of any contributions in accordance with Apache requirements for graduating from incubator status.

All contributions to the Jena codebase are under BSD-style license. The majority of copyright is held by HP. Some copyright is held by others, as noted in the codebase. This includes contributions from the initial committers below and any other contributions.

External Dependencies

Details of license of components used by Jena are available at: http://openjena.org/Licenses/index.html

The Jena GRDDL Reader has some additional dependencies: http://jena.sourceforge.net/grddl/license.html

We are heavily dependent on Xerces for both parsing and also for XML datatype support.

Cryptography

No specific cryptography.

Required Resources

Mailing lists

- jena-private (with moderated subscriptions)
- jena-dev
- jena-commits
- jena-user

Subversion Directory

- jena

Issue Tracking

- JIRA

Other Resources

- Hudson

Initial Committers

The intial committers are the currently active developers for Jena.

- Chris Dollin
- Paolo Castagna
- Damian Steer
- Jeremy Carroll
- Ian Dickinson
- Dave Reynolds
- Andy Seaborne

Affiliations

- Epimorphics Ltd: Dave Reynolds, Ian Dickinson, Chris Dollin, Andy Seaborne
- Talis Systems Ltd: Paolo Castagna
- University of Bristol: Damian Steer
- TopQuadrant Inc: Jeremy Carroll

Sponsors

Champion

Ross Gardler (rgardler .at. apache.org

Nominated Mentors

- Bertrand Delacretaz (bdelacretaz .at. apache.org)
- Leo Simons (leosimons .at. apache.org)
- Dave Johnson (snoopdave .at. gmail.com)
- Benson Margulies (bimargulies .at. gmail.com)

Sponsoring Entity

Incubator PMC