

Lucene.Net Proposal

FINAL

Submitted on Wednesday, January 12th, 2011

Lucene.Net - A .NET port of Lucene

Preface

Lucene.Net is a sub-project which is being spun off from the Lucene TLP but is not yet ready for graduation. We propose to address certain needs of the project by transitioning to an Incubator Podling.

Abstract

Lucene.Net will be a port of the Lucene search engine library, written in C# and targeted at .NET runtime users.

Proposal

Lucene.Net has three aims. First, it will maintain the existing line-by-line port from Java to C#, fully automating and commoditizing the process such that the project can easily synchronize with the Java Lucene release schedule. Second, it will be a high-performance C# search engine library. Third, it will maximize its usability and power when used within the .NET runtime. To that end, it will present a highly idiomatic, carefully tailored API that takes advantage of many of the special features of the .NET runtime.

Background

Lucene.Net, began as a independent project focused on creating a line-by-line, API for API port of Java Lucene to C#. It continued successfully in this way and eventually became a ASF Incubator project in April of 2006 and graduated as a sub-project of Lucene in October of 2009.

The last year has been challenging for the project. The committers who originally lead the project have stopped maintaining it and development has stagnated since June of 2010. The user community has spoken out requesting a change in philosophy and direction for the project, but those requests have been unheeded. This has led to a number of forks outside of the ASF. We would like to bring those forks back in as branches and be responsive to the needs of community without the need for multiple non-ASF forks.

The Lucene PMC wants to see the project continue to thrive and has indicated that a return to the Incubator is an appropriate step, with the end goal of building a new team of committers and maintaining a steady release cycle meeting the previously stated goals. Because Lucene is working to move away from being an "umbrella project", a long term goal of the Lucene.Net project is to graduate to an ASF TLP.

Rationale

There is great need for a search engine library in the mode of Lucene within the .NET runtime. Individuals naturally wish to code in their language of choice. Organizations which do not have significant Java expertise may not want to support Java strictly for the sake of running a Lucene installation. Developers may want to take advantage of C#'s unique language features and the .NET runtime's unique execution and interoperability model. Lucene.Net will meet all these demands.

Apache is a natural home for our project given the way it has always operated: user-driven innovation, lively and amiable mailing list discussions, strength through diversity, and so on. We feel comfortable here, and we believe that we will become exemplary Apache citizens.

Initial Goals (to be completed before Feb 2011)

- Build a new list of committers
- Make a 2.9.2 compatible release as quickly as possible (this already exists, it just needs to be packaged correctly)
- Update website, documentation, etc.
- Create a well documented repeatable and fully automated language porting process
- Start a ".NET style API" branch, either by incorporating some or all existing fork projects or by starting a new branch to this end

Current Status

Meritocracy

We understand meritocracy and will fully embrace this concept in our project management methodology. One of the proposed committers, DIGY, has been a committer on the current Lucene.Net project since November 2008. Prescott Nasser has been a contributor on the project, submitting patches, documentation, and website enhancements. Three of the other proposed initial committers, Troy Howard, Chris Currens and Sergey Mirvoda are both already actively involved in other open source projects, either as committers of code or in coordination roles. Troy, Chris, Sergey and Prescott are currently committers on a Lucene.Net fork known as Lucere, and as such are intimately familiar with the code base and share a vision for the future direction of the project. Scott Lombard and Michael Herndon are passionate about Lucene.Net as well and have already contributed significantly in terms of project organization and direction and discussions on the mailing list.

All of the proposed committers are familiar with the challenges faced with starting and maintaining a project over time. We also understand that opportunity is essential to an effective meritocracy and so will remain transparent, open and actively engage the community to find new contributors, include and review their contributions, and bring them on as committers as appropriate.

Community

There is already a well established, active and vibrant community surrounding the Lucene.Net project. This is primarily a users community, as the previous committers have not engaged the user community to find or leverage would-be contributors. There is a lot of talent available in the community to this end and many people have come forth offering their time as contributors.

