Semantics & Search Engine Optimisation

Jop Bröcker and Gerrit Jan van Ahee

August 29, 2008

Abstract

Although the Semantic Web is a concept of the late 1990s, it still has not been widely adopted due to lack of a killer application. This paper performs three case studies on search engines, to analyse whether these could offer the incentive to webmasters to add meta-data to their websites. This origins from the Search Engine Optimisation field, where webmasters are known to do ”anything” to get more traffic from search engines.

The results of the case studies shows Yahoo! takes the leads towards the Semantic Web. They have adopted RDF, RDFa and Microformats and offer an Application Programming Interface to webmasters to extend the Search Engine.

Google and Hakia have not entered the field of semantics yet. The conclusion of this paper is that the efforts of Yahoo! will encourge webmasters to add meta-data to their websites. As soon as Google enters the field, the Semantic Web will soon come within reach.

1 Introduction

This paper researches the field of the semantic web from a search engine optimisation perspective. The following paragraphs are an introduction to these concepts.

1.1 The Semantic Web

For years, the Semantic Web has been seen as the next step for the web, c.f [2]. This is a web where is meaningful to computers, which will take information retrieval to a higher level and makes it possible for automated agents to communicate and exchange data.

1.2 Progress in the Field of Semantics

At the current speed of development, the Semantic Web has a long way to go. Most websites and pages are still built for the human visitor and not for the machine to read, with little semantic information added to these pages. Even the most basic semantic HTML elements (title, description, keywords and meta) are often injected with lists of keywords and are identical for every page within a website. This makes them indistinguishable from one another, resulting in anything but useful material for the Semantic Web.

One possible cause for the lack of interest on the part of webmasters, is the absence of a killer application for the Semantic Web: there is no application that stimulates the webdesigner and webmasters to enrich their web-pages with semantic markup, let alone to add semantic documents to the web that can only be read by machines and agents (such as RDF). There is no incentive for website owners, e.g. businesses, charities etc, to spend money on it and to demand semantic markup.

1.3 Search Engine Optimisation

Search Engine Optimisation (SEO),[1] on the other hand, is big business. With the web and the power of search engines growing, site owners understand that a website not found in the Search Engine Result Pages (SERP), doesn't exist on the web. Webmasters spend a lot of time and money to improve their ranking. In this quest for top-rankings, their focus lies on the large search engines (Google and Yahoo!), as this will result in the most traffic.

The "killer app" needed to accelerate the advent of the semantic web, could very well be these search engines, as webmasters are willing to take effort into adding semantics to their website in return for more visitors from the Search Engines. At the moment, little is known about the use of semantics by the major search engines to improve search results. Smaller semantic aware search engines like Kartoo[2] and Hakia[3] have too little authority to convince the webmasters into semantification of their content.

Very recently however, Yahoo! announced the support for semantic web standards [11]. Since Search Engine Optimisation is so popular, semantic web enthusiasts hope this provides the "killer app" to motivate webmasters to enrich their documents.

[1] Search Engine Optimisation is the effort to optimise a website such that Search Engines, e.g. Google, index as much as possible of the information presented on the website.

[2] Kartoo is available at `http://www.kartoo.com`.

[3] Hakia is available at `http://www.hakia.com`.

1.4 Research Questions

For webmasters a new problem arises, to improve search engine results, understanding is needed of the use of semantics by search engines. This leads to our main research question: How does a semantic web page improve a websites ranking on search engine result pages? To answer this question, we will need to identify the different available semantic standards and we need to determine how this affects the search engine results within the different search engines. The subquestions concerned are What are the currently available semantic standards? and What semantics do Search Engines make use of?

This information will allow us to know which techniques are best applied to improve search engine results using semantic data on web-pages. Answers to the research questions will enable webmasters to embed semantic information on their websites, which the Search Engines will index. This will ultimately lead to improved rankings.

