ijecs
open access www.ijecs.in

International Journal Of Engineering And Computer Science ISSN:2319-7242

Volume 3 Issue 6 June, 2014 Page No. 6414-6417

RIAs AT ODDs WITH SEOs?

Baiju A S
Post-Graduate Student
Department of Computer Science and Systems Engineering
Geci Kerala, India

Abstract. Rich Internet Applications provides rich user experience owing to its advanced user interfaces and highly responsive design implementation. JavaScript, Ajax, Flash etc. are some of the widely used technologies to build such immense Rich internet applications. They offer great functionality to the applications in less lines of code. But on the flipside, it has the limitation, that Search Engines find it hard to index the contents of RIA.Temporary solutions does exists, but they are effective only up to some extent. This paper discuses some of the challenges faced by RIA developers to make their application Search engine friendly and some commonly followed workarounds to override these limitations. It also deeply examines these solutions and identify flaws or limitations of these methods, if any.

Keywords: Rich Internet Applications, Search Engine Optimization

INTRODUCTION

Rich Internet Applications have been around for a few years now. They are majorly focused on improving the client side functionality of websites making them more user friendly there by increasing the overall user experience and performance of the website .A variety of technologies and platforms are used to build these highly user interactive applications. Some of the most popular being Ajax, Curl, JavaFX, Microsoft Silverlight, Adobe Flash etc. It is a well-aware fact that, yes! RIAs do present challenges for Search Engines. Perhaps one might have spent over a few weeks and shed quite a fortune over building and maintaining such Rich Internet Applications, but optimizing it so that search engine crawlers can pick them up has always been sort of a challenge for RAI developers. But from the investor's point of view, it's a nightmare, because poor search engine ranking is going to affect the online business in a very bad way. Developers did come up with some solutions and workarounds, that they claim, has worked and will work, on live environments. Search engine bots can easily crawl HTML and other STATIC contents of a web page, but to the best of my understanding, there are no crawling techniques or tools that could efficiently and systematically crawl RIAs. However some academic tools does exists, but they have their own confinements and restrictions. So for the most of time RIA developers are forced to provide their content in a crawler friendly manner. This is a position paper to support my view that, RIAs has a long way to go, as far the google indexing and Search Engine Optimization is concerned.

• Challenges with RIA and search engines

In order to make the webpage load quickly and to improve the user interaction, modern web developers (RIA developers) adopt the SPA architecture (Single Page Application) .HTML or XML is used to build a lightweight static page which loads first and then JavaScript, Ajax, Flash, Silverlight etc. on the page builds up the rest of the page as it is being served to the browser .All modern web browsers heavily depend on these client-side code to fetch and deliver its contents. Data is dynamically loaded on the fly by making frequent requests to the server. In the good old golden days of programming, When a user types in a keyword and initiates a world wide web search, the search engine spiders crawls the HTML content of the website and check for matching contents and if doesn't find any match it moves on to the next URL. This is because search engine crawlers don't have the ability to interact with the Flash content or any dynamically loaded element for that matter.

• JAVASCRIPT CHALLENGE

JavaScript would be one of the most widely used and most preferred client-side scripting language owing to its ability to handle simple and complex task with great ease. An SPA cannot be directly indexed by search engines, because the search bots cannot execute client-side JavaScript. Without JavaScript the web site would be nothing but just a mere skeletal frame of markup language text, because data is read synchronically in a JSON format and the HTML content of the page is rendered on the client side itself

• Workarounds to get JavaScript contents indexed by search engines.

