Google Keeps Close Eye on Open Source

May 18 2007

Q&A: Chris DiBona, a programs manager for Google, talks about how the company uses open-source software and what it contributes to the open-source community.

Chris DiBona, open-source programs manager at Google, gave a talk called "A Year of Open Source at Google" for the Google New York speaker series. Prior to his talk, which was closed to press, DiBona spoke on May 16 with eWEEK Senior Editor Darryl K. Taft about a series of issues such as Microsoft's recent saber rattling over patents, Google's open-source development contributions and what GPLv3 means for Google.

What will you be talking about tonight?

I'm giving a modified form of one of my regular talks… about how Google uses open source, how we keep an eye on it internally, as well as what we do externally in terms of things like the Summer of Code, [and] code release—we release a great amount of code into open source. So I go around and try to talk about that.

Can you talk about the open-source components that are used in software development at Google?

One of the things I talk about is the kinds of projects that we patch, that we use internally. Those include things like the Linux kernel, the GNU compiler collection, Python, Wine, Derby, Aspell, DSpace, Autoconf,

MySQL, all kinds of great things like that.

What about in terms of open-source software that's used in deployment or production at Google?

We use the Linux kernel … every time you use Google, you're using a Linux machine. And then we have some fairly common open-source tools that we run on top of those, and then on top of those we run our proprietary software for serving Google, Gmail and all the different services.

What are the common tools you mentioned?

Things like the [GNU] binutils, like OpenSSL, OpenSSH, some network monitoring stuff… Basically things you would consider operating-system-level tools.

Are you involved with the Google Code (project hosting) project?

Yes, that's one of the Web sites that we master in our group.

How has that been going? How do you measure what's been going on there?

Well, there are a couple of facets of Google Code that are very important to us. One thing is we host a number of open-source projects that have nothing to do with Google on them. So in doing that we've become the No. 2 Web host for that kind of thing, after SourceForge. So that's been really fantastic.

Another thing that we do there is we turn on software and we have a bunch of documentation there about our APIs. It's sort of a way for coders and developers out in the world to learn more about Google technically, and how their programs can interact with Google technically. And it's been very successful at that. We're very happy with how that's gone.

From your perspective, what do you think has been the impact of the Google Summer of Code project?

Well, there have been a couple of things that are pretty important that have come out of the SoC. The first is that to date we've engaged somewhere around 2,000 developers now. This year it is 1,000 developers, last year it was 600 and the year before that it was 400. So there are 2,000 developers that we've introduced to open-source software development.

And on the other side of things, with open-source projects and having the opportunity to take on students in this manner, they've become very good at taking on new developers. So if you look at projects today, as opposed to when you looked at them three years ago, many of them now have processes and practices and ideas and ways of welcoming in new and inexperienced developers. And I think that's a very powerful thing and a very good thing for open-source software. So from both of those perspectives the project's been very successful.

That was part of what I was trying to get at—helping to bring more people who are interested in computer science into the world of open-source software…

Yeah, if you think about it, there's a lot of great software out there, but it's sort of a tough jump for somebody who's young to switch from being

a user of open source to being a developer of it. Because suddenly their code is out there for everybody to see, and they have to be able to interact with a lot of people who are pretty far along in their careers and people they often admire or are intimidated by. And so this is a nice way of making that happen, in my mind.

Do you have any data on how much open-source software Google gives back to the community?

Well, we've given over a million lines of code into open source. That's one way to measure it. And that's a good number; it's impressive, right? But I think more importantly, if you look at every major open-source software project out there—and a lot of minor ones—you'll find Googlers either patching or releasing new features or releasing code for them or right into those projects.

A good example of that is just recently we released a bunch of tools for enabling folks to use MySQL better, with replication and such. So that was really fun to do. And we've released all kinds of things, like incredibly minor changes to incredibly major things… Like the Google Web Toolkit, which is completely open source as well. So we think it's a really good way of sharing our level of innovation as a company with the rest of the world.

Are there any Google technologies that are currently in the pipeline to be open-sourced?