1.5 Outline of this Document

The outline of this paper is as follows: in Section 2 we will further explore the problem field and gain understanding of both the Semantic Web and Search Engine optimisation. The final subsection will discuss various methods of adding semantic data to a webpage. Section 3 discusses the case studies on Google, Yahoo! and Hakia and presents the results. Section 4 will provide some lessons learned for webmasters and webdevelopers based on the results found in the case studies. We will conclude in section 5.

2 Problem Analysis

2.1 The ever growing Web

The web is growing every day, its contained information doubling in size every year (source?). Although search engines do a fairly good job indexing this content and making it available to the public, they do not truly ”understand” the meaning of the information they index or the terms people search for. This is best illustrated with an example: When we search for ”book about hotels” it is clear to the reader we are not interested in booking a hotel, but are looking for a book. To be able to provide the right results, the search engine needs to ”understand” we are looking for a book (by processing it with a morphological analyzer which analyses the relation the words have to other words in the search phrase), but also needs to be able to determine whether a webpage contains information on a book, or information on bookings. Searching for ”book about hotels” on Google, shows us Google is clearly incapable of providing the right results, see Figure 1.

2.2 The Semantic Web

The semantic web is considered to be the solution to this problem: instead of building websites which are human-readable, the web should also be machine-readable, through the addition of semantics to web contents. Semantics consist of meta data (data about the data) and ontologies. For a machine to really “understand” what a document is about, agreements need to be made about how semantic value is added to a page. These argeements are called ontologies. Thus semantics add meaning to content with the use of meta-data and ontologies in such a way computer applications can process the information. Having

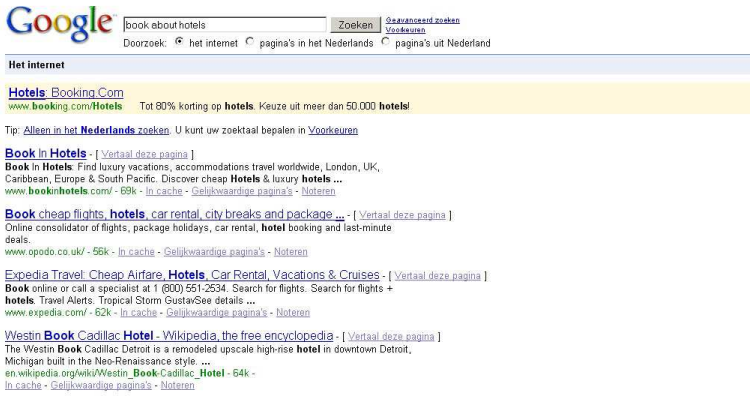

Figure 1: Screenshot google result page.

content enriched with meta-data, the possibilities seem endless to the semantic day-dreamer.

The first noticeable results will be in the field of Search Engines: search results will be more accurate (and be able to understand the meaning of the example search query above), visitors from search engines will be better targeted at the websites they visit. The consecutive progression will be web-browser add-ons which know how to process the information presented to the user. A basic example of this is RSS: if the webpage provides a link to a RSS feed (a standardised XML document), the browser gives the user the option to save this feed and to receive future updates. This will finally result in “agents”: automated, stand-alone software programs, capable of aggregating information found on various websites, able to interact with other agents and take over common tasks from the user.

Apart from the occasional RSS feed, site owners are reluctant to add more semantics to their websites. This may be caused by the lack of demand for semantic value, resulting in a low return on investment. Another reason could lie in the unprofessionalism in the webdesign sector, where websites are often built by people not educated in Computer Science, having little notion of the possibilities and advantages of the semantic web.

2.3 Search Engine Optimisation

Most website owners understand that if their website is not found by the mayor search engines, it may as well not exist. For this reason the Search Engine Optimisation industry flourishes. Search Engine Optimisation (SEO) is the effort of enhancing and changing a webpage in such a way that search engines show these pages at higher positions in the search result pages. Even webdesign-companies look into the material and take best effort into optimising the websites they produce, for the sake of additional visitors.

