

SpatialProposal

Apache SIS, A toolkit for constructing spatial information systems.

Abstract

Spatial information systems (SIS) (akin to Geographic Information Systems, or GIS) are rapidly growing as information has taken on a sense of location. This location context has allowed people to start exploring different ways of searching, clustering, and displaying information. Spatial queries such as:

- **point-radius**, e.g., show me all objects within X miles of point P, typically a lat/lon;
- **bounding box**, e.g., show me all objects within a box defined by south, east, north, west bounding coordinates; and
- **polygon**, an extension of bounding box to arbitrary shapes defined by arbitrary points

are becoming a part of everyday life, where some combination of the above is used to find a restaurant, determine sites of interest for climate research, for data reduction and subsetting, or demographic profiling, social networking, and a host of other applications. There exist a number of libraries, and frameworks written in Java, C/C++, and other P/Ls that deal with the aforementioned issues, however the one consistent homogeneity is that most of these software do not include ASF-friendly licensing. On the contrary, most of these software systems and tools are LGPL licensed, as their use is primarily to produce GIS software, which is then sold for a profit. What's more, even the standards organization the Open Geospatial Consortium (OGC) promotes the use of LGPL SIS/GIS software to implement its interfaces and specifications, leaving those interested in a more ASF-friendly solution with a major hole to fill, or having to deal with the license implications of leveraging LGPL open source software in their applications.

We propose to construct Apache SIS, an ASF 2.0 licensed toolkit that spatial information system builders or users can leverage to support the aforementioned activities, alleviating much of the software and potentially legal difficulties in implementing SIS/GIS systems. This project will look to expand on those concepts and serve as a place to store reference implementations of spatial algorithms, utilities, services, etc. as well as serve as a sandbox to explore new ideas. Further, the goal is to have Apache SIS grow into a thriving Apache top-level community, where a host of SIS/GIS related software (OGC datastores, REST-ful interfaces, data standards, etc.) can grow from and thrive under the Apache umbrella.

Proposal

The Internet is changing to the "local world" wide web, where information no longer exists in a digital vapor, but contains real world context. From news stories to tweets, location is a very powerful concern, evidenced by the proliferation of popular websites offering geo-referenced information for all relevant content ([Flickr](#), [Twitter](#), [Google Maps](#), etc). Besides the social utility of spatial data, there are also national interest related uses of prime importance. For example, from a national policy perspective, and federal agency perspective (e.g., NASA, NOAA, DoD), global climate concerns have underscored the importance of science data collected about our planet, all of which is location based. So-called "operational" and "actionable" data including climate models, weather forecasts as well as scientific, "offline" data (measurements of CO2 in the atmosphere, measurements of sea surface temperature, etc.) all provide some sense of where the data was created, where currently resides, and/or what it references. These are just a sampling of the spatially relevant information available – the list is growing as scientists, policy-makers and decision makers develop new downstream activities that leverage spatial data. As we move forward there is also no reason to restrict the focus of SIS/GIS to *just this planet* as a point of reference; other sciences (astrophysics, planetary science) have been collecting information about our universe and other celestial bodies for years, information that could be "spatial"-enabled. There has been a growing recent interest in data collected about the Earth's moon as in the case of NASA's [Lunar Reconnaissance Orbiter](#), its [Lunar CRater Observation and Sensing Satellite \(LCROSS\)](#) and its [Lunar Mapping and Modeling Project \(LMMP\)](#), as well as [Google Moon](#) and other such projects. Spatial data can offer substantial value added for consumers of data through the use of location-rich metadata, as well as through the use of layering, allowing users of spatial data to explore layers of data (points of interest, elevation and other parameters) in an interactive fashion. What's more, the algorithms that drive SIS/GIS can be leveraged to represent data which is not just geographical based, such as bio-informatics, fingerprints search, facial search etc., providing substantial reuse benefits if an ASF-friendly software system that provided SIS/GIS functionality existed. Apache SIS will provide a manner in which spatial data such as that described above can be represented and used with existing technologies. The proposed founders of Apache SIS all have relevant and experience either developing spatial software that can easily perform the above tasks, or have experience working on the domains containing the georeferenced data of interest. We will leverage this experience and data expertise to deliver an Apache SIS system of use to a broad community of interest, making Apache an ideal home for this important software.

