



> Semantic Web Use Cases and Case Studies

Case Study: Improving the Reliability of Internet Search Results Using Search Thresher

David Rooks, Segala

March 2007



General Description

The Problem

Search results are not as reliable as they can be. They don't provide any form of trust to end users to help them make informed decisions before entering Web sites. Some users may not wish to wade through search results before discovering a site that 'looks' like it 'can' be trusted, or until they stumble upon a site with a Trustmark.

Some users may only trust Web sites that have been vetted by an independent authority to guarantee that the information provided is trustworthy enough to rely on it. For example, if you conducted a search on treatment for a particular illness, how would you know which Web sites to trust?

Until now, it hasn't been possible to see from search results, which sites make conformance claims about suitability for mobile devices, disabled users or children. Furthermore, it hasn't been possible (until now) to see which sites follow codes of conduct which cover concerns such as privacy statements, returns policy and online advertising.

Companies like [Google](#) and [Yahoo!](#) are battling it out to be the first to enable relevance and reliability on the Web. They're failing miserably as they fight to build user profiles in the hope that it will help them deliver more meaningful search results. However, how often do you search for travel? Not often is likely to be the answer. So, what's the point in collecting our travel habits?!

Companies such as [Netscape](#) are making an attempt through its [Security Centre](#), first introduced in Netscape 8.1. [Mcafee](#) is trying with [SiteAdviser](#) and [GeoTrust](#) with [Trustwatch](#). Then you have [VeriSign](#) who use SSL certificates for security to enable trust on e-commerce sites and identification to help battle against phishing.

The problem with all of these implementations is that they're based on proprietary technologies that only those companies can make use of. They're not even compatible with each other. Furthermore, with the exception of SSL certificates, they're not scalable. Although, SSL certificates aren't without their own limitations too.

The Solution

Based on the principles of the Semantic Web, Content Labels are files that contain Metadata that enable search engines and browsers to provide more information about trust in search results. Like the title and description tags, Content Labels can be read and utilised by search engines and browsers to display more information about a Web site in search results.

"In its earliest days, W3C recognized a need to be able to describe content according to a defined vocabulary. This could be done for a variety of reasons including, but not limited to, child protection. The result was the *PICS* system which, despite early promise, has achieved limited support." Content Labels will be proposed as a replacement of PICS now that it has made it onto a [W3C Full Recommendation Track](#).

How To Demonstrate

Search Thresher is a Firefox extension that has been developed to empower users to search for Web sites based on reliability (trust).



Figure 1: Screen shot of search results containing label information for e-consultancy

How it Works

- User searches a keyword using Google
- Before Google returns the results page, Search Thresher checks to see if any of the Web sites contain a Content Label link tag (see below) in addition to the title and description tags.

```
<link rel="meta" xhref="http://www.segala.com/labels/tcuk_label_001.rdf"
mce_href="http://www.segala.com/labels/tcuk_label_001.rdf" type="application/rdf+xml"
title="Segala label"/>
```

- Pages that contain a link tag with Segala's name space are highlighted with a green tick, all other link tags are highlighted with an amber tick. Sites that don't contain a link tag are highlighted with a red box containing an X.

NB: Content Label Providers that independently verify assertions made for a particular standard or code of conduct in which they specialise, may wish to apply for a green tick to represent their conformance claims.

Each search annotation (icon beside each search result) is hyperlinked to a page, providing more information about the assertions, asserter, verifier (if applicable), data etc.

Key Benefits of Using Semantic Web Technology

For the end user

- Users can find what they are looking for and trust what they find
- Users will be warned if they are browsing to a site that has made fraudulent claims

For business

- Sites using Content Labels will get highlighted in search results. Depending on the user settings they may be the ONLY results displayed.
- Demonstrate to the end user that they care about standards and codes of conduct

¹ A Trustmark is normally awarded to a Web site to help the site own demonstrate their commitment and conformance to a standard or a code of conduct.