The SEO business is split up into two divisions: white hat SEO and black hat SEO. The former being the legitimate way of optimisation: trying to improve the quality of web-pages, which is rewarded by search engines with more traffic. The quality of web-pages is improved by writing good quality content and using valid HTML elements and CSS, separating content and design. “Plain old HTML” also contains some semantic elements: meta data about a page can be added through

the <title>, <meta name="description">, <meta name="keywords"> and <meta name="author"> elements. Since search engines have shown to use this information, site owners are more stimulated to supply these elements with correct and meaningfull values. Although this is meta-data for the website and usable by the search engine to match with the search terms, the search engine does not ”understand” the meaning of the meta data, as long as there is no ontology for that meta data.

Black hat SEO consists of reverse engineering the algorithms used by the search engines and finding the weak spots to exploit in order to achieve higher rankings on the search result pages. This is of course not appreciated by search engines and therefore results in a arms race between the black hat SEO specialist and the Search Engine. In this paper, we will be focussing on white hat SEO.

2.4 Adding Semantics to a Webpage

In this section we will adress different methods on how to add semantics to websites, starting from lightweight measures (using plain old HTML elements) to end with the more involving options (seperate documents for XHTML and RDF).

2.4.1 Keeping-XHTML-valid

As we noted in the previous section, web-page semantics start at the basic HTML elements intended for adding meta-data. Although the <title> and <meta> tags could help search engines and other computer programs, they still have no notion of the meaning of these words and the relation to other words used on the page. Search engines currently try to make sense of this with the use of linguistic models [12], but are having a hard time due to ambiguousness in languages (as demonstrated in Section 2.1 by the “book about hotels”).

Further meaning can be added through the proper use of title-attributes in HTML tags. This makes it possible to add words (information still not “understood” by search engines) to images, links or other elements. These words can again be interpreted by the linguistic analyzer to understand what a page is about.

What is missing from a semantic viewpoint are ontologies. These are not part of the standard XHTML recommendation, so this information either has to be placed in a seperate document, or somehow integrated in the XHTML recommendation.

Although there are no ontologies available to process XHTML meta-elements like Title, Description, Keywords Google has shown to use these elements to get a better understanding of the aboutness of the page. When a description contains a keyword searched for, it is displayed on the search result page. The same goes for the title, futhermore pages having a keyword contained in the title are preferred over pages that do not [15].

The above is however seen as common white hat SEO, already applied by many webdesigners and has yet been of little value to the success of the semantic web. We will therefore not analyse the searchengines for the use of XHTML meta data elements.

2.4.2 Microformats

Element attributes like `rel`, `for`, `class` and `id` can contain additional data on the element contents (as recommended by the XHTML specification). Microformats is an initiative to make conventions on the use of these attributes [22]. For example, if `class="vCard"` is used, a semantic parser knows this element contains

an address card of a person. This is all done without “breaking” the XHTML recommendation. This is accomplished by the agreement on the use of ”vCard” as an identifier for an address card, in which way it is comparable with FOAF (an RDF onthology). Currently, however, class-names are mostly used for document styles, adding anything but semantic value, e.g. `class="red-background"`.

2.4.3 Extending XHTML

As XHTML is part of the XML family, it is also extensible, although most validators do not accept this. Any XML not part of the XHTML recommendation is currently ignored by the webbrowsers, so enriching a document is not a problem. RDFa and eRDF provide a standard on how to include semantic data into XHTML documents. Using RDFa, it is possible to include ontologies like FOAF or Dublin Core to the XHTML documents. Although RDFa adds some attributes to XHTML (e.g. `property`), most meta data is added through the existing attributes like `class` and `rel`, drawing extra attention to these semantic features of XHTML and keeping it backwards compatible. RDFa is currently more widely adopted and a working draft of the W3C and will therefore be analysed for the different case studies in the next section.

2.4.4 Seperated from XHTML

A final option is to build the semantic web seperated from existing XHTML websites. RDF can be used to form these documents and they could contain FOAF or Dublin Core, or any chosen ontology. Two problems arise: For these documents to contribute to search engine results, they must be usefull to search engines or their visitors. Another problem is the difficulty for a human to generate machine-readable XML outside the tools they use to edit their website. Keeping these documents accurate and consistent is another problem when not edited with the website itself at the same time.

