LucyIncubatorProposal

This document has moved

The final version of the 2010 proposal to move Lucy to the Apache Incubator can be found at [http://wiki.apache.org/incubator/LucyProposal].

Preface

Lucy 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, and assimilating the KinoSearch codebase.

Abstract

Lucy will be a loose port of the Lucene search engine library, written in C and targeted at dynamic language users.

Proposal

Lucy has two aims. First, it will be a high-performance C search engine library. Second, it will maximize its usability and power when accessed via dynamic language bindings. To that end, it will present highly idiomatic, carefully tailored APIs for each of its "host" binding languages, including support for subclasses written entirely in the "host" language.

Background

Lucy, a "loose C" port of Java Lucene, began as an ambitious, from-scratch Lucene sub-project, with David Balmain (author of Ferret, a Ruby/C port of Lucene), Doug Cutting, and Marvin Humphrey (founder of KinoSearch, a Perl/C port) as committers. During an initial burst of activity, the overall architecture for Lucy was sketched out by Dave and Marvin. Unfortunately, Dave became unavailable soon after, and without a working codebase to release or any users, it proved difficult to replace him. Still, Marvin carried on their work throughout a period of seemingly low activity.

In the last year, that work has come to fruition: major technical milestones have been achieved and Lucy's underpinnings have been completed. Additionally, other developers from the KinoSearch community have taken an interest in Lucy and have begun to ramp up their contributions. The next steps for Lucy were articulated by the Lucene PMC in a recent review: make releases, acquire users, grow community.

To implement the Lucene PMC's recommendations and get to a release as quickly as possible, the Lucy community proposes to assimilate the KinoSearch codebase, which has been retrofitted to use Lucy's core. Lucy still lacks a number of important indexing and search classes; we wish to flesh these out via IP clearance work rather than software development.

Because Lucene is working to move away from being an "umbrella project", a long term goal of the Lucy project is to graduate to an ASF TLP. With that in mind, it seems more appropriate for the KinoSearch software grant to take place within the context of the Incubator, and that a Lucy podling and PPMC be established which will ultimately take responsibility for the codebase.

Rationale

There is great hunger for a search engine library in the mode of Lucene which is accessible from various dynamic languages, and for one accessible from pure C. 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 interoperability and fine-grained control. Lucy will meet all these demands.

Apache is a natural home for our project given the way it has always operated: user-driven innovation, security as a requirement, 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

* Make a 1.0 stable release as quickly as possible.
* Concentrate on community expansion.
* Expose a public C API.

Current Status

Meritocracy

Our initial committer list includes two individuals (Peter Karman and Nathan Kurz) who started off as KinoSearch users, demonstrated merit through constructive forum participation, adept negotiation, consensus building, and submission of high-quality contributions, and were invited to become committers. Peter now rolls most releases.

We look forward to continuing to operate as a meritocracy under the established traditions and rules of the ASF.

Community

Lucy's most active participants of late have been drawn from the KinoSearch and Lucene communities. Having been focused on features and technical goals for a long time, we are considerably overdue for a stable release, and anticipate rapid growth in its wake.

Core Developers

- Marvin Humphrey is the project founder of KinoSearch, and co-founded the existing Lucy sub-project. He is presently employed by Eventful, Inc.
- Peter Karman has contributed to several open source projects since 2001, including being a committer at http://swish-e.org/ (a search engine), http://code.google.com/p/rose/ (an ORM) and http://catalyst.perl.org/ (web framework). He is employed by American Public Media.
- Nathan Kurz is excited by the intersection of search and recommendations, and has been a KinoSearch committer since 2007. As the owner of Scream Sorbet (http://screamsorbet.com), he divides his time between code and fruit.

Alignment

One Apache value which is particularly cherished by the Lucy community is codebase transparency. We have developed institutions which enable us to measure and maximize usability (see http://wiki.apache.org/lucy/BrainLog), and we feel strongly that the bindings for Lucy must present APIs and documentation which are idiomatic to the host language culture so that end users can consume our work as easily as possible.

The controlled competition of meritocratic community development is also very important to us. There has been substantial cross-pollination of ideas between Lucene and Lucy, yielding considerable benefits for both projects. The Lucy developers envision that our host-language sub-communities will approach using and extending the library in distinct ways; we hope to harness the creative tension between them to drive innovation, building productive relationships akin to the one that Lucene and Lucy have today.

A third priority of ours is to be bound by existing Apache institutions, for the protection of all our stakeholders.