Well, as you know, we do not discuss things we have not released yet. And the reason we don't do that, by the way, is we like to make sure that when we launch something it's ready to be launched. So we're pushing pretty hard for some interesting things for Google Developer Day [May 31].

What impact will the GPLv3 have at Google?

Well, if you had asked me this nine months ago I would have said that it would mean that some GPL 3 programs, we wouldn't be able to adopt them because of the ASP [application service provider] provision in the original version. And I would have said at that time, and I still mean it, that that's not the end of the world, you know. We don't have to use every piece of open-source software out there.

But the most recent draft of GPL Version 3 has actually dropped that provision, so it makes it very easy for us to say it's likely that we'll welcome GPL Version 3 software into the company—even for things that may end up in production. Whereas before, if people opt to have that kind of restriction on [open-source software], we just couldn't use it in production and expose it to the end user.

It was sort of a thing that was like whatever they work on is fine with us, because we're very good at managing incoming code into the company. So it was never really a problem. The latest revision [of GPLv3] is actually pretty good.

Do you have any thoughts on Microsoft's recent claim that free and open-source software violates a large number of Microsoft patents?

Yeah, we saw that, and like most of the world we'd like to see them actually enumerate what [those patents] are. It's more of a wait and see. It's easy to say things like that, it's another thing to se what concrete actions come of it.

But if there is real meat to it then places like Google

would have to be concerned, I'd say…

You know, like I said, I don't know. There's just not enough information for us to know right now.

Does Sun's open-sourcing of Java have an impact on the way Google views Java as a development platform?

It doesn't change how we're looking at it, but it does increase the utility of Java for us. So before they had released Java as GPL, we had signed a source code agreement with them where we could give them patches and bugs and all this other stuff—because we have a lot of fairly advanced Java development going on at the company. We have folks like Joshua Bloch working for us and he's a very prominent Java developer and he's involved in the Java Community Process very heavily.

So we always had a way of getting patches in and some features developed. So that was fine for us. But with it being open source, it's actually better for us in a lot of ways, because we can access certain parts of the code in ways we couldn't before. And we can fix them and offer those fixes up without as much ceremony around submitting those patches and features. We can say, OK, it's an open-source project so we can just release this stuff. That's incredibly freeing for us. So we were very happy to see them go GPL there.

Do you, or have you done something like a Black Duck or Palamida assessment of your code?

No, and the reason why is we practice extremely tight control on how code comes into the company. And we're very, very good at training our

engineers. So, to give you an idea, I can look at any end binary in the company and I can tell you what open-source software is expressed within that—because of the way that we manage our code base.

So while those kinds of tools are interesting during an acquisition process—and we generally do not talk about our practices around acquisitions—they're not as interesting to us internally. Also, I think that expanding the utility of that code would be useful. Right now I'm not sure how incredibly useful that would be for us to run internally. They are good, quality projects, though.

Well, since you said you have a bunch of proprietary code running on top of a stack that consists of lots of open-source components, I was wondering how you could discern what was in there.

It's worth pointing out that it's much like if you're running an application on top of Linux. It's the same way we sort of run our Web servers, our Web applications. And then we have Linux as a kernel and as an operating system underneath it.

The way we actually bring code into the company when we're using an open-source library is extremely controlled. And the thing is, internally, Google as a company has always had a lot of discipline about how we bring code into the company.

Specifically, when you create a piece of code and you submit it, another Googler has to do a code review of your code before it ever gets into the code repository. And if somebody suddenly showed up and submitted 25,000 lines of code, well that would be questionable. And we have ways of dealing with that that are really very efficient. We tell people you want that to be inside this one directory, you want to tag it in a very

specific way so that we can track it… So we're actually quite facile at managing incoming code.

Copyright 2007 by Ziff Davis Media, Distributed by United Press International

Citation: Google Keeps Close Eye on Open Source (2007, May 18) retrieved 2 May 2024 from https://phys.org/news/2007-05-google-eye-source.html