2.5 Conclusion

There are different ways to add semantic information to a webpage, having different implications for the webmaster. In our case studies we will focus on the following above found levels: Microformats, RDFa and RDF (seperated from XHTML files).

3 Case Studies

This section will look at three different search engines and analyze whether they make use of semantic information found in websites and how this affects the ranking of the websites in the search results. This study is conducted through the analysis of literature on search engines and the documentation provided by the search engines. We will focus on the above described application semantics and how it affects the search engine results.

We have chosen three different search engines for these case studies: Google, Yahoo! and Hakia. Google was chosen as it is the biggest player in the search engine market. Yahoo! (and its experimental SearchMonkey) have recently announced to start supporting semantics, which makes it an interesting case, as is has also a large userbase in the search engine market. Finally Hakia was chosen because it claims to be the only true semantic search engine providing results

only on concept match rather than keyword match or popularity ranking [9], which got the company in the 2008 c—net webware top 100 [6].

3.1 Google

By providing the search results users are looking for, Google quickly became the largest search engine, leaving incumbant Altavista behind. With a strong focus on relevant search results, Google has taken measures to both understand the web pages it indexes, and to understand the search term a user types in [17]. The former (understanding web pages) would have been a lot easier if lots of webpages provide semantics. The latter (understanding the user) is not likely to ever change, since users do not conform to ontologies. This could change when the users of Google more and more become automated agents (machines) which perform search queries [16]. Although it is likely the number of agents will grow, it is uncertain whether these will be using Google or that there are better alternatives at that time.

3.1.1 Microformats

Google has started to implement a couple of Microformats in some of their services.[4] Google Maps provides hCards for addresses found on the map [14]. With a mouse click the visitor can add this information to their address book, or let a machine process it in another way. This is however not exactly what we are looking for: This does not provide an incentive to webmasters to use Microformats. What would have been useful is a function where Google would display a map besides search results, based on the hCards found on a website.

Blogger (weblog software provided by Google) implemented hAtom in their weblogs.[5] hAtom is part of the microformats and intends to structure a web page containing multiple news-messages (as is often the case with weblogs). This way an agent knows which information is related and where the articles start and end. Again, Blogger implementing hAtom does not necessarily encourage other webmasters to do the same, this is only achieved when search engines use the provided semantic information. In that respect, there is no evidence that Google actually uses hAtom in their search engine. The only known Microformat used by the Google search engine is `rel="nofollow"`, which is specifically aimed at search engines, telling them it is an untrusted link resource. Altough this is useful to search engines, it is not the cornerstone of the semantic web. It does show how big the influence is of search engines on webmasters: the `rel="nofollow"` is widely adopted.

3.1.2 RDFa

Besides an invited Tech Talk on the Google Complex on RDFa [3] by Mark Birbeck (Editor on the RDFa working draft), no evidence can be found that Google makes use of RDFa embedded in XHTML pages. A reason could be that RDFa is relativly new [7], it only being a working draft of the W3C, not yet a recommendation [1]. On the other hand, Google has always been very mysterical on their algorithms, so it may be possible they are experimenting with RDFa, without the webmasters knowing.

[4] For an up to date overview see `http://microformats.org/wiki/implementations#Google_Search`.

[5] For an example of hAtom, see `http://groups.google.com/group/bloggerDev/browse_thread/thread/69344c5cc35b472e?hl=en`.

3.1.3 RDF

Allthough the RDF recommendation dates back to 2004 (rdf-primer), there is little previous literature on the use of RDF by google. Google does provide a means to see how many documents ending on .rdf it has indexed,[6] which shows us there are 420.000 RDF documents in the index. However, Google treats these documents exactly the same as conventional HTML documents. As semantic agents currently do not use Google as their search engine, these search results are of little use, since the average human visitor does not understand the meaning of the RDF document and most browsers do not know how to display such documents.