There are a number of well established and significant .NET open source projects which are widely used by the larger .NET community which depend on Lucene.Net. There are also countless commercial products which use, and are dependent on this project. The mailing lists are active with numerous community members both asking and answering technical questions. The status and activities of the project are watched closely by the larger .NET development community and regularly commented on in blogs and other discussion forums.

Because of the size of the community and the fact that it's audience is largely developers, finding new committers over time should remain an easy task. The user base is also constantly growing, because Lucene.Net is one of the very few high quality products in this space (either commercial or open source). The ability to index and search content is an essential part of many web-based applications which are developed on the .NET framework, and Lucene.Net is widely used to support that scenario. This will only grow with time.

Community Projects and Commercial Products

A brief list of open source projects depending on Lucene.Net (alphabetical):

- iReaper - <http://ireaper.codeplex.com/>
- Linq to Lucene - <http://lingtolucene.codeplex.com/>
- Lucandra.NET - <http://lucandranet.codeplex.com/>
- MindTouch - <http://www.mindtouch.com/> (not sure why the wiki disallows this)
- Orchard - <http://orchard.codeplex.com/>
- RavenDB - <http://ravendb.net/>
- SubText - <http://subtextproject.com/>
- Umbraco - <http://umbraco.codeplex.com>

A brief list of known commercial products using Lucene.Net (alphabetical):

- Autodesk - <http://autodesk.com/>
- BBC UK Top Gear site - <http://www.topgear.com/>
- Epro, Bluewire Technologies - <http://www.epro.com/>
- IntelliEnterprise Intranet Suite - <http://www.adenin.com/>
- Koders.com, Black Duck Software - <http://www.koders.com/>
- Oxford Analytica - <http://www.oxan.com>
- Paradigm - <http://discover-e-legal.com/aboutparadigm.html>
- PowerSearch - <http://datassimilate.com/PSindex.html>

There are of course many, many more.

Contributors

In addition to wide adoption of the existing code base, there has been a strong interest from the community to contribute to the project. Beyond the initial committers list of seven (7) people, the following eight (8) people, who, despite being unable to fully commit to being committers, have come forth and offered to be contributors on the project (alphabetical):

- Alex Thompson <pierogitus AT hotmail DOT com>
- Ben Martz <benmartz AT gmail DOT com>
- Frank Yu <frank DOT yu AT farpoint DOT com>
- Glyn Darkin <glyn AT darkinsystems DOT com>
- Peter Mateja <peter DOT mateja AT gmail DOT com>
- Shashi Kant <skant AT sloan DOT mit DOT edu>
- Simone Chiaretta <simone DOT chiaretta AT gmail DOT com>
- Wyatt Barnett <wyatt DOT barnett AT gmail DOT com>
- Karell Ste-Marie <stemarie AT brain-bank DOT com>

This strong basis of support (fifteen initial contributors), if leveraged correctly, will represent a strong development team and create a spirit of productivity in the community which will encourage more contributors to appear and maintain the project.

Core Developers

The core developers are a diverse group of developers many of which are already very experienced open source developers. Here are some quick descriptions of the initial committers, their backgrounds and qualifications (alphabetical):

Individuals

- Chris Currens is a passionate developer who has worked on a number of open source projects.
- DIGY is one of the current committers on Lucene.Net and brings extensive experience with this specific codebase, as well as Java, C# and other language knowledge.
- Michael Herndon is the rare mix of a designer, software developer, & tech lead that has incorporated opensource software into both commercial, educational (uva's [blacklight](#)) & government (saic's [pathfinder](#) C# version & updated logo/images which is still in use.) and has worked on a few opensource projects including mitre's [pophealth](#).
- Prescott Nasser is a product manager and lead software developer focusing on back office workflow and process automation at a major US financial institution.
- Scott Lombard is an Automation Engineer who works in developing applications to allow users to manage and understand data. He has worked implementing various open source projects in websites and applications to show plant floor data.
- Sergey Mirvoda is a skilled developer with a strong interest in unit testing. He is a committer on the Lucene project.
- Troy Howard is an experienced developer and software project manager in the commercial world. He has founded and been a committer on numerous open source projects over the years, including most recently, Lucene, a fork of Lucene.Net which he hopes to integrate back into the main project.