Placing the contents inside the <script> tags,

```
        <script type=”text/javascript”>
         document.Write (“what up google!! Can u index
me... uh??Test”)
        </script>
```

Placing the contents inside the <noscript >tags,

```
        <script type=”text/javascript”>
        document.Write (“Indexing Test”)
        </script>
        <noscript> what up google!! Can u index me...
uh??</noscript>
```

Placing the contents using the innerHTML method

```
       <p id="demo"></p>
       <script type="text/javascript">
        var content = "This is the actual inner text";
       document.getElementById
("demo").innerHTML=what up Google!! Can u
    index me...uh??
       </script>
```

Placing the contents inside an external .js file

GOOGLE, BING and YAHOO were able to index inside<noscript>tag, inside <script>tag, innerHTML method but could read the contents placed in a separate external .js file. Forcing additional tags inside page results in increase of the page size which negatively affects its speed. Also <noscript> method is widely misused by spammers and hence chances are that good search engines might black list your website

• AJAX CHALLENGE

It’s a well-accepted fact that, Ajax applications are difficult for the search engines to process because Ajax loads the contents dynamically. Ajax gives the ability to update only a small portion of webpage without having to reload the entire contents. As far as my understanding goes about websites and web site development, almost every good websites in the world uses Ajax technology to increase the speed and improve the efficiency and performance of the website.

• Workarounds to get Ajax Contents Indexed by search engines.

HIJAX METHOD

First step would be to let the crawlers know that your site support Ajax crawling scheme buy including a special token harsh fragment (everything after the #) in the URL. Hash fragments begin with the ! Character. E.g.: www.example.com/ajax.html#!key=value .once the site adapts the harsh fragment scheme the crawlers can access its contents if site supplies HTML snapshots. When you supply the HTML snapshot of the URL the server needs to identify if it has to return an HTML snapshot (i.e. All the contents on page after executing the javascript) instead of a regular page. The crawler modifies the AJAX URL and the above URL will be temporally changed to www.example.com/ajac.html?_escaped_fragment_=key=value .there are 2 reasons for it

1. Harsh fragments can’t be sent to the server as part of the HTTP request because it a specification violation
2. To inform the server that it needs to return an HTML snapshot rather than the normal page sent to the browser.

The server now returns the HTML snapshot for www.example.com/ajax.html#!key=value to the crawler which is nothing but the original URL. In order to produce the HTML snapshot you can use a headless browser such as the HtmlUnit or you can use tools such as crawljaax or watij.com.

META TAG METHOD

For Pages without hash fragments, include a special tag called <meta> tag which takes the form

<meta name=”fragment” content=”!”> In the head section of the HTML.

When you place the <meta name="fragment" content="!"> Into the www.example.com page the crawler temporally maps this URL to www.example.com?_escaped_fragment_= and requests it from the server. The server then returns the HTML snapshot for www.example.com

Google estimates that about 70% of all web content is dynamically created and this figure would only keep on increasing. Hijax method is not the best nor the most efficient Ajax crawling method but it works fine if the site is small in terms of the actual contents dynamically loaded and also if the website is already configured with Hijax. Sometimes the escape-fragment request needs to be redirected to another controller which in turn generates HTML and give it as a response to the crawl engine. This URL will also be crawled which is not so recommended as it might redirect us to a wrong page

• FLASH CHALLENGES

Flash keeps the website alive. It is an RIA developer’s favorite tool if they want to build websites that are highly user interactive and extremely responsive. Flash animations helps the website communicate with the users in a highly expressive manner. However it is also ill famous for its extremely odd behavior towards search engine crawlers. A highly responsive website falling behind in the website's ranking list does not throw an element of surprise at all.

- **Workarounds to get FLASH indexed by search engines.**

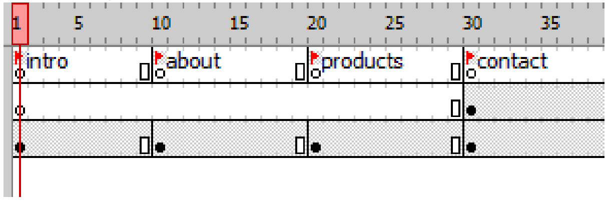
<u>SWFObject</u>

Swfobject is the common method of embedding flash movies into Html pages. HTML contents could be added in the placeholder <div> which can be accessed by Search Engines.

<u>ACCOMODATE DEEP LINKING</u>

Simple timeline-based flash website could be made crawler friendly by using a bit of PhP (*) and action Script. PHP is used to transfer the 'page url' into a flashvar and the SWFObject will create javascript code that can be accessed by search bots.

E.g.: let's say we have flash movie with a frame labeled 'contact' as given below.