As Google is very mysterious about their algorithm, it is unclear whether they use RDF documents in the background. Even if they do use RDF, this is not clear outside Google.

This is again not a stimulance to website owners to enrich their contents with RDF documents. The effect of such efforts will not result in a growth of traffic from Google.

3.1.4 Conclusion Google

As Google has the largest market share, its movements are closely followed by webmasters and webdesigners, who will try to get as much traffic from Google searches as possible. The fact that Google has not yet shown any effort in making use of meta-data, is a disappointing finding.

On the other hand, Google has a reputation for deploying new features at high rate, so they could be possibly working on semantics in the background. This is however pure speculation, there is no concrete evidence.

3.2 Yahoo!

Yahoo! is quite the contrary to Google in their level of openness on their algorithms and use of semantics. It is only very recently that the announced their focus on the Semantic Web [21]. Peter Mika of Yahoo! addresses exactly the issue this paper focusses on: webmasters need an incentive to add semantic data to their websites, search engines could provide this incentive by showing they make use of this information [13]. Yahoo! introduced WebMonkey as their sandbox for semantic search. With SearchMonkey both Yahoo! developers and external developers can programm plug-ins (or add-ons) to the Yahoo! Search engine which make use of semantic information available on indexed websites. This makes it possible, for example, to create a search engine extension searching for all friends of a certain person, making use of the FOAF onthology.[7] The result can be displayed as a search result page, or as a XML stream, which could be interpreted by machines [18].

The next step is to analyse which formats of metadata and which onthologies are supported.

[6] see `http://www.google.com/search?q=filetype%3Ardfforthesearchresultsof Google`

[7] see Appendix A for a glossary.

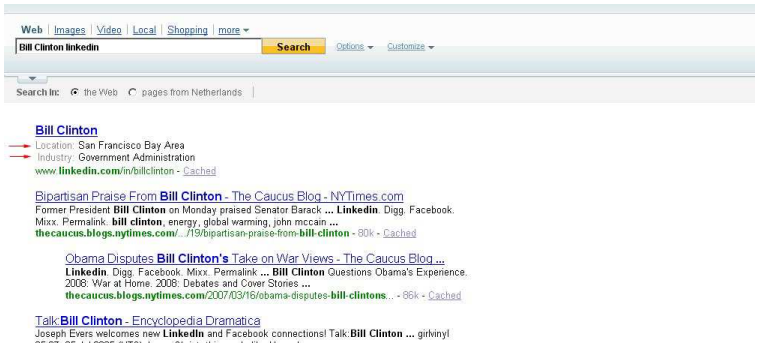

Figure 2: Searchmonkey in action.

3.2.1 Microformats

Yahoo! supports several Microformats: hCard,[8] hReview,[9] hCalendar[10] and XFN [11] [11] and will implement more in time, based on feedback they receive. An example is the LinkedIn website, where the hCard Microformat is used. When you search for "Bill Clinton LinkedIn" on Yahoo! (`http://search.yahoo.com/search?p=Bill+Clinton+linkedin`), the first result is the LinkedIn profile of Bill Clinton, see figure 2. Notice the "Location" and "Industry" information provided inside the search results. This is automatically extracted from the hCard on the profile webpage and shown on Yahoo!. This is just one example of a SearchMonkey application, developed by Yahoo! itself, anyone is free to develop and use more sophisticated extentions to the Yahoo! search engine.

3.2.2 RDFa

In the same article the support for Microformats are announced, Yahoo! also announces its support for various onthologies embedded through RDFa (and eRDF) [11, 18]. Currently Dublin Core, Creative Commons, FOAF and GeoRSS are supported onthologies, other onthologies will be added based on feedback. As with microformats, SearchMonkey provides the possibility to write extentions which make use of semantic data found in websites.

3.2.3 RDF

Yahoo! also indexes RDF pages containing RSS, GeoRSS, FOAF, Dublin Core or Creative Commons formatted data. The information is also accessible for the Search Monkey Apps, in the same way RDFa data is accessible. Yahoo! does not treat RDF different than RDFa, and uses both with their search results, as any developer could do when building SearchMonkey applications.

