

Blog Export: Rasmus' Toys Page, <http://toys.lerdorf.com/>

Thursday, May 8, 2008

SearchMonkey

One of the things I have been playing with lately is Yahoo!'s SearchMonkey project. It appeals to me on many different levels. The geeky name is a play on GreaseMonkey. But instead of writing plugins that run locally in the browser, SearchMonkey is a way to write plugins for the Yahoo! Search results page that change the appearance of the results themselves. Best explained with an example. Assume I am looking for a Japanese restaurant, and on my search results page I see:

That's ok, I guess. It tells me it is somewhere in Redwood City and that it is a neighborhood restaurant, whatever that means. Compare that to:

This gets me a real address and phone number plus a number of other useful bits of information. That is the first level SearchMonkey appeals to me on. The usefulness is obvious. My usefulness test is to see if I can explain it to my mother. Having her search for recipes and get pictures of dishes, ingredients and preparation times right on the search results page makes this an easy sell.

The second level this appeals to me on is the way it is implemented. Writing these SearchMonkey plugins becomes much simpler if the site you are writing the plugin for uses microformats of some sort. hCard, hCalendar, hReview, hAtom, xfn or generic structured eRDF or RDFa tags. The data can also be collected via a separate XML feed that can then be converted via XSLT in the SearchMonkey developer tool. The microformat data is collected and indexed and when you go to write a plugin and specify the url pattern you are writing the plugin for, it will find whatever indexed metadata it has for that url. If it doesn't have what you are looking for, you can still write a custom data scraper to get it, but that gets a bit more involved. I really like that the easy path is to add some sort of semantic markup to the pages. Yes, as Micah points out, this is not the (uppercase) Semantic Web, but it is still a push towards semantic markup. Having such a tangible and visible result of adding semantic tags is going to encourage people other than microformat geeks to do so. The more semantic markup we get, the better off the Web is.

The third part that appeals to me is the way the plugins are written. You write a little snippet of PHP. It is actually a method in a class you can't see, but its job is to return an associative array of data such as the title to display, the summary, extra links to show and whatever other key/value pairs you might want in the output. Because you have a full-featured scripting language available, you can write quite complicated logic in one of these plugins and pull whatever data you want from the site the plugin is written for.

You can also write an add-on to your plugin which is called an Infobar. It is a little bar that is shown below the plugin and from an Infobar you can access arbitrary external services. This example shows it well:

This one shows an OpenTable reservation link and a Yelp review, but almost anything can go there as long as you can squeeze it into the limited space you have.

The SearchMonkey is still in its infancy. It needs developer support. If you are in Silicon Valley, please come to the Developer Launch Party next week on Thursday May 15. See the link for details. If you aren't in the area, or even if you are, sign up for a developer account at <http://developer.yahoo.com/searchmonkey/preview.html> and help encourage the Web to become more semantic.

Posted by Rasmus in PHP at 14:50