Organizing the web with microformats

HIDDEN MEANING

Microformats are simple HTML tags that reveal information about web data. We'll show you how to take advantage of this handy technology.

BY DMITRI POPOV

A recent concept called the semantic web [1] is building new meaning into ordinary HTML text. Previous issues of *Linux Magazine* have discussed semantic web initiatives, such as the Simile Project [2]. One of the simplest and most mature semantic web technologies is *microformats*.

Microformats are simple HTML tags that reveal information about the meaning and context of web data. For instance, a microformat tag might specify that text is intended to be part of a resume or business card.

A browser designed to recognize the microformat then interprets that data accordingly. The browser might display the data to resemble a business card, or, perhaps more importantly, a client-side application could extract the contact information and then copy it to a user's address book.

Microformats are also useful in situations in which a single web client must integrate information from several sources. For instance, if you use a microformat tag to embed the latitude and longitude coordinates for a restaurant, a browser with the necessary plugins could automatically plot the restaurant location on a Google map.

Like other semantic web technologies, microformats do not require the web developer to know in advance how the client will use the data. A microformat specification merely defines a format for associating text with context.

The recent Operator plugin and an earlier plugin called Tail bring microformat support to Firefox. In this article, I will introduce you to the world of microformats and take a close look at some of the more popular microformat options, such as *hCard* [3], *hCalendar* [4], and

geo [5]. Some other common alternatives are shown in Table 1.

For a complete description of all microformat specifications, see the microformats wiki [6].

Introducing Microformats

To explain what microformats are and how they can help you, I'll start with a simple example. Take a look at the following text:

Linux Pro Magazine
719 Massachusetts Street
Lawrence, KS
USA

The human eye immediately recognizes this text as an address. You can see the street name, the city, and the country. You know that USA is a country and Kansas is a state. Unfortunately, computers don't (yet) have the ability to understand information just by looking at it; they need a set of markers that identify the data. Microformats can help in this situation. Unlike other similar technologies, microformats don't attempt to provide a universal solution to identifying and marking all sorts of data. Each mi-

croformat is designed for a specific situation. As I will explain later, you can use the *hCard* microformat to specify contact information such as this address.

If you have ever worked with HTML and CSS, you will have no trouble understanding how microformats work and how to use them. The three main building blocks of microformats are the *div*, *span*, and *class* elements, which are all-purpose tools for adding structure to documents.

The *div* and *span* elements define divisions or sections in a document. Whereas the *div* element defines a section of a document similar to a paragraph, the *span* element is used to mark a segment directly in the text body. When used with the *class* attribute, the *div* and *span* elements can be used to define types of information that can't be described by HTML. For example, in an article, *< div class = "author" > Dmitri Popov < /div >* may be used to indicate the author's name, and *< span class = "pubdate" > August 17, 2007 </ span >* may be used to indicate a publication date. Both *div* and *span* can be used to add semantic markers to a section or a text segment of a document.

Simpler microformats such as Rel-License and Rel-Tag use another element called *rel*. In HTML, the *rel* element describes the relationship from the current document to the anchor specified by the *href* attribute. For the sake of simplicity, you can say that the *rel* element describes the resource that the *href* link points to. Knowing that, you can easily figure out that, in the case of Rel-License, the *rel* element describes the link to a particular copyright license.

Rel-License and Rel-Tag are perfect examples of how easy it is to add microformats to your existing content. For example, suppose you have published an article under the Creative Commons Attribution-Noncommercial 3.0 license. In this case, you are most likely to provide a copyright note linking to the particular license:

```
<a ref="http://⏎
creativecommons.org/⏎
licenses/by-nc/3.0/"> ⏎
Creative Commons Attribution-⏎
Noncommercial 3.0</a>
```

This link is good old HTML, but you can easily microformat it by simply adding the *rel = "license"* attribute to it:

```
<a rel="license" href="http://⏎
creativecommons.org/licenses⏎
```


Figure 1: The Yahoo! Creative Commons search makes use of the Rel-License microformat.

```
/by-nc/3.0/"> Creative Commons ⏎
Attribution-Noncommercial ⏎
3.0</a>
```

As you can see, turning the link into a microformatted copyright notice is not particularly difficult. The immediate advantage is that by adding the Rel-License microformat, you make the content of a web page available for search on the basis of license type.