Groups

- Chris, Prescott, Sergey, and Troy are part of the core development team on the Lucene project. They are eager to bring their passion for Lucene.Net back to the main project, as well as integrate the Lucene source into a new API for Lucene.Net.
- Chris and Troy work together at their day jobs as well as having worked together on two previous open source projects. They make a solid and productive team no matter what project they are working on.

Alignment

Lucene.Net has been an ASF project since 2006 and has benefitted greatly from that affiliation. We appreciate the careful oversight and structure that Apache provides which ensures that the project stays on track and productive. We also appreciate being associated with the Lucene TLP and the sharing that provides.

Beyond that, a very practical concern is that we would like to continue developing Lucene.Net with its current name. The project has built a hard earned reputation, and leaving the ASF would mean a forced rebranding and losing that reputation in the process. This would not be good for the health of the project.

Known Risks

Orphaned products

The purpose of this proposal is to recover from the fact that Lucene.Net has been orphaned by its current list of committers. There are numerous reasons why that happened, such as; Project vision not aligned with community needs, Committers not taking advantage of contributors in the community, Committers not being upfront about their ability or interest in maintaining the project, Lack of effort to incorporate new committers from the community or engage non-committers in the development process, Abandonment by commercially-focused committers .

Because we will be coming from the perspective of recovering from orphaning, we will be strongly focused on building a community, team of committers, and process to ensure our long term stability. We will learn from the past and not repeat the mistakes of our predecessors.

Beyond that, there is significant commercial interest in this project which we believe can be converted to direct support in terms of on-the-clock work by developers working for companies that have software products which rely on Lucene.Net. The initial committers list includes two such developers and we hope to attract more of them. Because we understand that commercial support of this nature can be fickle, we will also ensure that the team remains diversified and seek out committers who are personally motivated.

We also hope to incorporate the three existing forks of Lucene.Net back into this project. Doing so would bring a large body of reliable committers and contributors into the fold of this project (Lucene has more than 10 active committers, Lucille and Aimee.Net are one-man projects, both of which are very committed individuals).

Inexperience with Open Source

The core developers all have significant experience with open source development. We recognize that we lack PMC experience and seek to address that deficiency by using the Incubator environment to educate ourselves and prepare for responsible self-governance.

Homogenous Developers

Our community is geographically dispersed, with members in many areas of the USA, Canada, Russia, UK, and other countries. We all work for different organizations.

Reliance on Salaried Developers

We incorporate both salaried and non-salaried developers, from multiple organizations. We feel this gives us the best of both worlds and will increase our viability as a long-term project.

Relationships with Other Apache Products

Lucene.Net's relationship with the Lucene TLP has been relatively unidirectional until now. Lucene.Net has simply been porting the Java Lucene code to C# using automated methods. We hope to change that and feed-back more into the Java Lucene community both on the conceptual level and in terms of API changes that we make. We have an interest in possibly integrating the work of the Lucy project into Lucene.Net at some point as well. There is also a strong interest in creating .NET ports of ASF's Solr, Tika, Hadoop, and others. While that would fall outside of the scope of this project, there may be overlap in terms of the committers between those projects and sharing of code and methodologies pioneered in the Lucene.Net project.

An Excessive Fascination with the Apache Brand

Our desire to maintain Lucene.Net's affiliation with Apache has less to do with the brand and more to do with our conviction that developing the project 'The Apache Way' under Apache institutions is in Lucene.Net's best interests. However, we have to acknowledge that during its time as a Lucene subproject, Lucene.Net has not always fulfilled certain key requirements for an Apache project. In particular, it has failed to "release early, release often". Also, despite making significant progress in expanding its user community, it has failed to engage with the community and remain responsive to community needs.