[8] See `http://microformats.org/wiki/hCard` for more information on hCard.
[9] See `http://microformats.org/wiki/hreview` for more information on hReview.
[10] See `http://microformats.org/wiki/hcalendar` for more information on hCalendar.
[11] See `http://microformats.org/wiki/xfn` for more information on XFN.

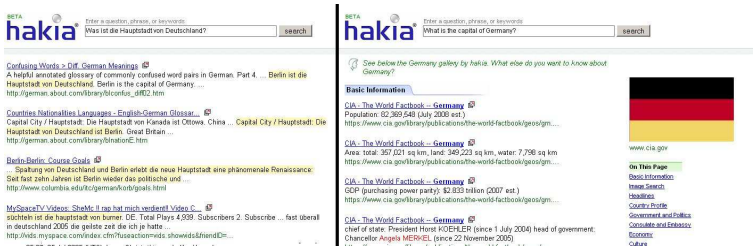

Figure 3: Different Hakia results for different languages, but the same query.

3.2.4 Conclusion Yahoo!

For Yahoo! the Semantic Web is the future and they clearly put everything at work to make progress in that field. Not only do they clearly communicate that they make use of semantic data, they also educate webmasters and webdesigners in the benefits of the Semantic Web and they even offer developers the opportunity to build their own Semantic Search engine.

It is clear Yahoo! is trying to solve the chicken and egg problem by taking the lead. They process meta-data in the search result pages to display extra information. It is unclear whether they also use the meta-data to rank search results. As long as they do not, or it remains unknown whether they do, this will not encourage webmasters to semantify their contents.

On the other hand, the meta-data enhanced search results stand out more between other results, so as Yahoo! adds more SearchMoney apps, the effort of adding meta-data results in more visitors from Yahoo!.

3.3 Hakia

Hakia aims to provide search results that have greater quality rather than popularity. During the indexing phase, the search engine focusses on the age of the web content and the credibility of the source, e.g. in the field of medicine, sources that are used by the Medical Library Association are considered to be credible. Using their algorithm of Semantic Rank, which relies heavily on onthological semantics and computational linguistics, and QDEX, Hakia's algorithm for extracting (potentially) meaningful information from content, the search engine will match a query based on it's semantic meaning, rather than the actual individual words. Often the search engine will engage in a dialogue to acquire a more specific query to narrow the search results.

Due to the relative scarcity of semantic metadata on the internet, such as microformats and RDF(a), as discussed above, Hakia have developed their own onthology. Rather than extracting information from RDF documents or other sources, they aim to derive the semantics implicit in the content. This is done using a natural language processor, making it quite hard to match queries to results in other languages, as grammars for each language will be needed. See the different results for the questions "What is the capital of Germany?"[12] and "Was ist die Hauptstadt von Deutschland?".[13] Figure 3 shows the visual difference in results. It is also nice to note, that the more grphically descriptive english version does not provide the answer.

[12] `http://hakia.com/search.aspx?q=What+is+the+capital+of+Germany.`
[13] `http://hakia.com/search.aspx?q=Was+ist+die+Hauptstadt+von+Deutschland.`

3.3.1 Hakia Conclusion

Hakia deliberately decided to develop their own ontology, since little websites provide semantic data through standardised ontologies. Allthough it should conflict with their own ontology, Hakia decided not to adopt any other forms of semantic information and ignores any Microformats, RDFa and external RDF. For this reason we will not analyse these forms for Hakia. The result is however, that even though Hakia is a supporter of the semantic web, it in no way encourages webmasters to enrich their websites with semantics. For our research, this is a rather disappointing conclusion for the search engine which sounded most promising from the Semantic community. This is however not disasterous, since the market share of Hakia is not of any influence.

3.4 Case Studies Conclusion

The above case studies have shown us that Google has not shown any interest yet in the Semantic Web, other than listening to a presentation. There is no evidence that Google uses meta-data in their ranking algorithm, not even rumors. Since Google is market leader, this has significant impact on the results of our research: Without the support of Google, the Semantic Web has a long way to go, as it has the largest influence on webmasters. In countries where its market share is largest (up to 94% market share in The Netherlands [5]), the efforts of other search engines will have little impact.