Both Yahoo! and Google are aware of the Rel-License format and allow you to search for content on the basis of its license type. For example, Yahoo! has a dedicated Creative Commons search page [7] (Figure 1), whereas Google allows you to specify a license type in its

Table 1: Microformats

Format	Description
hCalendar	Calendar and event information
hCard	Contact information and other data describing a person or organization
Rel-License	License type for the content on the page; lets you search for content by license
XFN	Tags for representing human relationships (e.g., *friend, colleague, crush*)
adr	Address information
geo	Latitude and longitude data for integration with mapping utilities
hResume	A format for publishing resumes and CVs
hReview	Indicates that the content on the page is a product review
RelPayment	Indicates the presence of an online payment system

Listing 1: Example of the hCard Microformat

```
01 <div class="fn org">Linux Pro
   Magazine</div>
02
03  <div class="adr">
04
05   <div
    class="street-address">719
    Massachusetts Street</div>
06
07   <div>
08
09   <span class="locality">Lawre
    nce</span>,
10
11   <abbr="region"
    title="Kansas">KS</abbr> <span
    class="postal-code">66044</
    span>
12
13 </div>
14
15   <div><div
    class="country-name">USA</div>
16
17 </div>
18
19   <div>Phone: <span
    class="tel">+1-785-856-3081</
    span></div>
20
21  <div>Email: <span class="emai
    l">info@linuxpromagazine.com</
    span></div>
22
23 </div>
```
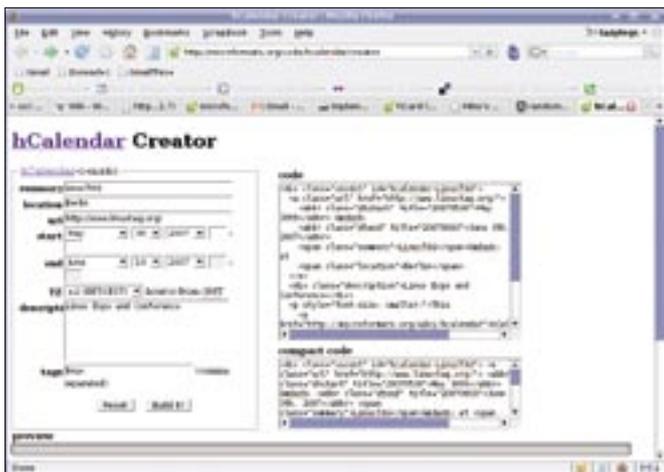
Figure 2: You can use hCalendar Creator to schedule an event and publish it.

Advanced Search section. The Rel-Tag microformat works in a similar way.

If your article describes wiki basics, for example, it makes sense to tag it with the "wiki" tag as follows:

```
<a href="http://
en.wikipedia.org/wiki
/Wiki" rel="tag">wiki</a>
```

This microformatted link consists of two parts. The destination of the link is called the tag space, and the part of the link after the last forward slash (/) is called the tag value.

The tag space is "a place that collates or defines tags" [8], which means that the tag space should link to a place that provides a specific meaning of the tag. In the example shown above, the link to the Wikipedia article provides the best possible (from the author's point of view) explanation of the wiki tag.

hCard

You can think of the *hCard* micro-format as an XHTML represen-tation of the vCard format, a widely accepted format for exchanging contact informa-tion between ap-plications. Al-though *hCard* is more complex than Rel-License and Rel-Tag, it is still easy enough to un-derstand. Here is what the previously mentioned address looks like in an *hCard*-formatted form (see Listing 1).

All the formatting in this example should be obvious, and if you don't feel like formatting your existing contact info by hand, the handy hCard Creator [9] can do this for you (Figure 2).

hCalendar

The *hCalendar* microformat handles cal-endaring data. Like *hCard*, *hCalendar* uses self-explanatory formatting (see Listing 2).

A microformat-ready browser that en-counters the *hCalendar* format can seamlessly integrate the data with other calendar and event information. The so-called *abbr* design pattern is used in the formatting to embed data without mak-ing it visible on the page. In this case,

abbr is used to embed the start and end dates of the event.

Geo

The *geo* microformat allows you to en-code latitude and longitude data into your web content:

```
<span class="geo">

<span class="latitude">39.00505
</span>

<span class="longitude">
-95.23297</span>

</span>
```

If you add this data to a web page, the Firefox Operator plugin, which I describe later in this article, will map the location with Google Maps. Another way to use the *geo* microformat is to embed geo-graphic data (geodata) directly into the web content with the *abbr* element. For example, if you are blogging about your recent trip to Berlin, your blog post might look something like this,

```
<abbr class="geo"
```

Listing 2: Example of the hCalendar Microformat

```
01 <div class="vevent"
   id="hcalendar-LinuxTAG">

02

03 <a class="url" href="http://
   www.linuxtag.org/">

04

05 <abbr class="dtstart"
   title="20070530">May 30th</
   abbr> &mdash; <abbr
   class="dtend"
   title="20070603">June 3rd,
   2007</abbr>

06

07 <span class="summary">LinuxTA
   G</span>, <span
   class="location">Berlin</
   span> </a>

08

09 <div class="description">Linu
   x Expo and Conference</div></
   div>
```
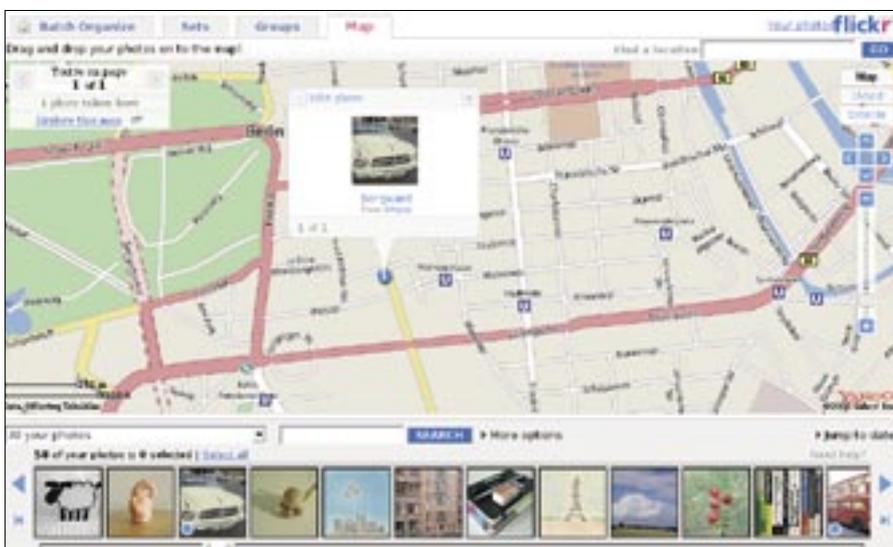