The using the Accommodate Deep linking technique, the SWFObject will create a javascript code that looks like

```
var so = new SWFObject("movie.swf", "mymovie", "200", "100", "7", "#336699");
so.addVariable ("page", <? php echo $_GET['page']; ?>);
so.write ("flashcontent");
```

Which can be used to link to a specific page and also could be used in the google site map.

<u>SCALABLE INMAN FLASH REPLACEMENT</u>

This technique (sIFR) uses the javascript to read in HTML text and renders it in flash. This ensure that the HTML content and the FLASH content are identical which is very helpful for the crawlers to render the text in an anti-aliased format

Adobe made the first breakthrough with the release of SEO Technology center which uses "headless" version of flash players that can change states of the SWF content and gain access to the text contained in them. Though it could work and it did work to some extend it is not a permanent solution to this problem. More over the sIFR technique was found to work for only small swaths of text like headlines, pull quotes etc. Overuse of this trick makes the page load much slower. The SWFObject doesn't give a guarantee that HTML and the content in the FLASH are same and they doesn't reference the HTML text at all but simply runs a pre-compiled flash movie in the place of HTML. This is an open invitation for spammers to misuse this technique and generate spams by showing one type of content in FLASH and faking an entirely different type of content in another HTML.

• MICROSOFT SILVERLIGHT CHALLENGE

Silverlight is Microsoft's solution for building Rich Internet Applications. It is based on the WPF animation model**,** which is time-based which means that once you define the start and end conditions, we don't need to deal with matrices nor need to calculate the position of various objects in the frame, it automatically figures out how to do it. Search engines can crawl HTML and TEXT contents of the page and but Silverlight consists of dynamic and nonstandard elements like script , style ,objects and embedded tags which are not accessible by the crawl engines.

- **Workaround to get Silverlight indexed by search engines.**

<u>Combine HTML with Silverlight:</u>

Mixing HTML text with Silverlight content in the same page increases the functionality of the application, and at the same time, native HTML content could be indexed by crawl engines. Silverlight content could be placed inside, or around, a block of text.

<u>Putting Strong <H1> title in text above the Silverlight application:</u>

The text contents like, titles, page and section headers, body content, and exchange content etc. Are easily indexed by search engines.

<u>Placing Captions under the Silverlight application with an <H2> title tag:</u>

Adding H2 captions to integrate the target keywords will give the search bots an idea of what the page is about.

<u>Incorporating keywords in the Silverlight application name:</u>

for example if you want your page to be indexed when user types "web hosting", then name your Silverlight application "webhosting.xap".

<u>Putting understandable text under the Silverlight application using a Silverlight Content Div ID:</u>

The Silverlight content <div id> should include the precise same content in the Silverlight application.

All these methods are nothing but different ways of embedding the indelible TEXT in the Silverlight applications, which cannot be considered as solution or solid method to resolve the indexing issue. This works for simple Silverlight application but get reduce the performance as the size of the application increases.

- ## CONCLUSION

We discussed some of the commonly used and working solutions for making RIA applications indexible by the Search Engine. Majority of the techniques are centered on embedding TEXT or KEYWORDS inside the code or along with the code, in one way or the other. They found to work well on simple applications, but as the complexity level increases, they tend to become less effective. There is a bound on the use of such techniques, as excessive use of them makes the website load slower and reduce their performance. Some of these techniques are extensively misused by spammers to flood the internet with spam contents. Security is another concern, even though not in the context of search engine indexing, but in general, JavaScript and Ajax has always been vulnerable to cross-site scripting (XSS) attacks and cross-site request forgery (CSRF) exploitation etc. No doubt that RIAs can build powerful magnificent websites, but the cloud of doubt still surrounds them: can RIAs improve the traffic to your website? As long as search engine tools can't crawl dynamic contents of a web site, and since Rich Internet Applications are dynamic by virtue, I would like to force myself to believe that RIAs has way to improve when it comes to making its contents visible to the outside world through search engines.