Yahoo! (which still has a 20% market share in the US [20]) does provide the nessesary incentives for webmasters. The question remains whether Yahoo!'s impact is large enough to convince many webdesigners, but chances are that Google will follow the initiative in the future.

Hakia decided not to analyse provided meta-data, but tries to generate their own meta-data using an Natural Language Processor. Although this could result in a semantic search engine, it will never result in the semantic web.

4 Semantics for Webmasters

Webmasters interested in the Semantic Web should take a close look at the developments at Yahoo!, as this is currently the pioneer of the search engines working on the semantic web. Adding meta-data in one of the supported onthologies, or one of the supported Microformats, could result in extra search engine traffic, as the meta-data enhanced search results stand out more than others. One question remains how webmasters can easily add meta-data to their website?

For less technical webmasters, who are using a WYSIWYG (What You See Is What You Get) XHTML editor to maintain a website, adding microformats or even RDFa can be difficult, as there is often no interface for these meta-data. A sollution would be the installation of plugins to the editor which then provide an interface.[14] Most Content Management System suppliers will be able to install such plugins.

However, a lot of information on a website is already entered in a structured way. Many Content Management Systems (CMS) offer a calendar view to enter events. For these events all meta-data needed for a hCalendar Microformat are already available. The CMS could be programmed to output calendar events on

[14] An example of a plugin can be found for the TinyMCE editor at
`http://undergroundwebdesigns.com/blog/how-to-use-the-tinymce-hcalendar-and-hcard-plugins-with-modx`

the website, including hCalendar meta-data. This way a webmaster is not asked to perform any other actions.

The same can be done for address information (hCard Microformat) and news items (RDF RSS). Although it is not a very hard thing to do for Content Management System suppliers, as long as webmasters do not explicitly ask for it, they unlikely to implement it.

5 Conclusion

Our case studies show that of Google, Yahoo! and Hakia, Yahoo! has recently started to focus on the Semantic Web. Meta-data in Microformats, RDFa and RDF is collected by Yahoo! and is used when displaying search results. Third party developers can build applications on top of the Yahoo! search engine which make use of the aggregated meta data. These application can also output their results as XML, will could be used by other applications, indicating a startingpoint for “agents”.

Google and Hakia on the other hand have shown no progress in the field of meta-data and semantics. Hakia claims to be a semantic search engine, but instead of using available meta-data and onthologies, they abstract their own meta-data and store it in their own onthology, rendering all semantic efforts of webmasters useless.

Hopefully Google will follow the initiatives of Yahoo!, as it has the largest market share and therefore the most influence on webmasters. Search engine aware webmasters will still be motivated by the efforts of Yahoo! and will probably not wait for Google. As soon as Google enters the field of semantics, those webmasters will be in advantage.

References

[1] Ben Adida and Mark Birbeck. *RDFa Primer, 12-07-2008*. W3C, http://www.w3.org/TR/xhtml-rdfa-primer/, June 2008. 7, 15

[2] Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web: Scientific american. *Scientific American*, May 2001. 2

[3] Mark Birbeck. *Google Tech Talk: All The Information In The World, The Way You Want It, 25-05-2008*. http://webbackplane.com/node/50. 7

[4] Dan Brickley and Libby Miller. *FOAF Vocabulary Specification, 30-05-2008*. http://xmlns.com/foaf/spec/, November 2007. 15

[5] citexxje. *Marktaandeel Google 94 procent*. http://www.techzine.nl/nieuws/12276/Marktaandeel-Google-94-procent.html. 11

[6] C—net, http://www.webware.com/html/ww/100/2008/winners.html. *The 100 top Web apps for 2008, 17-05,2008*. 7

[7] Bob DuCharme. *Introducing RDFa, 19-07-2008*. http://www.xml.com/lpt/a/1691. 7