Background

There are several projects of different spatial capabilities available today, the two most common are:

- [GeoTools](#)
- [PostGIS](#)

Apache SIS goal is not aiming to compete with these tools but, instead, to provide a spatial framework that enables better representation of coordinates for searching, data clustering, archiving, or any other relevant spatial needs. By developing a toolkit framework that is independent of underlying implementation we hope to also reduce duplication of both software and effort with a published interface which other software projects can simply tie it into their own frameworks. The initial concept behind Apache SIS comes from [LocalLucene](#), an extension to [Apache Lucene](#) that provided a Geographical filter on top of the Lucene search library. [LocalLucene](#) went on to become [LocalSolr](#), and has since been included in many frameworks from [Spring](#) to [Hibernate](#), to [Hbase](#), and to [Compass](#). The [LocalLucene](#) framework has also been contributed to Apache Lucene under the moniker "Spatial Lucene", and currently exists as a contrib module within the Lucene project, version 2.9 and later. From January 2009-Dec 2009, while working on building out spatial capabilities in Apache SOLR for oceans-data and lunar-data related projects at NASA JPL, Chris Mattmann stumbled across [LocalLucene](#) and [LocalSOLR](#), and eventually discussed its limitations and benefits with Patrick O'Leary, along with the rest of the proposed committers in this effort. The consensus was there was a significant lack of a generic spatial data focused library out there in Apache land, and if present, such a library would present a unique contribution to the folks who were working with GIS data, that weren't only interested in search. In other words, there are a host of activities besides search (visualization, data reduction, statistical analysis) where a generic SIS/GIS library would be of prime importance. Both Chris, and Patrick, as well as the other committers had been stung by the issues in dealing with LGPL libraries and there was a difficult time finding any SIS library that was useful, and also ASF licensed. From these conversations, Patrick and Chris approached Ian Holman, and asked for his support in championing this proposal and helping to get this effort started. From there, we all agreed that the general community at large would be best served by establishing a top level project that focused primarily on solving spatial problems including search, visualization, data reduction and the aforementioned use cases.

- Apache SIS will also be the first known spatial project of this nature to be licensed under Apache License v2.0, [the vast majority of other GIS projects are LGPL](#). Further Apache SIS will be the first known (to our knowledge) Apache top level project focused on implementing spatial standards, and focused on building an Apache-based community in this thriving area.

Initial Goals

- The initial goals of the proposed project are:
- Viable community around the Apache SIS codebase
- Active relationships and possible cooperation with related projects and communities such as OGC
- Provide a geo-spatial coordinate system, with planetary plugins.
- Provide a polygon and line string coordinate comparison system.
- Build a Java framework to start out, but look to develop other P/L support (Python, Ruby, as a start).

Current Status

Meritocracy

All the initial committers are familiar with the meritocracy principles of Apache, and have already worked on the various source code bases (incl. Lucene Contrib, Tika, Nutch, and SOLR), providing issue comments, patches, and in some cases, committing (O'Leary & Mattmann) and participating as PMC members (Mattmann). We will follow the normal meritocracy rules also with other potential contributors.

Community

That Apache SIS community will be a co-mingling of several other communities that depend on Spatial & Geo Spatial solutions for their projects, the expectation is there will be members from the original [LocalLucene](#) project, the strong [LocalSolr](#) project, as well as Compass, Lucene and Solr at very early if not immediate stages. We will also look to garner support and contributions from other projects that are working in spatial, e.g., PostGIS, and other OGC efforts as well. There is already a growing number of folks at NASA who are also interested in spatial systems and work in the area. We will approach those people as well and attempt to bring them into the Apache SIS community. The idea would be for Apache SIS to grow into a top-level project that allows for sub projects based on SIS focus (visualization, data reduction/algorithms, OGC standards, etc.)

Core Developers

The initial developers come from a diverse set of backgrounds ranging from software architecture, search, academic, research/practice, to data mining. All of the proposed initial developers require the functionality of Apache SIS (Ramirez - LMMP, [McCleese](#) - oceans data, Mattmann -lunar/oceans, O'Leary - local search) in a compatible way.