By rebooting the project with a new list of motivated and enthusiastic committers, we expect to avoid the trap that ensnared Lucene.Net's first incarnation: we will release early, release often, accumulate users, nurture contributors, and grow our community.

Documentation

- Current Lucene.Net website: <http://lucene.apache.org/lucene.net/>
- Current Lucene.Net Subversion repository: <http://svn.apache.org/repos/asf/lucene/lucene.net/>
- Current Lucene.Net mailing lists: <http://lucene.apache.org/lucene.net/> (See info under "Mailing Lists" heading)

Initial Source

We will continue working with the existing Lucene.Net codebase located at: <http://svn.apache.org/repos/asf/lucene/lucene.net/>

We will also attempt to contact the coordinators of the following Lucene.Net forks and incorporate their work into the Lucene.Net project:

- Lucere - <http://lucere.codeplex.com/>
- Lucille - <http://lucille.codeplex.com/>
- Aimee.Net - <http://aimee.codeplex.com/>

Source and Intellectual Property Submission Plan

All source code referred to in this project (existing codebase and that of forks) is already licensed under the Apache 2.0 License. There should be no conflicts in this regard.

External Dependencies

The only external dependencies represented in any of the proposed code are on unit testing and mocking frameworks, all of which have ASF compatible licenses.

Required Resources

Mailing lists

- lucene-net-dev
- lucene-net-commits
- lucene-net-users

Lucene.Net already has lucene-net-dev, lucene-net-users, and lucene-net-commits mailing lists under lucene.apache.org. While these could be deactivated and the memberships migrated to the appropriate lists under incubator.apache.org, leaving the lucene.apache.org archives as read-only, we would prefer to keep the mailing lists the same, rather than moving to incubator.apache.org. The purpose of that would be to remain engaged with our community, with minimal disruption.

Subversion Directory

Lucene.Net already has a Subversion directory at <https://svn.apache.org/repos/asf/lucene/lucene.net>. In keeping with naming conventions, it could be moved to <http://svn.apache.org/repos/asf/incubator/lucene.net>.

Issue Tracking

Lucene.Net already has a JIRA tracker: Lucene.Net (LUCENENET)

Other Resources

Lucene.Net already has a [MoinMoin](http://wiki.apache.org/jakarta-lucene/lucene.Net) wiki at <http://wiki.apache.org/jakarta-lucene/lucene.Net>. It can be moved to standard Incubator wiki placement. There is currently no content of value in the wiki.

Initial Committers

Name	Email	CLA
Chris Currens	currens.chris AT gmail DOT com	Yes
DIGY	digydigy AT gmail DOT com	Yes
Michael Herndon	mherndon AT wickedsoftware DOT net	Yes
Prescott Nasser	prescott.nasser AT hotmail DOT com	Yes
Scott Lombard	lombardenator AT gmail DOT com	Yes
Sergey Mirvoda	sergey AT mirvoda DOT com	Yes
Troy Howard	thoward37 AT gmail DOT com	Yes

Affiliations

- Troy Howard and Chris Currens both work for discover-e Legal, LLC and will work on Lucene.Net as part of their paid work. discover-e Legal uses the current Lucene.Net build in their products and so, has a vested interest in seeing the project continue. Beyond that, the reason Lucene.Net was chosen by Troy initially for the discover-e Legal products is due to his strong interest in the project. This interest was well established before working for discover-e Legal, and is independent of its needs. The same is true for Chris Currens. So, even though there is commercial support for their work on this project it is not the primary reason or motivator for their interest.

Sponsors

Champion

- Grant Ingersoll (gsingers AT apache DOT org)

Nominated Mentors

- Gianugo Rabellino
- Stefan Bodewig
- Benson Margulies

Sponsoring Entity

Lucene.Net is currently sponsored by Lucene as a sub-project. This proposal advocates changing Lucene.Net's relationship with Apache from existing as a Lucene sub-project, to existing under the sponsorship of the Incubator.