[8] foaf. *The Friend of a Friend (FOAF) project, 30-05-2008*. http://www.foaf-project.org/. 15

[9] Hakia, http://company.hakia.com/technology.html. *Technology, 17-05-2008*. 7

[10] "Dublin Core Metadata Initiative". *Dublin Core Metadata Initiative (DCMI), 06-08-2008*. http://dublincore.org/, 08 2008. 15

[11] Amit Kumar. *The Yahoo! Search Open Ecosystem*. Yahoo!, http://www.ysearchblog.com/archives/000527.html, March 2008. 2, 9

[12] Udi Manber. *Introduction to Google Search Quality, 25-05-2008*. Google, http://googleblog.blogspot.com/2008/05/introduction-to-google-search-quality.html, May 2008. 5

[13] Peter Mika. *Semantic Search Arrives at the Web*. http://www.devx.com/semantic/Article/38595/0/page/1. 8

[14] Gregor J. Rothfuss and Kevin Marks. *Microformats in Google Maps, 15-07-2008*. Google, http://googlemapsapi.blogspot.com/2007/06/microformats-in-google-maps.html. 7

[15] Darren Rowse. *The importance of Title Tags in Search Engine Optimization, 25-05-2008*. http://www.problogger.net/archives/2005/05/21/the-importance-of-title-tags-in-search-e May 2008. 5

[16] Catherine Seda. *How to Win Sales and Influence Spiders: Boosting Your Business and Buzz on the Web (Voices That Matter)*. New Riders Publishing, Thousand Oaks, CA, USA, 2007. 7

[17] Amit Singhal. *Technologies behind Google ranking, 19-07-2008*. Google, http://googleblog.blogspot.com/2008/07/technologies-behind-google-ranking.html. 7

[18] Mattt Thompson. *RDF, XSLT, & the Monkey Make 3*. http://developer.yahoo.net/blog/archives/2008/08/rdf_xslt_and_the_monkey_makes_3.html. 8, 9

[19] W3C, http://www.w3.org/RDF/. *Resource Description Framework (RDF) / W3C Semantic Web Activity, 26-05-2008*. 15

[20] Jelle Wijkstra. *Google breidt Amerikaans marktaandeel gestaag uit, 25-08-2008*. http://www.automatiseringgids.nl/Markt-monitor/Markt%20in%20cijfers/2008/35/Google%20br 11

[21] Yahoo! Search Blog team. *Paving the way to Semantic Search, 06-08-2008*. http://www.ysearchblog.com/archives/000604.html. 8

[22] Tantek elik, Ali Diab, Ian McAllister, John Panzer, and Adam Rifkin andMichael Sippey. *hreview - Microformats, 12-07-2008*. http://microformats.org/wiki/hreview. 5

A Glossary

Dublin Core The Dublin Core metadata initiative maintains a vocabulary for describing library-related metadata which has been suggested by the W3C for use on the Web in stating authorship, copyright, and so on, using RDF. See also [10].

FOAF, Friend-Of-A-Friend FOAF is an XML namespace enabling people to incorporate personal information and relations to their web-pages. In order to accomplish this, FOAF information is best placed in a seperate file, which can be pointed to in the head of the web-page. Using FOAF, a web can be constructed of people linking to each other, as friends, sharing interests or living in similar areas.
See also [8] for more information on FOAF and [4] for the namespace specifications.

RDF, Resource Description Framework RDF is an XML standard for describing resources on the web. As this is a rather general and potentially ambiguous description, RDF is often extended by more specific projects, such as FOAF, RDFa and eRDF. RDF aims to make the management and navigation of web content easier to automate by providing structured metadata complementing the content.
See also [19].

RDFa RDFa supplies a number of HTML attributes that provide machine-readable information about the content of the HTML element. Using these attributes, web-content providers may mark up their human-readable content with machine-readable indicators, e.g. a title, personal information about the author (e.g. linking to a FOAF file) or an entire social network. RDFa can easily work together with other vocabularies, such as FOAF or Dublin Core.
See also [1].