Alignment

Existing Apache projects currently rely on the proposed starting point for Apache SIS, such as Lucene and Solr. We will begin by refactoring the [LocalLucene](#) contribution into a library independent of any underlying substrate (e.g., independent of Lucene). We will then look to add in functionality for calculating distances, functionality for persisting spatial data (to DBMS'es, search indexes, key/value stores, to [Hadoop](#)/etc.) We will follow by then focusing on data models and export of spatial data, culminating in an initial release that includes all of the basic functionality to at a minimum compute on spatial data, and store/export it.

Known Risks

Orphaned products

Several projects currently contain implementations of the initial code basis for Apache SIS, these projects can continue with the existing code base without impact, or adopt Apache SIS and reap the benefits of a common code base. Our goal is to provide value-added, shared ASL-licensed spatial software that is easy to adapt and adopt in any of the existing Apache (and external communities) developing SIS/GIS. Our initial focus will be on building a Java library but we will look at means for extending the Java library into additional P/Ls and frameworks.

Inexperience with Open Source

All the initial developers have worked on open source before and many are committers (O'Leary, Mattmann) and PMC members (Mattmann) within other Apache projects. [McCleese](#) and Ramirez are recent Apache committers on the soon to be initiated [OODT project that was accepted into the Incubator](#).

Homogenous Developers

The initial developers come from a variety of backgrounds and with a variety of needs for the proposed toolkit. Further, the developers consist of folks from at least two widely diverse companies, AT&T Interactive and NASA's Jet Propulsion Laboratory, spanning industry and government/research.

Relationships with Other Apache Products

Apache SIS is related to the following projects, non of the projects are direct competitors, but contain some functionality provided by Apache SIS

- Lucene Java, contains Spatial Lucene. We will look to leverage this code, combined with updates present at Local Lucene at Sourceforge as a starting point for the refactoring activity.

- Apache Solr, uses functionality from Spatial Lucene and may have some inspiration for how to perform some of the spatial computations we would like to have present in Apache SIS. Once Apache SIS matures, Solr could rely on SIS as a library component.
- Apache HBase - can index spatial reference id's and incorporate SIS query methodology to extend it to providing Spatial services once Apache SIS matures.

Initial Source

Apache SIS is an amalgamation of Spatial Lucene, and [LocalSolr](#) components.

- Spatial Lucene contains the original Spatial Coordinate system
- [LocalSolr](#) provides polygon and line string builders and comparator features.
- Local Lucene at Sourceforge contains a number of updates that we will merge into Apache SIS

The above code sources will serve as a basis for a fundamental generalization and refactoring activity that will result in an Apache SIS system focused on: spatial computation, and spatial data storage/export to start out. Activities such as visualization, reduction, and standards will occur downstream of this initial activity once the code base becomes stable.

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. Local Lucene is ASL-licensed, as is the other code bases (Local SOLR, and Spatial Lucene). All proposed committers have CLAs on file and are familiar with the code contribution process in Apache.

External Dependencies

At the moment, we will build Apache SIS so that it has no external dependencies, and is self contained. If we do require common dependencies, such as libraries for computation, or for storage/persistence, we will ensure that they leverage an ASL or compatible license. For example, to support persistence, we may leverage other libraries (e.g., Derby, K/V stores, etc.), and in these cases, we will focus on those libraries with a compatible license.

Cryptography

There is no cryptography required in Apache SIS at present time.

Required Resources

- Mailing lists
- sis-dev@incubator.apache.org
- sis-user@incubator.apache.org
- sis-commits@incubator.apache.org
- sis-private@incubator.apache.org

Subversion Directory

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

Issue Tracking

- JIRA SIS (SIS)

Other Resources

none

Initial Committers

Name	Email	Institution	CLA
Patrick O'Leary	pjaol@apache.org	AT&T Interactive	yes
Chris A. Mattmann	mattmann@apache.org	NASA Jet Propulsion Laboratory	yes
Sean McCleese	smccleese@jpl.nasa.gov	NASA Jet Propulsion Laboratory	yes
Paul Ramirez	pramirez@jpl.nasa.gov	NASA Jet Propulsion Laboratory	yes

Sponsors

- **Champion**
- Ian Holsman (ianh@apache.org)

Nominated Mentors

- Chris Mattmann (mattmann at apache dot org)
- Ian Holsman (ianh at apache dot org)
- Greg Reddin (greddin at apache dot org)
- Joe Schaefer (joes at apache dot org)
- Kevan Miller (kevan at apache dot org)

Sponsoring Entity

Apache Incubator