Known Risks

Orphaned products

All core developers have been associated with the project for several years across multiple jobs. However, at this time, the project would probably not survive the departure of Marvin Humphrey, so there is a risk of being orphaned. Marvin has no plans to leave, but we have been actively working to disperse his knowledge of the code base and administrative responsibilities in order to make him dispensable. Having staggered badly after Dave Balmain's departure, we are keenly aware of this vulnerability and highly motivated to eliminate it.

Inexperience with Open Source

The core developers all have significant experience with open source development, and include one present Apache committer. 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 San Diego, Oakland, and Minneapolis. We all work for different organizations.

Reliance on Salaried Developers

Marvin Humphrey has a great job at Eventful working primarily on this project and supporting applications that use it. Nevertheless, he is extremely dedicated to Lucy and is determined to see it through to the point where it becomes self-sustaining, regardless of work circumstances.

Relationships with Other Apache Products

Lucy's relationship with Lucene of cordial "coopetition" has produced benefits for Lucene users in terms of indexing speed, near-real-time search support, and more. We expect this dynamic to continue delivering improvements for all parties involved.

An Excessive Fascination with the Apache Brand

Our desire to maintain Lucy'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 Lucy's best interests. However, we have to acknowledge that during its time as a Lucene subproject, Lucy has not always fulfilled certain key requirements for an Apache project. In particular, it has failed to "release early, release often", and it has made minimal progress in expanding its community.

We attribute some of our difficulties to the what may have been excess ambition in the original Lucy plan, given the scope of the project and the size of the initial committer list:

- http://www.apache.org/foundation/how-it-works.html#incubator
- The basic requirements for incubation are:
  - a working codebase – over the years and after several failures, the foundation came to understand that without an initial working codebase, it is generally hard to bootstrap a community.

By rebooting the project with a working codebase, we expect to avoid the trap that ensnared Lucy's first incarnation: we will release early, release often, accumulate users, nurture contributors, and grow our community.

Documentation

- Current Lucy website: http://lucene.apache.org/lucy/
- Current Lucy Subversion repository: http://svn.apache.org/repos/asf/lucene/lucy/
- Current Lucy mailing lists: http://lucene.apache.org/lucy/mailing_lists.html
- KinoSearch Subversion repository: http://www.rectangular.com/svn/kinosearch/
- KinoSearch Perl API documentation: http://www.rectangular.com/kinosearch/docs/devel/
- KinoSearch Discussion list: http://www.rectangular.com/mailman/listinfo/kinosearch/

Initial Source

The initial source will be a snapshot from the KinoSearch subversion repository.

Source and Intellectual Property Submission Plan

KinoSearch is currently under a GPL/Artistic license. There are five individuals who have made multiple significant contributions to the codebase and whose participation is either essential or would be very helpful: Marvin Humphrey, Peter Karman, Nathan Kurz, Chris Nandor, and Father Chrysostomos. All have been contacted and are amenable to re-licensing their work and contributing it to Apache. We will contact as many other contributors as possible; if there are any that we cannot obtain permission from, we will refactor to expunge their work.

External Dependencies

The Perl bindings for KinoSearch currently depend on a few CPAN modules which do not have Apache-compatible licenses. It will be possible to eliminate all such dependencies if necessary.

Required Resources

Mailing lists

- lucy-dev
- lucy-private (with moderated subscriptions)
- lucy-commits
- lucy-users

Lucy already has lucy-dev, lucy-users, and lucy-commits mailing lists under lucene.apache.org. Perhaps 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.

Subversion Directory

Lucy already has a Subversion directory at http://svn.apache.org/repos/asf/lucene/lucy. In keeping with naming conventions, it could be moved to http://svn.apache.org/repos/asf/incubator/lucy.

Issue Tracking

Lucy already has a JIRA tracker: Lucy (LUCY)

Other Resources

Lucy already has a MoinMoin wiki at wiki.apache.org/lucy. It would be convenient to keep it, especially since its current location is also where it would end up upon TLP graduation, but we will defer to the wishes of the Incubator PMC if standard Incubator wiki placement is recommended.

Initial Committers

Name	Email	Affiliation	CLA
Marvin Humphrey	marvin AT apache DOT org	Eventful	yes
Peter Karman	peter AT peknet DOT com	American Public Media	yes
Nathan Kurz	nate@verse.com	Scream Sorbet	yes

Sponsors

Champion

- Chris Hostetter (hossman AT apache DOT org)

Nominated Mentors

- Chris Mattmann (mattmann AT apache DOT org)

Sponsoring Entity

Lucy is currently sponsored by Lucene as a sub-project. This proposal advocates changing Lucy's relationship with Apache from developing all new code as a Lucene sub-project, to instead assimilating existing code (KinoSearch) under the sponsorship of the Incubator.