Figure 3: Flickr uses the geo microformat for geocoding photos.

Figure 4: Operator is a must-have Firefox extension that can manipulate microformatted content.

Figure 5: Tails is another Firefox extension that can come in handy when dealing with microformats.

```
title="52.51191;13.38519">

Mohren strasse</abbr>
```

which embeds a reference to my favorite street in Berlin.

Operator

To see what you can do with the tagged web page, you need the Operator extension for Firefox. You can download Operator from the Add-ons section of the Firefox website [10]. Operator was developed by Michael Kaply, who describes it as "an extension for Firefox that provides interoperability between microformats and various web services." In other words, the Operator extension is the tool that actually puts microformats to some practical use by acting as a mediator between microformatted content and web-based services that can process it. For example, Operator can feed *hCard*-formatted data to the Google Maps service, which uses it to locate the address on the map.

When installed, Operator adds a toolbar containing different tools. Point your browser to a page containing microformats and you can use Operator to perform different actions on the microformatted content. For example, if you open a blog post containing tags, Operator will automatically detect them and activate the related tools. You can use Operator to search Flickr photos, del.icio.us bookmarks, and blogs on Technorati containing a specific tag.

Previously, I showed you how to embed geodata into a web page. Because Operator can handle the *geo* microfor-

mat, you can use it to display the specified geodata on Google Maps, but that is not all. For example, the Flickr photo gallery site lets you use the *geo* microformat to add your photos to the map (Figure 3). Click on the *Map* tab in *Organizr* and place your photos on the map. This automatically adds geocoding to the photos. Now if you view the mapped photo, Operator will allow you to map it on Google Maps with the use of embedded geocoding (Figure 4).

If you have microformatted contact information embedded into your web page or blog, users can extract it easily. To export the contact information, you can use Operator or Tails Export [11], which is another Firefox extension that can process microformatted content. Tails is not as flexible as Operator, but it is quite useful. When you click on the Tails icon in the browser status bar (the icon turns orange when Tails detects microformatted content on the page), you can see a nicely formatted list of all the available contacts and events (Figure 5, left column). You can then export and add them to your address book or calendar.

Going Further

These examples are just a few samples of what you can do with microformats, but you don't have to stop here. Other microformats, including *hReview* and *hResume*, could also be useful for your particular needs, and additional tools can help you manage your microformatted content. For example, there is an *hReview* plugin for WordPress [12] and a few user scripts that can make the Operator extension even more useful [13].

Finally, the book *Microformats: Empowering Your Markup for Web 2.0* [14] covers everything you would want to know about microformats and how to use them. ∎

INFO

[1] Semantic web: *http://en.wikipedia. org/wiki/Semantic_Web*

[2] "Tomorrow's Toolbox: Semantic web tools of the Simile Project," by Oliver Frommel, *Linux Pro Magazine*, August 2007, pg. 56

[3] *hCard*: *http://microformats.org/wiki/hcard*

[4] *hCalendar*: *http://microformats.org/ wiki/hcalendar*

[5] *geo*: *http://microformats.org/wiki/geo*

[6] Microformats wiki: *http:// microformats.org/wiki/Main_Page*

[7] Yahoo! Creative Commons search page: *http://search.yahoo.com/cc*

[8] Rel-Tag wiki page: *http://microformats.org/wiki/rel-tag*

[9] hCard Creator: *http://microformats. org/code/hcard/creator*

[10] Firefox Operator add-on: *https://addons.mozilla.org/en-US/ firefox/addon/4106*

[11] Firefox Tails Export add-on: *https://addons.mozilla.org/en-US/ firefox/addon/2240*

[12] *hReview* plugin for WordPress: *http://www.aes.id.au/?page_id=28*

[13] Operator user scripts: *http://www.kaply.com/weblog/ operator-user-scripts/*

[14] Allsopp, John. *Microformats: Empowering Your Markup for Web 2.0.* Friendsof, 2007: *http://www.friendsofed.com/book. html?isbn=1590598148*