1. Bianchi, D. 2014. Search optimization techniques for RIAs | Adobe Developer Con nection. [Online] Available at: http://www.adobe.com/devnet/seo/articles/techniques_ria.html [Accessed: 9 Mar 2014]

2. Advanced Web Ranking Blog. 2013. How to Make JavaScript Indexible by Google. [Online] Available at: ttp://www.advancedwebranking.com/blog/make-javascript-indexable-by-google/ [Accessed: 9 Mar 2014].

3. Agileworks.ca. 2014. Single Page Application Architecture Tech Stacks: Examples and Characteristics Agile Works. [Online] Available at: http://www.agileworks.ca/blog/single-page-application-architecture/ [Accessed: 9 Mar 2014].

4. Amadeus Consulting. 2011. Are Rich Internet Applications SEO Friendly? [Online] Available at: http://blog.amadeusconsulting.com/are-rich-internet-applications-seo-friendly/ [Accessed: 9 Mar 2014].

5. Application, H. 2014. knockout.js - How to improve SEO for single page application - Stack Overflow. [Online] Available at: http://stackoverflow.com/questions/20048476/how-to-improve-seo-for-single-page-application [Accessed: 9 Mar 2014].

6. Arrow.dit.ie. 2014. [Online] Available at: http://arrow.dit.ie/cgi/viewcontent.cgi?article=1101&context=scschcomcon [Accessed: 9 Mar 2014].

7. Bing.com. 2014. How to Verify Bingbot - Bing Webmaster Tools. [Online] Available at: http://www.bing.com/webmaster/help/how-to-verify-bingbot-3905dc26 [Accessed: 9 Mar 2014].

8. Blogs.msdn.com. 2014. Search Engine Optimization for Silverlight Applications - Sine of the times - Site Home - MSDN Blogs. [Online] Available at: http://blogs.msdn.com/b/waldred/archive/2009/03/23/search-engine-optimization-for-silverlight-applications.aspx [Accessed: 9 Mar 2014].

9. Crawlable, M. 2014. Making AJAX Applications Crawlable - Webmasters — Google Developers. [Online] Available at: https://developers.google.com/webmasters/ajax-crawling/ [Accessed: 9 Mar 2014].

10. Enge, E., Enge, E. and Enge, S. 2008. How Flash Can Be Search Engine Friendly. [Online] Available at: http://searchengineland.com/how-flash-can-be-search-engine-friendly-13754 [Accessed: 9 Mar 2014].

11. Fragment, A. 2014. Ajax - angualrjs: google crawling with escape fragment - Stack Overflow. [Online] Available at: http://stackoverflow.com/questions/21936032/angualrjs-google-crawling-with-escape-fragment [Accessed: 9 Mar 2014].

12. Kumar, A. and Kumar, A. 2013. All indexible Content Should Be in HTML. [Online] Available at: http://www.futurewhisper.com/305/all-indexable-content-should-be-in-html/ [Accessed: 9 Mar 2014].

13. Pagetrafficbuzz.com. 2014. Adobe's Launches SEO Technology Center for RIA! - Search, Social News Page Traffic Buzz. [Online] Available at: http://www.pagetrafficbuzz.com/adobes-launches-seo-technology-center-ria/6112/ [Accessed: 9 Mar 2014].

14. Plaveb.com. 2014. How to Optimize Silverlight Based Web Applications for Search Engines. [Online] Available at: http://www.plaveb.com/blog/how-to-optimize-silverlight-based-web-applications-for-search-engines [Accessed: 9 Mar 2014].

15. Roytanck.com. 2014. Making Flash websites search engine friendly – Roy Tanck's weblog. [Online] Available at: http://www.roytanck.com/2007/02/14/making-flash-websites-search-engine-friendly [Accessed: 9 Mar 2014].

16. Tricksbag.com. 2013. Can Google access content inside JavaScript & js files? [Online] Available at: http://www.tricksbag.com/seo/can-google-access-javascript_content.html [Accessed: 9 Mar 2014].

17. Zafra, A. 2009. Adobe Launches SEO Technology Center for RIAs - Search Engine Journal. [Online] Available at: http://www.searchenginejournal.com/adobe-launches-seo-technology-center-for-rias/9311/ [Accessed: 9 Mar 2014].